# Real Analysis: A Complete Guide

Real Analysis

December 21, 2021


Elijah Thompson,
Physics and Math Honors


*Solo Pursuit of Learning*

# Contents

# Part I

# Single Variable Analysis

# Chapter 1

# Topology and Construction of the Real Line

## 1.1.0 Peano Arithmetic

We begin by forming our number system from the ground up, starting wtih Giuseppe Peano's (1858-1932) axiomatized system for the natural numbers.

**Axiom 1.1.1 (Peano's Axioms).** *Peano's system for the naturals consists of two central axioms:*

1. *Assume that there exists a set* $\mathbb{N}$ *and an element* $0 \notin \mathbb{N}$*. Define, notationally,* $\tilde{\mathbb{N}} = \mathbb{N} \cup \{0\}$*. Then, assume there exists an injective map* $s : \tilde{\mathbb{N}} \to \mathbb{N}$ *called the* **successor function**
2. **Mathematical Induction***: Whenever a subset* $S \subseteq \tilde{\mathbb{N}}$ *satisfies* $0 \in S$*, and if* $k \in S$ *then* $s(k) \in S$*, then this implies* $S = \tilde{\mathbb{N}}$*. Notationally we have*

$$(0 \in S \wedge (k \in S \implies s(k) \in S)) \implies S = \tilde{\mathbb{N}}$$

Given these axioms, we can define an addition operation on $\tilde{\mathbb{N}}$:

**Definition 1.1.2.** *We define the binary operation* $+ : \tilde{\mathbb{N}} \times \tilde{\mathbb{N}} \to \tilde{\mathbb{N}}$ *inductively for* $x, y \in \tilde{\mathbb{N}}$ *by stating*

1. $x + 0 = x$
2. $x + s(y) = s(x + y)$

The definition is performed inductively, or recursively. Fix $x \in \tilde{\mathbb{N}}$ and let $S = \{y \in \tilde{\mathbb{N}} : x + y \text{ is defined}\}$. By definition $0 \in S$. Further, if $y \in S$, $x+y$ is defined in $\tilde{\mathbb{N}}$, so $x+s(y) := s(x+y)$. Since $x + y \in \tilde{\mathbb{N}}$, by axiom 1 of Peano's arithmetic $s(x + y) \in \mathbb{N} \subset \tilde{\mathbb{N}}$. Thus, $s(x + y)$ is defined so $s(y) \in S$. Hence by mathematical induction $S = \tilde{\mathbb{N}}$ so $+$ is a well defined binary operation for all $x, y \in \tilde{\mathbb{N}}$.

We can similarly define multiplication:

**Definition 1.1.3.** *We define the binary operation* $\cdot : \tilde{\mathbb{N}} \times \tilde{\mathbb{N}} \to \tilde{\mathbb{N}}$ *inductively for* $x, y \in \tilde{\mathbb{N}}$ *by stating*

1. $x \cdot 0 = 0$
2. $x \cdot s(y) = x \cdot y + x$

This defines $x \cdot y$ inductively as in the case of $+$. Now, we define our unit:

**Definition 1.1.4.** *We define* $1 \in \mathbb{N}$ *by* $1 := s(0)$.

Now we can derive many of the standard properties of the naturals.

**Proposition 1.1.1.** *For all* $x \in \tilde{\mathbb{N}}$, $x + 1 = s(x)$.

*Proof.* Let $x \in \tilde{\mathbb{N}}$. Then $x + 1 = x + s(0)$, but then by our inductive definition for $+$, $x + s(0) = s(x + 0)$, and $x + 0 = x$ by definition, so

$$x + 1 = s(x)$$

■

**Proposition 1.1.2.** *For all* $x \in \tilde{\mathbb{N}}$, $0 + x = x$.

*Proof.* We proceed by induction on $x \in \tilde{\mathbb{N}}$. If $x = 0$, then $0 + 0 = 0 = x$, by definition. Suppose inductively that we have $x \in \tilde{\mathbb{N}}$ such that $0 + x = x$. Then $0 + s(x) = s(0 + x)$. But, $0 + x = x$ by the induction hypothesis, so $0 + s(x) = s(x)$ and our result holds for $s(x)$. Thus, by mathematical induction we conclude that $0 + x = x$ for all $x \in \tilde{\mathbb{N}}$. ■

**Proposition 1.1.3.** *For all* $x, y \in \tilde{\mathbb{N}}$, $s(y) + x = s(y + x)$.

*Proof.* Let $y \in \tilde{\mathbb{N}}$, and we proceed by induction on $x \in \tilde{\mathbb{N}}$. If $x = 0$ then $s(y) + 0 = s(y)$, and $s(y + 0) = s(y)$, so the base case holds. Now, suppose the proposition holds for some $x \in \tilde{\mathbb{N}}$. Then $s(y) + s(x) = s(s(y) + x)$ by definition, and by the induction hypothesis $s(y) + x = s(y + x) = y + s(x)$ by definition. Thus, $s(y) + s(x) = s(y + s(x))$, as desired. Thus, by mathematical induction we conclude taht $s(y) + x = s(y + x)$ for all $x, y \in \tilde{\mathbb{N}}$. ■

**Proposition 1.1.4.** *For all* $x, y \in \tilde{\mathbb{N}}$, $x + y = y + x$ *(Commutative Law).*

*Proof.* Fix $x \in \tilde{\mathbb{N}}$, and we proceed by induction on $y \in \tilde{\mathbb{N}}$. If $y = 0$ then $x + 0 = x$ by definition, and $0 + x = x$ by Proposition 1.1.2, so the base case holds. Now, suppose the proposition holds for some $y \in \tilde{\mathbb{N}}$. Then $x + s(y) = s(x + y)$ by definition, and by Proposition 1.1.3 we have $s(y) + x = s(y + x)$. By the induction hypothesis $x + y = y + x$, so $s(x + y) = s(y + x)$, and it follows that $s(y) + x = x + s(y)$. Hence, as $x$ was arbitrary, we have by mathematical induction that $x + y = y + x$ for all $x, y \in \tilde{\mathbb{N}}$. ■

**Proposition 1.1.5.** *For all $x, y, z \in \tilde{\mathbb{N}}$, $x + (y + z) = (x + y) + z$ (Associative Law).*

*Proof.* Fix $x, y \in \tilde{\mathbb{N}}$, and proceed by mathematical induction on $z \in \tilde{\mathbb{N}}$. If $z = 0$, $x + (y + 0) = x + y = (x + y) + 0$, so the base case holds. Suppose the proposition holds for some $z \in \tilde{\mathbb{N}}$. Then it follows that

$$\begin{aligned} x + (y + s(z)) &= x + s(y + z) = s(x + (y + z)) \\ &= s((x + y) + z) && \text{(by Induction Hypothesis)} \\ &= (x + y) + s(z) \end{aligned}$$

as desired. Thus by the axiom of mathematical induction we have our result. ■

**Proposition 1.1.6.** *For all $x \in \tilde{\mathbb{N}}$, $x \cdot 1 = x$.*

*Proof.* Fix$x \in \tilde{\mathbb{N}}$. Then $x \cdot 1 = x \cdot s(0) := x \cdot 0 + x = 0 + x = x$, by Proposition 1.1.4. ■

**Proposition 1.1.7.** *For all $x \in \tilde{\mathbb{N}}$, $0 \cdot x = 0$.*

*Proof.* We proceed by induction on $x \in \tilde{\mathbb{N}}$. If $x = 0$, $0 \cdot 0 = 0$ by definition. If $0 \cdot x = 0$ for some $x \in \tilde{\mathbb{N}}$, then $0 \cdot s(x) = 0 \cdot x + 0 = 0$, so by mathematical induction we have our result. ■

**Proposition 1.1.8.** *For all $x, y \in \tilde{\mathbb{N}}$, $s(x) \cdot y = x \cdot y + y$.*

*Proof.* Fix $x \in \tilde{\mathbb{N}}$, and proceed by induction on $y \in \tilde{\mathbb{N}}$. If $y = 0$, $s(x) \cdot 0 = 0 = x \cdot 0 + 0$. Suppose the result holds for some $y \in \tilde{\mathbb{N}}$. Then $s(x) \cdot s(y) = s(x) \cdot y + s(x)$ by definition, and $s(x) \cdot y = x \cdot y + y$ by the induction hypothesis. By Proposition 1.1.5, $(x \cdot y + y) + s(x) = x \cdot y + (y + s(x)) = x \cdot y + s(y + x)$. By Proposition 1.1.4, $x \cdot y + s(y + x) = x \cdot y + s(x + y) = x \cdot y + (x + s(y))$. Finally, by Proposition 1.1.5 again we have $x \cdot y + (x + s(y)) = (x \cdot y + x) + s(y) = x \cdot s(y) + s(y)$, as desired. The result follows by the principle of mathematical induction. ■

**Proposition 1.1.9.** *For all $x, y \in \tilde{\mathbb{N}}$, $x \cdot y = y \cdot x$.*

*Proof.* Fix $x \in \tilde{\mathbb{N}}$, and proceed by mathematical induction on $y \in \tilde{\mathbb{N}}$. If $y = 0$, $x \cdot 0 = 0 = 0 \cdot x$ by Proposition 1.1.7, so the base case holds. Suppose it holds for some $y \in \tilde{\mathbb{N}}$. Then $x \cdot s(y) = x \cdot y + x = y \cdot x + x$ by the induction hypothesis, and by Proposition 1.1.8, $y \cdot x + x = s(y) \cdot x$. Thus by mathematical induction we have commutivity of multiplication. ■

**Proposition 1.1.10.** *For all $x, y, z \in \tilde{\mathbb{N}}$, $(x + y) \cdot z = x \cdot z + y \cdot z$ (Distributivity).*

*Proof.* Fix $x, z \in \tilde{\mathbb{N}}$ and proceed by induction on $y \in \tilde{\mathbb{N}}$. If $y = 0$, $(x + 0) \cdot z = x \cdot z = x \cdot z + 0 \cdot z$ by Proposition 1.1.7. Suppose it holds for some $y \in \tilde{\mathbb{N}}$. Then $(x + s(y)) \cdot z = s(x + y) \cdot z = (x + y) \cdot z + z$

by Proposition 1.1.8. By the induction hypothesis, associativity, and the same proposition again we have

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

as desired. Thus we have distributivity of multiplication by mathematical induction. ∎

**Proposition 1.1.11.** *For all $x, y, z \in \tilde{\mathbb{N}}$, if $x + z = y + z$, then $x = y$.*

*Proof.* Let $x, y \in \tilde{\mathbb{N}}$, and we proceed by induction on $z \in \tilde{\mathbb{N}}$. If $z = 0$, $x + 0 = y + 0$ implies $x = y$, so the base case holds. If it holds for some $z \in \tilde{\mathbb{N}}$, then $x + s(z) = y + s(z)$ implies $s(x + z) = s(y + z)$. But $s$ is an inductive function, so $x + z = y + z$ which implies $x = y$ by the induction hypothesis, and by mathematical induction we have our result. ∎

**Proposition 1.1.12.** *If $x \cdot z = y \cdot z$ and $z \neq 0$, then $x = y$.*

*Proof.* Suppose $x, y, z \in \tilde{\mathbb{N}}$ such that $x\cdot z = y\cdot z$. We argue by contrapositive and suppose $x \neq y$. Then by Trichotomy (to be shown) $x < y$ or $y < x$. Without loss of generality suppose $x < y$. Then $y = x + u$ for some $u \in \mathbb{N}$. Then $y \cdot z = x \cdot z + u \cdot z$, and so $u \cdot z = 0$ by the cancellation property. If $z \neq 0$, $z = s(w), w \in \tilde{\mathbb{N}}$, so $0 = u \cdot z = u \cdot w + u$. As $u \in \mathbb{N}$ this implies $u \cdot w < 0$, but $0 \leqslant k$ for all $k \in \tilde{\mathbb{N}}$ which contradicts trichotomy. ∎

As noted in the previous proof we used properties of the standard order relation on the naturals which we shall now define and prove.

**Definition 1.1.5.** *If $x, y \in \tilde{\mathbb{N}}$ we say*

1. *$x < y$ if $y = x + u$ for some $u \in \mathbb{N}$*
2. *$x \leqslant y$ if $y = x + v$ for some $v \in \tilde{\mathbb{N}}$*

We could also say $x \leqslant y$ if and only if $y \in Rx = \{x + v : v \in \tilde{\mathbb{N}}\}$. We define $y > x \iff x < y$ and $y \geqslant x \iff x \leqslant y$ for all $x, y \in \tilde{\mathbb{N}}$.

**Proposition 1.1.13.** *For $x, y \in \tilde{\mathbb{N}}$, if $x \leqslant y$ and $y \leqslant x$, then $x = y$.*

*Proof.* As $y = x + v, v \in \tilde{\mathbb{N}}$, and $x = y + u$ for $u \in \tilde{\mathbb{N}}$ by definition, $x = x + v + u$, so by the cancellation property $v + u = 0$. Towards a contradiction suppose $v$ or $u$ is not $0$. Without loss of generality suppose $v \neq 0$, so $v \in \mathbb{N}$ and there exists $m \in \tilde{\mathbb{N}}$ such that $v = s(m)$. Then $0 = u + s(m) = s(u + m) \in \mathbb{N}$, which contradicts our axiom that $0 \notin \mathbb{N}$. Thus, $v = u = 0$, so $x = y$. ∎

**Proposition 1.1.14 (Trichotomy).** *If $x, y \in \tilde{\mathbb{N}}$, then one and only one of the following hold:*

$$x < y \text{ or } x = y \text{ or } x > y$$

*Proof.* Let $x, y \in \tilde{\mathbb{N}}$. If $x < y$ then $y = x + u$ for some $u \in \mathbb{N}$. If $x = y$ then $u = 0$, but $u \in \mathbb{N}$ so $u \neq 0$, and $x \neq y$. If $x > y$ then $x = y + v$, $v \in \mathbb{N}$, but by the proof of Proposition 1.1.13 this implies $u = v = 0$ contradicting the fact $u, v \in \mathbb{N}$. By similar arguments $x = y \implies x \not< y, x \not> y$, and $x > y$ follows from the first case. Let $y \in \tilde{\mathbb{N}}$ and proceed by induction on $x \in \tilde{\mathbb{N}}$. If $x = 0$, $y = y + 0$ so $y \geqslant x$. If $y = 0$, $x = y$, and if $y \neq 0$, $y \in \mathbb{N}$ so $y > x$. Suppose the claim holds for some $x \in \tilde{\mathbb{N}}$. If $x < y$, then $s(x) \leqslant y$. Then either $s(x) = y$ or $s(x) + u = y$ for some $u \neq 0$, so $u \in \mathbb{N}$ and $s(x) < y$. If $x = y$, $s(x) = x + 1 = y + 1$, so $y < s(x)$. A similar argument holds if $y < x$, since $y < s(x)$, completing the induction. ∎

We now define the partial function of subtraction on $\tilde{\mathbb{N}}$:

**Definition 1.1.6.** *If $x, y \in \tilde{\mathbb{N}}$ with $x \leqslant y$, then we define $z := y - x \iff y = x + z$, where $z \in \tilde{\mathbb{N}}$.*

Notice $y - x$ is well defined by the cancellation property of addition.

**Proposition 1.1.15.** *If $x, y, u \in \tilde{\mathbb{N}}$, with $x \leqslant y$, then $(y - x)u = yu - xu$.*

*Proof.* Let $x, y \in \tilde{\mathbb{N}}$ with $x \leqslant y$ and let $u \in \tilde{\mathbb{N}}$. Then there exists $w \in \tilde{\mathbb{N}}$ such that $y = x + w$, so $yu = xu + wu$ by distributivity, Then by definition $yu - xy = wu = (y - x)u$. ∎

Next we move on to a central property of the natural numbers which is equivalent to the axiom of mathematical induction:

### Theorem 1 (Well-Ordering Property of $\tilde{\mathbb{N}}$).

*If $T \subseteq \tilde{\mathbb{N}}$ is non-empty, then $T$ has a smallest element.*

*Proof.* We proceed by contrapositive. Suppose $T \subseteq \tilde{\mathbb{N}}$ and $T$ has no smallest element. Then $0 \notin T$, since for all $x \in \tilde{\mathbb{N}}$, $0 \leqslant x$, as either $x = 0$ or $x \in \mathbb{N}$ so $x = x + 0$ and $x > 0$. Let $S = \{x \in \tilde{\mathbb{N}} : x < y, \forall y \in T\}$, so $0 \in S$. Inductively suppose $x \in S$. If $s(x) \in S$, we're done, so suppose $s(x) \geqslant y$ for some $y \in T$. But $x < y$, so $y = x + s(w)$, for some $w \in \tilde{\mathbb{N}}$, and $y = s(x) + w$. Thus $s(x) \leqslant y$, so by Proposition 1.1.13, $s(x) = y \in T$. But $s(x) \leqslant t$ for all $t \in T$, so $s(x)$ is a minimal element of $T$, contradicting the hypothesis. Thus $s(x) \in S$, so by mathematical induction $S = \tilde{\mathbb{N}}$. Then as $T \subseteq \tilde{\mathbb{N}} \backslash S$, $T = \varnothing$ as desired. ∎

In the next section we perform an arithmetic closure of the naturals to obtain the rational field, $\mathbb{Q}$.

## 1.2.0 Construction of The Rational Field

First we need the notion of an equivalence relation for our constructions:

**Definition 1.2.1 (Equivalence Relation).** *An equivalence relation on a set $S$ is a subset $E \subseteq S \times S$ such that*

1. *For all $x \in S$, $xEx$ (reflexivity)*
2. *For all $x, y \in S$, if $xEy$ then $yEx$ (symmetry)*
3. *For all $x, y, z \in S$, if $xEy$ and $yEz$, then $xEz$ (transitivity)*

An important property of equivalence relations is there relation to partitions of a set: in particular, we have a bijection between partitions of a set and equivalence relations.

**Definition 1.2.2.** *For an equivalence relation $\sim$ on a set $S$, and $x \in S$, the* **equivalence class** *for $x$ is defined by*

$$[x]_{\sim} := \{y \in S : x \sim y\}$$

Note that $[y]_{\sim} = [x]_{\sim}$ if and only if $x \sim y$. Further, the equivalence classes for $\sim$ form a partition on $S$. That is $S = \cup_{x \in S}[x]_{\sim}$, and if $[y]_{\sim} \neq [x]_{\sim}$, $[y]_{\sim} \cap [x]_{\sim} = \varnothing$.

**Definition 1.2.3.** *Define $\sim$ on $\tilde{\mathbb{N}} \times \tilde{\mathbb{N}}$ by $(a, b) \sim (x, y)$ if and only if $a + y = x + b$.*

We consider $(a, b)$ to be $a - b$. We note that this defines an equivalence relation, we the proof left to the reader.

**Definition 1.2.4.** *We define the* **Integers***, $\mathbb{Z}$, to be the set*

$$\mathbb{Z} := \{[(x, a)]_{\sim} \in \mathcal{P}(\tilde{\mathbb{N}} \times \tilde{\mathbb{N}}) : x, a \in \tilde{\mathbb{N}}\} = \tilde{\mathbb{N}} \times \tilde{\mathbb{N}} / \sim$$

*and we have the natural injection*

$$\begin{aligned} \iota : \tilde{\mathbb{N}} &\to \mathbb{Z} \\ x &\mapsto [(x, 0)] \end{aligned}$$

We can define operations of addition and multiplication on $\mathbb{Z}$ inherited from $\tilde{\mathbb{N}}$.

**Definition 1.2.5.** *For $[(x, a)], [(y, b)] \in \mathbb{Z}$, we define*

$$\begin{aligned} [(x, a)] + [(y, b)] &= [(x + y, a + b)] \\ [(x, a)] \cdot [(y, b)] &= [(xy + ab, xb + ay)] \end{aligned}$$

To show these definitions are well defined we must show that the operation is independent of the choice of representative of each equivalence class. This is a routine check left to the reader.

**Definition 1.2.6.** *We define $0 = [(0, 0)]$ and $-1 = [(0, 1)]$ in $\mathbb{Z}$, and if $m = [(x, a)]$, we define $-m := [(a, x)]$.*

**Proposition 1.2.1.** *For all $m \in \mathbb{Z}$, $m \cdot (-1) = -m$.*

The proof is a quick calculation:

$$m \cdot (-1) = [(x, a)][(0, 1)] = [(x \cdot 0 + a, x + a \cdot 0)] = [(a, x)] = -m$$

Similarly, we have all the properties we derived for $\tilde{\mathbb{N}}$ for the operations on $\mathbb{Z}$, such as $m \cdot 0 = 0$, $m(n + k) = mn + mk$, $m + n = m + k \implies n = k$, and $m \cdot n = k \cdot n \implies m = k$ if $n \neq 0$. Now we have closed the $\tilde{\mathbb{N}}$ under ring operations, obtaining the integral domain $\mathbb{Z}$.

Next we perform a similar construction to obtain our field - this process is known as constructing a fraction field for an integral domain.

**Definition 1.2.7.** *Define an equivalence relation $\sim$ on $\mathbb{Z} \times \mathbb{Z}\backslash\{0\}$ by $(x/a) \sim (y/b)$ if and only if $xb = ya$, for $(x/a), (y/b) \in \mathbb{Z} \times \mathbb{Z}\backslash\{0\}$.*

It is routine to show that this is indeed an equivalence relation on the set. Next, we can define addition and multiplication operations:

**Definition 1.2.8.** *We define the rationals to be*

$$\mathbb{Q} = \mathbb{Z} \times \mathbb{Z}\backslash\{0\}/\sim$$

*For $[m/n], [a/b] \in \mathbb{Q}$, we define*

$$[m/n] + [a/b] = [(mb + an)/(nb)]$$

*and*

$$[m/n] \cdot [a/b] = [(ma)/(nb)]$$

It is a routine varification that these operations are well-defined and independent of the representative. If $x = [(a/b)] \in \mathbb{Q}$, and $x \neq 0$ so $a \neq 0$, then we can define

$$x^{-1} = \frac{1}{x} := [(b/a)] \in \mathbb{Q}$$

We also define $0 := [0/1], 1 := [1/1]$, and $-1 := [-1/1]$. So far we have the chain

$$\tilde{\mathbb{N}} \hookrightarrow \mathbb{Z} := \tilde{\mathbb{N}} \times \tilde{\mathbb{N}}/\sim \hookrightarrow \mathbb{Z} \times (\mathbb{Z}\backslash\{0\})/\sim$$

## 1.3.0 Divisibility

We now go over some fundamental theorems of number theory involving divisibility.

**Definition 1.3.1.** *We say $x \in \mathbb{N}$ is* <u>**composite**</u> *if $a, b \in \mathbb{N}$ such that $x = ab$ and $a, b \neq 1$. If $x$ is not composite, and $x > 1$, then $x$ is said to be* <u>**prime**</u>*. That is, $x$ is prime if and only if $x = ab$ implies $a = 1$ or $b = 1$.*

**Definition 1.3.2.** *If $x = ab$, $x, a, b \in \mathbb{N}$, then we say $a$ **divides** $x$, or $a$ is a **divisor** of $x$, and we write $a \mid x$.*

**Definition 1.3.3.** *Given $x \in \mathbb{N}$, define the collection of non-trivial divisors as*

$$D_x := \{a \in \mathbb{N}\backslash\{1\} : a \mid x\} \subseteq \mathbb{N}$$

Note that if $a \mid x$, then $a \leqslant x$. As $D_x \subseteq \mathbb{N}$, and $x \in D_x$ so it is non-empty, $D_x$ has a smallest element $p_1 \in D_x$. Then $x = p_1 x_1$ for some $x_1 \in \mathbb{N}$. Then $p_1$ is prime since if note it can be written as $p_1 = ab$ for $1 < a < p_1$, and then $a \mid x$ with $a < p_1$, contradicting its minimality. If $x_1 > 1$, then we can obtain $x_1 = p_2 x_2$, for $p_2$ prime and $p_2 \geqslant p_1$. Repeating in this fashion, since $x$ is finite there must exist $N \in \mathbb{N}$ such that $x_N = 1$ and $x_{N-1} = p_N \cdot 1$. Then $x = p_1 \cdot \ldots \cdot p_N$. This is the existence portion of the following result.

## Theorem 2 (Fundamental Theorem of Arithmetic).

*Every natural number $x > 1$ has a unique factorization, up to reordering, into a product of prime numbers.*

*Proof.* If $x = p_1 \ldots p_N = q_1 \ldots q_M$, then for each $1 \leqslant i \leqslant N$, $p_i \mid q_j$ for some $1 \leqslant j \leqslant M$. But, $p_i$ and $q_j$ are prime, so $p_i = q_j$, and after reordering $p_1 \ldots p_N = p_1 \ldots p_N q_{N=1} \ldots q_M$. By cancellation $q_{N+1} \ldots q_M = 1$. Thus, $N = M$ and the terms are equal up to reordering. ■

**Definition 1.3.4.** *We say $x, y \in \mathbb{N}$ are **coprime** if $x$ and $y$ have no common prime factors.*

**Proposition 1.3.1.** *If $x, y \in \mathbb{N}$ are coprime, then there exists $m, n \in \mathbb{Z}$ such that*

$$xm + ny = 1$$

# 1.4.0 Reals in terms of Cauchy Sequences

To construct the reals we use the standard notion of a completion of metric spaces using equivalence classes of Cauchy sequences. But first we must define what a sequence is, and what it means for one to be Cauchy.

**Definition 1.4.1.** *We define the **absolute value function** by*

$$|x| = \begin{cases} x & \text{if } x \geqslant 0 \\ -x & \text{if } x < 0 \end{cases}$$

It is a standard proof by cases, that the absolute value function defines a norm on $\mathbb{Q}$.

**Definition 1.4.2.** *A sequence is a function $a : \mathbb{N} \to \mathbb{Q}$, denoted $a(j) = a_j$ and $(a_j)_{j=1}^{\infty}$.*

**Definition 1.4.3.** *We say a sequence $(a_j)$ converges to a number $a \in \mathbb{Q}$, and write $a_j \to a$, if for every $n \in \mathbb{N}$, there exists an index $K(n) \in \mathbb{N}$ such that if $j \geqslant K(n)$, then $|a_j - a| < \frac{1}{n}$.*

**Definition 1.4.4.** *A sequence $(a_j)$ is said to be* **Cauchy** *if for all $n \in \mathbb{N}$, there exists $K(n) \in \mathbb{N}$ such that if $j, k \geqslant K(n)$, then*

$$|a_j - a_k| < \frac{1}{n}$$

**Proposition 1.4.1.** *If $a_j \to a$, then $(a_j)$ is Cauchy.*

*Proof.* Since $a_j \to a$, for $n \in \mathbb{N}$ there exists $K(2n) \in \mathbb{N}$ such that if $j \geqslant K(2n)$, $|a_j - a| < \frac{1}{2n}$. Thus, if $k, j \geqslant K(2n)$, then

$$|a_j - a_k| \leqslant |a_j - a| + |a - a_k| < \varepsilon$$

as desired. ■

**Proposition 1.4.2.** *If $(a_j)$ is Cauchy then $(a_j)$ is bounded.*

*Proof.* Suppose $(a_j)$ is Cauchy. Then there exists $K(1) \in \mathbb{N}$ such that for $k, j \geqslant K(1)$, $|a_k - a_j| < 1$. Then for all $j \geqslant K(1)$, $|a_j| < 1 + |a_{K(1)}|$. Letting $M = \max\{|a_1|, ..., |a_{K(1)-1}|, 1 + |a_{K(1)}|\}$, we have that $a_n \leqslant M$ or all $n \in \mathbb{N}$, so the sequence is bounded. ■

We now have some standard results about convergence of sequences:

**Proposition 1.4.3.** *If $a_j \to a$ and $b_j \to b$, then*

$$a_j + b_j \to a + b, \quad \text{and} \quad a_j b_j \to ab$$

*If $b \neq 0$, and $b_j \neq 0$ for all $j$, then*

$$a_j / b_j \to a/b$$

More generally we have

**Definition 1.4.5.** *If $(a_j), (b_j)$ are Cauchy sequences, then $(a_j + b_j)$ is Cauchy, $(a_j b_j)$ is cauchy, and if there exists $n \in \mathbb{N}$ such that $|b_j| > \frac{1}{n}$ for all $j$, then $(a_j/b_j)$ is Cauchy.*

Although for general metric spaces we have the inclusion

$$\left\{ \begin{matrix} Convergent \\ Sequences \end{matrix} \right\} \subseteq \left\{ \begin{matrix} Cauchy \\ Sequences \end{matrix} \right\}$$

the other inclusion is not in general true.

**Example 1.4.1.** Let $a_j = \sum_{l=0}^{j} \frac{1}{l!}$, in $(\mathbb{Q}, d)$, $d(x, y) = |x - y|$. $a_j$ is a Cauchy sequence, as $|a_j - a_k| = \left|\sum_{l=k+1}^{j} \frac{1}{l!}\right| = \sum_{l=k+1}^{j} \frac{1}{l!} \to 0$. For $l \geqslant 2$ we have $\frac{1/(l+1)!}{1/l!} = \frac{1}{l} \leqslant \frac{1}{2}$. Then $\frac{1}{(2+j)!} \leqslant \frac{1}{2^j}\frac{1}{2}$. Then for $j > k \geqslant 2$,

$$\sum_{l=k+1}^{j} \frac{1}{l!} = \sum_{l=k-2}^{j-2} \frac{1}{(l+2)!} \leqslant \sum_{l=k-2}^{l-2} \frac{1}{2^l}\frac{1}{2} < \frac{1}{2}\frac{1}{1 - 1/2} = 1$$

Thus, $a_j$ is a bounded increasing sequence and hence Cauchy. Now observe

$$\begin{aligned} a_{n+j} - a_n &= \frac{1}{(n+1)!} + \dots + \frac{1}{(n+j)!} \\ &\leqslant \frac{1}{n!}\left(\frac{1}{n+1} + \frac{1}{(n+1)^2} + \dots + \frac{1}{(n+1)^j}\right) \\ &< \frac{1}{n!}\sum_{k=0}^{j-1} \frac{1}{(n+1)^{k+1}} \\ &= \frac{1}{(n+1)!}\frac{1-\frac{1}{(n+1)^{j-1}}}{1-\frac{1}{n+1}} < \frac{1}{n!n} \end{aligned}$$

So if we fix $N \in \mathbb{N}$, $N + j > N + k \geqslant N$, then $a_{N+j} - a_{N+k} < \frac{1}{N!N} < \frac{1}{N}$. Hence $a_j$ is Cauchy. Since it is Cauchy, in the complete metric space $\mathbb{R}$ it is convergent, so let $a = \lim a_j$, so we observe $a = \sum_{l=0}^{\infty} \frac{1}{l!} = e$. But $e \notin \mathbb{Q}$, so this limit cannot be in the rationals and hence the rationals is not complete.

The following is a very important series known as the ***geometric series***:

**Proposition 1.4.4.** *If $a \in \mathbb{Q}$, with $|a| < 1$, then $\sum_{j=0}^{\infty} a^j = \frac{1}{1-a}$.*

**Proposition 1.4.5.** *If $|a| < 1$, then $|a|^n \to 0$.*

*Proof.* If $a = 0$, then $|a|^n = 0$ for all $n$, so the result holds. Hence, suppose $a \neq 0$. Then $|a|^{j+1} < |a|^j$, so $|a|^n$ is a bounded decreasing sequence, and hence converges in $\mathbb{R}$. Hence $\lim\limits_{n\to\infty} |a|^n = k$ for some $k \in \mathbb{R}$. Then $k = \lim\limits_{n\to\infty} |a|^n = \lim\limits_{n\to\infty} |a|^{n+1} = |a|k$. But $|a| \neq 1$, so $k = 0$. Thus, $|a|^n \to 0$, as desired. ■

**Proposition 1.4.6** (Bolzono-Weierstrass (Cauchy))**.** *If $(a_j)$ is a bounded sequence, then there exists a Cauchy subsequence.*

*Proof.* Since $(a_j)$ is bounded, there exists $M > 0$ such that $|a_j| \leqslant M$ for all $j$. In particular, $a_j \in I_0 = [-M, M]$ for all $j$. Then either $[-M, 0]$ or $[0, M]$ contains an infinite number of $a_j$. Let $I_1$ be the one with such. Inductively, suppose there exists $k \in \tilde{\mathbb{N}}$ such that an infinite number of $a_j$ are in $I_k$, for all $0 \leqslant l \leqslant k-1$ $I_{l+1} \subseteq I_l$, and $\ell(I_l) = \frac{2M}{2^l}$. Then, we have a sequence $I_j$ of closed intervals containing infinitely many terms of $a_j$. Let $b_1 = a_1$, and let $b_k = a_{j(k)}$, where $j(k) = \min\{m \in \mathbb{N} : a_m \in I_k, m > j(k-1)\}$, which exists and is well defined by the construction of $I_k$ and the well-ordering of $\mathbb{N}$. Then $b_k = a_j(k)$ is a subsequence of $j$, as $j(k) < j(k+1)$ for all $k$. Now, fix $n \in \mathbb{N}$. As $2^{-j} \to 0$, there exists $K(2Mn) \in \mathbb{N}$ such that for $j \geqslant K(2Mn)$, $\frac{1}{2^j} < \frac{1}{2Mn}$. Then, for $k, l \geqslant K(2Mn)$, $b_k, b_l \in I_{K(2Mn)}$, so

$$|b_k - b_l| \leqslant \ell(I_{K(2Mn)}) = \frac{2M}{2^{K(2Mn)}} < \frac{2M}{2Mn} = \frac{1}{n}$$

Thus $b_j$ is Cauchy. ■

**Corollary 1.4.7.** *Each bounded Monotone sequence is Cauchy.*

*Proof.* Let $(a_j)$ be a bounded monotone sequence. Then we have a Cauchy subsequence $(a_{j_n})$. Fix $n \in \mathbb{N}$. Then there exists $K(n) \in \mathbb{N}$ such that if $k, l \geqslant K(n)$, $|a_{j_k} - a_{j_l}| < \frac{1}{n}$. Let $K'(n) = j_{K(n)}$. Then for $k, l \geqslant K'(n)$, let $m \in \mathbb{N}$ such that $j_m \geqslant k, l$ and $m \geqslant K(n)$. Then as $a_j$ is an increasing (decreasing) sequence, so

$$a_{j_{K(n)}} \leqslant a_k \leqslant a_l \leqslant a_{j_m} \text{ (respectively } a_{j_{K(n)}} \geqslant a_k \geqslant a_l \geqslant a_{j_m})$$

Then $0 \leqslant a_l - a_k \leqslant a_{j_m} - a_{j_{K(n)}}$, so $|a_l - a_k| < \frac{1}{n}$, and similarly for a decreasing sequence. Hence $a_j$ is Cauchy. ∎

We now begine defining the reals using equivalence relations on our Cauchy sequences:

**Definition 1.4.6.** *Let $\mathcal{S} = \{(a_j) \subseteq \mathbb{Q} : (a_j) \text{ is Cauchy}\}$. Define an equivalence relation $\sim$ on $\mathcal{S}$ by*

$$(a_j) \sim (b_j) \iff a_j - b_j \to 0$$

It is a routine check that $\sim$ is an equivalence relation on $\mathcal{S}$.

**Definition 1.4.7.** *We define $\mathbb{R} := \mathcal{S}/\sim$, so $x \in \mathbb{R}$ if and only if $x = [(a_j)]$ for some Cauchy sequence $(a_j)$ in $\mathbb{Q}$.*

**Definition 1.4.8.** *If $x = [(a_j)], y = [(b_j)] \in \mathbb{R}$, we define*

$$x + y := [(a_j + b_j)] \quad xy := [(a_j b_j)]$$

*and $-x := [(-a_j)]$.*

As with the notion of an equivalence relation, it is a routine check using the boundedness of Cauchy sequences to prove that these operations are well defined. We can then define a natural injection $\mathbb{Q} \hookrightarrow \mathbb{R}$ by $a \mapsto [(a, a, a, ...)]$. In particular, $0 := [(0, 0, 0, ...)]$ in $\mathbb{R}$.

Now, note that if $x = [(a_j)], y = [(b_j)] \in \mathbb{R}$, then $x \neq y$ if and only if $a_j - b_j$ does not converge to 0, so there exists $n \in \mathbb{N}$ such that for all $j \in \mathbb{N}$, there exists $k \geqslant j$ such that

$$|a_k - b_k| \geqslant \frac{1}{n}$$

Specializing to the case of $y = 0 = [(0, 0, 0, ...)]$, as $(a_j)$ is Cauchy, there exists $K(2n) \in \mathbb{N}$ such that $k, l \geqslant K(2n)$, $|a_k - a_l| < \frac{1}{2n}$, so in particular $|a_j| \geqslant |a_k| - |a_j - a_k| > \frac{1}{2n}$ for all $j \geqslant K(2n)$. It follows that either $a_j > \frac{1}{2n} > 0$ for all $j \geqslant K(2n)$, or $a_j < \frac{-1}{2n} < 0$ for all $j \geqslant K(2n)$.

Thus, if $x \neq 0$, then $x = [(a_j)] = [(\alpha_j)]$ such that there exists $n \in \mathbb{N}$ such that either $\alpha_j \geqslant \frac{1}{2n}$ for all $j$, or $\alpha_j \leqslant \frac{-1}{2n}$ for all $j$. Then we can define $x^{-1} = \frac{1}{x} := [(\alpha_j^{-1})]$.

**Definition 1.4.9.** *We define the following subsets of $\mathbb{R}$:*

$$\mathbb{R}^+ = \{x = [(a_j)] : \exists n, K \in \mathbb{N}; a_j \geqslant \frac{1}{2n}, \forall j \geqslant K\}$$

$$\mathbb{R}^- = \{x = [(a_j)] : \exists n, K \in \mathbb{N}; a_j \leqslant \frac{-1}{2n}, \forall j \geqslant K\}$$

We have shown that if $x \neq 0$ then either $x \in \mathbb{R}^+$ or $x \in \mathbb{R}^-$. Thus

$$\mathbb{R} = \mathbb{R}^+ \sqcup \{0\} \sqcup \mathbb{R}^-$$

**Proposition 1.4.8.** *For $x \in \mathbb{R}$, $x \in \mathbb{R}^+$ if and only if $-x \in \mathbb{R}^-$, and $x \in \mathbb{R}^-$ if and only if $-x \in \mathbb{R}^+$.*

**Definition 1.4.10.** *We define a total order $<$ on $\mathbb{R}$ by*

$$x < y \iff y - x \in \mathbb{R}^+ \iff x = [(a_j)], y = [b_j)], \exists n, K \in \mathbb{N}; b_j - a_j \geqslant \frac{1}{2n} \forall j \geqslant K$$

We have a few standard results about the order relation on $\mathbb{R}$:

**Proposition 1.4.9.** *Let $x_1, x_2, y_1, y_2 \in \mathbb{R}$. Then*

- $x_1 < y_1, x_2 < y_2 \implies x_1 + x_2 < y_1 + y_2$
- $x_1 < y_1 \implies -y_1 < -x_1$
- $0 < x_1 < y_1, c > 0$, *then* $0 < cx < cy$
- $0 < x < y \iff 0 < \frac{1}{y} < \frac{1}{x}$

Note that in $\mathbb{N}$ we have well-ordering, but under the standard orders on $\mathbb{Q}$ and $\mathbb{R}$ this property does not hold.

**Definition 1.4.11.** *For $S \subseteq \mathbb{R}$, we say $x$ is an* **upper bound** *of $S$ if $s \in S$ implies $s \leqslant x$. Dually, we say $y$ is a* **lower bound** *for $S$ is $s \in S$ implies $s \geqslant y$.*

**Definition 1.4.12.** *For $S \subseteq \mathbb{R}$, the* **least upper bound***, denoted $\sup S$, is an upper bound for $S$ such that if $y$ is any other upper bound for $S$ then $\sup S \leqslant y$. Dually, the* **greatest lower bound***, denoted $\inf S$, is a lower bound for $S$ such that if $y$ is any other lower bound for $S$, then $y \leqslant \inf S$.*

**Theorem 1.4.10** (Completeness of $\mathbb{R}$)**.** *If $(x_j)$ is a Cauchy sequence of real numbers, then there exists $x = [(a_j)] \in \mathbb{R}$ such that $x_j \to x$, of $x_j \sim a_j$, extending the equivalence relation to $\mathbb{R}$.*

**Proposition 1.4.11.** *If $S$ is a non-empty subset of $\mathbb{R}$ that has an upper bound, then there exists $x \in \mathbb{R}$ such that $x = \sup S$.*

*Proof.* By hypothesis, there exists $x_0 \in \mathbb{R}$ such that for all $s \in S$, $s \leqslant x_0$. As $S$ is non-empty, there exists $s_0 \in S$. Define an interval $I_0 = [s_0, x_0]$, and divide it into 2 subintervals, $I_0^l, I_0^r$. If $I_0^r \cap S \neq \emptyset$ let $I_1^* = I_0^r$, and otherwise let $I_1^* = I_0^l$. In either case $I_1 = [s_1, x_1]$ is such that $x_1$ is an upper bound of $S$, and where we choose $s_1 \in I_1^*$ such that $s_1 \in S$. Further, $s_0 \leqslant s_1 \leqslant x_1 \leqslant x_0$, and letting $x_0 - s_0 = L$, $\ell(I_1) \leqslant \frac{L}{2}$. Proceeding inductively we find sequences $s_0 \leqslant s_1 \leqslant s_2 \leqslant ...$ in $S$ and $x_0 \geqslant x_1 \geqslant x_2 \geqslant ...$, with $I_j = [s_j, x_j]$, $\ell(I_j) \leqslant \frac{L}{2^j}$, with $x_j \geqslant s$ for all $s \in S$, and all

$j \in \tilde{\mathbb{N}}$. Note $x_j$ is a decreasing bounded sequence, so it converges to some $x \in \mathbb{R}$, as $\mathbb{R}$ is complete. As $x_j \geqslant s$ for all $s \in S$, $x \geqslant s$ so $x$ is an upper bound. Further, if $\varepsilon > 0$, there exists $K \in \mathbb{N}$ such that for all $j \geqslant K$, $\frac{1}{2^j} < \frac{\varepsilon}{L}$, so $0 \leqslant x_j - s_j < \frac{L}{2^j} < \varepsilon$ for all $j \geqslant K$. Then $x - \varepsilon < s_j$ for all $j \geqslant K$. Thus, $x - \varepsilon$ is not an upper bound of $S$, and hence $x$ must be the least upper bound. ■

## 1.5.0 Axiomatized Reals

**Definition 1.5.1 (The Reals).** *The real number system $\mathbb{R}$ is an* **ordered field** *which contains $\mathbb{Q}$ as a subfield, which satisfies the* **axiom of choice**. *In particular, the real numbers is a set $\mathbb{R}$ with two binary operations $+$ and $\cdot$, two distinct elements $0$ and $1$, and a subset $\mathbb{P}$ of positive numbers satisfying the following $13$ postulates:*

1. *Addition is associative:* $\forall a, b, c \in \mathbb{R}, a + (b + c) = (a + b) + c$
2. *The number $0$ is an additive identity:* $\forall a \in \mathbb{R}, a + 0 = 0 + a = a$
3. *Additive inverses exist:* $\forall a \in \mathbb{R}; \exists (-a) \in \mathbb{R}$ *s.t.* $a + (-a) = (-a) + a = 0$
4. *Addition is commutative:* $\forall a, b \in \mathbb{R}, a + b = b + a$
5. *Multiplication is associative:* $\forall a, b, c \in \mathbb{R}, a \cdot (b \cdot c) = (a \cdot b) \cdot c$
6. *The number $1$ is a multiplicative identity:* $\forall a \in \mathbb{R} a \cdot 1 = 1 \cdot a = a$
7. *Multiplicative inverses exist:* $\forall a \neq 0; \exists a^{-1} \in \mathbb{R}$ *s.t.* $a \cdot a^{-1} = a^{-1} \cdot a = 1$
8. *Multiplication is commutative:* $\forall a, b \in \mathbb{R}, a \cdot b = b \cdot a$
9. *The distributive law:* $\forall a, b, c \in \mathbb{R}, a \cdot (b + c) = a \cdot b + a \cdot c$
10. *The trichotomy of $\mathbb{P}$: for every $a \in \mathbb{R}$, exactly one of the following holds:* $a = 0, a \in \mathbb{P}, (-a) \in \mathbb{P}$
11. *Closure under addition: if $a \in \mathbb{P}$ and $b \in \mathbb{P}$, then $a + b \in \mathbb{P}$*
12. *Closure under multiplication: if $a \in \mathbb{P}$ and $b \in \mathbb{P}$, then $a \cdot b \in \mathbb{P}$*
13. *(to be added)*

*From positive postulates we can define the order relations $>, <, \geqslant, \leqslant$ on $\mathbb{R}$ for $a, b \in \mathbb{R}$ by*

1. $a > b$ *if* $a - b \in \mathbb{P}$
2. $a < b$ *if* $b > a$
3. $a \geqslant b$ *if* $a > b$ *or* $a = b$
4. $a \leqslant b$ *if* $a < b$ *or* $a = b$

*Note in particular* $a > 0$ *if and only if* $a \in \mathbb{P}$.

**Remark 1.5.1.** A few points which follow from the postulates are:

1. Finite sums such as $a_1 + a_2 + ... + a_n$ are well defined
2. The additive identity is unique (also multiplicative)
3. Additive inverses are unique (also multiplicative)
4. Subtraction can be defined
5. $a \cdot b = b \cdot c \iff a = 0 \vee b = c$
6. $a \cdot b = \iff a = 0 \vee b = 0$
7. $a - b = b - a \iff a = b$
8. A "well behaved" order relation can be defined.
9. The "absolute value" function $a \mapsto |a|$ can be defined by

   $$|a| = \begin{cases} a, & a \geqslant 0 \\ -a, & a \leqslant 0 \end{cases}$$

   and for all $a, b \in \mathbb{R}$, the triangle inequality

   $$|a + b| \leqslant |a| + |b|$$

   holds

**Axiom 1.5.2 (Axiom of Completeness).** *Every non-empty subset of the real numbers that is bounded above has a least upper bound.*

## Upper and Lower Bounds

**Definition 1.5.3 (Bounds).** *A set* $A \subseteq \mathbb{R}$ *is* **bounded above** *if there exists a number* $b \in \mathbb{R}$ *such that* $a \leqslant b$ *for all* $a \in A$. *The number* $b$ *is called an* **upper bound** *for* $A$.

*Similarly, the set* $A$ *is* **bounded below** *if there exists a* **lower bound** $l \in \mathbb{R}$ *satisfying* $l \leqslant a$ *for every* $a \in A$.

**Definition 1.5.4 (Least Upper Bound).** *A real number* $s$ *is the* **least upper bound** *for a set* $A \subseteq \mathbb{R}$ *if it meets the following two criteria:*

1. $s$ *is an upper bound for* $A$;
2. *if* $b$ *is any upper bound for* $A$, *then* $s \leqslant b$.

*The least upper bound of a set* $A$ *is also called the* **supremum** *of* $A$, *and denoted by* $\sup A$.

**Definition 1.5.5 (Greatest Lower Bound).** *A real number $i$ is the* **Greatest Lower bound** *for a set $A \subseteq \mathbb{R}$ if it meets the following two criteria:*

1. *$i$ is a lower bound for $A$;*
2. *if $b$ is any lower bound for $A$, then $b \leqslant i$.*

*The greatest lower bound of a set $A$ is also called the* **infemum** *of $A$, and denoted by* $\inf A$.

**Remark 1.5.2.** From the definitions we assert that the least upper bound and greatest lower bound of a set, if they exist, are unique.

**Example 1.5.1.** Consider the set

$$A = \left\{ \frac{1}{n} : n \in \mathbb{N} \right\} = \left\{ 1, \frac{1}{2}, \frac{1}{3}, \dots \right\}$$

The set $A$ is bounded above and below. Moreover, the least upper bound of $A$ is $\sup A = 1$, which is in $A$, while $\inf A = 0$, which is not in $A$.

**Definition 1.5.6 (Max and Min).** *A real number $a_0$ is a* **maximum** *of a set $A$ if $a_0$ is an element of $A$ and $a_0 \geqslant a$ for all $a \in A$. Similarly, a number $a_1$ is a* **minimum** *of $A$ if $a_1 \in A$ and $a_1 \leqslant a$ for all $a \in A$.*

**Example 1.5.2.** Consider the open interval $(0, 2)$, and the closed interval $[0, 2]$. Note both sets are bounded above and below, and both have the same infimum and supremum, namely $\inf = 0$ and $\sup = 2$. However, $[0, 2]$ has both a maximum and a minimum, namely its infimum and supremum, while $(0, 2)$ has neither.

**Example 1.5.3.** Let $A \subseteq \mathbb{R}$ be a non-empty and bounded above set, and let $c \in \mathbb{R}$. Define the set $c + A$ by

$$c + A := \{c + a : a \in A\}$$

Then I claim that $\sup(c + A) = c + \sup A$.

*Proof.* Let $\alpha = c + \sup A$. First let us show that $\alpha$ is an upper bound of $c + A$. Indeed, for $x \in c + A$ we can write $x = c + a$ for some $a \in A$ by definition. Then, by definition we have that $a \leqslant \sup A$. Thus, adding $c$ to both sides we obtain

$$x = c + a \leqslant c + \sup A = \alpha$$

Therefore, as $x$ was arbitrary $\alpha$ is indeed an upper bound of $c + A$. Now, suppose $b$ is an upper bound of $c + A$. Then, $c + a \leqslant b$ for all $c + a \in c + A$, so in particular $a \leqslant b - c$ for all $a \in A$. Then, $b - c$ is an upper bound for $A$, so as $\sup A$ is the least upper bound of $A$ we have that $\sup A \leqslant b - c$. Hence, we conclude that $\alpha = c + \sup A \leqslant b$. Thus $\alpha$ satisfies the axioms of a least upper bound for $c + A$, and we conclude that $\sup(c + A) = c + \sup A$. ∎

**Lemma 1.5.1.** *Assume $s \in \mathbb{R}$ is an upper bound for a set $A \subseteq \mathbb{R}$. Then $s = \sup A$ if and only if, for every choice $\epsilon > 0$, there exists an element $a \in A$ satisfying $s - \epsilon < a$.*

*Proof.* Let $s \in \mathbb{R}$ be an upper bound for a set $A \subseteq \mathbb{R}$.

( $\Longrightarrow$ ) First, suppose that $s = \sup A$, and choose $\epsilon \in \mathbb{R}$ with $\epsilon > 0$. Then $s - \epsilon$ is not an upper bound of $A$. Indeed, if $s - \epsilon$ was an upper bound then $s \leqslant s - \epsilon$ which implies that $\epsilon \leqslant 0$, but by assumption $\epsilon > 0$. Thus, there must exist $a \in A$ such that $s - \epsilon < a$, satisfying the implication.

( $\Longleftarrow$ ) Conversely, suppose that for all $\epsilon > 0$ there exists $a \in A$ such that $s - \epsilon < a$. Now, suppose that $b$ is an upper bound of $A$, and towards a contradiction suppose $s > b$. Then $s - b > 0$, so there exists $a \in A$ such that $s - (s - b) < a$. In particular, $b < a$. However, $a \in A$ and $b$ is an upper bound of $A$ by assumption, so $b < a$ is a contradiction. Therefore we conclude that $s \leqslant b$, so $s$ is the supremum of $A$ by definition. ■

**Theorem 1.5.2** (Archimedean Property for the Reals)**.** *For all $x, y > 0$ in $\mathbb{R}$, there exists $n \in \mathbb{N}$ such that $nx > y$.*

*Proof.* Towards a contradiction suppose $nx \leqslant y$ for all $n \in \mathbb{N}$. Then the set $\{nx : n \in \mathbb{N}\}$ is bounded above by $y$. Thus, by the least upper bound property of $\mathbb{R}$ we have a supremum $\alpha \in \mathbb{R}$. Then for all $n \in \mathbb{N}$ $\alpha \geqslant nx$. In particular, $\alpha \geqslant (n+1)x$ for all $n \in \mathbb{N}$, so $\alpha - x \geqslant nx$ for all $n \in \mathbb{N}$. But this implies that $\alpha - x$ is also an upper bound of the set, which contradicts the fact that $\alpha$ is the least upper bound. Thus, we must have that $nx > y$ for some $n \in \mathbb{N}$, as claimed. ■

**Corollary 1.5.3.** *$\mathbb{N}$ is not bounded above.*

**Corollary 1.5.4.** *For any $\epsilon > 0$ there is a natural number $n$ with $1/n < \epsilon$.*

*Proof.* Consider $0 < 1/\epsilon \in \mathbb{R}$ and $1 \in \mathbb{R}$. Then by the Archimedean Property of $\mathbb{R}$ there exists $n \in \mathbb{N}$ such that $1 \cdot n > 1/\epsilon$. In particular, we have that $n > 0$, so $\epsilon > 1/n$, completing the proof. ■

## 1.6.0 Metric Properties of the Reals

First we extend our definition of sequences to the reals:

**Definition 1.6.1.** *A sequence $(p_j)$ in $\mathbb{R}$ converges to a point $p \in \mathbb{R}$ if and only if for every $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that if $j \geqslant N$, then $|p_j - p| < \varepsilon$.*

Using sequences we can define notions of our topology, such as closed and open sets, and limit points:

**Definition 1.6.2.** *$S \subseteq \mathbb{R}$ is said to be* **closed** *if and only if whenever $(p_j) \subseteq S$ is a sequence in $S$ which converges to a point $p \in \mathbb{R}$, then $p \in S$.*

**Definition 1.6.3.** *A point $p \in \mathbb{R}$ is said to be a* **limit point** *of $S$ if there exists $(p_j) \subseteq S$ such that $p_j$ converges to $p$, and $p_j \neq p$ for all $j \in \mathbb{N}$.*

Note that $S$ is closed if and only if $S$ contains all of its limit points.

**Definition 1.6.4.** *$U \subseteq \mathbb{R}$ is said to be **open** if and only if $U^c = \mathbb{R} \backslash U$ is closed.*

**Definition 1.6.5.** *For $S \subseteq \mathbb{R}$, the **closure** of $S$, $\overline{S}$, is defined as*

$$\overline{S} := S \cup S'$$

*where*

$$S' = \{p \in \mathbb{R} : p \text{ is a limit point of } S\}$$

**Proposition 1.6.1.** *For all $S \subseteq \mathbb{R}$, $\overline{\overline{S}} = \overline{S}$*

*Proof.* Let $(p_j) \subseteq \overline{S}$ which converges to some point $p \in \mathbb{R}$. Then for each $j$ we have $(b_{jk})$ in $S$ which converges to $p_j$. For each $j \in \mathbb{N}$, there exists $K(j) \in \mathbb{N}$ such that for $k \geqslant K(j)$, $|b_{jk} - p| < \frac{1}{j}$. Define $(c_j)$ by $c_j = b_{jK(j)}$. Fix $\varepsilon > 0$. Then there exists $N \in \mathbb{N}$ such that $j \geqslant N$ implies $|p_j - p| < \frac{\varepsilon}{2}$. By the Archimedian property there exists $n \in \mathbb{N}$ such that $\frac{1}{n} < \frac{\varepsilon}{2}$. Then for a $j \geqslant \max\{n, N\}$,

$$|c_j - p| \leqslant |c_j - p_j| + |p_j - p| < \frac{1}{n} + \frac{\varepsilon}{2} < \varepsilon$$

Thus $(c_j) \subseteq S$ and $c_j \to p$, so $p \in \overline{S}$. Hence, $\overline{S} \supseteq \overline{\overline{S}}$, and by definition $\overline{S} \subseteq \overline{\overline{S}}$, so $\overline{S} = \overline{\overline{S}}$. ■

**Theorem 1.6.2.** *Every Cauchy sequence in $\mathbb{R}$ has a limit point in $\mathbb{R}$.*

Note if $(x_j) = ([(a_{jk})])$ is Cauchy in $\mathbb{R}$, then $(a_{jj})$ is Cauchy in $\mathbb{Q}$ with $a_{jj} - x_j \to 0$, so then $x_j$ converges to $[(a_{jj})]$.

**Proposition 1.6.3** (Density of the Rationals)**.** *For all $x \in \mathbb{R}$ and $\varepsilon > 0$, there exists $y \in \mathbb{Q}$ such that $|y - x| < \varepsilon$.*

In particular, for all $a < b$ in $\mathbb{R}$, there exists $c \in \mathbb{Q}$ such that $a < c < b$. Indeed, for $x = [(a_j)]$, $a_j \to x$ so there exists $N \in \mathbb{N}$ such that $|a_N - x| < \varepsilon$ for any $\varepsilon > 0$.

**Definition 1.6.6.** *A subset $K \subseteq \mathbb{R}$ is **sequentially compact** if and only if for every **infinite sequence** $(p_j) \subseteq K$, there exists a subsequence which converges to a point in $K$.*

## Theorem 3 (Bolzano-Weierstrass Property).

*Every bounded sequence of real numbers has a convergent subsequence.*

**Theorem 1.6.4.** *If $K \neq \varnothing$, $K \subseteq \mathbb{R}$, and $K$ is closed and bounded, then $K$ is sequentially compact.*

*Proof.* If $K \subseteq \mathbb{R}$, $K \neq \varnothing$, is bounded and $(p_j) \subseteq K$, $(p_j)$ has a convergent subsequence $(p_{j_k})$ by Bolzano-Weierstrass, so $p_{j_k} \to p$ for some $p \in \mathbb{R}$. But $K$ is closed so $(p_{j_k}) \subseteq K$, so $p \in K$. ■

Note that if $K \subseteq \mathbb{R}$ is compact, then $K$ is closed since all subsequences of a convergent sequence converge to the same point. Additionally, $K$ is bounded as otherwise we can construct $p_1 \in K$, $p_2 \in K$ such that $|p_2| > |p_1| + 1$, and for $p_k \in K$, choose $p_{k+1} \in K$ such that $|p_{k+1}| > |p_k| + 1$. Thus, for all $j, k \in \mathbb{N}$, $|p_j - p_k| > 1$, so $(p_j)$ has no convergent subsequence.

## Theorem 4 (Heine-Borel).

*If $K \neq \emptyset$, $K \subseteq \mathbb{R}$, the following are equivalent:*

- *$K$ is sequentially compact*
- *$K$ is closed and bounded*

If $K$ is compact, $K \neq \emptyset$, in $\mathbb{R}$, then there exists $a, b \in K$ such that

$$a = \min K := \inf K \text{ and } b = \max K := \sup K$$

which is to say $K$ contains its infimum and supremum.

**Definition 1.6.7.** *A function $f : S \to \mathbb{R}$, $\emptyset \neq S \subseteq \mathbb{R}$, is said to be* **continuous** *at a point $p \in S$ if whenever $(p_j) \subseteq S$ such that $p_j \to p$, then $f(p_j) \to f(p)$*

**Definition 1.6.8.** *A point $p \in S$ is said to be an* **isolated point** *of $S$ if there exists some $\varepsilon > 0$ such that $(p - \varepsilon, p + \varepsilon) \cap S = \{p\}$*

Every function is continuous at isolated points of its domain.

**Definition 1.6.9.** *If $f : S \to \mathbb{R}$ is continuous at every point $p \in S$, we say $f$ is* **continuous** *on $S$.*

**Proposition 1.6.5.** *If $K \subseteq \mathbb{R}$, $K \neq \emptyset$, is a compact subset of $\mathbb{R}$, and $f : K \to \mathbb{R}$ is continuous, then $f(K)$ is compact.*

*Proof.* Let $(q_k) \subseteq f(K)$. Then we have $(p_k) \subseteq K$ such that $f(p_k) = q_k$ for all $k$. Then as $K$ is sequentially compact there exists $p \in K$ and a subsequence $(p_{k_j}) \subseteq K$ such that $p_{k_j} \to p$. As $f$ is continuous we have $q_{k_j} = f(p_{k_j}) \to f(p)$, where $f(p) \in f(K)$. Thus $f(K)$ is sequentially compact as claimed. ■

**Proposition 1.6.6.** *If $\emptyset \neq K \subseteq \mathbb{R}$ is sequentially compact and $f : K \to \mathbb{R}$ is continuous on $K$, then there exist $q, p \in K$ such that*

$$f(p) = \max_K f, \quad f(q) = \min_K f$$

## Theorem 5 (Intermediate Value Theorem).

*If $f : [a, b] \to \mathbb{R}$ is continuous on $[a, b]$ and $f(a) < c < f(b)$ (or $f(a) > c > f(b)$), then there exists $x \in (a, b)$ such that $f(x) = c$.*

*Proof.* Let $S = \{y \in [a,b] : f(y) \leqslant c\}$. Without loss of generality suppose $f(a) < c < f(b)$ (if the other inequality holds, replace $f$ with $-f$). THen $a \in S$ and $b \notin S$. Further, if $(y_j) \in S$, $y_j \to y$, then by continuity $f(y_j) \to f(y)$ and as $f(y_j) \leqslant c$ for all $j$, $f(y) \leqslant c$. Thus $y \in S$, so $S$ is closed. As $S \subseteq [a,b]$, $S$ is bounded, so by Heine-Borel $S$ is compact. Let $x = \max S \in S$. Then $f(x) \leqslant c$. If $f(x) < c$, then for some $\varepsilon > 0$, $f(y) < c$ for all $y \in (x-\varepsilon, x+\varepsilon)$. Let $\varepsilon = c - f(x) > 0$. As $f$ is continuous, there exists $\delta > 0$ such that if $|x-y| < \delta$, $|f(x) - f(y)| < \frac{\varepsilon}{2}$. Then $c - f(y) = c - f(x) - (f(y) - f(x)) \geqslant |c - f(x)| - |f(y) - f(x)|$, which is greater than or equal to $\varepsilon/2$, so $f(y) < c$. Taking $y = x + \frac{\delta}{2}$, $f(y) \leqslant c$, so $y \in S$, but $y > x = \max S$ which is a contradiction. Thus, $f(x) \not< c$, so $f(x) = c$. ∎

**Proposition 1.6.7.** *Suppose $K$ is sequentially compact and $X_1 \supseteq X_2 \supseteq X_3 \supseteq ...$ is a sequence of closed subsets of $K$ so all $X_i$ are compact by Heine-Borel. If $X_m \neq \varnothing$, for all $m$, then $\bigcap_{m \geqslant 1} X_m \neq \varnothing$.*

*Proof.* Let $x_m \in X_m$. As $K$ is compact and $x_m \in K$, there exists a convergent subsequence $x_{m_j} \to x \in K$. Since $X_1 \supseteq X_2 \supseteq ...$,

$$\{x_{m_l} : l \geqslant j\} \subseteq X_{m_j}$$

But $X_{m_j}$ is closed, so $x \in X_{m_j}$, for all $j$. Then for all $m \in \mathbb{N}$, $x \in X_m$ as for all $n \in \mathbb{N}$ there exists $m_j \geqslant n$ so $x \in X_{m_j} \subseteq X_n$. Thus $x \in \bigcap_{m \geqslant 1} X_m$,a s claimed. ∎

**Corollary 1.6.8.** *If $K$ is sequentially compact and $U_1 \subseteq U_2 \subseteq ...$ is a sequence of open sets such that $K \subseteq \bigcup_{j \geqslant 1} U_j$, then there exists $M \in \mathbb{N}$ such that $K \subseteq U_M$.*

Use Proposition 1.6.7 with $X_j = K \backslash U_j$.

We now discuss some general results on open and closed sets:

**Proposition 1.6.9.** *If $\{A_\alpha\}_{\alpha \in J}$ is a family of closed sets in $\mathbb{R}$, then $\bigcap_{\alpha \in J} A_\alpha$ is closed. If $A$ and $B$ are closed, then $A \cup B$ is closed.*

*Proof.* Suppose $A_\alpha, \alpha \in J$ are as in the hypothesis. If $\varnothing = \bigcap_{\alpha \in J} A_\alpha$ then the claim vauously holds. Otherwise, let $(p_j) \subseteq \bigcap_{\alpha \in J} A_\alpha$ such that $p_j \to p \in \mathbb{R}$. As $A_\alpha$ is closed for all $\alpha \in J$, it follows that $p \in A_\alpha$, so $p \in \bigcap_{\alpha \in J} A_\alpha$.

Next, let $A$ and $B$ be closed, and take $(p_j) \subseteq A \cup B$ such that $p_j \to p \in \mathbb{R}$. Either infinitely many points of the sequence are in $A$ or infinitely many are in $B$. Without loss of generality suppose infinitely many are in $A$. Then there exists a subsequence $(p_{j_k}) \subseteq (p_j)$ contained in $A$, which will converge to $p$ so as $A$ is closed $p \in A$. Thus $p \in A \cup B$, so the union is closed. ∎

**Corollary 1.6.10.** *If $\{U_\alpha\}_{\alpha \in J}$ is a family of open sets in $\mathbb{R}$, then $\bigcup_{\alpha \in J} U_\alpha$ is open. If $A$ and $B$ are open, then $A \cap B$ is open.*

Conversely, we have $I_n = [-1 + \frac{1}{n}, 1 - \frac{1}{n}]$, a collection of closed intervals, who's union is $\bigcup_{n \geqslant 1} I_n = (-1, 1)$ is not closed. Further, if $U_n = (-\frac{1}{n}, 1 + \frac{1}{n})$, then $\bigcap_{n \geqslant 1} U_n = [0, 1]$ is not open.

**Definition 1.6.10.** *The ball of radius $r > 0$ centered at $x \in \mathbb{R}$ is defined by*

$$B_r(x) := \{y \in \mathbb{R} : |x - y| < r\}$$

**Proposition 1.6.11.** *$U \subseteq \mathbb{R}$ is open if and only if for all $x \in U$, there exists $r_x > 0$ such that $B_{r_x}(x) \subseteq U$.*

*Proof.* First we prove the reverse implication:

$\Longleftarrow$ If for all $x \in U$ there exists $r_x > 0$ such that $B_{r_x}(x) \subseteq U$, then $U = \bigcup_{x \in U} B_{r_x}(x)$. Each $B_{r_x}(x) = (x - r_x, x + r_x)$ is open as $(-\infty, x - r_x] \cup [x + r_x, \infty)$ is a finite union of closed sets. Thus $U$ is open, being the union of open sets.

$\Longrightarrow$ Towards a contradiction there exists $x \in U$ such that for all $r > 0$, $B_r(x) \cap U^c \neq \emptyset$. Then by the axiom of choice there exists $f : X \to \bigcup X$ with $X = \{B_{1/n}(x) \cap U^c : n \in \mathbb{N}\}$, and we can define a sequence $a(n) = f(B_{1/n}(x) \cap U^c)$ which converges to $x$, and $(a_n) \subseteq U^c$ which is closed. But $x \in U$ implies $x \notin U^c$ contradicting the closedness of $U^c$, so $U$ must satisfy the hypothesis.

■

**Definition 1.6.11.** *A topological space $(X, \tau)$ is called* **separable** *if it has a* **countable dense subset***.*

$\mathbb{R}$ is an example of a separable space, with countable dense subset $\mathbb{Q}$.

**Definition 1.6.12.** *A topological space $(X, \tau)$ is called* **Lindelöf** *if and only if every open cover of $X$ has a countable subcover.*

**Definition 1.6.13.** *A topological space $(X, \tau)$ is called* **second countable** *if $\tau$ has a countable base.*

All metrizable spaces are second countable if and only if they are separable, as we can take the rational balls around points of the countably dense subset. THus, $\tau_{\mathbb{R}} = \bigcup \mathcal{B}$, where $\mathcal{B} = \{B_p(q) : p, q \in \mathbb{Q}, p > 0\}$ is a countable base.

**Definition 1.6.14.** *A* **covering** *of a set $F$ is any family of sets $\{X_\alpha\}_{\alpha \in \mathcal{A}}$, such that $F \subseteq \bigcup_{\alpha \in \mathcal{A}} X_\alpha$.*

**Definition 1.6.15.** *An* **open covering** *of a topological space is a covering by open sets.*

**Proposition 1.6.12.** *If $K \subseteq \mathbb{R}$ is sequentially compact, then every open covering of $K$ has a finite subcovering.*

*Proof.* Suppose $\emptyset \neq K \subseteq \mathbb{R}$ is sequentially compact. Let $K \subseteq \bigcup_{\alpha \in J} U_\alpha$ be an open covering.

Note $U_\alpha = \bigcup_{x \in U_\alpha} B_{p_{\alpha,x}}(q_{\alpha,x})$ such that $x \in B_{p_{\alpha,x}}(q_{\alpha,x})$, as $U_\alpha$ is open. Hence

$$K \subseteq \bigcup_{\alpha \in J} U_\alpha = \bigcup_{\alpha \in J} \bigcup_{x \in U_\alpha} B_{p_{\alpha,x}}(q_{\alpha,x})$$

but the right side consists of countably many distinct open sets. Now consider $K \subseteq \bigcup_{j \geqslant 1} V_j$ is countable. Let $J_n = \bigcup_{j=1}^{n} V_j$. Then $K \subseteq \bigcup_{j \geqslant 1} J_j$, and $J_1 \subseteq J_2 \subseteq \ldots$ so by Proposition **??** there exists $M \in \mathbb{N}$ such that $K \subseteq J_M = \bigcup_{j=1}^{M} V_j$. Thus, in particular there exist $B_{p_{\alpha,x_1}}(q_{\alpha,x_1}), \ldots, B_{p_{\alpha,x_M}}(q_{\alpha,x_M})$ such that $K \subseteq \bigcup_{j=1}^{M} B_{p_{\alpha,x_j}}(q_{\alpha,x_j})$. Thus $K \subseteq \bigcup_{j=1}^{M} U_{\alpha_j}$, so we have a finite subcovering of $K$ as desired. ■

**Proposition 1.6.13.** *If $K \subseteq \mathbb{R}$ is topologically compact, then $K$ is sequentially compact.*

*Proof.* We prove the equivalent notion of limit point compactness (which is equivalent for metric spaces). Then we argue by contrapostive, supposing $S$ has no accumulation points. Then $S$ and $S_x = S \backslash \{x\}$ is closed for all $x \in S$. Setting $U_x = \mathbb{R} \backslash S_x$, we have that $K = (\bigcup_{x \in S} U_x) \cup \mathbb{R} \backslash S$ is an open cover, and as $K$ is topologically compact we have a finite subcover $U_{x_1}, \ldots, U_{x_N}, \mathbb{R} \backslash S$. Then $U_{x_1}, \ldots, U_{x_N}$ covers $S$, so $S = \bigcup_{i=1}^{N} U_{x_i} \cap S = \bigcup_{i=1}^{N} \{x_i\} = \{x_1, \ldots, x_N\}$, so $S$ is finite. Thus, $K$ is limit point compact as desired. ■

Thus, we have the equivalence between sequentially compact, topologically compact, and limit point compact, for subsets of $\mathbb{R}$.

## 1.7.0 Limits

**Remark 1.7.1** (Motivating Definition)**.** The function $f$ approaches the limit $l \in \mathbb{R}$ near $a \in \mathbb{R}$, if we can make $f(x)$ as "close as we like" to $l$ by requiring that $x$ be "sufficiently close to," but unequal to, $a$.

**Definition 1.7.1 (Limit).** *A real valued function $f : \mathbb{R} \to \mathbb{R}$ **approaches the limit $l$ near** $a$ if for every $\epsilon > 0$ there is some $\delta > 0$ such that, for all $x \in \mathbb{R}$, if $0 < |x - a| < \delta$, then $|f(x) - l| < \epsilon$.*

**Notation 1.7.2.** We denote the number $l$ which a function $f$ approaches near $a \in \mathbb{R}$ by $\lim_{x \to a} f(x)$, read ***the limit of** $f(x)$ **as** $x$ **approaches** $a$*.

**Lemma 1.7.1.**

1. *If*

$$|x - x_0| < \frac{\epsilon}{2} \text{ and } |y - y_0| < \frac{\epsilon}{2}$$

*then*

$$|(x + y) - (x_0 + y_0)| < \epsilon$$

2. *If*
$$|x - x_0| < \min\left(1, \frac{\epsilon}{2(|y_0| + 1)}\right) \text{ and } |y - y_0| < \frac{\epsilon}{2(|x_0 + 1)}$$
*then*
$$|xy - x_0y_0| < \epsilon$$

3. *If* $y_0 \neq 0$ *and*
$$|y - y_0| < \min\left(\frac{|y_0|}{2}, \frac{\epsilon|y_0|^2}{2}\right)$$
*then* $y \neq 0$ *and*
$$\left|\frac{1}{y} - \frac{1}{y_0}\right|$$

*Proof.* (1) Suppose $|x - x_0| < \frac{\epsilon}{2}$ *and* $|y - y_0| < \frac{\epsilon}{2}$. Then it follows that

$$\begin{aligned} |(x + y) - (x_0 + y_0)| &= |(x - x_0) + (y - y_0)| \\ &\leqslant |x - x_0| + |y - y_0| \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2} \\ &= \epsilon \end{aligned}$$

as desired.

(2) Next, suppose $|x - x_0| < \min\left(1, \frac{\epsilon}{2(|y_0|+1)}\right)$ *and* $|y - y_0| < \frac{\epsilon}{2(|x_0+1)}$. Note that as $|x - x_0| < 1$ we have that $|x| - |x_0| < 1$ so $|x| < 1 + |x_0|$. It follows that

$$\begin{aligned} |xy - x_0y_0| &= |xy - xy_0 + xy_0 - x_0y_0| \\ &\leqslant |x||y - y_0| + |y_0||x - x_0| \\ &< (|x_0| + 1)\frac{\epsilon}{2(|x_0| + 1)} + (|y_0| + 1)\frac{\epsilon}{2(|y_0| + 1)} \\ &= \frac{\epsilon}{2} + \frac{\epsilon}{2} \\ &= \epsilon \end{aligned}$$

(3) Suppose $y_0 \neq 0$ and $|y - y_0| < \min\left(\frac{|y_0|}{2}, \frac{\epsilon|y_0|^2}{2}\right)$. Then note that as $|y - y_0| < \frac{|y_0|}{2}$ $-\frac{|y_0|}{2} < y - y_0 < \frac{|y_0|}{2}$. If $y_0 > 0$ we have that $y_0 = |y_0|$ so $\frac{|y_0|}{2} < y < \frac{3|y_0|}{2}$. On the other hand if $y_0 < 0$ then $y_0 = -|y_0|$ so $-\frac{3|y_0|}{2} < y < -\frac{|y_0|}{2}$. In either case we have that $|y| > \frac{|y_0|}{2} > 0$, so $y \neq 0$. Then it follows that

$$\begin{aligned} \left|\frac{1}{y} - \frac{1}{y_0}\right| &= \left|\frac{y - y_0}{yy_0}\right| \\ &< \frac{\epsilon|y_0|^2}{2} \cdot \frac{1}{|y_0|} \cdot \frac{2}{|y_0|} \\ &= \epsilon \end{aligned}$$

as claimed. ■

**Theorem 1.7.2** (Limit Laws). *If $\lim_{x\to a} f(x) = l$ and $\lim_{x\to a} g(x) = m$, then*

1. $\lim_{x\to a}(f+g)(x) = l+m$;
2. $\lim_{x\to a}(f\cdot g)(x) = l\cdot m$;
3. *Moreover, if $m \neq 0$, then* $\lim_{x\to a}\left(\frac{1}{g}\right)(x) = \frac{1}{m}$

*Proof.* Suppose that $\lim_{x\to a} f(x) = l$ and $\lim_{x\to a} g(x) = m$. Let $\epsilon > 0$.

(1) Then since $\lim_{x\to a} f(x) = l$ and $\lim_{x\to a} g(x) = m$ there exist $\delta_1, \delta_2 > 0$ such that $|f(x)-l| < \frac{\epsilon}{2}$ if $0 < |x-a| < \delta_1$ and $|g(x)-m| < \frac{\epsilon}{2}$ if $0 < |x-a| < \delta_2$. Choose $\delta = \min(\delta_1, \delta_2)$. Then it follows that for $0 < |x-a| < \delta$:

$$\begin{aligned} |(f+g)(x)-(l+m)| &= |(f(x)-l)+(g(x)-m)| \\ &\leqslant |f(x)-l) + |g(x)-m| \\ &< \frac{\epsilon}{2}+\frac{\epsilon}{2} \\ &= \epsilon \end{aligned}$$

Thus, we have that by definition $\lim_{x\to a}(f+g)(x) = l+m$.

(2) Now, fix $\epsilon_1 = \min\left(1, \frac{\epsilon}{2(|m|+1)}\right)$ and $\epsilon_2 = \frac{\epsilon}{2(|l|+1)}$. Then since $\lim_{x\to a} f(x) = l$ and $\lim_{x\to a} g(x) = m$ there exist $\delta_1, \delta_2 > 0$ such that if $0 < |x-a| < \delta_1$ then $|f(x)-l| < \epsilon_1$ and if $0 < |x-a| < \delta_2$ then $|g(x)-m| < \epsilon_2$. Choose $\delta = \min(\delta_1, \delta_2)$. It follows that for $0 < |x-a| < \delta$, we have $|f(x)g(x)-lm| < \epsilon$, so by definition $\lim_{x\to a}(f\cdot g)(x) = l\cdot m$.

(3) Now, suppose $m \neq 0$. Fix $\epsilon_1 = \min\left(\frac{|m|}{2}, \frac{\epsilon|m|^2}{2}\right)$. Then as $\lim_{x\to a} g(x) = m$, there exists $\delta > 0$ such that if $|x-a| < \delta$, $|g(x)-m| < \epsilon_1$. By the previous Lemma we have that if $|x-a| < \delta$, $g(x) \neq 0$ and $\left|\frac{1}{g(x)} - \frac{1}{m}\right| < \epsilon$. Hence, by definition $\lim_{x\to a}\left(\frac{1}{g}\right)(x) = \frac{1}{m}$, as desired. ■

**Definition 1.7.2 (Limits from Above and Below).** *The* **limit from above** *for a function $f$ as $x$ goes to $a$ is denoted by $\lim_{x\to a^+} f(x) = l$, which means for every $\epsilon > 0$ there is a $\delta > 0$ such that for all $x$,*

$$if\ 0 < x-a < \delta,\ then\ |f(x)-l| < \epsilon$$

*where $0 < x-a < \delta$ is equivalent to $0 < |x-a| < \delta$ and $x > a$.*

*The* **limit from below** *for $f$ as $x$ goes to $a$ is denoted by $\lim_{x\to a^-} f(x) = l$, and means that for every $\epsilon > 0$ there is a $\delta > 0$ such that, for all $x$,*

$$if\ 0 < a-x < \delta,\ then\ |f(x)-l| < \epsilon$$

**Remark 1.7.3.** For a function $f : \mathbb{R} \to \mathbb{R}$, $\lim_{x\to a} f(x)$ exists if and only if $\lim_{x\to a^+} f(x)$ and $\lim_{x\to a^-} f(x)$ both exist and are equal.

**Definition 1.7.3 (Limits at Infinity).** *A* **limit at infinity** *is denoted by* $\lim_{x\to\infty} f(x) = l$*, and means that for every* $\epsilon > 0$ *there is* $M \in \mathbb{R}$ *such that for all* $x$*,*

$$if\ x > M,\ then\ |f(x) - l| < \epsilon$$

*A limit at negative infinity is defined analogously, replacing* $x > M$ *with* $x < M$*.*

**Definition 1.7.4.** *We define* $\lim_{x\to a} f(x) = \infty$ *to mean that for all* $N \in \mathbb{R}$*, there exists* $\delta > 0$ *such that, for all* $x \in \mathbb{R}$*, if* $0 < |x - a| < \delta$*, then* $f(x) > N$*. (the case for* $-\infty$ *is defined similarly)*

## 1.8.0 Continuous Functions

**Definition 1.8.1 (Continuity).** *Let* $f : \mathbb{R} \to \mathbb{R}$ *be a real function. Then* $f$ *is said to be* **continuous at a point** $a$ *if*

$$\lim_{x\to a} f(x) = f(a) \tag{1.8.1}$$

**Theorem 1.8.1.** *If* $f$ *and* $g$ *are continuous at a point* $a$*, then*

1. $f + g$ *is continuous at* $a$
2. $f \cdot g$ *is continuous at* $a$
3. *Moreover, if* $g(a) \neq 0$*, then* $1/g$ *is continuous at* $a$*.*

*Proof.* Suppose $f$ and $g$ are continuous at a point $a$. Then by the 1.7.2 theorem, we have that as $\lim_{x\to a} f(x) = f(a)$ and $\lim_{x\to a} g(x) = g(a)$,

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

Hence, $f + g$ is continuous at $a$. Similarly, again by the 1.7.2 theorem, we have that

$$\lim_{x\to a}(f \cdot g)(x) = f(a) \cdot g(a) = (f \cdot g)(a)$$

Thus, $f \cdot g$ is continuous at $a$. Finally, if $g(a) \neq 0$, the

$$\lim_{x\to a}(1/g)(x) = 1/g(a) = (1/g)(a)$$

so $1/g$ is continuous at $a$. ∎

**Theorem 1.8.2.** *If* $g$ *is continuous at* $a$*, and* $f$ *is continuous at* $g(a)$*, then* $f \circ g$ *is continuous at* $a$*.*

*Proof.* Let $\epsilon > 0$. Then by continuity of $f$ there exists $\delta_1 > 0$ such that if $|g(x) - g(a)| < \delta_1$, then

$$|f(g(x)) - f(g(a))| < \epsilon$$

Then, by the continuity of $g$, there exists $\delta > 0$ such that if $|x-a| < \delta$, then $|g(x)-g(a)| < \delta_1$. Thus, if $|x-a| < \delta$ we have that

$$|(f \circ g)(x) - (f \circ g)(a)| = |f(g(x)) - f(g(a))| < \epsilon$$

proving continuity of $f \circ g$ at $a$. ∎

**Definition 1.8.2.** *A function $f$ is called* **continuous on** *an open interval $(a,b)$, if $f$ is continuous at $x$ for all $x \in (a,b)$.*

*A function $f$ is called* **continuous on** *a closed interval $[a,b]$ if*

1. *$f$ is continuous at $x$ for all $x \in (a,b)$*
2. *$\lim_{x \to a^+} f(x) = f(a)$ and $\lim_{x \to b^-} f(x) = f(b)$*

*In general, a function $f$ is* **continuous** *if it is continuous at $x$ for all $x$ in its domain.*

**Theorem 1.8.3.** *Suppose $f$ is continuous at $a$, and $f(a) > 0$. Then $f(x) > 0$ for all $x$ in some interval containing $a$; more precisely, there is a number $\delta > 0$ such that $f(x) > 0$ for all $x$ satisfying $|x-a| < \delta$. Similarly, if $f(a) < 0$, then there is a number $\delta > 0$ such that $f(x) < 0$ for all $x$ satisfying $|x-a| < \delta$.*

*Proof.* Suppose $f$ is continuous at $a$. Let $\epsilon = \frac{f(a)}{2} > 0$. Then by continuity there exists $\delta > 0$ such that if $|x-a| < \delta$, $|f(x) - f(a)| < \epsilon$. Then we have that $-\epsilon < f(x) - f(a) < \epsilon$, so $f(x) > \epsilon > 0$, satisfying the claim. The case for $f(a) < 0$ is proved analogously. ∎

## Important Theorems and Results on Continuity

**Theorem 1.8.4.** *If $f$ is continuous on a closed interval $[a,b]$ (a compact set) and $f(a) < 0 < f(b)$, then there is some $x \in [a,b]$ such that $f(x) = 0$.*

*Proof.* Consider an interval $[a,b]$ such that $f(a) < 0 < f(b)$. Define the set

$$a := \{x \in \mathbb{R} : a \leqslant x \leqslant b, \text{and } f \text{ is negative on } [a,x]\}$$

Clearly $A \neq \varnothing$ as $a \in A$. In fact, there exists some $\delta > 0$ such that $A$ contains all points $x \in \mathbb{R}$ satisfying $a \leqslant x < a+\delta$, since $f$ is continuous on $[a,b]$ and $f(a) < 0$. Similarly, $b$ is an upper bound for $A$ and, in fact, there is a $\delta > 0$ such that all points satisfying $b-\delta < x \leqslant b$ are upper bounds for $A$.

Thus, applying the Least Upper Bound property of $\mathbb{R}$, $A$ has a least upper bound $\alpha$ and $a < \alpha < b$. We wish to show that $f(\alpha) = 0$. First, if $f(\alpha) < 0$, then by a previous result there is a $\delta > 0$ such that $f(x) < 0$ for $\alpha - \delta < x < \alpha + \delta$. In particular, there is some number $x_0 \in A$ satisfying $\alpha - \delta < x < \alpha$ since $\alpha$ is the supremum of $A$. Thus $f$ is negative on the whole interval $[a, x_0]$. But, if $x_1 \in (\alpha, \alpha + \delta$, then $f$ is also negative on the whole interval $[x_0, x_1]$.

Therefore $f$ is negative on the interval $[a, x_1]$ so $x_1 \in A$. But, this contradicts the fact that $\alpha$ is an upper bound for $A$, so $f(\alpha) < 0$ must be false.

Suppose, on the other hand, that $f(\alpha) > 0$. Then there is a number $\delta > 0$ such that $f(x) > 0$ for all $\alpha - \delta < x < \alpha + \delta$. Now there is some number $x_0 \in A$ such that $\alpha - \delta < x_0 < \alpha$ as $\alpha$ is presumed to be the supremum of $A$. This means that $f$ is negative on the whole interval $[a, x_0]$, which is impossible since $f(x_0) > 0$. Thus, the assumption $f(\alpha) > 0$ leads to a contradiction, leaving $f(\alpha) = 0$ as the only possible alternative. ∎

**Lemma 1.8.5.** *If $f$ is continuous at $a$, then there is a number $\delta > 0$ such that $f$ is bounded on the interval $(a - \delta, a + \delta)$.*

*Proof.* Since $f$ is continuous at $a$ we have that $\lim_{x \to a} f(x) = f(a)$. Then, fix $\epsilon = 1$. By continuity it follows that there exists $\delta > 0$ such that for all $x \in (a - \delta, a + \delta)$, $|f(x) - f(a)| < 1$. In particular, we have that $f(x) < f(a) + 1$, so $f$ is bounded above by $f(a) + 1$ on $(a - \delta, a + \delta)$. Moreover, $f(x) > f(a) - 1$, so $f$ is bounded below by $f(a) - 1$ on $(a - \delta, a + \delta)$. Thus $f$ is bounded on the interval $(a - \delta, a + \delta)$ as claimed. ∎

**Corollary 1.8.6.** *If $\lim_{x \to a^+} f(x) = f(a)$ then there exists $\delta > 0$ such that $f$ is bounded on the interval $[a, a+\delta)$. Moreover, if $\lim_{x \to b^-} f(x) = f(b)$ then there exists $\delta > 0$ such that $f$ is bounded on the interval $(b - \delta, b]$.*

**Theorem 1.8.7.** *If $f$ is continuous on a closed interval $[a, b]$ (a compact set), then $f$ is bounded above on $[a, b]$, that is, there is some number $M \in \mathbb{R}$ such that $f(x) \leqslant M$ for all $x \in [a, b]$ (consequence of the continuous image of a compact set being compact and the Heine-Borel Theorem).*

*Proof.* Define the set

$$A := \{x \in [a, b] : f \text{ is bounded above on } [a, x]\}$$

Clearly $A \neq \emptyset$ as $a \in A$, and $A$ is bounded above by $B$, so $A$ has a least upper bound $\alpha \in \mathbb{R}$. We wish to show that $\alpha = b$. Suppose towards a contradiction that $\alpha < b$. Then there exists $\delta > 0$ such that $f$ is bounded on $(\alpha - \delta, \alpha + \delta)$ since $f$ is continuous on $[a, b]$, so in particular $f$ is continuous at $\alpha$. Since $\alpha$ is the least upper bound of $A$ there is some $x_0 \in A$ satisfying $\alpha - \delta < x_0 < \alpha$. This implies that $f$ is bounded on $[a, x_0]$. But, if $x_1$ is any number with $\alpha < x_1 < \alpha + \delta$, then $f$ is also bounded on $[x_0, x_1]$. Therefore $f$ is bounded on $[a, x_1]$ so $x_1 \in A$, contradicting the fact that $\alpha$ is an upper bound for $A$. This contradiction shows that $\alpha = b$. Now, there is a $\delta > 0$ such that $f$ is bounded on $(b - \delta, b]$. There is $x_0 \in A$ such that $b - \delta < x_0 < b$, since $\alpha = b$. Thus $f$ is bounded on $[a, x_0]$, and also on $[x_0, b]$, so $f$ is bounded on $[a, b]$, completing the proof. ∎

**Theorem 1.8.8.** *If $f$ is continuous on a closed interval $[a, b]$, then there is some number $y \in [a, b]$ such that $f(y) \geqslant f(x)$ for all $x \in [a, b]$.*

*Proof.* From the previous theorem we know that $f$ is bounded on $[a, b]$, so the set $\{f(x) : x \in [a, b]\}$ is bounded. This set is obviously non-empty, so it has a least upper bound $\alpha \in \mathbb{R}$. Since

$\alpha \geqslant f(x)$ for all $x \in [a,b]$, it suffices to show that $\alpha = f(y)$ for some $y \in [a,b]$. Suppose instead that $\alpha \neq f(y)$ for all $y \in [a,b]$. Then the function $g$ defined by

$$g(x) = \frac{1}{\alpha - f(x)}, x \in [a,b]$$

is continuous on $[a,b]$ since the denominator is never zero and is the sum of continuous functions. On the other hand, $\alpha$ is the least upper bound of $\{f(x) : x \in [a,b]\}$ so for every $\epsilon > 0$ there exists $x \in [a,b]$ such that $\alpha - \epsilon < f(x)$, so $\alpha - f(x) < \epsilon$. This in turn implies that for every $\epsilon > 0$ there exists $x \in [a,b]$ with $g(x) > 1/\epsilon$. But, this implies that $g$ is not bounded on $[a,b]$, contradicting the previous theorem as $g$ is assumed to be continuous. Hence, $g$ is not continuous, and i particular $\alpha - f(y) = 0$ for some $y \in [a,b]$. ■

## Theorem 6 (Intermediate Value Theorem).

*If $f$ is continuous on $[a,b]$ and $f(a) < c < f(b)$, then there is some $x \in [a,b]$ such that $f(x) = c$ (continuous image of a connected set is connected).*

*Moreover, if $f(a) > c > f(b)$, then there is some $x \in [a,b]$ such that $f(x) = c$.*

**Theorem 1.8.9.** *If $f$ is continuous on $[a,b]$, then $f$ is bounded below on $[a,b]$, that is, there is some number $M \in \mathbb{R}$ such that $f(x) \geqslant M$ for all $x \in [a,b]$.*

**Theorem 1.8.10.** *If $f$ is continuous on $[a,b]$, then there is some $y \in [a,b]$ such that $f(y) \leqslant f(x)$ for all $x \in [a,b]$.*

**Corollary 1.8.11.** *For all $\alpha \in \mathbb{P}$, so $\alpha > 0$, there exists $x \in \mathbb{R}$ such that $x^2 = \alpha$.*

*Proof.* Consider the function $f(x) = x^2$, which is certainly continuous over $\mathbb{R}$. Consider $\alpha \in \mathbb{P}$. Then there exists $b > 0$ such that $f(b) > \alpha$. Indeed, if $\alpha > 1$ we can take $b = \alpha$, and if $\alpha < 1$ we can take $b = 1$. Then, $f$ is defined on the closed interval $[0,b]$ and $f(0) = 0 < \alpha < f(b)$. Therefore, by the 6 there exists $c \in [0,b]$ such that $f(c) = \alpha$. In particular, $c^2 = \alpha$. ■

**Corollary 1.8.12.** *If $n$ is odd, then any equation*

$$x^n + a_{n-1}x^{n-1} + ... + a_0 = 0 \tag{1.8.2}$$

*has a solution, or root.*

*Proof.* Consider the function $f(x) = x^n + a_{n-1}x^{n-1} + ... + a_0$. Write

$$f(x) = x^n + a_{n-1}x^{n-1} + ... + a_0 = x^n \left(1 + \frac{a_{n-1}}{x} + ... + \frac{a_0}{x^n}\right)$$

Then note that

$$\left|\frac{a_{n-1}}{x} + \frac{a_{n-2}}{x^2} + ... + \frac{a_0}{x^n}\right| \leqslant \frac{|a_{n-1}|}{|x|} + ... + \frac{|a_0|}{|x^n|}$$

Choose $x$ such that

$$|x| > 1, 2n|a_{n-1}|, ..., 2n|a_0|$$

so $|x^k| > |x|$ for all $k > 1$, and

$$\frac{|a_{n-k}|}{|x^k|} < \frac{|a_{n-k}|}{|x|} < \frac{|a_{n-k}|}{2n|a_{n-k}|} < \frac{1}{2n}$$

Thus, we have that

$$\left|\frac{a_{n-1}}{x} + \frac{a_{n-2}}{x^2} + ... + \frac{a_0}{x^n}\right| \leqslant \frac{|a_{n-1}|}{|x|} + ... + \frac{|a_0|}{|x^n|} < \underbrace{\frac{1}{2n} + ... + \frac{1}{2n}}_{n \text{ times}} = \frac{1}{2}$$

In other words,

$$-\frac{1}{2} < \frac{a_{n-1}}{x} + \frac{a_{n-2}}{x^2} + ... + \frac{a_0}{x^n} < \frac{1}{2}$$

which implies that

$$\frac{1}{2} < 1 + \frac{a_{n-1}}{x} + \frac{a_{n-2}}{x^2} + ... + \frac{a_0}{x^n}$$

Choosing $x_1 > 0$ which satisfies our condition, we have

$$\frac{x_1^n}{2} \leqslant x_1^n \left(1 + \frac{a_{n-1}}{x} + \frac{a_{n-2}}{x^2} + ... + \frac{a_0}{x^n}\right) = f(x_1)$$

so that $f(x_1) > 0$. On the other hand, choosing $x_2 < 0$ satisfying our condition, $x_2^n < 0$ as $n$ is odd and

$$\frac{x_2^n}{2} \geqslant x_2^n \left(1 + \frac{a_{n-1}}{x} + \frac{a_{n-2}}{x^2} + ... + \frac{a_0}{x^n}\right) = f(x_2)$$

so that $f(x_2) < 0$. Applying the 6 to the interval $[x_2, x_1]$ we conclude that there exists $c \in [x_2, x_1]$ such that $f(c) = 0$. ∎

**Theorem 1.8.13.** *If $n$ is even and $f(x) = x^n + a_{n-1}x^{n-1} + ... + a_0$, then there is a number $y$ such that $f(y) \leqslant f(x)$ for all $x \in \mathbb{R}$.*

*Proof.* Choose $M$ such that

$$M = \max(1, 2n|a_{n-1}|, ..., 2n|a_0|)$$

Then for all $x$ with $|x| \geqslant M$ we have

$$\frac{1}{2} \leqslant 1 + \frac{a_{n-1}}{x} + \frac{a_{n-2}}{x^2} + ... + \frac{a_0}{x^n}$$

Since $n$ is even, $x^n \geqslant 0$ for all $x$, so

$$\frac{x^n}{2} \leqslant x^n \left(1 + \frac{a_{n-1}}{x} + \frac{a_{n-2}}{x^2} + ... + \frac{a_0}{x^n}\right) = f(x)$$

provided that $|x| \geqslant M$. Now consider the number $f(0)$. Let $b > 0$ be a number such that $b^n \geqslant 2f(0)$ and also $b > M$. Then if $x \geqslant b$, we have

$$f(x) \geqslant \frac{x^n}{2} \geqslant \frac{b^n}{2} \geqslant f(0)$$

The same holds for $x \leqslant -b$. In particular, if $x \geqslant b$ or $x \leqslant -b$, then $f(x) \geqslant f(0)$. Applying the extreme value theorem to $f$ on the interval $[-b, b]$, we conclude that there is a number $y \in [-b, b]$ such that if $-b \leqslant x \leqslant b$, then $f(y) \leqslant f(x)$. In particular, $f(y) \leqslant f(0)$. Thus, if $x \leqslant -b$ or $x \geqslant b$, then $f(x) \geqslant f(0) \geqslant f(y)$. Combining these results we find that $f(y) \leqslant f(x)$ for all $x \in \mathbb{R}$. ∎

**Corollary 1.8.14.** *Consider the equation*

$$x^n + a_{n-1}x^{n-1} + ... + a_0 = c \tag{1.8.3}$$

*for $n$ even. Then there is a number $m$ such that the equation has a solution for $c \geqslant m$ and has no solution for $c < m$.*

*Proof.* Let $f(x) = x^n + a_{n-1}x^{n-1} + ... + a_0$. According to our previous theorem there exists $y \in \mathbb{R}$ such that $f(y) \leqslant f(x)$ for all $x \in \mathbb{R}$. Let $m = f(y)$. If $c < m$ then the equation above has no solutiion, since the left hand side has a value $\geqslant m$ always. If $c = m$, then $y$ is a solution of the equation. Finally, for $c > m$, let $b > y$ such that $f(b) > c$. Then the 6 applied to the interval $[y, b]$ states that there exists $x \in [y, b]$ such taht $f(x) = c$, so $x$ is a solution of the equation. ∎

## Uniform Continuity

**Definition 1.8.3.** *A function $f : \mathbb{R} \to \mathbb{R}$ is* **uniformly continuous on an interval** *$I$ if for every $\epsilon > 0$ there is some $\delta > 0$ such that, for all $x, y \in I$, if $|x - y| < \delta$ then $|f(x) - f(y)| < \delta$.*

**Lemma 1.8.15.** *Let $a < b < c$ and let $f$ be continuous on the interval $[a, c]$. Let $\epsilon > 0$ and suppose that*

1. *if $x, y \in [a, b]$ and $|x - y| < \delta_1$, then $|f(x) - f(y)| < \epsilon$*
2. *if $x, y \in [b, c]$ and $|x - y| < \delta_2$, then $|f(x) - f(y)| < \epsilon$*

*Then there is a $\delta > 0$ such that if $x, y \in [a, c]$ and $|x - y| < \delta$, then $|f(x) - f(y)| < \epsilon$.*

*Proof.* Fix $\epsilon > 0$. Since $f$ is continuous at $b$ there exists $\delta_3 > 0$ such that if $|x - b| < \delta_3$, then $|f(x) - f(b)| < \frac{\epsilon}{2}$. It follows that if $|x - b| < \delta_3$ and $|y - b| < \delta_3$ then $|f(x) - f(y)| < \epsilon$. Choose $\delta = \min(\delta_1, \delta_2, \delta_3)$. Let $x, y \in [a, c]$ with $|x - y| < \delta$. If $x$ and $y$ are both in $[a, b]$, then $|f(x) - f(y)| < \epsilon$ by assumption. Similarly, if $x, y \in [b, c]$, then again $|f(x) - f(y)| < \epsilon$ by assumption. Finally, without loss of generality suppose $x < b < y$. Since $|x - y| < \delta$ we have that $|x - b| = |b - x| = b - x < y - x = |y - x| < \delta$ and similarly $|y - b| < \delta$. Thus, we have that $|f(x) - f(y)| < \epsilon$, completing the proof. ∎

**Theorem 1.8.16 (Uniform Continuity Theorem).** *If $f$ is continuous on $[a, b]$, then $f$ is uniformly continuous on $[a, b]$.*

*Proof.* Consider $\epsilon > 0$. Define the set

$$A(\epsilon) := \{x \in [a, b] : \exists \delta > 0; \forall y, z \in [a, x]; |y - z| < \delta \implies |f(y) - f(z)| < \epsilon\}$$

Then $A(\epsilon) \neq \varnothing$ since $a \in A(\epsilon)$, and $A(\epsilon)$ is bounded above by $b$, so $A(\epsilon)$ has a least upper bound $\alpha_\epsilon \in \mathbb{R}$. Suppose towards a contradiction that $\alpha < b$. Since $f$ is continuous at $\alpha$, there is some $\delta_0$ such that if $|y - \alpha| < \delta_0$, then $|f(y) - f(\alpha)| < \epsilon/2$. Consequently, if $|y - \alpha| < \delta_0$ and $|z - \alpha| < \delta_0$, then $|f(y) - f(z)| < \epsilon$. So, $f$ surely satisfies the condition for containment in $A(\epsilon)$ on the interval $[\alpha - \delta_0, \alpha + \delta_0$. On the other hand, since $\alpha$ is the least upper bound of $A$ it is also clear that the condition is satisfies on $[a, \alpha - \delta_0]$, namely $\alpha - \delta_0 \in A$. Then, the Lemma implies that $f$ satisfies the condition on $[a, \alpha + \delta_0]$ since it satisfies it on $[a, \alpha - \delta_0]$ and $[\alpha - \delta_0, \alpha + \delta_0]$. Hence, $\alpha + \delta_0 \in A$, contradicting the fact that $\alpha$ is an upper bound.

To complete the proof we must show that $\alpha = b$ is in $A$. Since $f$ is continuous at $b$, there is some $\delta_0 > 0$ such that if $y \in (b - \delta_0, b)$, then $|f(y) - f(b)| < \epsilon/2$. So, for any $x, y \in [b - \delta_0, b]$, $|f(y) - f(x)| < \epsilon$. But, $f$ satisfies the condition for $A(\epsilon)$ on $[a, b - \delta_0]$ since $b$ is the least upper bound of $A(\epsilon)$, so the Lemma implies that $f$ satisfies the condition on $[a, b]$. Therefore, as $\epsilon > 0$ was arbitrary, we conclude that $f$ is uniformly continuous on $[a, b]$, completing the proof. ■

## 1.9.0 Metric Properties of the Reals Summarized

First we need to define a sequence

## 1.A.0 Cardinality

In this appendix we investigate the notion of the size of a set. First, we define the set

$$I_n := \{j \in \mathbb{N} : j \leqslant n$$

the prototypical set of size $n$.

**Lemma 1.A.1.** *$I_1 = \{1\}$, and $I_{n+1} = I_n \cup \{n + 1\}$.*

*Proof.* If $n = 1$, $I_1 = \{j \in \mathbb{N} : j \leqslant 1\} = \{1\}$, so the base case holds. Further, $I_2 = \{j \in \mathbb{N} : j \leqslant 2\} = \{1, 2\} = I_1 \cup \{2\}$. Now, suppose for some $j \in \mathbb{N}$, $I_{j+1} = I_j \cup \{j + 1\}$. Then

$$I_{j+2} = \{k \in \mathbb{N} : k \leqslant j + 2\} = \{k \in \mathbb{N} : k \leqslant j + 1 \vee k = j + 2\} = I_{j+1} \cup \{j + 2\}$$

■

**Definition 1.A.1.** *A non-empty set $S$ has $n$ elements if and only if there exists a bijective map $\varphi : S \to I_n$.*

**Proposition 1.A.2.** *For $m, n \in \mathbb{N}$, if there exists an injection $\varphi : I_m \to I_n$, then $m \leqslant n$.*

*Proof.* If $n = 1$, then $I_n = I_1 = \{1\}$. Now, suppose $\varphi : I_m \to I_1$ is an injection. If $x, y \in I_m$, then $\varphi(x) = 1 = \varphi(y)$, so by injectivity $x = y$. Thus, $I_m$ has only one element, and since $1 \in I_m$, we must have $I_m = \{1\} = I_1$. Hence, $m = 1 \leqslant 1 = n$. Assume the result is true for some $1 \leqslant n < N$. Then let $\varphi : I_m \to I_N$ be an injection. If $m = 1$ the result is immediately satisfied, so suppose $m \geqslant 2$.

(1) Suppose there exist $j \in I_m$ such that $\varphi(j) = N$. Then define $\psi : I_{m-1} \to I_{N-1}$ by $\psi(l) = \varphi(l)$ for $l < j$, and $\psi(l) = \varphi(l+1)$ for $j \leqslant l \leqslant m-1$. Then $\psi$ is injective because $\varphi$ is injective, so the the induction hypothesis $m-1 \leqslant N-1$, so $m \leqslant N$.

(2) If there does not exist $j \in I_m$ such that $\varphi(j) = N$, then we can restrict the codomain of $\varphi$ to obtain $\varphi : I_m \to I_{N-1}$. By the induction hypothesis $m \leqslant N-1 < N$, as desired.

■

**Corollary 1.A.3.** *If there exists $\varphi : I_m \to I_n$ bijective, then $m = n$.*

**Corollary 1.A.4.** *Suppose $S$ is a set, $n, m \in \mathbb{N}$, and there exist bijections $\varphi : S \to I_n$ and $\psi : S \to I_m$, then $n = m$.*

The result follows from considering $\varphi \circ \psi^{-1} : I_m \to I_n$.

**Definition 1.A.2.** *If either $S = \varnothing$ or $S$ has $n$ elements for some $n \in \mathbb{N}$, then we say $S$ is* **finite**. *Otherwise, $S$ is said to be* **infinite**.

**Proposition 1.A.5.** *If $n \in \mathbb{N}$ and $S \subseteq I_n$, then there exists $m \leqslant n$ and $\varphi : S \to I_m$ bijective.*

*Proof.* If $n = 1$, then $I_1 = \{1\}$ and the only non-empty subset is $S = \{1\} = I_1$ itself. Then $\varphi : S \to I_1$ given by $\varphi(1) = 1$ is a bijection so the base case holds. Suppose for $k \geqslant 1$, if $S \subseteq I_k$, then there exists $m \leqslant k$ such that $\varphi_k : S \to I_m$, bijective. Suppose $S \subseteq I_{k+1}$. If $S = I_{k+1}$, then $\mathrm{id} : I_{k+1} \to I_{k+1}$ is the desired bijection. Otherwise, there exists $j \in I_{k+1}$ such that $j \notin S$. If $j = k+1$, define $\varphi : S \to I_k$ by $\varphi(s) = s$ for all $s \in S$. Then $\varphi$ is an injection and $\varphi(S) \subseteq I_k$ so by the induction hypothesis there exists $m \leqslant k$ and a bijection $\psi : \varphi(S) \to I_m$. Then $\psi \circ \varphi : S \to I_m$ is bijective. If $j \neq k+1$, then define $\varphi(S) \to I_k$ by $\varphi(l) = l$ for $l \leqslant k$, and $\varphi(k+1) = j$. Then $\varphi$ is injective, and by the induction hypothesis there exists $m \leqslant k$ and a bijection $\psi : \varphi(S) \to I_m$, so $\psi \circ \varphi : S \to I_m$ is the desired bijection. ■

**Proposition 1.A.6.** *$\mathbb{N}$ is not finite.*

*Proof.* If $\mathbb{N}$ was finite then there would exist $n \in \mathbb{N}$ and a bijection $\varphi : I_n \to \mathbb{N}$. As $I_{n+1} \subseteq \mathbb{N}$, if we restrict $\varphi$ to $\varphi^{-1}(I_{n+1}) = S$, we have $\psi : S \to I_{n+1}$, which is bijective. But $S \subseteq I_n$, so by Proposition 1.A.5 $S$ has $m$ elements with $m \leqslant n$, and as $n+1 = m \leqslant n$, a contradiction. Thus, $\mathbb{N}$ must be infinite. ■

**Proposition 1.A.7.** *If $S$ is not finite, then there exists an injection $\varphi : \mathbb{N} \to S$.*

*Proof.* As $S$ is non-empty, there exists $s_1 \in S$, so define $\varphi_1(1) = s_1$. Suppose there exists $k \in \mathbb{N}$ such that for $\varphi_k(l) = s_l \in S \backslash \{s_1, ..., s_{l-1}$ for all $1 \leqslant l \leqslant k$. Then as $S$ is infinite, there exists $s_{k+1} \in S$ with $s_{k+1} \notin \{s_1, ..., s_k\}$, so we define $\varphi_{k+1}(k+1) = s_{k+1}$ and $\varphi_{k+1}(l) = \varphi_k(l)$ for all $1 \leqslant l \leqslant k$. Thus, for all $n \in \mathbb{N}$ we have $\varphi_n : I_n \to S$ injective, and for all $m \leqslant n$, $\varphi_n|_{I_m} = \varphi_m : I_m \to S$. Define $\Phi(m) = \varphi_n(m)$ for all $n \geqslant m$. Then $\Phi : \mathbb{N} \to S$ is well defined and injective. ■

**Definition 1.A.3.** *We say that a set $S$ is* **countably infinite** *if there exists a bijection $\varphi : S \to \mathbb{N}$.*

**Definition 1.A.4.** *Two sets $S$ and $T$ have the same cardinality if there exists a bijection between them, and we write $Card(S) = Card(T)$.*

**Definition 1.A.5.** *If $S$ is finite:*

- *$Card(S) = 0$ if $S = \varnothing$*
- *$Card(T) = n$ if $S$ has $n$ elements.*

We now move on to a fundamental theorem on cardinalities of sets:

## Theorem 7 (Schröder-Bernstein Theorem).

*If $S$ and $T$ are two non-empty sets such that there exist injective maps*

$$\varphi : S \hookrightarrow T \quad \text{and} \quad \psi : T \hookrightarrow S$$

*then there exists a bijection $\Phi : S \to T$.*

*Proof.* Let $\varphi : S \to T$ and $\psi : T \to S$ be the injections in the hypothesis. If $s \in S$ such that $\varphi(s) = t$, we say $s$ is a parent of $t$. Similarly, if $t' \in T$ such that $\psi(t') = s' \in S$, then we say $t'$ is a parent of $s'$. For elements in $S$ and $T$ there are three disjoint cases:

a) The set of elements who have an infinite number of ancestors

b) The set of elements whose last ancestor is an element of $S$

c) The set of elements whose last ancestor is an element of $S$

Define $S_a, T_a, S_b, T_b, S_c, T_c$ be the corresponding sets, so $S = S_a \sqcup S_b \sqcup S_c, T = T_a \sqcup T_b \sqcup T_c$. I claim that $\varphi|_{S_a} : S_a \to T_a$ is bijective. Indeed, $\varphi$ is injective. If $s \in S_a$, $\varphi(s) \in T_a$ as all ancestors of $s$ are ancestors of $\varphi(s)$. If $t \in T_a$, then $t$ has infinitely many ancestors, so in particular there exists $s \in S$ such that $\varphi(s) = t$. But, all ancestors of $t$, except $s$, are ancestors of $s$ so $s \in S_a$. Hence $\varphi|_{S_a}$ is indeed well defined and bijective.

Next we claim that $\varphi|_{S_b} : S_b \to T_b$ is bijective. If $s \in S_b$, then $s$'s ancestors terminate in $S$, so $\varphi(s) \in T_b$. If $t \in T_b$, then $t$ has at least one ancestor $s \in S$, and as $\varphi(s) \in T_b$, $s \in S_b$. Thus $\varphi|_{S_b}$ is well defined and bijective. Dually, we have that $\psi|_{T_c} : T_c \to S_c$ is bijective. Then define $\Phi : S \to T$ by $\Phi(s) = \varphi(s)$ for $s \in S_a \cup S_b$, and $\Phi(s) = \psi|_{T_c}^{-1}(s)$ for $s \in S_c$, which is by construction bijective. ■

A classical application of this result is in the following example:

**Example 1.A.1.** $\text{Card}(\mathbb{N}) = \text{Card}(\mathbb{N} \times \mathbb{N})$. Define $\varphi : \mathbb{N} \to \mathbb{N} \times \mathbb{N}$ by $\varphi = \Delta$ is the diagonal, so $\varphi$ is injective. Conversely, define $\psi : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ by $\psi(n, m) = 2^n 3^m$, so by the fundamental theorem of arithmetic $\psi$ is injective. Then by Schröder-Bernstein there exists a bijection $\Phi : \mathbb{N} \to \mathbb{N} \times \mathbb{N}$, as desired.

The following are the axioms founding the Zermelo-Frankel axioms of set theory with Choice:

**Axiom 1.A.6 (ZFC Axioms).** *The following constitute the axioms of ZFC:*

1. *(Emptyset) There is a set, denoted by $\varnothing$, which has no members:*
$$\exists x \forall t\ \neg t \in x$$
2. *(Pairset) For any two sets $x$ and $y$ there is a set $p$ with the property that $t \in p$ if and only if $t = x$ or $t = y$. This set $p$ is usually denoted $\{x, y\}$:*
$$\forall x \forall y \exists p (t \in p \iff (t = x \vee t = y))$$
3. *(Extensionality) For any two sets $x$ and $y$, $x = y$ if and only if $x$ and $y$ have exactly the same members:*
$$\forall x \forall y (x = y \iff \forall t (t \in x \iff t \in y))$$
4. *(Union Set) For any set $x$ there exists a set denote by $\bigcup x$ whose members are exactly the members of the members of $x$:*
$$\forall x \exists y \forall t (t \in y \iff \exists y (y \in x \wedge t \in y))$$
5. *(Infinity) There exists a set $I$ which contains $0 = \varnothing$ as well as the successor of each of its members; that is, if $x \in I$, then $S(x) := x \cup \{x\} \in I$:*
$$\exists x (\varnothing \in I \wedge \forall x (x \in I \implies \bigcup \{x, \{x\}\}))$$
6. *(Powerset) For each set $A$ there exists a set $\mathcal{P}(A)$ whose members are the subsets of $A$:*
$$\forall x \exists y \forall t (t \in y \iff \forall z (z \in t \implies z \in x))$$
7. *(Separation) Suppose $P$ is a definite condition. For each set $A$ there exists a set $B$ whose members are exactly the members of $A$ that satisfy $P$:*
$$\forall x \exists y \forall t (t \in B \iff (t \in x \wedge P(t)))$$
8. *(Replacement) Suppose $P$ is a definite binary condition such that for each set $x$ there is a unique set $y$ for which $P(x, y)$ holds. Given a set $A$ there exists a set $B$ with the property that $y \in B$ if and only if there exists $x \in A$ such that $P(x, y)$:*
$$\forall x \exists y \forall t (t \in y \iff \exists z (z \in x \wedge P(z, t)))$$
9. *(Regularity) Every non-empty set $A$ contains an element that is disjoint from $A$:*
$$\forall x (\neg x = \varnothing \implies \exists t (t \in x \wedge \forall y (y \in t \implies \neg y \in x)))$$
10. *(Choice) Every non-empty set $X$ whose members are all non-empty sets, there exists a function $f : X \to \bigcup X$ such that $f(A) \in A$ for all $A \in X$.*

**Lemma 1.A.8.** *If $\varphi : S \to T$ is onto, then there exists $\psi : T \to S$ which is injective.*

*Proof.* Let $\varphi : S \to T$ be onto. Then let $X = \{\varphi^{-1}(t) : t \in T\}$, so $S = \bigcup X$, and all members of $X$ are non-empty sets since $\varphi$ is onto. By the axiom of choice there exists a function $f : X \to \bigcup X$ such that $f(\varphi^{-1}(t)) \in \varphi^{-1}(t)$ for all $t \in T$. Since $\varphi$ is a well defined function $f$ is injective. Let $g : T \to X$ by $g(t) = \varphi^{-1}(t)$. Then $g$ is also injective as $\varphi$ is well-defined, so their composite $f \circ g : T \to \bigcup X = S$ is an injection. ∎

As an application of our results, I claim that $\mathrm{Card}(\mathbb{R}) \neq \mathrm{Card}(\mathbb{N})$:

*Proof.* First, $\iota : (0,1) \to \mathbb{R}$, being the natural inclusion, is an injection, and $\varphi : \mathbb{R} \to (0,1)$ given by $\varphi(x) = \frac{e^x}{e^x+1}$ is an injection, so by Schröder Bernstein there exists a bijection $\Phi : (0,1) \to \mathbb{R}$. Towards a contradiction suppose $\mathrm{Card}(\mathbb{N}) = \mathrm{Card}(\mathbb{R}) = \mathrm{Card}((0,1))$, so we have a bijection $f : \mathbb{N} \to (0,1)$. Expand the terms in their decimal expansion, so $f(j) = \sum_{n=1}^{\infty} a_{jn} 10^{-n}$, $a_{jn} \in \{0,1,2,...,9\}$. Define $x$ by $x = \sum_{n=1}^{\infty} b_n 10^{-1}$ where $b_n = 2$ if $a_{nn} \neq 2$ and $b_n = 3$ if $a_{nn} = 2$. Thus $x \neq f(j)$ for all $j \in \mathbb{N}$, but by assumption $f$ is bijective, and hence onto, which is a contradiction since $x \in (0,1)$. Thus, no such bijection can exist. ∎

**Definition 1.A.7.** *For sets $S$ and $T$ we define $\leqslant$ on the cardinals by $\mathrm{Card}(S) \leqslant \mathrm{Card}(T)$ if and only if there exists an injection $\varphi : S \to T$, and $\mathrm{Card}(S) < \mathrm{Card}(T)$ if and only if $\mathrm{Card}(S) \leqslant \mathrm{Card}(T)$ and $\mathrm{Card}(S) \neq \mathrm{Card}(T)$.*

Then it follows that $\mathrm{Card}(\mathbb{N}) < \mathrm{Card}(\mathbb{R})$.

**Continuum Hypothesis.** *There exists no set with cardinality strictly between $\aleph_0 = \mathrm{Card}(\mathbb{N})$ and $\aleph_1 = \mathrm{Card}(\mathbb{R})$.*

This hypothesis cannot be proven and taking it to be true or false in your system for set theory leads to no contradictions.

**Proposition 1.A.9.** *For any set $S$, $\mathcal{P}(S) \cong 2^S$.*

*Proof.* Let $\varphi : \mathcal{P}(S) \to 2^S$ by $\varphi(A)(s) = 1$ if and only if $s \in A$, and 0 otherwise, for all $A \in \mathcal{P}(S)$. This is an injection, and a surjection as we can take $X_f = \{s \in S : f(s) = 1\}$ for all $f \in 2^S$, so $\varphi(X_f) = f$. ∎

**Proposition 1.A.10** (Cantor's Theorem). *For any set $S$, $\mathrm{Card}(S) < \mathrm{Card}(\mathcal{P}(S))$*

*Proof.* The inclusion $\iota : S \to \mathcal{P}(S)$ by $\iota(s) = \{s\}$ for all $s \in S$, is an injection so $\mathrm{Card}(S) \leqslant \mathrm{Card}(\mathcal{P}(S))$. Towards a contradiction we have $f : S \twoheadrightarrow \mathcal{P}(S)$. Let $B = \{s \in S : s \notin f(s)\}$. Then $f(s) = B$ for some $s \in S$. But then

$$s \notin B \iff s \notin f(s) \iff s \notin B$$

which is a contradiction, so no such $f$ can exist. ∎

Due to Cantor's theorem we obtain an infinite chain of infinite cardinals

$$\mathrm{Card}(\mathbb{N}) < \mathrm{Card}(\mathcal{P}(\mathbb{N})) < \mathrm{Card}(\mathcal{P}(\mathcal{P}(\mathbb{N}))) < ...$$

An important, but non-trivial result, is $\mathrm{Card}(\mathbb{R} \times \mathbb{R}) = \mathrm{Card}(\mathbb{R})$.

# Chapter 2

# Differentiation

## 2.1.0 Introduction to Derivatives

**Definition 2.1.1 (Differentiability).** *A function $f : \mathbb{R} \to \mathbb{R}$ is said to be* **differentiable at** *$a$ if*

$$\lim_{h \to 0} \frac{f(a+h) - f(a)}{h} \tag{2.1.1}$$

*exists. In this case the limit is denoted by $f'(a)$ and is called the* **derivative of $f$ at** *$a$. We also say that $f$ is* **differentiable** *if $f$ is differentiable at $a$ for all $a$ in its domain.*

**Definition 2.1.2.** *We define the* **tangent line** *to the graph of $f$ at $(a, f(a))$ to be the line through $(a, f(a))$ with slope $f'(a)$. That is, the tangent line at $(a, f(a))$ is well defined if and only if $f$ is differentiable at $a$.*

**Remark 2.1.1.** Given a function $f$, we denote by $f'$ the function whose domain is the set of all numbers $a \in \mathbb{R}$ such that $f$ is differentiable at $a$, and whose value at such a number $a$ is

$$\lim_{h \to 0} \frac{f(a+h) - f(a)}{h} \tag{2.1.2}$$

The function $f'$ is called the ***derivative*** of $f$.

**Notation 2.1.2.** For a given function $f : \mathbb{R} \to \mathbb{R}$, the derivative $f'$ is often denoted by

$$\frac{df(x)}{dx} \tag{2.1.3}$$

and the number $f'(a)$ is denoted by

$$\left.\frac{df(x)}{dx}\right|_{x=a} \tag{2.1.4}$$

**Theorem 2.1.1.** *If $f$ is differentiable at $a$, then $f$ is continuous at $a$.*

*Proof.* Suppose $f$ is differentiable at a point $a$. Then we have that the limit

$$\lim_{h \to 0} \frac{f(a+h) - f(a)}{h}$$

exists. It follows by 1.7.2 that

$$\begin{aligned}\lim_{h\to 0} f(a+h)-f(a) &= \lim_{h\to 0}\frac{f(a+h)-f(a)}{h}\cdot h\\ &= \lim_{h\to 0}\frac{f(a+h)-f(a)}{h}\cdot\lim_{h\to 0} h\\ &= f'(a)\cdot 0\\ &= 0\end{aligned}$$

Thus, by 1.7.2 the result that $\lim\limits_{h\to 0} f(a+h)-f(a) = 0$ is equivalent to $\lim\limits_{h\to 0} f(a+h) = \lim\limits_{h\to 0} f(a) = f(a)$. Thus, $f$ is continuous at $a$, replacing $a+h$ with $x$ and $h \to 0$ with $x \to a$. ■

**Definition 2.1.3 (Higher Order Derivatives).** *Since the derivative of a function $f$ is also a function, we can take its derivative to obtain the function $(f')' = f''$. In general, we denote the $k+1$-th derivative of $f$ inductively by*

$$\begin{aligned}f^{(1)} &= f'\\ f^{(k+1)} &= (f^{(k)})'\end{aligned}$$

*These are called* **higher order derivatives of $f$**. *We also define $f^{(0)} = f$. In Leibnitzian notation we write*

$$\frac{d^k f(x)}{dx} = f^{(k)} \tag{2.1.5}$$

## 2.2.0 Differentiation Results

**Theorem 2.2.1.** *If $f$ is a constant function, $f(x) = c$, then $f'(a) = 0$ for all $a \in \mathbb{R}$.*

*Proof.* Observe that for $a \in \mathbb{R}$,

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

as desired. ■

**Theorem 2.2.2.** *If $f$ is the identity function, $f(x) = x$, then $f'(a) = 1$ for all $a \in \mathbb{R}$.*

*Proof.* Observe that for $a \in \mathbb{R}$,

$$f'(a) = \lim_{h\to 0}\frac{f(a+h)-f(a)}{h} = \lim_{h\to 0}\frac{a+h-a}{h} = \lim_{h\to 0} 1 = 1$$

as desired. ■

**Theorem 2.2.3** (Linearity)**.** *If $f$ and $g$ are differentiable at $a$, then $f + cg$ is differentiable for all $c \in \mathbb{R}$*

*Proof.* Observe that

$$\begin{aligned}
(f+cg)'(a) &= \lim_{h\to 0} \frac{(f+cg)(a+h)-(f+cg)(a)}{h} \\
&= \lim_{h\to 0} \frac{f(a+h)+cg(a+h)-[f(a)+cg(a)]}{h} \\
&= \lim_{h\to 0} \frac{[f(a+h)-f(a)]+c[g(a+h)-g(a)]}{h} \\
&= \lim_{h\to 0} \left(\frac{f(a+h)-f(a)}{h}+c\frac{g(a+h)-g(a)}{h}\right) \\
&= \lim_{h\to 0} \frac{f(a+h)-f(a)}{h}+\lim_{h\to 0} c\frac{g(a+h)-g(a)}{h} \\
&= f'(a)+c\lim_{h\to 0} \frac{g(a+h)-g(a)}{h} \\
&= f'(a)+cg'(a)
\end{aligned}$$

as desired. ■

**Theorem 2.2.4** (Product Rule)**.** *If $f$ and $g$ are differentiable at $a$, then $f\cdot g$ is also differentiable at $a$ and*

$$(f\cdot g)'(a) = f'(a)\cdot g(a)+f(a)\cdot g'(a)$$

*Proof.* Observe that

$$\begin{aligned}
(f\cdot g)'(a) &= \lim_{h\to 0} \frac{(f\cdot g)(a+h)-(f\cdot g)(a)}{h} \\
&= \lim_{h\to 0} \frac{f(a+h)g(a+h)-f(a+h)g(a)+f(a+h)g(a)-f(a)g(a)}{h} \\
&= \lim_{h\to 0} \frac{f(a+h)[g(a+h)-g(a)]}{h}+\lim_{h\to 0} \frac{g(a)[f(a+h)-f(a)]}{h} \\
&= \lim_{h\to 0} f(a+h)\cdot \lim_{h\to 0} \frac{g(a+h)-g(a)}{h}+\lim_{h\to 0} \frac{f(a+h)-f(a)}{h}\cdot \lim_{h\to 0} g(a) \\
&= f(a)\cdot g'(a)+f'(a)\cdot g(a)
\end{aligned}$$

as claimed, where $\lim_{h\to 0} f(a+h) = f(a)$ since $f$ is differentiable at $a$, which implies it is also continuous at $a$. ■

**Theorem 2.2.5** (Power Rule)**.** *IF $f(x) = x^n$ for some natural number $n$, then*

$$f'(a) = na^{n-1}$$

*for all $a$.*

*Proof.* For the proof we will proceed by induction on $n$. For $n = 1$ we have shown that $f'(a) = 1 = 1\cdot a^0$, satisfying the base case. Assume that there exists $k\in\mathbb{N}$ such that if $n = k$, $f'(a) = ka^{k-1}$. Then, for the case of $n = k+1$ we may write $g(x) = x\cdot x^k = I(x)\cdot f(x)$. Hence, by the product rule we have that for all $a$

$$g'(a) = (I\cdot f)'(a)$$

$$
\begin{aligned}
&= I'(a)\cdot f(a) + I(a)\cdot f'(a)\\
&= 1\cdot a^k + a\cdot ka^{k-1}\\
&= (k+1)a^k
\end{aligned}
$$

as claimed. Hence, by mathematical induction we conclude that if $f(x) = x^n$ for $n \in \mathbb{N}$, then $f'(a) = na^{n-1}$ for all $a \in \mathbb{R}$. ∎

**Theorem 2.2.6** (Derivative of a Quotient). *If $g$ is differentiable at $a$, and $g(a) \neq 0$, then $1/g$ is differentiable at $a$ and*

$$\left(\frac{1}{g}\right)'(a) = \frac{-g'(a)}{|g(a)|^2}$$

*Proof.* Note that since $g$ is differentiable at $a$ it is continuous at $a$. Moreover, since $g(a) \neq 0$, there exists $\delta > 0$ such that $g(a+h) \neq 0$ for $|h| < \delta$. Therefore, $(1/g)(a+h)$ is well defined for small enough $h$, and we can write

$$
\begin{aligned}
\lim_{h\to 0}\frac{(1/g)(a+h)-(1/g)(a)}{h} &= \lim_{h\to 0}\frac{1/g(a+h)-1/g(a)}{h}\\
&= \lim_{h\to 0}\frac{g(a)-g(a+h)}{h[g(a)\cdot g(a+h)]}\\
&= \lim_{h\to 0}\frac{-[g(a+h)-g(a)]}{h}\cdot\lim_{h\to 0}\frac{1}{g(a)\cdot g(a+h)}\\
&= -g'(a)\cdot\frac{1}{|g(a)|^2}
\end{aligned}
$$

where $\lim_{h\to 0} 1/g(a+h) = 1/g(a)$ by continuity of $g$. ∎

**Theorem 2.2.7** (Quotient Rule). *If $f$ and $g$ are differentiable at $a$ and $g(a) \neq 0$, then $f/g$ is differentiable at $a$ and*

$$(f/g)'(a) = \frac{g(a)\cdot f'(a) - f(a)\cdot g'(a)}{|g(a)|^2}$$

*Proof.* Note that $f/g = f\cdot(1/g)$, so we have

$$
\begin{aligned}
(f/g)'(a) &= (f\cdot 1/g)'(a)\\
&= f'(a)\cdot(1/g)(a) + f(a)\cdot(1/g)'(a) && \text{(Product Rule)}\\
&= \frac{f'(a)}{g(a)} - \frac{f(a)g'(a)}{|g(a)|^2} && \text{(Quotient Derivative)}\\
&= \frac{f'(a)g(a) - f(a)g'(a)}{|g(a)|^2}
\end{aligned}
$$

as claimed. ∎

**Theorem 2.2.8** (General Product Rule). *If $f_1, f_2, \ldots, f_n$ are differentiable at $a$ for some $n \in \mathbb{N}$, then $f_1\cdot f_2\cdot\ldots\cdot f_n$ is differentiable at $a$ and*

$$(f_1\cdot\ldots\cdot f_n)'(a) = \sum_{i=1}^{n} f_1(a)\cdot\ldots\cdot f_{i-1}(a)\cdot f_i'(a)\cdot f_{i+1}(a)\cdot\ldots\cdot f_n(a)$$

*Proof.* We proceed by induction on $n$. If $n = 1$ then $f_1'(a) = f_1'(a)$, so the base case holds. Now, suppose the claim is true for some $k \in \mathbb{N}$. Then it follows that if $n = k + 1$

$$\begin{aligned}(f_1 \cdot ... \cdot f_k \cdot f_{k+1})'(a) &= (f_1 \cdot ... \cdot f_k)'(a) f_{k+1}(a) + (f_1 \cdot ... \cdot f_k)(a) f_{k+1}'(a) && \text{(Product Rule)}\\ &= \left[\sum_{i=1}^{k} f_1(a) \cdot ... \cdot f_{i-1}(a) \cdot f_i'(a) \cdot f_{i+1}(a) \cdot ... \cdot f_k(a)\right] f_{k+1}(a)\\ &+ f_1(a) \cdot ... \cdot f_k(a) \cdot f_{k+1}'(a) && \text{(by Induction Hypothesis)}\\ &= \sum_{i=1}^{k+1} f_1(a) \cdot ... \cdot f_{i-1}(a) \cdot f_i'(a) \cdot f_{i+1}(a) \cdot ... \cdot f_{k+1}(a)\end{aligned}$$

as desired. Thus by mathematical induction we conclude that the formula holds for all $n \in \mathbb{N}$. ∎

**Theorem 2.2.9** (Chain Rule). *If $g$ is differentiable at $a$ and $f$ is differentiable at $g(a)$, then $f \circ g$ is differentiable at $a$ and*

$$(f \circ g)'(a) = f'(g(a)) \cdot g'(a)$$

*Proof.* Define a function $\phi$ as follows:

$$\phi(h) = \begin{cases} \frac{f(g(a+h))-f(g(a))}{g(a+h)-g(a)}, & \text{if } g(a+h) - g(a) \neq 0 \\ f'(g(a)), & \text{if } g(a+h) - g(a) = 0 \end{cases} \tag{2.2.1}$$

Note that by differentiability of $g$ at $a$, $g$ is continuous at $a$ as well so as $h \to 0$, $g(a+h)-g(a) \to 0$, so if $g(a + h) - g(a)$ is not zero, then $\phi(h)$ will approach $f'(g(a))$ as $h$ goes to zero. If it is zero then $\phi(h)$ is exactly $f'(g(a))$. Note that as $f$ is differentiable at $g(a)$ we have

$$\lim_{k\to 0} \frac{f(g(a) + k) - f(g(a))}{k} = f'(g(a))$$

Thus, if $\epsilon > 0$ there is some number $\delta' > 0$ such that, for all $k$,

$$(1) \text{ if } 0 < |k| < \delta', \text{ then } \left|\frac{f(g(a) + k) - f(g(a))}{k} - f'(g(a))\right| < \epsilon$$

Now, $g$ is differentiable at $a$, hence continuous at $a$, so there is $\delta > 0$ such that for all $h$,

$$(2) \text{ if } |h| < \delta, \text{ then } |g(a + h) - g(a)| < \delta'$$

Consider now any $h$ with $|h| < \delta$. If $k = g(a + h) - g(a) \neq 0$, then

$$\phi(h) = \frac{f(g(a + h)) - f(g(a))}{g(a + h) - g(a)} = \frac{f(g(a) + k) - f(g(a))}{k}$$

it follows from $(2)$ that $|k| < \delta'$, and hence from $(1)$ that

$$|\phi(h) - f'(g(a))| < \epsilon$$

On the other hand, if $g(a + h) - g(a) = 0$, then $\phi(h) = f'(g(a))$, so it is surely true that

$$|\phi(h) - f'(g9a))| < \epsilon$$

We therefore have proved that

$$\lim_{h\to 0}\phi(h) = f'(g(a))$$

so $\phi$ is continuous at 0. If $h \neq 0$, then we have

$$\frac{f(g(a+h)) - f(g(a))}{h} = \phi(h) \cdot \frac{g(a+h) - g(a)}{h}$$

even if $g(a+h) - g(a) = 0$. Therefore, we have that

$$\begin{aligned}(f \circ g)'(a) &= \lim_{h\to 0}\frac{f(g(a+h)) - f(g(a))}{h} \\ &= \lim_{h\to 0}\phi(h) \cdot \lim_{h\to 0}\frac{g(a+h) - g(a)}{h} \\ &= f'(g(a)) \cdot g'(a)\end{aligned}$$

by continuity of $\phi(h)$ at 0. ∎

## 2.3.0 Applications of Derivatives

**Definition 2.3.1 (Extrema).** *Let $f$ be a function and $A$ a set of numbers contained in the domain of $f$. A point $x \in A$ is* **maximum point** *for $f$ on $A$ if*

$$f(x) \geqslant f(y) \forall y \in A \tag{2.3.1}$$

*The number $f(x)$ is itself called the* **maximum value** *of $f$ on $A$.*

*A point $x \in A$ is a* **minimum point** *for $f$ on $A$ if*

$$f(x) \leqslant f(y) \forall y \in A \tag{2.3.2}$$

*The number $f(x)$ is itself called the* **minimum value** *of $f$ on $A$.*

**Theorem 2.3.1.** *Let $f$ be any function defined on $(a, b)$. If $x$ is an extremum point for $f$ on $(a, b)$, and $f$ is differentiable at $x$, then $f'(x) = 0$.*

*Proof.* Consider the case where $f$ has a maximum at $x$. If $h$ is any number such that $x + h \in (a, b)$, then

$$f(x) \geqslant f(x+h)$$

since $f$ has a maximum on $(a, b)$ at $x$. This implies that

$$f(x+h) - f(x) \leqslant 0$$

Thus, if $h > 0$ we have that

$$\frac{f(x+h) - f(x)}{h} \leqslant 0$$

and consequently

$$\lim_{h\to 0^+}\frac{f(x+h) - f(x)}{h} \leqslant 0$$

as otherwise $\frac{f(x+h)-f(x)}{h} > 0$ for some $h$, contradicting our initial assumptions. Similarly, if $h < 0$ we have

$$\frac{f(x+h)-f(x)}{h} \geqslant 0$$

so

$$\lim_{h\to 0^-} \frac{f(x+h)-f(x)}{h} \geqslant 0$$

By hypothesis $f$ is differentiable at $x$, so these two limits must be equal, so in fact $f'(x) \leqslant 0$ and $f'(x) \geqslant 0$. Thus, $f'(x) = 0$.

On the other hand, suppose $f$ has a minimum at $x$. Then $-f$ has a maximum at $x$. Indeed, for all $y \in (a,b)$ we have $f(y) \geqslant f(x)$, so $-f(y) \leqslant -f(x)$. Then, from our above argument and the differentiability of $f$ at $x$, we have $-f'(x) = 0$, which implies that $f'(x) = 0$. ■

**Definition 2.3.2 (Local Extrema).** *Let $f$ be a function, and $A$ a set of numbers contained in the domain of $f$. A point $x$ in $A$ is a* **local maximum [minimum] point** *for $f$ on $A$ if there is some $\delta > 0$ such that $x$ is a maximum [minimum] point for $f$ on $A \cap (x-\delta, x+\delta)$.*

**Definition 2.3.3.** *A* **critical point** *of a function $f$ is a number $x$ such that*

$$f'(x) = 0 \tag{2.3.3}$$

*The number $f(x)$ itself is called a* **critical value** *of $f$.*

**Remark 2.3.1.** Give a function continuous $f$, if $x$ is an extrumum of $f$ on $[a,b]$, then one of the following must be satisfied:

1. $x$ is a critical point of $f$ in $[a,b]$
2. $x = a$ or $x = b$ so $x$ is an endpoint of $[a,b]$
3. $x$ is a point in $[a,b]$ such that $f$ is not differentiable at $x$

## Theorem 8 (Rolle's Theorem).

*If $f$ is continuous on $[a,b]$ and differentiable on $(a,b)$, and $f(a) = f(b)$, then there is a number $x \in (a,b)$ such that $f'(x) = 0$.*

*Proof.* It follows from continuity of $f$ on $[a,b]$ that $f$ has a maximum or minimum value on $[a,b]$ (by the Extreme Value Theorem).

Suppose first that the maximum value occurs at a point $x \in (a, b0$. Then $f'(x) = 0$ by Theorem 2.3.1. On the other hand suppose that the minimum value of $f$ occurs at some point $x$ in $(a,b)$. Then, again, $f'(x) = 0$ by Theorem 2.3.1.

Finally, suppose the maximum and minimum values both occur at the end points. Since $f(a) = f(b)$, the maximum and minimum values of $f$ are equal, so $f$ is a constant function, and for a constant function we can choose any $x \in (a,b)$ and have $f'(x) = 0$, completing the proof. ■

## Theorem 9 (The Mean Value Theorem).

*If $f$ is continuous on $[a,b]$ and differentiable on $(a,b)$, then there is a number $x \in (a,b)$ such that*

$$f'(x) = \frac{f(b) - f(a)}{b - a} \tag{2.3.4}$$

*Proof.* Let

$$h(x) = f(x) - \left[\frac{f(b) - f(a)}{b - a}\right](x - a)$$

Evidently, $h$ is continuous on $[a,b]$ and differentiable on $(a,b)$ as it is the sum of correspondingly continuous and differentiable functions. Moreover,

$$\begin{aligned} h(a) &= f(a) \\ h(b) &= f(b) - \left[\frac{f(b) - f(a)}{b - a}\right](b - a) \\ &= f(a) \end{aligned}$$

Consequently, we may apply 8 to $h$ and conclude that there exists $x \in (a,b)$ such that

$$0 = h'(x) = f'(x) - \frac{f(b) - f(a)}{b - a}$$

so that

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

as desired. ∎

**Corollary 2.3.2.** *If $f$ is defined on an interval and $f'(x) = 0$ for all $x$ in the interval, then $f$ is constant on the interval.*

*Proof.* Let $a$ and $b$ be any two points in the interval with $a \neq b$. Then there is some $x \in (a,b)$ such that

$$0 = f'(x) = \frac{f(b) - f(a)}{b - a}$$

so $f(b) - f(a) = 0$ and consequently $f(a) = f(b)$. Thus the value of $f$ at any two points in the interval is the same, so $f$ is constant on the interval. ∎

**Corollary 2.3.3.** *If $f$ and $g$ are defined on the same interval, and $f'(x) = g'(x)$ for all $x$ in the interval, then there is come number $c$ such that $f = g + c$.*

*Proof.* For all $x$ in the interval we have $(f - g)'(x) = f'(x) - g'(x) = 0$, so by the previous corollary there is some number $c$ such that $f - g = c$. ∎

**Definition 2.3.4.** *A function is **increasing** on an interval $I$ if $f(a) < f(b)$ whenever $a, b \in I$ with $a < b$. The function $f$ is **decreasing** on an interval $I$ if $f(a) > f(b)$ for all $a, b \in I$ with $a < b$.*

**Corollary 2.3.4.** *If $f'(x) > 0$ for all $x$ in an interval, then $f$ is increasing on the interval; if $f'(x) < 0$ for all $x$ in the interval, then $f$ is decreasing on the interval.*

*Proof.* Consider the case where $f'(x) > 0$. Let $a, b \in I$ with $a < b$. Then by 9 there exists $x \in (a, b)$ such that

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

But, $f'(x) > 0$ for all $x \in (a, b)$, so

$$\frac{f(b) - f(a)}{b - a} > 0$$

Since $b - a > 0$ we conclude that $f(b) > f(a)$ so $f$ is increasing.

Next, consider the case for $f'(x) < 0$. Then $-f'(x) > 0$ for all $x \in I$, so by the first case we have that for all $a, b \in I$ with $a < b$, $-f(a) < -f(b)$. Multiplying both sides by $-1$ we have that $f(a) > f(b)$ for all $a, b \in I$ such that $a < b$, so $f$ is decreasing, as desired. ∎

**Theorem 2.3.5** (Second Derivative Test)**.** *Suppose $f'(a) = 0$. If $f''(a) > 0$, then $f$ has a local minimum at $a$; if $f''(a) < 0$ then $f$ has a local maximum at $a$.*

*Proof.* By definition

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

Since $f'(a) = 0$ by assumption, we can write

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

Suppose now that $f''(a) > 0$. Then there exists $\delta > 0$ such that if $|h| < \delta$ $f'(a+h)/h > 0$. Thus, for $|h| < \delta$, if $h < 0$ we must have $f'(a+h) < 0$ and if $h > 0$ we must have $f'(a+h) > 0$. This means by our previous corollary that $f$ is increasing in the interval $(a, a+\delta)$, and decreasing in $(a - \delta, a)$. Thus, as $f'(a) = 0$, $f(a)$ must be a local minimum.

If $f''(a) < 0$, then $-f''(a) > 0$ so $-f(a)$ is must be a local minimum. That is, there exists $\delta > 0$ such that if $x \in (a - \delta, a + \delta)$, then $-f(x) \geqslant -f(a)$. Hence, it follows that $f(x) \leqslant f(a)$ for all $x \in (a - \delta, a + \delta)$, so $f(a)$ is a local maximum of $f$. ∎

**Theorem 2.3.6.** *Suppose $f''(a)$ exists. If $f$ has a local minimum at $a$, then $f''(a) \geqslant 0$; if $f$ has a local maximum at $a$, then $f''(a) \leqslant 0$.*

*Proof.* Suppose $f$ has a local minimum at $a$. If $f''(a) < 0$ then by our previous result $f$ would have a local maximum at $a$. But, this implies that $f$ would be constant in some interval containing $a$, so that $f''(a) = 0$, which is a contradiction. Thus, we must have that $f''(a) \geqslant 0$.

The case for a local maximum is analogous. ∎

**Theorem 2.3.7.** *Suppose that $f$ is continuous at $a$, and that $f'(x)$ exists for all $x$ in some interval containing $a$, except perhaps for $x = a$. Suppose, moreover, that $\lim_{x\to a} f'(x)$ exists. Then $f'(a)$ also exists and*

$$f'(a) = \lim_{x\to a} f'(x) \tag{2.3.5}$$

*Proof.* By definition

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

For sufficiently small $h > 0$ the function $f$ will be continuous on $[a, a+h]$, and differentiable on $(a, a+h)$, by assumption (similarly for sufficiently small $h < 0$). By 9 there is a number $\alpha_h \in (a, a+h)$ such that

$$\frac{f(a+h) - f(a)}{h} = f'(\alpha_h)$$

Now, $\alpha_h$ approaches $a$ as $h$ approaches 0, because $\alpha_h$ is in $(a, a+h)$. Since $\lim_{x\to a} f'(x)$ exists, it follows that

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

For this last equality write $\lim_{x\to a} f'(x) = L \in \mathbb{R}$. Fix $\epsilon > 0$. Then there exists $\delta > 0$ such that for all $x \in (a-\delta, a+\delta)$, $|f'(x) - L| < \epsilon$. It follows that for $|h| < \delta$, if $h > 0$ and $\alpha_h \in (a, a+h) \subset (a-\delta, a+\delta)$ we have $|f'(\alpha_h) - L| < \epsilon$ and if $h < 0$ and $\alpha_h \in (a+h, a) \subset (a-\delta, a+\delta)$, then $|f'(\alpha_h) - L| < \epsilon$. Thus, by definition we have that $\lim_{h\to 0^+} f'(\alpha_h) = \lim_{h\to 0^-} f'(\alpha_h) = L$, so in particular $\lim_{h\to 0} f'(\alpha_h) = L = \lim_{x\to a} f'(x)$, completing the proof. ■

## Theorem 10 (The Cauchy Mean Value Theorem).

*If $f$ and $g$ are continuous on $[a, b]$ and differentiable on $(a, b)$, then there is a number $x \in (a, b)$ such that*

$$[f(b) - f(a)]g'(x) = [g(b) - g(a)]f'(x) \tag{2.3.6}$$

*Proof.* Let

$$h(x) = f(x)[g(b) - g(a)] - g(x)[f(b) - f(a)]$$

Then $h$ is continuous on $[a, b]$, differentiable on $(a, b)$, and

$$h(a) = f(a)g(b) - g(a)f(b) = h(b)$$

It follows by 8 that $h'(x) = 0$ for some $x \in (a, b)$, which implies that

$$0 = h'(x) = f'(x)[g(b) - g(a)] - g'(x)[f(b) - f(a)]$$

completing the proof. ■

## Theorem 11 (L'Hôpital's Rule).

*Suppose that*

$$\lim_{x\to a} f(x) = 0 \text{ and } \lim_{x\to a} g(x) = 0 \tag{2.3.7}$$

*and suppose also that $\lim_{x\to a} f'(x)/g'(x)$ exists. Then $\lim_{x\to a} f(x)/g(x)$ exists, and*

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \lim_{x\to a} \frac{f'(x)}{g'(x)} \tag{2.3.8}$$

*Proof.* The hypothesis that $\lim_{x \to a} f'(x)/g'(x)$ exists contains two implicit assumptions:

1. there is an interval $(a - \delta, a + \delta)$ such that $f'(x)$ and $g'(x)$ exist for all $x \in (a - \delta, a + \delta)$, except, perhaps, $x = a$,
2. in this interval $g'(x) \neq 0$, with the possible exception of $x = a$

If we define $f(a) = g(a) = 0$, then $f$ and $g$ are continuous at $a$. If $x \in (a, a + \delta)$, then 9 and 10 apply to $f$ and $g$ on $[a, x]$ (a similar statement holds for $x \in (a - \delta, a)$). First, applying the 9 to $g$, we see that $g(x) \neq 0$, for if $g(x) = 0$ there would exist $x_1 \in (a, x)$ with $g'(x_1) = 0$, contradicting 2.. Now, applying 10 to $f$ and $g$, we see that there is a number $\alpha_x \in (a, x)$ such that

$$[f(x) - 0]g'(\alpha_x) = [g(x) - 0]f'(\alpha_x)$$

or

$$\frac{f(x)}{g(x)} = \frac{f'(\alpha_x)}{g'(\alpha_x)}$$

Now, let $\lim_{y \to a} f'(y)/g'(y) = L \in \mathbb{R}$. Fix $\epsilon > 0$. Then there exists $\delta' > 0$ such that if $y \in (a - \delta', a + \delta')$ then $|f'(y)/g'(y) - L| < \epsilon$. Then, for $x \in (a, a + \delta)$ (or $x \in (a - \delta, a)$) we have $(a, x) \subset (a - \delta, a + \delta)$ (or $(x, a) \subset (a - \delta, a + \delta)$). Thus, for $|x - a| < \delta$ we have $\alpha_x \in (a, x) \subset (a - \delta, a + \delta)$ (or $\alpha_x \in (x, a) \subset (a - \delta, a + \delta)$), so $|f'(\alpha_x)/g'(\alpha_x) - L| < \epsilon$. Therefore, we conclude that

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

so in particular

$$\lim_{x \to a} \frac{f(x)}{g(x)} = \lim_{x \to a} \frac{f'(\alpha_x)}{g'(\alpha_x)} = \lim_{y \to a} \frac{f'(y)}{g'(y)}$$

completing the proof. ∎

## Convexity

**Definition 2.3.5.** *A function $f$ is* **convex** *on an interval $I$, if for all $a, b \in I$, the line segment joining $(a, f(a))$ and $(b, f(b))$ lies above the graph of $f$.*

*This is equivalent to stating that for all $x \in (a, b)$,*

$$\frac{f(x) - f(a)}{x - a} < \frac{f(b) - f(a)}{b - a} \tag{2.3.9}$$

**Definition 2.3.6.** *A function $f$ is* **concave** *on an interval $I$, if for all $a, b \in I$, the line segment joining $(a, f(a))$ and $(b, f(b))$ lies below the graph of $f$.*

*This is equivalent to stating that for all $x \in (a, b)$,*

$$\frac{f(x) - f(a)}{x - a} > \frac{f(b) - f(a)}{b - a} \tag{2.3.10}$$

**Theorem 2.3.8.** *Let $f$ be convex. If $f$ is differentiable at $a$, then the graph of $f$ lies above the tangent line through $(a, f(a))$, except at $(a, f(a))$ itself. If $a < b$ and $f$ is differentiable at $a$ and $b$, then $f'(a) < f'(b)$.*

*Proof.* If $0 < h_1 < h_2$, then $a < a + h_1 < a + h_2$, and applying $f$'s convexity we have that

$$\frac{f(a+h_1) - f(a)}{h_1} < \frac{f(a+h_2) - f(a)}{h_2}$$

This implies that the values of $[f(a+h) - f(a)]/h$ decrease as $h \to 0^+$. Consequently,

$$f'(a) < \frac{f(a+h) - f(a)}{h}, h > 0$$

In fact, $f'(a)$ is the infimum of these numbers. Similarly, for $h$ negative, if $h_2 < h_1 < 0$, then

$$\frac{f(a+h_1) - f(a)}{h_1} > \frac{f(a+h_2) - f(a)}{h_2}$$

This shows that the slope of the tangent line is greater that $[f(a+h) - f(a)]/h$ for $h < 0$. In fact, $f'(a)$ is the supremum of all these numbers, so $f(a+h)$ lies above the tangent line if $h < 0$. This satisfies the first part of the theorem. Now, suppose $a < b$. Then we have that

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

since $b - a > 0$ and

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

since $a - b < 0$. Combining these inequalities we obtain $f'(a) < f'(b)$, as desired. ∎

**Lemma 2.3.9.** *Suppose $f$ is differentiable and $f'$ is increasing. If $a < b$ and $f(a) = f(b)$, then $f(x) < f(a) = f(b)$ for $a < x < b$.*

*Proof.* Suppose towards a contradiction that $f(x) \geqslant f(a) = f(b)$ for some $x \in (a, b)$. Then the maximum of $f$ on $[a, b]$ occurs at some point $x_0 \in (a, b)$ with $f(x_0) \geqslant f(a)$ and, of course, $f'(x_0) = 0$. On the other hand, applying 9 to the interval $[a, x_0]$, we find that there is $x_1$ with $a < x_1 < x_0$ and

$$f'(x_1) = \frac{f(x_0) - f(a)}{x_0 - a} \geqslant 0$$

contradicting the fact that $f'$ is increasing (since $f'(x_0) = 0$ and $x_1 < x_0$). ∎

**Theorem 2.3.10.** *If $f$ is differentiable and $f'$ is increasing, then $f$ is convex.*

*Proof.* Let $a < b$. Define $g$ by

$$g(x) = f(x) - \frac{f(b) - f(a)}{b-a}(x - a)$$

It is easy to see that $g'$ is also increasing; moreover, $g(a) = g(b) = f(a)$. Applying the lemma to $g$ we conclude that

$$a < x < b \implies g(x) < f(a)$$

In other words, if $a < x < b$, then

$$f(x) - \frac{f(b) - f(a)}{b - a}(x - a) < f(a)$$

or

$$\frac{f(x) - f(a)}{x - a} < \frac{f(b) - f(a)}{b - a}$$

Hence, $f$ is convex. ∎

**Theorem 2.3.11.** *If $f$ is differentiable and the graph of $f$ lies above each tangent line except at the point of contact, then $f$ is convex.*

*Proof.* Let $a < b$. Since the tangent lien at $(a, f(a))$ is the graph of the function

$$g(x) = f'(a)(x - a) + f(a)$$

and since $(b, f(b))$ lies above the tangent line, we have

$$(1)\ \ f(b) > f'(a)(b - a) + f(a)$$

Similarly, since the tangent line at $(b, f(b))$ is the graph of $h(x) = f'(b)(x - b) + f(b)$, and $(a, f(a))$ lies above the tangent line at $(b, f(b))$, we have

$$(2)\ \ f(a) > f'(b)(a - b) + f(b)$$

It follows from $(1)$ and $(2)$ that $f'(a) < f'(b)$. Then, from our previous theorem we have that $f$ is convex. ∎

## 2.4.0 Inverse Functions

**Definition 2.4.1.** *For any function $f$. the* **inverse image** *of $f$, denoted by $f^{-1}$, is the set of all pairs $(a, b)$ such that $(b, a) \in f$.*

**Remark 2.4.1.** $f^{-1}$ is a function if and only if $f$ is one-to-one.

**Theorem 2.4.1.** *If $f$ is increasing (decreasing) on an interval $I$, then $f$ is injective on $I$ so $f^{-1}$ is a function and in fact $f^{-1}$ is increasing (decreasing).*

*Proof.* Consider the case that $f$ is increasing. Then suppose $a, b \in I$ with $a \neq b$. Without loss of generality suppose $a < b$. Then since $f$ is increasing $f(a) < f(b)$ so in particular $f(a) \neq f(b)$. Therefore, $f$ is injective as claimed, so $f^{-1}$ is a well-defined function on $I$. Now, consider $a' < b'$ in $f(I) = I'$. Then there exist $x, y \in I$ such that $f(x) = a'$ and $f(y) = b'$, so in particular $f^{-1}(a') = x$ and $f^{-1}(b') = y$. Since $f$ is increasing and $f(x) = a' < b' = f(y)$ we must have that $x < y$. Thus, $f^{-1}(a') = x < y = f^{-1}(b')$, so $f^{-1}$ is increasing as claimed.

Consider the case that $f$ is decreasing. Then $-f$ is increasing so it is injective and $-f^{-1}$ is increasing by the first case. Hence, we have that $f^{-1}$ is decreasing as desired. ∎

**Theorem 2.4.2.** *If $f$ is continuous and one-to-one on an interval $I$, then $f$ is either increasing or decreasing on $I$.*

*Proof.* We proceed in three steps:

(1) If $a < b < c$ are three points in $I$, then I claim either $f(a) < f(b) < f(c)$ or $f(a) > f(b) > f(c)$. Indeed, suppose that $f(a) < f(c)$. If we have $f(b) < f(a)$, then the 6 applied to $[b, c]$ gives an $x \in (b, c)$ such that $f(x) = f(a)$, contradicting the fact that $f$ is injective on $[a, c]$. Similarly, if $f(b) > f(c)$ we would find a contradiction, so $f(a) < f(b) < f(c)$. Similar argumentation leads to the result that $f(a) > f(b) > f(c)$ in the second case.

(2) If $a < b < c < d$ are four points in $I$, then I claim that either $f(a) < f(b) < f(c) < f(d)$ or $f(a) > f(b) > f(c) > f(d)$. Indeed we can apply (1) to $a < b < c$ and then to $b < c < d$.

(3) Take any $a < b$ in $I$, and suppose $f(a) < f(b)$. Then $f$ is increasing, for if $c, d \in I$ are any two points, we can apply (2) to the collection $\{a, b, c, d\}$ after arranging them in increasing order. ∎

**Theorem 2.4.3.** *If $f$ is continuous and one-to-one on an interval, then $f^{-1}$ is also continuous.*

*Proof.* Since $f$ is continuous and injective on the interval, it is either increasing or decreasing. Consider the case that $f$ is increasing. We must show that

$$\lim_{x \to b} f^{-1}(x) = f^{-1}(b)$$

for each $b$ in the domain of $f^{-1}$. Such a number $b$ is of the form $f(a)$ for some $a$ in the domain of $f$. For any $\epsilon > 0$, we want to find a $\delta > 0$ such that for all $x$, if $x \in (f(a) - \delta, f(a) + \delta)$, then $|f^{-1}(x) - a| < \epsilon$, as $a = f^{-1}(b) = f^{-1}(f(a))$. Now, since $a - \epsilon < a < a + \epsilon$ we have that $f(a - \epsilon) < f(a) < f(a + \epsilon)$ since $f$ is presumed increasing. Let $\delta = \min(f(a + \epsilon) - f(a), f(a) - f(a - \epsilon))$. Our choice of $\delta$ ensures that

$$f(a - \epsilon) \leqslant f(a) - \delta \text{ and } f(a) + \delta \leqslant f(a + \epsilon)$$

Consequently, if

$$f(a) - \delta < x < f(a) + \delta$$

then

$$f(a - \epsilon) < x < f(a + \epsilon)$$

SInce $f$ is increasing, $f^{-1}$ is also increasing, and we obtain

$$f^{-1}(f(a - \epsilon)) < f^{-1}(x) < f^{-1}(f(a + \epsilon))$$

so $a - \epsilon < f^{-1}(x) < a + \epsilon$, which is precisely $|f^{-1}(x) - a| < \epsilon$, as desired. ∎

**Theorem 2.4.4.** *If $f$ is a continuous one-to-one function defined on an interval $I$, and $f'(f^{-1}(a)) = 0$, then $f^{-1}$ is not differentiable at $a$.*

*Proof.* We have $f(f^{-1}(x)) = x$. If $f^{-1}$ were differentiable at $a$, then the chain rule would imply that

$$f'(f^{-1}(a)) \cdot (f^{-1})'(a) = 1$$

hence

$$0 \cdot (f^{-1})'(a) = 1$$

which is impossible. ■

**Theorem 2.4.5.** *Let $f$ be a continuous one-to-one function defined on an interval $I$, and suppose that $f$ is differentiable at $f^{-1}(b)$, with derivative $f'(f^{-1}(b)) \neq 0$. Then $f^{-1}$ is differentiable at $b$, and*

$$(f^{-1})'(b) = \frac{1}{f'(f^{-1}(b))} \tag{2.4.1}$$

*Proof.* Let $b = f(a)$. Then

$$\lim_{h\to 0} \frac{f^{-1}(b+h) - f^{-1}(b)}{h} = \lim_{h\to 0} \frac{f^{-1}(b+h) - a}{h}$$

Now, every number $b+h$ in the domain of $f^{-1}$ can be written in the form $b+h = f(a+k)$ for a unique $k(h)$. Then

$$\begin{aligned} \lim_{h\to 0} \frac{f^{-1}(b+h) - a}{h} &= \lim_{h\to 0} \frac{f^{-1}(f(a+k(h))) - a}{f(a+k(h)) - b} \\ &= \lim_{h\to 0} \frac{k(h)}{f(a+k(h)) - f(a)} \end{aligned}$$

Since $b+h = f(a+k(h))$ we have $f^{-1}(b+h) = a + k(h)$, or $k(h) = f^{-1}(b+h) - f^{-1}(b)$. Now, since $f$ is continuous on $I$, $f^{-1}$ is also continuous on its domain, and in particular it is continuous at $b$. This means that $\lim_{h\to 0} k(h) = 0$, so $k(h)$ goes to zero as $h$ goes to 0. Hence, as

$$\lim_{k\to 0} \frac{f(a+k) - f(a)}{k} = f'(a) = f'(f^{-1}(b)) \neq 0$$

this implies that $f^{-1}$ is differentiable at $b$ and

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

■

## 2.A.0 Alternative Differentiation Formulation

**Definition 2.A.1.** *A function $f : (a,b) \subseteq \mathbb{R} \to \mathbb{R}$ (or $\mathbb{C}$) is said to be* **differentiable** *at $x \in (a,b)$ with derivative $f'(x)$ if the limit*

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

*exists.*

If $f$ is differentiable at $x \in (a,b)$, it is immediate that $f$ is continuous at $x$. Indeed, let $\varepsilon > 0$. Then there exists $\delta > 0$ such that if $0 < |h| < \delta$, $\left|\frac{f(x+h)-f(x)}{h} - f'(x)\right| < \varepsilon$. Then let $\delta' = \min\{\delta, |f'(x)| + \varepsilon\}$. Then for $x + h \in B_{\delta'}(x)$, $h \neq 0$, we ahve

$$|f(x+h) - f(x)| = |h| \left|\frac{f(x+h) - f(x)}{h}\right| < \delta'(|f'(x)| + \varepsilon) \leqslant \varepsilon$$

Now a useful equivalent condition to differentiability at $x \in (a,b)$ is the existence of a function $r(x,h)$ for $h$ close to 0 such that $f(x+h) = f(x) + Dh + r(x,h)$ such that $\frac{r(x,h)}{h} \to 0$ as $h \to 0$, and then $D = f'(x)$. Indeed if $f$ is differentiable set $r(x,h) = f(x+h) - (f(x) + f'(x)h)$. On the other hand, if such an $r(x,h)$ exists, then

$$\frac{f(x+h) - f(x)}{h} = \frac{r(x,h)}{h} + D \to D$$

so the limit exists and is $D$.

**Definition 2.A.2.** *We say that $f$ is differentiable on $(a,b)$ if it is differentiable at all $x \in (a,b)$.*

**Proposition 2.A.1.** *Let $f$ and $g$ be differentiable at $x$. Then*

- *$f \pm g$ is differentiable at $x$ with $(f \pm g)'(x) = f'(x) \pm g'(x)$ (additivity)*
- *$fg$ is differentiable at $x$ with $(fg)'(x) = f'(x)g(x) + f(x)g'(x)$ (Liebnitz's rule)*
- *If $g(y) \neq 0$ in a neighborhood of $x$, then $(1/g)$ is differentiable at $x$ with $(1/g)'(x) = -\frac{g'(x)}{g(x)^2}$*
- *For all $c \in \mathbb{R}$ $cf$ is differentiable at $x$ with $(cf)'(x) = cf'(x)$*

*Proof.* The first bullet is by linearity of limits. The second bullet follows from the computation

$$\frac{(fg)(x+h) - (fg)(x)}{h} = \frac{(f(x+h) - f(x))g(x+h)}{h} + \frac{f(x)(g(x+h) - g(x)}{h}$$

and the fact that $g$ is continuous at $x$ and the product of limits is the limit of the product, if the limits involved all exist. The quotient also follows by a similar computation

$$\frac{\frac{1}{g(x+h)} - \frac{1}{g(x)}}{h} = \frac{-(g(x+h) - g(x))}{hg(x+h)g(x)}$$

and the same properties of limits and $g$ at $x$. Finally, bullet follows simply by the calculation

$$\frac{cf(x+h) - cf(x)}{h} = c\frac{f(x+h) - f(x)}{h}$$

∎

Some basic derivatives from the definition are $f'(x) = 0$ if $f(x) = c \in \mathbb{R}$ is a constant, and later we shall show the converse also holds. Additionally, $id'(x) = 1$, for $id(x) = x$. Then by induction $\frac{d}{dx}x^n = nx^{n-1}$ for all $n = 0, 1, 2, 3, \ldots$.

**Proposition 2.A.2.** *If $f : (a,b) \to (\alpha,\beta)$ is differentiable at $x \in (a,b)$ and $g : (\alpha,\beta) \to \mathbb{R}$ is differentiable at $f(x) \in (\alpha,\beta)$, then $g \circ f$ is differentiable at $x$ with*

$$(g \circ f)'(x) = g'(f(x))f'(x)$$

*Proof.* We use the remainder definition of differentiability to prove the claim. As $f$ is differentiable at $x$ we have $r_f(x,h)$ such that $f(x+h) = f(x)+f'(x)h+r_f(x,h)$, and as $g$ is differentiable at $f(x)$ we have $r_g(x,h)$ such that $g(f(x)+h) = g(f(x)) + g'(f(x))h + r_g(f(x),h)$. Then

$$\begin{aligned} g \circ f(x+h) &= g(f(x) + f'(x)h + r_f(x,h)) \\ &= g(f(x)) + g'(f(x))(f'(x)h + r_f(x,h)) + r_g(f(x),\tilde{h}) \quad (\tilde{h} := f'(x)h + r_f(x,h)) \end{aligned}$$

Let $r_{g\circ f}(x,h) = g'(f(x))r_f(x,h) + r_g(f(x),\tilde{h})$. Then we have $\lim\limits_{h\to 0} g'(f(x))\frac{r_f(x,h)}{h} = 0$ and $\lim\limits_{h\to 0}\tilde{h} = 0$, so

$$\lim_{h\to 0}\frac{r_g(f(x),\tilde{h})}{h} = \lim_{h\to 0}\frac{r_g(f(x),\tilde{h})}{\tilde{h}}\frac{\tilde{h}}{h} = 0 \cdot f'(x) = 0$$

Thus, we have that $g \circ f$ is differentiable at $x$ with $(g \circ f)'(x) = g'(f(x))f'(x)$, as desired. ∎

**Proposition 2.A.3.** *If $f : (a,b) \to \mathbb{R}$ and $x \in (a,b)$ such that*

$$f(x) \geqslant f(y)(\text{or } f(x) \leqslant f(y))\forall y \in (a,b)$$

*then if $f$ is differentiable at $x$ it follows that*

$$f'(x) = 0$$

*Proof.* First, note that

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

But for $h$ such that $x+h \in (a,b)$, and $h > 0$, we have

$$\frac{f(x+h)-f(x)}{h} \leqslant 0(\text{respectively } \geqslant 0)$$

while if $h < 0$,

$$\frac{f(x+h)-f(x)}{h} \geqslant 0(\text{respectively } \leqslant 0)$$

Thus, by order properties of limits $f'(x) \leqslant 0$ and $f'(x) \geqslant 0$, so $f'(x) = 0$. ∎

Next we explore a funcdamental result for differentiable functions on an interval, which we will use to prove the fundamental theorem of calculus.

## Theorem 12 (Mean Value Theorem).

*Suppose $f$ is continuous on $[a,b]$ and is differentiable on $(a,b)$. Then there exists $\xi \in (a,b)$ such that*

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

*Proof.* Consider $f(x) = f(x) - (x-a)\frac{f(b)-f(a)}{b-a}$. Then $g$ is continuous on $[a,b]$, differentiable on $(a,b)$, and $g(a) = f(a)$ while $g(b) = f(a)$ as well, so $\frac{g(b)-g(a)}{b-a} = 0$. Since $g$ is continuous on the compact set $[a,b]$, $g$ attains a maximum on $[a,b]$. If it occurs at $\xi \in (a,b)$, then $g'(\xi) = 0$. On the other hand, if the maximum occurs at $a$ or $b$, then since $g(a) = g(b) = \max_{[a,b]} g$, $g$ must attain its minimum in $(a,b)$. Thus, there exists $\zeta \in (a,b)$ such that $g(\zeta)$ is a minimum and $g'(\zeta) = 0$. Thus in either case we have a $\xi \in (a,b)$ such that $f'(\xi) - \frac{f(b)-f(a)}{b-a} = 0$, so $f'(\xi) = \frac{f(b)-f(a)}{b-a}$ as desired. ∎

## Theorem 13 (Inverse Function Theorem).

*Suppose $f$ is continuous on $[a,b]$ and differentiable on $(a,b)$, and there exists $\gamma_0, \gamma_1 \in \mathbb{R}$ such that $0 < \gamma_0 \leqslant f'(x) \leqslant \gamma_1 < \infty$ (or $-\infty < \gamma_0 \leqslant f'(x) \leqslant \gamma_1 < 0$) for all $x \in (a,b)$. Let $\alpha = f(a)$ and $\beta = f(b)$. Then there exists an inverse function $g : [\alpha,\beta] \to [a,b]$ (or $g : [\beta,\alpha] \to [a,b]$) which is continuous on $[\alpha,\beta]$ and differentiable on $(\alpha,\beta)$, with derivative*

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

*for all $y \in (\alpha,\beta)$.*

*Proof.* Let $x_1, x_2 \in [a,b]$, with $a \leqslant x_1 < x_2 \leqslant b$. By the mean value theorem there exists $\xi \in (x_1,x_2)$ such that

$$f'(\xi) = \frac{f(x_2)-f(x_1)}{x_2-x_1}$$

This implies

$$0 < \gamma_0 \leqslant \frac{f(x_2)-f(x_1)}{x_2-x_2} \leqslant \gamma_1 < \infty$$

so

$$0 < \gamma_0(x_2-x_1) \leqslant f(x_2)-f(x_1) \leqslant \gamma_1(x_2-x_1)$$

for all $a \leqslant x_1 < x_2 \leqslant b$. This implies that $f$ is strictly increasing, and so injective. Further, $f(x) \in [\alpha,\beta]$ for all $x \in [a,b]$ since $\alpha = f(a)$ and $\beta = f(b)$. By the intermediate value theorem it follows that $f$ is surjective onto $[\alpha,\beta]$. Since $f : [a,b] \to [\alpha,\beta]$ is a continuous bijection on a compact set, $f$ is a homeomorphism and $f^{-1} : [\alpha,\beta] \to [a,b]$ is also a homeomorphism. As $f$ is differentiable on $(a,b)$, there exists $r_f(x,h)$ satisfying certain properties discussed previously. Then for $y \in (\alpha,\beta)$, there exists $x \in (a,b)$ such that $f^{-1}(y) = x$, so $f(x) = y$. Then

$$\begin{aligned} f^{-1}(y) + h = x + h = f^{-1}(f(x+h)) &= f^{-1}(f(x) + f'(x)h + r_f(x,h)) \\ &= f^{-1}(y + f'(x)h + r_f(x,h)) \end{aligned}$$

Let $\tilde{h} = f'(x)h + r_f(x,h)$, so $h = \frac{\tilde{h}}{f'(x)} - \frac{r_f(x,h)}{f'(x)}$. Then

$$f^{-1}(y+\tilde{h}) = f^{-1}(y) + \frac{1}{f'(x)}\tilde{h} - \frac{r_f(x,h)}{f'(x)}$$

Recall $r_f(x,h) = f(x+h) - f(x) - f'(x)h$. Then let $r_g(x,\tilde{h}) = -\frac{[f(x+h)-f(x)-f'(x)h]}{f'(x)}$. Note $\gamma_0 h + r_f(x,h) \leqslant \tilde{h} \leqslant \gamma_1 h + r_f(x,h)$, so

$$\gamma_0 + \frac{r_f(x,h)}{h} \leqslant \frac{\tilde{h}}{h} \leqslant \gamma_1 + \frac{r_f(x,h)}{h}$$

which implies $\frac{\tilde{h}}{h}$ and $\frac{h}{\tilde{h}}$ are bounded. Thus

$$\frac{r_g(x,\tilde{h})}{\tilde{h}} = -\frac{1}{f'(x)}\frac{h}{\tilde{h}}\frac{r_f(x,h)}{h}$$

which goes to 0 as $\tilde{h}$ goes to 0, since $\tilde{h} \to 0$ implies $h \to 0$. Thus, $g = f^{-1}$ is differentiable at $y = f(x)$, and

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

■

The result for a negative derivative follows by replacing $f$ with $-f$.

Note that if $f$ is differentiable in $(a,b)$ then $f'(x)$ is a function on $(a,b)$. If $f'(x)$ is differentiable at $x_0$ we may write

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

and in general $f^{(k+1)}(x) = \frac{d}{dx}f^{(k)}(x)$, for all $k \in \mathbb{N} \cup \{0\}$.

**Proposition 2.A.4.** *If $f$ is differentiable on $(a,b)$ and $x_0 \in (a,b)$, $f'(x_0) = 0$ but $f''(x_0) > 0$, then there exists $\delta > 0$:*

$$f(x_0) < f(x), \quad \forall x \in (x_0-\delta, x_0+\delta)\backslash\{x_0\}$$

*We say that $f$ has a local minimum at $x_0$.*

*Proof.* Note

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

Then there exists $\delta > 0$ such that $\frac{f'(x_0+h)-f'(x_0)}{h} > 0$ for all $h \in [-\delta,\delta]$. In particular, as $f'(x_0) = 0$, $f'(x_0+h) > 0$ for $h \in (0,\delta]$ and $f'(x_0+h) < 0$ for $h \in [-\delta, 0)$. Consider $h \in (0,\delta]$. By the mean value theorem there exists $c \in (x_0, x_0+h)$ such that

$$f(x_0+h) - f(x_0) = hf'(c) > 0$$

so $f(x_0) < f(x_0+h)$. If $h \in [-\delta, 0)$, by the mean value theorem there exists $c \in (x_0+h, x_0)$ such that

$$f(x_0+h) - f(x_0) = hf'(c) < 0$$

so $f(x_0) < f(x_0+h)$ again. Thus, for all $x \in (x_0-\delta, x_0+\delta)\backslash\{x_0\}$, we have $f(x_0) < f(x)$. ■

The result for a local maximum is then given by replacing $f$ by $-f$ in the previous result.

**Proposition 2.A.5.** *Suppose $f$ is twice differentiable on $(a,b)$ and $f''(x) > 0$ on $(a,b)$. Then for all $a < x_0 < x_1 < b$ and $\lambda \in (0,1)$,*

$$f(\lambda x_0 + (1-\lambda)x_1) < \lambda f(x_0) + (1-\lambda)f(x_1)$$

*Proof.* Let $g(s) = sf(x_0) + (1-s)f(x_1) - f(sx_0 + (1-s)x_1)$. Then $g(0) = f(x_1) - f(x_1) = 0$ and $g(1) = f(x_0) - f(x_0) = 0$. Note $g$ is also twice differentiable. Towards a contradiction suppose there exists $c \in (0,1)$ such that $g(c) < 0$. Since $g$ is continuous it attains its minimum, so there exists $s_0 \in (0,1)$ such that $g(s_0) \leqslant g(c) < 0$. Further, $g'(s_0) = 0$, where $g'(s) = f(x_0) - f(x_1) - f'(sx_0 + (1-s)x_1)(x_0 - x_1)$. Then

$$f'(s_0x_0 + (1-s_0)x_1) = \frac{f(x_1) - f(x_0)}{x_1 - x_0}$$

and $g''(s) = -f''(sx_0 + (1-s)x_0)(x_0 - x_1)^2 < 0$. In particular, $g''(s_0) < 0$, contradicting the fact that $g(s_0)$ is a minimum and Proposition 2.A.4 ■

# Chapter 3

# Integration

## 3.1.0 Introduction to Definite Integrals

**Definition 3.1.1.** *Let $a < b$. A* **partition** *of the interval $I = [a, b]$ is a finite collection of points in $[a, b]$, one of which is $a$, and one of which is $b$. Equivalently, a partition of $I$ into $N$ subintervals consists pf endpoints $x_0 = a < x_1 < ... < x_N = b$ with subinterval $k$ being $J_k = [x_k, x_{k+1}]$ for $0 \leqslant k \leqslant N-1$.*

The points in a partition can be numbered $t_0, ..., t_n$ so that

$$a = t_0 < t_1 < ... < t_{n-1} < t_n = b \tag{3.1.1}$$

we shall always assume that such a numbering has been assigned.

**Definition 3.1.2.** *Suppose $f : I \to \mathbb{R}$ is bounded on $I = [a, b]$ so there exists $M \in \mathbb{R}^+$ such that $||f||_I \leqslant M$. Let $P = \{t_0, ..., t_N\} \in \prod([a, b])$ be a partition of $[a, b]$. The* **lower Riemann sum** *of $f$ for $P$, denoted $L(f, P)$, is defined as*

$$L(f, P) := \sum_{k=0}^{N-1} \inf_{J_k}(f)\ell(J_k) \tag{3.1.2}$$

*The* **upper Riemann sum** *of $f$ for $P$, denoted $U(f, P)$, is defined as*

$$U(f, P) = \sum_{i=0}^{N-1} \sup_{J_k}(f)\ell(J_k) \tag{3.1.3}$$

*where $\ell(J_k) = t_{k+1} - t_k > 0$. Note*

$$-M \leqslant \sup_{J_k}(f) \leqslant M$$

*and*

$$-M \leqslant \inf_{J_k}(f) \leqslant M$$

**Remark 3.1.1.** If $P$ is any partition, then

$$L(f,P) \leqslant U(f,P) \tag{3.1.4}$$

because

$$L(f,P) = \sum_{k=0}^{N-1} \inf_{J_k}(f)\ell(J_k)$$
$$U(f,P) = \sum_{i=0}^{N-1} \sup_{J_k}(f)\ell(J_k)$$

and for each $i$ we have $\inf_{J_k}(f)\ell(J_k) \leqslant \sup_{J_k}(f)\ell(J_k)$.

**Definition 3.1.3.** *Given two partitions $P$ adn $Q$ of $I$, we say $P$ is a* **refinement** *of $Q$ if every endpoint in $Q$ belongs to $P$, and we write*

$$P \succ Q$$

**Lemma 3.1.1.** *If $P$ is a partition of $[a,b]$ which contains $Q$, that is $P \succ Q$, then*

$$L(f,Q) \leqslant L(f,P) \text{ and } U(f,Q) \qquad\qquad \geqslant U(f,P)$$

*Proof.* Consider first the special case in which $Q$ contains just one more point than $P$;

$$P = \{t_0, ..., t_n\}$$
$$Q = \{t_0, ..., t_{k-1}, u, t_k, ..., t_n\}$$

where

$$a = t_0 < t_1 < ... < t_{k-1} < u < t_k < ... < t_n = b$$

Let

$$m' = \inf\{f(x) : t_{k-1} \leqslant x \leqslant u\}$$
$$m'' = \inf\{f(x) : u \leqslant x \leqslant t_k\}$$

Then

$$L(f,P) = \sum_{i=1}^{n} m_i(t_i - t_{i-1})$$
$$L(f,Q) = \sum_{i=1}^{k-1} m_i(t_i - t_{i-1}) + m'(u - t_{k-1}) + m''(t_k - u) + \sum_{i=k+1}^{n} m_i(t_i - t_{i-1})$$

To prove that $L(f,P) \leqslant L(f,Q)$ it therefore suffices to show that

$$m_k(t_k - t_{k-1}) \leqslant m'(u - t_{k-1}) + m''(t_k - u)$$

Now, the set $\{f(x) : t_{k-1} \leqslant x \leqslant t_k\}$ contains all the numbers in $\{f(x) : t_{k-1} \leqslant x \leqslant u\}$ and possibly some smaller ones, so the greatest lower bound of the first set is less than or equal to the greatest lower bound of the second; thus

$$m_k \leqslant m'$$

Similarly,

$$m_k \leqslant m''$$

Therefore,

$$m_k(t_k - t_{k-1}) = m_k(t_k - u) + m_k(u - t_{k-1}) \leqslant m''(t_k - u) + m'(u - t_{k-1})$$

This proves, in this special case that $L(f,P) \leqslant L(f,Q)$. Now, let

$$\begin{aligned} M' &= \sup\{f(x) : t_{k-1} \leqslant x \leqslant u\} \\ M'' &= \sup\{f(x) : u \leqslant x \leqslant t_k\} \end{aligned}$$

Then

$$\begin{aligned} U(f,P) &= \sum_{i=1}^{n} M_i(t_i - t_{i-1}) \\ U(f,Q) &= \sum_{i=1}^{k-1} M_i(t_i - t_{i-1}) + M'(u - t_{k-1}) + M''(t_k - u) + \sum_{i=k+1}^{n} M_i(t_i - t_{i-1}) \end{aligned}$$

Hence, to prove that $U(f,Q) \leqslant U(f,P)$ it suffices to show that

$$M'(u - t_{k-1}) + M''(t_k - u) \leqslant M_k(t_k - t_{k-1})$$

As before, the set $\{f(x) : t_{k-1} \leqslant x \leqslant t_k\}$ contains all the numbers in $\{f(x) : t_{k-1} \leqslant x \leqslant u\}$ and possibly some larger ones, so the smallest upper bound of the first set is greater than or equal to the smallest upper bound of the second; thus

$$M_k \geqslant M'$$

Similarly,

$$M_k \geqslant M''$$

Therefore,

$$M_k(t_k - t_{k-1}) = M_k(t_k - u) + M_k(u - t_{k-1}) \geqslant M''(t_k - u) + M'(u - t_{k-1})$$

This proves, in this special case that $U(f,P) \geqslant U(f,Q)$.

The general case can now be deduced quite easily. The partition $Q$ can be obtained from $P$ by adding one point at a time; in otherwords, there is a sequence of partition

$$P = P_1 \subsetneq P_2 \subsetneq P_3 \subsetneq ... \subsetneq P_\alpha = Q$$

such that $P_{j+1} = P_j \cup \{u_{j+1}\}$ for some $u_{j+1} \in [a,b] - P_j$. Then

$$L(f,P) = L(f,P_1) \leqslant L(f,P_2) \leqslant ... \leqslant L(f,P_\alpha) = L(f,Q)$$

and

$$U(f,P) = U(f,P_1) \geqslant U(f,P_2) \geqslant ... \geqslant U(f,P_\alpha) = U(f,Q)$$

completing the proof. ∎

**Theorem 3.1.2.** *Let $P_1$ and $P_2$ be partitions of $[a,b]$, and let $f$ be a function which is bounded on $[a,b]$. Then*

$$L(f,P_1) \leqslant U(f,P_2) \tag{3.1.5}$$

*Proof.* There is a partition $P$ which contains both $P_1$ and $P_2$ (let $P = P_1 \cup P_2$). According to the lemma

$$L(f,P_1) \leqslant L(f,P) \leqslant U(f,P) \leqslant U(f,P_1)$$

■

**Definition 3.1.4.** *Let $\prod(I)$ denote the set of all partitions of $I$. We define the* **upper** *and* **lower Riemann integrals** *by*

$$U_I(f) = \inf_{P\in\prod(I)} U(f,P) \geqslant L_I(f) = \sup_{P\in\prod(I)} L(f,P)$$

**Definition 3.1.5 (Definite Integral).** *A function $f$ which is bounded on $[a,b]$ is said to be* **Riemann integrable** *on $[a,b]$ if and only if*

$$L_I(f) = U_I(f)$$

*In this case, this common number is called the* **integral** *of $f$ on $[a,b]$ and is denoted by*

$$\int_I f = \int_a^b f(x)dx = L_I(f) = U_I(f) \tag{3.1.6}$$

*The integral $\int_I f$ is also called the* **area** *of $R(f,a,b)$ when $f(x) \geqslant 0$ for all $x \in [a,b]$.*

**Theorem 3.1.3.** *If $f$ is bounded on $[a,b]$, then $f$ is integrable on $[a,b]$ if and only if for every $\epsilon > 0$ there is a partition $P$ of $[a,b]$ such that*

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

*Proof.* Suppose first that for every $\epsilon > 0$ there is such a partition $P$. Since

$$U_I(f) \leqslant U(f,P)$$
$$L_I(f) \geqslant L(f,P)$$

it follows that

$$U_I(f) - L_I(f) \leqslant U(f,P) - L(f,P) < \epsilon$$

Since this is true for all $\epsilon > 0$, it follows that

$$U_I(f) = L_I(f)$$

so by definition, then, $f$ is integrable. Conversely, if $f$ is integrable then

$$U_I(f) = L_I(f)$$

Let $M$ denote the value of this. Then for each $\epsilon > 0$ there exist partitions $P'$ and $P''$ such that $|U(f,P') - M| < \epsilon/2$ and $|L(f,P'') - M| < \epsilon/2$. Then as $U(f,P') \geqslant L(f,P'')$ from the previous theorem, we have that

$$U(f,P') - L(f,P'') = |U(f,P') - L(f,P'')| \leqslant |U(f,P') - M| + |M - L(f,P'')| < \epsilon$$

Let $P = P' \cup P''$ be a common refinement. Then, according to the lemma $U(f, P) \leqslant U(f, P')$ and $L(f, P) \geqslant L(f, P'')$ so

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

■

**Example 3.1.1.** Define $f(x) = \begin{cases} 1 & x \in \mathbb{Q} \cap [0,1] \\ 0 & x \notin \mathbb{Q} \cap [0,1] \end{cases}$ Since both $S = \mathbb{Q} \cap [0,1]$ and $[0,1] \backslash S$ are dense in $[0,1]$, for any subinterval $J_k$ we have $\sup_{J_k}(f) = 1$ and $\inf_{J_k}(f) = 0$ so $L_I(f) = 0 < 1 = U_I(f)$. Thus $f$ is not Riemann integrable, but it is Lebesque integrable since $\mathbb{Q}$ is a set of measure 0.

**Theorem 3.1.4.** *If $f \in C(I)$ is continuous on $I = [a, b]$, then $f \in \mathcal{R}(I)$ is integrable on $I = [a, b]$.*

*Proof.* Notice, first, that $f$ is bounded on $[a, b]$, because it is continuous on $[a, b]$. To prove that $f$ is integrable on $[a, b]$, we want to use our previous theorem, and show that for every $\epsilon > 0$ there is a partition $P$ of $[a, b]$ such that

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

Now we know, by our result on uniform continuity, that $f$ is uniformly continuous on $[a, b]$. So there is some $\delta > 0$ such that for all $x, y \in [a, b]$, if $|x-y| < \delta$, then $|f(x)-f(y)| < \epsilon/[2(b-a)]$. We choose a partition $P = \{t_0, ..., t_N\}$ such that each $|t_{k+1} - t_k| < \delta$. Then for each subinterval $J_k$ we have

$$|f(x) - f(y)| < \frac{\epsilon}{2(b-a)}$$

for all $x, y \in J_k$. Then, $f(x) < \frac{\epsilon}{2(b-a)} + f(y)$ for all $x \in J_k$, so $\sup_{J_k}(f) \leqslant \frac{\epsilon}{2(b-a)} + f(y)$. Further, $f(y) \geqslant \sup_{J_k}(f) - \frac{\varepsilon}{2(b-a)}$ for all $y \in J_k$, so $\inf_{J_k}(f) \geqslant \sup_{J_k}(f) - \frac{\varepsilon}{2(b-a)}$, so $\sup_{J_k}(f) - \inf_{J_k}(f) \leqslant \frac{\varepsilon}{2(b-a)}$. Since this is true for all $k$, we have that

$$\begin{aligned} U(f, P) - L(f, P) &= \sum_{k=0}^{N-1} (\sup_{J_k}(f) - \inf_{J_k}(f)) \ell(J_k) \\ &< \frac{\epsilon}{b-a} \sum_{k=0}^{N-1} \ell(J_k) \\ &= \frac{\epsilon}{b-a} \ell(I) \\ &= \epsilon \end{aligned}$$

Thus, by our previous theorem $f$ is integrable. ■

**Proposition 3.1.5.** *$\mathcal{R}(I)$ is a $\mathbb{R}$-vector space. If $f, g \in \mathcal{R}(I), a \in \mathbb{R}$, then $af + g \in \mathcal{R}(I)$, and*

$$\int_I (af + g) = a \int_I f + \int_I g$$

*Proof.* First we do additivity. Let $J_k \subseteq I$. Then $\sup_{J_k}(f+g) \leqslant \sup_{J_k}(f) + \sup_{J_k}(g)$ and $\inf_{J_k}(f+g) \geqslant \inf_{J_k}(f) + \inf_{J_k}(f)$. So it follows that $U(f+g,P) \leqslant U(f,P) + U(g,P)$ for any $P \in \prod(I)$ and $L(f+g,P) \geqslant L(f,P) + L(g,P)$. Then we have

$$L_I(f) + L_I(g) \leqslant L_I(f+g) \leqslant U_I(f+g) \leqslant U_I(f) + U_I(g)$$

where by assumption $L_I(f) = U_I(f)$ and $L_I(g) = U_I(g)$, so $U_I(f+g) = L_I(f+g)$, and $f+g \in \mathcal{R}(I)$. Further, $\int_I (f+g) = U_I(f+g) = U_I(f) + U_I(g) = \int_I f + \int_I g$.

Next, let $a \in \mathbb{R}$. If $a = 0$, then $af = 0$ so $U(af,P) = 0 = L(af,P)$ for all partitions $P$ of $[a,b]$, and so $U_I(af) = 0 = L_I(af)$. Thus, $af \in \mathcal{R}(I)$, and $\int_I af = 0 = a\int_I f$. Next, if $a > 0$, then $U(af,P) = aU(f,p)$ and $L(af,P) = aL(f,P)$. Then $U_I(af) = aU_I(f) = aL_I(f) = L_I(af)$, and we have our desired result. On the other hand, if $a < 0$, $U(af,P) = aL(f,P)$, and $L(af,P) = aU(f,P)$. Further, $U_I(af) = aL_I(f) = aU_I(f) = L_I(af)$, so we again obtain our desired result. ■

**Theorem 3.1.6.** *Let $a < c < b$. If $f$ is integrable on $[a,b]$, then $f$ is integrable on $[a,c]$ and one $[c,b]$. Conversely, if $f$ is integrable on $[a,c]$ and on $[c,b]$, then $f$ is integrable on $[a.b]$. Finally, if $f$ is integrable on $[a,b]$, then*

$$\int_a^b f = \int_a^c f + \int_c^b f \tag{3.1.7}$$

*Proof.* (1) Suppose $f$ is integrable on $[a,b]$. Then $f$ is bounded on $[a,b]$, so it is bounded on $[a,c]$ and $[c,b]$. Indeed, $f$ being bounded implies that there exists $M \in \mathbb{R}$ such that for all $x \in [a,b]$ $|f(x)| \leqslant M$. Thus, as this applies for all $x \in [a,b]$ and $[a,c],[c,b] \subset [a,b]$, we have that it holds for all $x \in [a,c]$ and all $x \in [c,b]$. Now fix $\epsilon > 0$. Then there exists a partition $P$ of $[a,b]$ such that

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

Without loss of generality suppose $c = t_j$ for some $t_j \in P = \{t_0, t_1, ..., t_n\}$. Then we have partitions $P' = \{t_0, ..., t_j\}$ and $P'' = \{t_j, ..., t_n\}$ for $[a,c]$ and $[c,b]$ respectively. Moreover,

$$U(f,P) = U(f,P') + U(f,P'')$$
$$L(f,P) = L(f,P') + L(f,P'')$$

Hence, we have that

$$[U(f,P') - L(f,P')] + [U(f,P'') - L(f,P'')] = U(f,P) - L(f,P) < \epsilon$$

But $U(f,P') \geqslant L(f,P')$ and $U(f,P'') \geqslant L(f,P'')$, so

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

Therefore, $f$ is integrable on $[a,c]$ and $[c,b]$

(2) Suppose $f$ is integrable on $[a,c]$ and $[c,b]$. Thus, there exists $M_1, M_2 \in \mathbb{R}$ such that for all $x \in [a,c]$ $|f(x)| \leqslant M_1$ and for all $x \in [c,b]$ $|f(x)| \leqslant M_2$. Let $M = \max(M_1, M_2)$. Then

for all $x \in [a,b]$ we have $|f(x)| \leqslant M$, so $f$ is bounded on $[a,b]$. Let $\epsilon > 0$. Then there exist partitions $P_1, P_2$ of $[a,c]$ and $[c,b]$ respectively such that

$$U(f,P_1) - L(f,P_1) < \epsilon/2$$
$$U(f,P_2) - L(f,P_2) < \epsilon/2$$

Let $P = P_1 \cup P_2$, where $P_1 \cap P_2 = \{c\}$. Then we have that

$$U(f,P) - L(f,P) = [U(f,P_1) - L(f,P_1)] + [U(f,P_2) - L(f,P_2)] < \epsilon/2 + \epsilon/2 = \epsilon$$

Therefore, by definition $f$ is integrable on $[a,b]$.

(3) Suppose $f$ is integrable on $[a,b]$, so by the previous results $f$ is integrable on $[a,c]$ and $[c,b]$. Let $\int_a^b f = R$, $\int_a^c f = R_1$, and $\int_c^b f = R_2$. Let $P$ be a partition of $[a,b]$, and without loss of generality suppose $c \in P = \{t_0, ..., t_j = c, ..., t_n\}$. Then let $P_1 = \{t_0, ..., t_j\}$ and $P_2 = \{t_j, ..., t_n\}$ be partitions of $[a,c]$ and $[c,b]$. It then follows that

$$L(f,P_1) \leqslant R_1 \leqslant U(f,P_1)$$
$$L(f,P_2) \leqslant R_2 \leqslant U(f,P_2)$$

Hence, we have that

$$L(f,P) = L(f,P_1) + L(f,P_2) \leqslant R_1 + R_2$$
$$U(f,P) = U(f,P_1) + U(f,P_2) \geqslant R_1 + R_2$$

Thus $L(f,P) \leqslant R_1 + R_2 \leqslant U(f,P)$. Note that this holds for all partitions $P$, as if $P'$ is a partition, then considering the partition $P'_c = P' \cup \{c\}$ we have that

$$L(f,P') \leqslant L(f,P'_c) \leqslant R_1 + R_2 \leqslant U(f,P'_c) \leqslant U(f,P')$$

Therefore, this holds for all partitions of $[a,b]$, but $R$ is the unique number which does this so we must have that $R = R_1 + R_2$. Thus

$$\int_a^b f = \int_a^c f + \int_c^b f$$

∎

**Definition 3.1.6.** *Using the previous theorem, we define*

$$\int_a^a f := 0 \quad and \quad \int_a^b f := -\int_b^a f, \; for\ a > b \tag{3.1.8}$$

**Theorem 3.1.7.** *Suppose $f$ is integrable on $[a,b]$ and that*

$$m \leqslant f(x) \leqslant M \forall x \in [a,b] \tag{3.1.9}$$

*Then*

$$m(b-a) \leqslant \int_a^b f \leqslant M(b-a) \tag{3.1.10}$$

*Proof.* It is clear that $M \geqslant \sup\{f(x) : x \in [a,b]\}$ and $m \leqslant \inf\{f(x) : x \in [a,b]\}$, so

$$m(b-a) \leqslant L(f,P) \text{ and } M(b-a) \geqslant U(f,P)$$

for every partition $P$. Since $\int_a^b f = \sup\{L(f,P)\} = \inf\{U(f,P)\}$ we have that

$$m(b-a) \leqslant \sup\{L(f,P)\} = \int_a^b f = \inf\{U(f,P)\} \leqslant M(b-a)$$

∎

**Remark 3.1.2.** If $f$ is integrable on $[a,b]$, we can define a new function $F$ on $[a,b]$ by

$$F(x) = \int_a^x f = \int_a^x f(t)dt \tag{3.1.11}$$

**Theorem 3.1.8.** *If $f$ is integrable on $[a,b]$ and $F$ is defined on $[a,b]$ by*

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

*then $F$ is continuous on $[a,b]$.*

*Proof.* Suppose $c \in [a,b]$. Since $f$ is integrable on $[a,b]$ it is, by definition, bounded on $[a,b]$; let $M$ be a number such that

$$|f(x)| \leqslant M, \forall x \in [a,b]$$

If $h > 0$ for $c+h \in [a,b]$, then

$$F(c+h) - F(c) = \int_a^{c+h} f - \int_a^c f = \int_c^{c+h} f$$

Since $-M \leqslant f(x) \leqslant M$ for all $x \in [a,b]$ it follows from the Theorem 3.1.7 that

$$-M \cdot h \leqslant \int_c^{c+h} f \leqslant M \cdot h$$

In other words

$$-M \cdot h \leqslant F(c+h) - F(c) \leqslant M \cdot h$$

If $h < 0$, with $c+h \in [a,b]$ we find the inequality

$$-M \cdot h \geqslant F(c+h) - F(c) \geqslant M \cdot h$$

In either case we have that

$$|F(c+h) - F(c)| \leqslant M \cdot |h|$$

Therefore, if $\epsilon > 0$, we have

$$|F(c+h) - F(c)| < \epsilon$$

provided that $|h| < \epsilon/M$. This proves that

$$\lim_{h \to 0} F(c+h) = F(c)$$

so $F$ is continuous at $c$. ∎

## 3.2.0 Reimann Sums

**Definition 3.2.1.** *Suppose $P = \{t_0, ..., t_N\}$ is a partition of $[a, b]$, and for each $k$ choose $\xi_k \in J_k$. Then we clearly have that*

$$L(f, P) \leqslant \sum_{k=0}^{N-1} f(\xi_k)\ell(J_k) \leqslant U(f, P) \tag{3.2.1}$$

*Any sum of the form*

$$\sum_{k=0}^{N-1} f(\xi_k)\ell(J_k) \tag{3.2.2}$$

*is called a* **Reimann sum** *of $f$ for $P$.*

Before proving an important result related to Riemann sums and definite integrals, we handle the following proposition:

**Proposition 3.2.1.** *Suppose $f : [a, b] \to \mathbb{R}$ is bounded by $M \in \mathbb{R}^+$. If $P$ and $Q$ are partitions of $[a, b]$ such that for some $k > 1$*

$$maxsize(P) \leqslant \frac{minsize(Q)}{k}$$

*then*

$$U(f, P) \leqslant U(f, Q) + \frac{2M\ell([a, b])}{k}$$

*and*

$$L(f, P) \geqslant L(f, Q) - \frac{2M\ell([a, b])}{k}$$

*Proof.* Let $P_1 = P \cup Q$ be the common refinement of $P$ and $Q$. The intervals in $P$, denoted $\tilde{J}_i$, can be separated into two classes,

(i) $\tilde{J}_i \subseteq J_k$ for some $k$, where $J_k$ are the intervals in $Q$

(ii) $\tilde{J}_i \not\subseteq J_k$ for all $k$

Note that in the case of (i), $\tilde{J}_i$ also belongs to $P_1$. In the second case, $\tilde{J}_i$ is split into two intervals in $P_1$. Then

$$|U(f, P) - U(f, P_1)| \leqslant \sum_{\text{intervals in (ii)}} |\sup_{\tilde{J}_i}(f)\ell(\tilde{J}_i) - (\sup_{\tilde{J}_i^+}(f)\ell(\tilde{J}_i^+) + \sup_{\tilde{J}_i^-}(f)\ell(\tilde{J}_i^-))|$$

where the intervals in class (i) cancel since they belong to both partitions. If $\tilde{J}_i$ is in class (ii), then $\tilde{J}_i = \tilde{J}_i^+ \cup \tilde{J}_i^-$ in $P_1$. Then proceeding with the inequality

$$|U(f, P) - U(f, P_1)| \leqslant M \sum_{\in(ii)} \ell(\tilde{J}_i) + \ell(\tilde{J}_i^+) + \ell(\tilde{J}_i^-)$$

$$= 2M \sum_{\in(ii)} \ell(\tilde{J}_i)$$

If $\tilde{J}_i$ is in class (ii), then there exists a unique endpoint $x_l$ in $Q$ such that $x_l \in \tilde{J}_i^\circ$ (the interior of the interval) so

$$\ell(\tilde{J}_i) \leqslant \frac{\ell(J_l)}{k}$$

Now

$$|U(f,P) - U(f,P_1)| \leqslant 2M \sum_{\in(ii)} \ell(\tilde{J}_i) \leqslant 2M \sum_{\in(ii)} \frac{\ell(J_l)}{k} \leqslant \frac{2M\ell(I)}{k}$$

As $P_1$ refines $Q$, $U(f,P_1) \leqslant U(f,Q)$, so

$$U(f,P_1) \leqslant U(f,P) \leqslant U(f,P_1) + \frac{2M\ell(I)}{k} \leqslant U(f,Q) + \frac{2M\ell(I)}{k}$$

Similarly, $L(f,P_1)$ refines $Q$ so $L(f,P_1) \geqslant L(f,Q)$, and by the same chain as above $|L(f,P) - L(f,P_1)| \leqslant \frac{2M\ell(I)}{k}$, so

$$L(f,P_1) \geqslant L(f,P) \geqslant L(f,P_1) - \frac{2M\ell(I)}{k} \geqslant L(f,Q) - \frac{2M\ell(I)}{k}$$

as desired. ■

Now we move on to our fundamental result:

## Theorem 14 (Darboux's Theorem).

*Let $f : [a,b] \to \mathbb{R}$ be bounded. Let $P_\nu \in \prod([a,b])$ be a sequence of partitions of $I = [a,b]$ such that*

$$\lim_{\nu\to\infty} maxsize(P_\nu) = 0$$

*Then*

$$U_I(f) = \lim_{\nu\to\infty} U(f,P_\nu)$$

*and*

$$L_I(f) = \lim_{\nu\to\infty} L(f,P_\nu)$$

*In particular, $f \in \mathbb{R}([a,b])$ if and only if*

$$\int_I f = \lim_{\nu\to\infty} \sum_{k=0}^{N_\nu - 1} f(\xi_{\nu,k})\ell(J_{\nu,k})$$

*where $P_\nu = (a = x_{\nu,0} < \ldots < x_{\nu,N_\nu} = b)$, $J_{\nu,k} = [x_{\nu,k}, x_{\nu,k+1}]$, and where $\xi_{\nu,k}$ is an arbitrary point of $J_{\nu,k}$.*

*Proof.* Let $P_\nu \in \prod([a,b])$ be a sequence of partition such that $\lim\limits_{\nu\to\infty} \text{maxsize} P_\nu = 0$. Note $U_I(f) = \inf_{P\in\prod([a,b])} U(f,P)$, which exists and is well defined if $f$ is bounded, and similarly for $L_I(f) = \sup_{P\in\prod([a,b])} L(f,P)$. Given $\varepsilon > 0$, there exists $Q \in \prod(I)$ such that $U_I(f) + \varepsilon \geqslant$

$U(f,Q) \geqslant U_I(f)$ and taking a refinement if needed, $L_I(f)-\varepsilon \leqslant L(f,Q) \leqslant L_I(f)$. Let $\nu,\mu \geqslant N$ such that

$$\text{maxsize}P_\nu \leqslant \varepsilon\text{minsize}Q,\ \forall\nu \geqslant N$$

In particular, we can choose $\varepsilon = \frac{1}{k}$, $k \in \mathbb{N}$, $k > 1$. By the previous proposition

$$U(f,P_\nu) \leqslant U(f,Q) + \varepsilon 2M\ell(I)$$
$$L(f,P_\nu) \geqslant L(f,Q) - \varepsilon 2M\ell(I)$$

Then $U(f,P_\nu) \leqslant U_I(f) + \varepsilon(2M\ell(I)+1)$ and $L(f,P_\nu) \geqslant L_I(f) - \varepsilon(2M\ell(I)+1)$. Then

$$|U(f,P_\nu) - U_I(f)| \leqslant \varepsilon(2M\ell(I)+1)$$

and

$$|L_I(f) - L(f,P_\nu)| \leqslant \varepsilon(2M\ell(I)+1)$$

for all $\nu \geqslant N$. As we can make $\varepsilon$ as small as we wish,

$$\lim_{\nu\to\infty} U(f,P_\nu) = U_I(f) \ \text{ and } \ \lim_{\nu\to\infty} L(f,P_\nu) = L_I(f)$$

Then $f \in \mathcal{R}(I)$ if and only if $U_I(f) = L_I(f)$, which occurs if and only if $\lim\limits_{\nu\to\infty} U(f,P_\nu) = \lim\limits_{\nu\to\infty} L(f,P_\nu)$. We observe that these last limits are equal if and only if $\lim\limits_{\nu\to\infty} \sum_{k=0}^{N_\nu-1} f(\xi_{\nu,k})\ell(J_{\nu,k})$ exists whenever maxsize$P_\nu \to 0$ for any choices of $\xi_{\nu,k} \in J_{\nu,k}$, as

$$L(f,P_\nu) \leqslant \sum_{k=0}^{N_\nu-1} f(\xi_{\nu,k})\ell(J_{\nu,k}) \leqslant U(f,P_\nu)$$

■

## 3.3.0 The Fundamental Theorem of Calculus

### Theorem 15 (The Fundamental Theorem of Calculus Part 1).

*Let $f \in \mathcal{R}([a,b])$ be integrable on $[a,b]$, and define $F$ on $[a,b]$ by*

$$F(x) = \int_a^x f = \int_a^x f(t)dt \tag{3.3.1}$$

*If $f$ is continuous at $c$ in $[a,b]$, then $F$ is differentiable at $c$, and*

$$F'(c) = f(c) \tag{3.3.2}$$

*(if $c = a$ or $b$, then $F'(c)$ is understood to mean the right or left hand derivative of $F$).*

*Proof.* First, consider $c \in (a,b)$. By definition,

$$F'(c) = \lim_{h\to 0} \frac{F(c+h) - F(c)}{h}$$

Suppose first that $h > 0$. Then

$$F(c+h) - F(c) = \int_c^{c+h} f$$

Define $m_h$ and $M_h$ as follows:

$$m_h = \inf\{f(x) : c \leqslant x \leqslant c+h\}$$
$$M_h = \sup\{f(x) : c \leqslant x \leqslant c+h\}$$

It follows from Theorem 3.1.7 that

$$m_h \cdot h \leqslant \int_c^{c+h} f \leqslant M_h \cdot h$$

Therefore,

$$m_h \leqslant \frac{F(c+h) - F(c)}{h} \leqslant M_h$$

If $h < 0$, let

$$m_h = \inf\{f(x) : c+h \leqslant x \leqslant c\}$$
$$M_h = \sup\{f(x) : c+h \leqslant x \leqslant c\}$$

It follows from Theorem 3.1.7 that

$$m_h \cdot (-h) \leqslant \int_{c+h}^{c} f \leqslant M_h \cdot (-h)$$

Since

$$F(c+h) - F(c) = \int_c^{c+h} f = -\int_{c+h}^{c} f$$

this yields

$$m_h \cdot h \geqslant F(c+h) - F(c) \geqslant M_h \cdot h$$

Since $h < 0$, we have that

$$m_h \leqslant \frac{F(c+h) - F(c)}{h} \leqslant M_h$$

This inequality is true for any integrable function, continuous or not. Since $f$ is continuous at $c$, however,

$$\lim_{h \to 0} m_h = \lim_{h \to 0} M_h = f(c)$$

and this proves that

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

Now, if $c = a$ we need only look at when $h > 0$, and in this case we still have

$$m_h \leqslant \frac{F(a+h) - F(a)}{h} \leqslant M_h$$

and from our previous limits,

$$\lim_{h\to 0^+} m_h = \lim_{h\to 0^+} m_h = f(a)$$

thus we have that

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

Similarly, if $c = b$ we need only look at $h < 0$, so we have that

$$\lim_{h\to 0^-} m_h = \lim_{h\to 0^-} m_h = f(b)$$

and

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

completing the proof. ∎

We may consider

$$F(x) = \int_a^x f \tag{3.3.3}$$

when $x < a$. In this case we have that

$$F(x) = -\int_x^a f = -\left(\int_b^a f - \int_b^x f\right) \tag{3.3.4}$$

so for $c \in [a,b]$,

$$F'(c) = -(-f(c)) = f(c) \tag{3.3.5}$$

as before.

## Theorem 16 (Fundamental Theorem of Calculus Part 2).

*Suppose $G$ is continuous in $[a,b]$ and differentiable in $(a,b)$, with $G' \in \mathcal{R}([a,b])$. Then*

$$\int_a^b G'(t)dt = G(b) - G(a)$$

*Proof.* Let $G$ be as described. Let $P_n$ be a partition with endpoints $a + \frac{b-a}{n}k$ for $0 \leqslant k \leqslant n$. By Darboux's theorem, for arbitrary $\xi_{n,k} \in J_{n,k} = \left[a + \frac{b-a}{n}k, a + \frac{b-a}{n}(k+1)\right]$,

$$\int_a^b G'(t)dt = \lim_{n\to\infty} \sum_{k=0}^{n-1} G'(\xi_{n,k})\ell(J_{n,k}) = \lim_{n\to\infty} \sum_{k=0}^{n-1} G'(\xi_{n,k})\frac{b-a}{n}$$

Now observe we have the telescopic sum

$$G(b) - G(a) = G(x_n) - G(x_0) = \sum_{k=0}^{n-1} G(x_{k+1}) - G(x_k)$$

As $G$ is continuous on $[a,b]$ and differentiable on $(a,b)$, we have by the mean value theorem that there exists $\xi^*_{n,k} \in (x_k, x_{k+1})$ such that

$$G(x_{k+1}) - G(x_k) = G'(\xi^*_{n,k})(x_{k+1} - x_k)$$

Then let the arbitrary $\xi_{n,k}$ be the $\xi^*_{n,k}$, so

$$\begin{aligned}\int_a^b G'(t)dt &= \lim_{n\to\infty}\sum_{k=0}^{n-1} G'(\xi^*_{n,k})\frac{b-a}{n}\\ &= \lim_{n\to\infty}\sum_{k=0}^{n-1} G(x_{k+1}) - G(x_k)\\ &= \lim_{n\to\infty}(G(b)-G(a)) = G(b)-G(a)\end{aligned}$$

as desired. ∎

It is important to note that this is merely a useful result for certain functions $f$, ***not*** a definition.

If $f$ is any bounded function on $I = [a,b]$, then

$$L_I(f) \;\; and \;\; U_I(f) \tag{3.3.6}$$

will both exist. These numbers are called the ***lower integral*** of $f$ on $[a,b]$ and the ***upper integral*** of $f$ on $[a,b]$, respectively, and will sometimes be denoted by

$$\mathbf{L}\int_a^b f \;\; and \;\; \mathbf{U}\int_a^b f \tag{3.3.7}$$

If $a < c < b$, then

$$\mathbf{L}\int_a^b f = \mathbf{L}\int_a^c f + \mathbf{L}\int_c^b f \;\; and \;\; \mathbf{U}\int_a^b f = \mathbf{U}\int_a^c f + \mathbf{U}\int_c^b f \tag{3.3.8}$$

and if $m \leqslant f(x) \leqslant M$ for all $x \in [a,b]$, then

$$m(b-a) \leqslant \mathbf{L}\int_a^b f \leqslant \mathbf{U}\int_a^b f \leqslant M(b-a) \tag{3.3.9}$$

$f$ is integrable precisely when

$$\mathbf{L}\int_a^b f = \mathbf{U}\int_a^b f \tag{3.3.10}$$

We shall now demonstrate an alternate proof for the following theorem stated previously.

**Theorem 3.3.1.** *If $f$ is continuous on $[a,b]$, then $f$ is integrable on $[a,b]$.*

*Proof.* Define function $L$ and $U$ on $[a,b]$ by

$$L(x) = \mathbf{L}\int_a^x f \;\; and \;\; U(x) = \mathbf{U}\int_a^x f$$

Let $x \in (a,b)$. If $h > 0$ and

$$m_h = \inf\{f(t) : x \leqslant t \leqslant x+h\}$$

$$M_h = \sup\{f(t) : x \leqslant t \leqslant x+h\}$$

then

$$m_h \cdot h \leqslant \mathbf{L}\int_x^{x+h} f \leqslant \mathbf{U}\int_x^{x+h} f \leqslant M_h \cdot h$$

so

$$m_h \cdot h \leqslant L(x+h) - L(x) \leqslant U(x+h) - U(x) \leqslant M_h \cdot h$$

or

$$m_h \leqslant \frac{L(x+h)-L(x)}{h} \leqslant \frac{U(x+h)-U(x)}{h} \leqslant M_h$$

If $h < 0$ and

$$m_h = \inf\{f(t) : x+h \leqslant t \leqslant x\}$$
$$M_h = \sup\{f(t) : x+h \leqslant t \leqslant x\}$$

one obtains the same inequality, precisely as in the proof of 15.

Since $f$ is continuous at $x$, we have

$$\lim_{h\to 0} m_h = \lim_{h\to 0} M_h = f(x)$$

and this proves that

$$L'(x) = U'(x) = f(x), \forall x \in (a,b)$$

THis means that there is a number $c$ such that

$$U(x) = L(x) + c, \forall x \in [a,b]$$

Since $U(a) = L(a) = 0$, the number $c$ must be equal to 0, so

$$U(x) = L(x) \forall x \in [a,b]$$

In particular,

$$\mathbf{U}\int_a^b f = U(b) = L(b) = \mathbf{L}\int_a^b f$$

so $f$ is integrable on $[a,b]$. ∎

# 3.4.0 Content of Sets

Using properties of sets, we can investigate the collection of Riemann integrable functions more carefully.

**Definition 3.4.1.** *Given $S \subseteq I$ we define the* **characteristic function for $S$** *by*

$$\chi_S(x) = \begin{cases} 1 & x \in S \\ 0 & x \notin S \end{cases}$$

*and the* **upper content of $S$** *and the* **lower content of $S$** *by*

$$cont^+(S) := U_I(\chi_S) \text{ and } cont^-(S) = L_I(\chi_S)$$

*If $cont^+(S) = cont^-(S)$ we say that $S$ has* **content**

$$m(S) = cont^+(S) = cont^-(S)$$

*and say $S$ is* **contented**.

We observe by Darboux's theorem $\text{cont}^+(S) = U_I(\chi_S) = \lim_{\nu\to\infty} U(\chi_S, P_\nu)$ if maxsize$P_\nu \to 0$. Since $\sup_J \chi_S = 1$ if $S \cap J \neq \varnothing$, $\sup_J \chi_S = 0$ if $S \cap J = \varnothing$, $\inf_J(\chi_S) = 1$ if $S \supseteq J$, and $\inf_J(\chi_S) = 0$ if $S \not\supseteq J$, we can formulate the upper and lower contents by

$$\text{cont}^+(S) = \inf\left\{\sum_{k=1}^N \ell(J_k) : S \subseteq \bigcup_{k=1}^N J_k\right\}$$

and

$$\text{cont}^-(S) = \sup\left\{\sum_{k=1}^N \ell(J_k) : S \supseteq \sqcup_{k=1}^N J_k\right\}$$

(Note we need disjoint sets for the lower content)

We now define the Lebesque measure, which extends the upper content to what is known as an outer measure.

**Definition 3.4.2.** *We define the* **Lebesque measure** *by*

$$m^*(S) = \inf\left\{\sum_{k\geqslant 1} \ell(J_k) : S \subseteq \bigcup_{k\geqslant 1} J_k\right\}$$

*which is an* **outer measure**, *and we can define the associated* **inner measure** *by*

$$m_*(S) = \sup\left\{\sum_{k\geqslant 1} \ell(J_k) : S \supseteq \sqcup_{k\geqslant 1} J_k\right\}$$

**Example 3.4.1.** We observe that $\text{cont}^+(\mathbb{Q}\cap[0,1]) = 1$ and $\text{cont}^-(\mathbb{Q}\cap[0,1]) = 0$, but $m^*(\mathbb{Q}\cap[0,1]) = 0$. Indeed, as $\mathbb{Q}\cap[0,1] = \{q_1, q_2, ...\}$ is countable, if $\varepsilon > 0$ we can choose $J_k = \left[q_k - \frac{\varepsilon}{2^{k+2}}, q_k + \frac{\varepsilon}{2^{k+2}}\right] \cap [0,1]$. Then $\mathbb{Q}\cap[0,1] \subseteq \bigcup_{k\geqslant 1} J_k$ and

$$\sum_{k\geqslant 1} \ell(J_k) \leqslant \sum_{k\geqslant 1} \frac{\varepsilon}{2^{k+1}} \leqslant \frac{\varepsilon}{2} < \varepsilon$$

In general, if $S$ is countable then $m^*(S) = 0$. We can now investigate a necessary and sufficient condition for Riemann integrability:

**Proposition 3.4.1.** *If $f : I = [a,b] \to \mathbb{R}$ is bounded, and $S$ is the set of points of discontinuity of $f$ in $[a,b]$, then $m^*(S)0$ implies $f \in \mathcal{R}(I)$, where*

$$m^*(S) = \inf\left\{\sum_{k\geqslant 1} \ell(J_k) : S \subseteq \bigcup_{k\geqslant 1} J_k\right\}$$

*Proof.* As $f$ is bounded, there exists $M > 0$ such that $||f||_\infty \leqslant M$. Let $\varepsilon > 0$. As $m^*(S) = 0$, there exists $J_k$, $k \geqslant 1$, such that

$$S \subseteq \bigcup_{k\geqslant 1} J_k \text{ and } \sum_{k=1}^{\infty} \ell(J_k) < \varepsilon$$

If $x \in I \backslash S$, then there exists an open interval $K_x$ containing $x$ such that

$$0 \leqslant \sup_{K_x}(f) - \inf_{K_x}(f) < \varepsilon$$

as $f$ is continuous at $x$. Then $I \subseteq \left(\bigcup_{k\geqslant 1} J_k\right) \cup \left(\bigcup_{x\in I\backslash S} K_x\right)$. But $I$ is compact so there exists a finite covering

$$I \subseteq \left(\bigcup_{k=1}^{N} J_k\right) \cup \left(\bigcup_{i=1}^{M} K_i\right)$$

Let $P \in \prod(I)$ be the partition conformed by all the endpoints of intervals $J_l$, $1 \leqslant l \leqslant N$, and $K_i$, $1 \leqslant i \leqslant M$. Write $P = \{L_1, ..., L_\mu\}$. We have two cases:

1. $L_j \subseteq K_i$ for some $i$ or
2. $L_j \subseteq J_l$ for some $l$

Let $A = \{1 \leqslant j \leqslant \mu : \exists i; L_j \subseteq K_i\}$ and $B = \{1 \leqslant j \leqslant \mu : \exists l; L_j \subseteq J_l\}$. Note $\sum_{j\in B} \ell(L_j) \leqslant \sum_{k=1}^{N} \ell(J_k) < \varepsilon$ and

$$\begin{aligned}
0 \leqslant U(f,P) - L(f,P) &= \sum_{j\in A}\left(\sup_{L_j}(f) - \inf_{L_j}(f)\right)\ell(L_j) + \sum_{j\in B}\left(\sup_{L_j}(f) - \inf_{L_j}(f)\right)\ell(L_j) \\
&< \varepsilon \sum_{j\in A} \ell(L_j) + 2M \sum_{j\in B} \ell(L_j) \\
&\leqslant \varepsilon \ell(I) + 2M\varepsilon \\
&= \varepsilon(\ell(I) + 2M)
\end{aligned}$$

which goes to 0 as we can make $\varepsilon$ arbitrarily small. Thus $f \in \mathcal{R}([a,b])$ as desired. ∎

## 3.A.0 Trigonometric Functions

**Definition 3.A.1.** *We define the mathematical constant $\pi$ as the area of the unit circle, or in this case, twice the area of a semi-circle:*

$$\pi := 2 \cdot \int_{-1}^{1} \sqrt{1 - x^2}dx \tag{3.A.1}$$

**Definition 3.A.2.** *If $-1 \leqslant x \leqslant 1$, then the area of the sector bounded between the upper unit circle from $[x, 1]$ and the $x$-axis and radial arm is*

$$A(x) := \frac{x\sqrt{1-x^2}}{2} + \int_{x}^{1} \sqrt{1 - t^2}dt \tag{3.A.2}$$

**Remark 3.A.1.** For $-1 < x < 1$, $A$ is differentiable at $x$ and

$$\begin{aligned}
A'(x) &= \frac{1}{2}\left[\sqrt{1-x^2} + x \cdot \frac{-2x}{2\sqrt{1-x^2}}\right] - \sqrt{1-x^2} \\
&= \frac{1}{2}\frac{1-x^2-x^2}{\sqrt{1-x^2}} - \frac{1-x^2}{\sqrt{1-x^2}} \\
&= \frac{1}{2}\frac{-1}{\sqrt{1-x^2}} \\
&= \frac{-1}{2\sqrt{1-x^2}}
\end{aligned}$$

Note that on $[-1, 1]$, the function $A$ decreases from $A(-1) = \frac{\pi}{2}$ to $A(1) = 0$.

**Definition 3.A.3.** *If $0 \leqslant x \leqslant \pi$, then $\cos x$ is the unique number in $[-1, 1]$ such that*

$$A(\cos x) = \frac{x}{2} \tag{3.A.3}$$

*and*

$$\sin x := \sqrt{1 - (\cos x)^2} \tag{3.A.4}$$

*Note that such a $\cos x$ exists as $A$ is continuous on $[-1, 1]$, and $A(-1) = \frac{\pi}{2}$ while $A(1) = 0$. Hence, by 6 there exists $y \in [-1, 1]$ such that $A(y) = \frac{x}{2}$ for all $x \in [0, \pi]$.*

**Theorem 3.A.1.** *If $0 < x < \pi$, then*

$$\begin{aligned}
\cos'(x) &= -\sin x \\
\sin'(x) &= \cos x
\end{aligned}$$

*Proof.* If $B = 2A$, then the definition $A(\cos x) = x/2$ can be written

$$B(\cos x) = x$$

in other words, cos is just the inverse of $B$. We have already computed taht

$$A'(x) = -\frac{1}{2\sqrt{1-x^2}}$$

from which we conclude

$$B'(x) = -\frac{1}{\sqrt{1-x^2}}$$

Consequently we have that

$$\begin{aligned}
\cos'(x) &= (B^{-1})'(x) \\
&= \frac{1}{B'(B^{-1}(x))} \\
&= \frac{1}{-\frac{1}{\sqrt{1-[B^{-1}(x)]^2}}} \\
&= -\sqrt{1-(\cos x)^2}
\end{aligned}$$

$$= -\sin x$$

Then, since $\sin x = \sqrt{1-(\cos x)^2}$ we also obtain

$$\begin{aligned}\sin'(x) &= \frac{1}{2}\cdot\frac{-2\cos x\cdot\cos'(x)}{\sqrt{1-(\cos x)^2}} \\ &= \frac{-\cos x\cdot(-\sin x)}{\sin x} \\ &= \cos x\end{aligned}$$

■

**Definition 3.A.4.** *Now, to define* $\sin$ *and* $\cos$ *on* $\mathbb{R}$*, we proceed as follows*

1. *If* $\pi \leqslant x \leqslant 2\pi$*, the*

$$\begin{aligned}\sin x &= -\sin(2\pi - x)\\ \cos x &= \cos(2\pi - x)\end{aligned}$$

2. *If* $x = 2\pi k + x'$ *for some integer* $k$ *and some* $x' \in [0, 2\pi]$*, then*

$$\begin{aligned}\sin x &= \sin x'\\ \cos x &= \cos x'\end{aligned}$$

**Lemma 3.A.2.** *Suppose* $f$ *has a second derivative everywhere and that*

$$\begin{aligned}f'' + f &= 0\\ f(0) &= 0\\ f'(0) &= 0\end{aligned}$$

*Then* $f = 0$

*Proof.* Multiplying both sides of the first equation by $f'$ yields

$$f'f'' + ff' = 0$$

Thus

$$[(f')^2 + f^2]' = 2(f'f'' + ff') = 0$$

so $(f')^2 + f^2$ is a constant function. From $f(0) = 0$ and $f'(0) = 0$ it follows that the constant is 0; thus

$$[f'(x)]^2 + [f(x)]^2 = 0 \forall x$$

This implies that

$$f(x) = 0 \forall x$$

■

**Theorem 3.A.3.** *If $f$ has a second derivative everywhere and*

$$\begin{aligned} f'' + f &= 0 \\ f(0) &= a \\ f'(0) &= b \end{aligned}$$

*then*

$$f = b \cdot \sin + a \cdot \cos$$

*Proof.* Let

$$g(x) = f(x) - b\sin x - a\cos x$$

Then

$$\begin{aligned} g'(x) &= f'(x) - b\cos x + a\sin x \\ g''(x) &= f''(x) + b\sin x + a\cos x \end{aligned}$$

Consequently,

$$\begin{aligned} g'' + g &= 0 \\ g(0) &= 0 \\ g'(0) &= 0 \end{aligned}$$

which shows by the previous lemma that

$$0 = g(x) = f(x) - b\sin x - a\cos x, \forall x$$

■

**Theorem 3.A.4.** *If $x$ and $y$ are any two numbers, then*

$$\begin{aligned} \sin(x+y) &= \sin x \cos y + \cos x \sin y \\ \cos(x+y) &= \cos x \cos y - \sin x \sin y \end{aligned}$$

*Proof.* For any particular $y \in \mathbb{R}$, we can define a function $f$ by

$$f(x) = \sin(x+y)$$

Then $f'(x) = \cos(x+y)$ and $f''(x) = -\sin(x+y)$. Consequently,

$$\begin{aligned} f'' + f &= 0 \\ f(0) &= \sin y \\ f'(0) &= \cos y \end{aligned}$$

It follows from the previous theorem that

$$f = (\cos y) \cdot \sin + (\sin y) \cdot \cos$$

that is

$$\sin(x+y) = \cos y \sin x + \sin y \cos x, \forall x$$

Since any number $y$ could have been chosen to begin with, this proves the first formula for $x$ and $y$.

Similarly, for any $y \in \mathbb{R}$ define $f(x) = \cos(x+y)$, so $f'(x) = -\sin(x+y)$ and $f''(x) = -\cos(x+y)$. Then $f'' + f = 0$, $f(0) = \cos y$ and $f'(0) = -\sin y$. Then we have that

$$\cos(x+y) = \cos y \cos x - \sin y \sin x$$

proving the second formula. ∎

**Remark 3.A.2.** Since

$$\arcsin'(x) = \frac{1}{\sqrt{1-x^2}}, -1 < x < 1$$

it follows from 16 that

$$\arcsin x = \arcsin x - \arcsin 0 = \int_0^x \frac{1}{\sqrt{1-t^2}} dt$$

Using this definition of arcsin we could define sin as $\arcsin^{-1}$, and the formula for the derivative of an inverse function would show that

$$\sin'(x) = \sqrt{1-\sin^2 x}$$

which could be defined as $\cos x$.

## 3.B.0 The Logarithm and Exponential Functions

**Definition 3.B.1.** *If $x > 0$, then define*

$$\log x := \int_1^x \frac{1}{t} dt \tag{3.B.1}$$

**Theorem 3.B.1.** *If $x.y > 0$, then*

$$\log(xy) = \log x + \log y \tag{3.B.2}$$

*Proof.* Notice first that $\log'(x) = 1/x$, by 15. Now, choose a number $y > 0$ and let

$$f(x) = \log(xy)$$

Then

$$f'(x) = \log'(xy) \cdot y = \frac{1}{xy} \cdot y = \frac{1}{x}$$

Thus, $f' = \log'$. This means that there is a number $c$ such that $f(x) = \log(x) + c$ for all $x > 0$, that is,

$$\log(xy) = \log x + c, \; \forall x > 0$$

The number $c$ can be evaluated by noting that $\log(1) = 0$, so $\log(1 \cdot y) = c$. Thus

$$\log(xy) = \log x + \log y$$

for all $x$. Since this is true for all $y > 0$, the theorem is proved. ∎

**Corollary 3.B.2.** *If $n$ is a natural number and $x > 0$, then*

$$\log(x^n) = n \log x \tag{3.B.3}$$

*Proof.* We proceed by induction on $n \in \mathbb{N}$. If $n = 1$ we simply have $\log(x^1) = 1 \cdot \log x$, so the base case holds. Now suppose inductively that there exists $k \geqslant 1$ such that if $n = k$,

$$\log(x^k) = k \log x$$

Then, observe that by the previous theorem

$$\begin{aligned}
\log(x^{k+1}) &= \log(x^k x) \\
&= \log(x^k) + \log x \\
&= k \log x + \log x && \text{(by the Induction Hypothesis)} \\
&= (k+1) \log x
\end{aligned}$$

as desired. Thus by mathematical induction we conclude that for all $n \geqslant 1$, $\log(x^n) = n \log x$. ■

**Corollary 3.B.3.** *If $x, y > 0$, then*

$$\log\left(\frac{x}{y}\right) = \log x - \log y$$

*Proof.* This result follows from the equation

$$\log x = \log\left(\frac{x}{y} \cdot y\right) = \log\left(\frac{x}{y}\right) + \log y$$

■

**Definition 3.B.2.** *The* **exponential function**, exp, *is defined as* $\log^{-1}$.

**Theorem 3.B.4.** *For all numbers $x$,*

$$\exp'(x) = \exp(x)$$

*Proof.* Observe that

$$\begin{aligned}
\exp'(x) &= (\log^{-1})'(x) \\
&= \frac{1}{\log'(\log^{-1}(x))} \\
&= \frac{1}{\frac{1}{\log^{-1}(x)}} \\
&= \log^{-1}(x) = \exp(x)
\end{aligned}$$

■

**Theorem 3.B.5.** *If $x$ and $y$ are any two numbers, then*

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

*Proof.* Let $x' = \exp(x)$ and $y' = \exp(y)$, so that $x = \log x'$ and $y = \log y'$. Then

$$x + y = \log x' + \log y' = \log(x'y')$$

This means that

$$\exp(x+y) = x'y' = \exp(x) \cdot \exp(y)$$

∎

**Definition 3.B.3.** *We define*

$$e := \exp(1) \tag{3.B.4}$$

*and this is equivalent to the equation*

$$1 = \log e = \int_1^e \frac{1}{t} dt \tag{3.B.5}$$

*Then, we note that* $\exp(x) = [\exp(1)]^x = e^x$ *for rational $x$, so we define for any $x \in \mathbb{R}$,*

$$e^x = \exp(x) \tag{3.B.6}$$

**Definition 3.B.4.** *If $a > 0$, then, for any real number $x$,*

$$a^x := e^{x \log a} \tag{3.B.7}$$

*If $a = e$ this definition agrees with our previous one.*

**Theorem 3.B.6.** *If $a > 0$, then*

$$(1) \;\; (a^b)^c = a^{bc}, \;\; \forall a, b \in \mathbb{R}$$

*and*

$$(2) \;\; a^1 = a \text{ and } a^{x+y} = a^x \cdot a^y, \;\; \forall x, y \in \mathbb{R}$$

*Proof.* First, observe that

$$\begin{aligned}
(a^b)^c &= e^{c \log a^b} \\
&= e^{c \log e^{b \log a}} \\
&= e^{cb \log a} \\
&= a^{bc}
\end{aligned}$$

Next, observe that

$$a^1 = e^{1 \log a} = e^{\log a} = a$$

and

$$\begin{aligned}
a^{x+y} &= e^{(x+y) \log a} \\
&= e^{x \log a + y \log a} \\
&= e^{x \log a} \cdot e^{y \log a} \\
&= a^x \cdot a^y
\end{aligned}$$

∎

**Theorem 3.B.7.** *If $f$ is differentiable and*

$$f'(x) = f(x), \ \forall x \in \mathbb{R}$$

*then there is a number $c$ such that*

$$f(x) = ce^x, \ \forall x \in \mathbb{R}$$

*Proof.* Let $g(x) = f(x)/e^x$, which is possible as $e^x \neq 0$ for all $x$. Then

$$g'(x) = \frac{e^x f'(x) - f(x)e^x}{(e^x)^2} = 0$$

THerefore, there is a number $c$ such that

$$g(x) = \frac{f(x)}{e^x} = c, \ \forall x$$

■

**Theorem 3.B.8.** *For any natural number $n$,*

$$\lim_{x \to \infty} \frac{e^x}{x^n} = \infty \tag{3.B.8}$$

*Proof.* Step 1. We claim that $e^x > x$ for all $x$, and consequently $\lim_{x \to \infty} e^x = \infty$.

For $x \leqslant 0$ this is immediate. Now, it suffices to show $x > \log x$ for all $x > 0$. If $x < 1$ this clearly holds since $\log x < 0$. If $x > 1$, then $x - 1$ is an upper sum for $f(t) = \frac{1}{t}$ on $[1, x]$, so $\log x < x - 1 < x$.

Step 2. We claim $\lim_{x \to \infty} \frac{e^x}{x} = \infty$. First, note that

$$\frac{e^x}{x} = \frac{e^{x/2} \cdot e^{x/2}}{\frac{x}{2} \cdot 2} = \frac{1}{2}\left(\frac{e^{x/2}}{\frac{x}{2}}\right) \cdot e^{x/2}$$

By Step 1. the expression in parentheses is greater than 1, and $\lim_{x \to \infty} e^{x/2} = \infty$; this shows that $\lim_{x \to \infty} e^x/x = \infty$.

Step 3. To prove the main claim note that

$$\frac{e^x}{x^n} = \frac{(e^{x/n})^x}{\left(\frac{x}{n}\right)^n \cdot n^n} = \frac{1}{n^n} \cdot \left(\frac{e^{x/n}}{\frac{x}{n}}\right)^n$$

The expression in parentheses becomes arbitrarily large, by Step 2., so the $n$th power certainly becomes arbitrarily large. ■

# Chapter 4

# Sequences and Series

## 4.1.0 Approximation by Polynomial Functions

**Definition 4.1.1.** *Given a function $f$ that is $n$ times differentiable in a neighborhood of a point $a$, let*

$$a_k = \frac{f^{(k)}(a)}{k!}, \quad 0 \leqslant k \leqslant n,$$

*and define*

$$P_{n,a}(x) = a_0 + a_1(x-a) + \ldots + a_x(x-a)^n$$

*The polynomial $P_{n,a}$ is called the* **Taylor polynomial of degree $n$ for $f$ at $a$**.

**Remark 4.1.1.** The Taylor polynomial has been defined so that

$$P_{n,a}^{(k)}(a) = f^{(k)}(a) \quad \text{for } \ 0 \leqslant k \leqslant n$$

in fact, it is evidently the only polynomial of degree $\leqslant n$ with this property.

**Example 4.1.1.** Consider the sin function. We have

$$\begin{aligned} \sin(0) &= 0 \\ \sin'(0) &= \cos 0 = 1 \\ \sin''(0) &= -\sin 0 = 0 \\ \sin'''(0) &= -\cos 0 = -1 \\ \sin^{(4)}(0) &= \sin 0 = 0 \end{aligned}$$

From this point on, the derivatives repeat modulo 4. The coefficients become

$$a_k = \frac{\sin^{(k)}(0)}{k!} = \begin{cases} 0 & \text{if } \exists l \in \mathbb{N}; k = 2l \\ \frac{(-1)^l}{(2l+1)!} & \text{if } \exists l \in \mathbb{N}; k = 2l+1 \end{cases}$$

Therefore, the Taylor polynomial $P_{2n+1,0}$ of degree $2n+1$ for sin at 0 is

$$P_{2n+1,0}(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \ldots + (-1)^n \frac{x^{2n+1}}{(2n+1)!}$$

Via a similar derivation we find that the Taylor polynomial $P_{2n,0}$ of degree $2n$ for cos at 0 is

$$P_{2n,0} = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \ldots + (-1)^n \frac{x^{2n}}{(2n)!}$$

**Example 4.1.2.** Note that for all $k \geqslant 0$, $\exp^{(k)}(0) = \exp(0) = 1$, so the Taylor polynomial of degree $n$ for exp at 0 is

$$P_{n,0}(x) = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots + \frac{x^n}{n!}$$

For log, observe that

$$\begin{aligned} \log'(x) &= \frac{1}{x}, & \log'(1) &= 1 \\ \log''(x) &= -\frac{1}{x^2}, & \log''(1) &= -1 \\ \log'''(x) &= \frac{2}{x^3}, & \log'''(1) &= 2 \end{aligned}$$

in general

$$\log^{(k)}(x) = \frac{(-1)^{k-1}(k-1)!}{x^k}, \quad \log^{(k)}(1) = (-1)^{k-1}(k-1)!$$

for $k \geqslant 1$, and $\log(1) = 0$. Therefore, the Taylor polynomial of degree $n$ for log at 1 is

$$P_{n,1}(x) = (x-1) - \frac{(x-1)^2}{2} + \frac{(x-1)^3}{3} - \ldots + \frac{(-1)^{n-1}(x-1)^n}{n}$$

If we consider the function $f(x) = \log(1+x)$, then the Taylor polynomial of degree $n$ of $f$ at 0 is

$$P_{n,0}(x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \ldots + \frac{(-1)^{n-1}x^n}{n}$$

**Theorem 4.1.1.** *Suppose that $f$ is a function and $a \in \mathbb{R}$ such that*

$$f'(a), ..., f^{(n)}(a)$$

*all exist. Let*

$$a_k = \frac{f^{(k)}}{k!}, 0 \leqslant k \leqslant n$$

*and define*

$$P_{n,a}(x) = a_0 + a_1(x-a) + \ldots + a_n(x-a)^n$$

*Then*

$$\lim_{x \to a} \frac{f(x) - P_{n,a}(x)}{(x-a)^n} = 0$$

*Proof.* Weiting out $P_{n,a}(x)$ expliticly we obtain

$$\frac{f(x) - P_{n,a}(x)}{(x-a)^n} = \frac{f(x) - \sum_{i=0}^{n-1} \frac{f^{(i)}(a)}{i!}(x-a)^i}{(x-a)^n} - \frac{f^{(n)}}{n!}$$

Let us introduce the new functions

$$Q(x) = \sum_{i=0}^{n-1} \frac{f^{(i)}(a)}{i!}(x-a)^i \text{ and } g(x) = (x-a)^n;$$

now we must prove that

$$\lim_{x\to a} \frac{f(x) - Q(x)}{g(x)} = \frac{f^{(n)}(a)}{n!}$$

Note that $Q^{(k)}(a) = f^{(k)}(a)$ for $k \leqslant n-1$, and $g^{(k)}(x) = n!\frac{(x-a)^{n-k}}{(n-k)!}$. By the continuity of $f^{(k)}$ and $Q^{(k)}$ for $k \leqslant n-1$, we have that

$$\lim_{x\to a} \left[f^{(k)}(x) - Q^{(k)}(x)\right] = f^{(k)}(a) - Q^{(k)}(a) = 0$$

and

$$\lim_{x\to a} g^{(k)}(x) = 0$$

Then applying l'Hopital's Rule $n-1$ times, we obtain

$$\lim_{x\to a} \frac{f(x) - Q(x)}{(x-a)^n} = \lim_{x\to a} \frac{f^{(n-1)}(x) - Q^{(n-1)}(x)}{n!(x-a)}$$

But the $n-1$st derivative of $Q$ is constant, and in fact $Q^{(n-1)}(x) = f^{(n-1)}(a)$, so

$$\lim_{x\to a} \frac{f(x) - Q(x)}{(x-a)^n} = \lim_{x\to a} \frac{f^{(n-1)}(x) - f^{(n-1)}(a)}{n!(x-a)} = \frac{f^{(n)}(a)}{n!}$$

by definition of $f^{(n)}(a)$, as desired. ■

**Theorem 4.1.2.** *Suppose that*

$$f'(a) = ... = f^{(n-1)}(a) = 0, \quad \textit{and} \quad f^{(n)}(a) \neq 0$$

1. *If $n$ is even and $f^{(n)}(a) > 0$, then $f$ has a local minimum at $a$.*
2. *If $n$ is even and $f^{(n)}(a) < 0$, then $f$ has a local maximum at $a$.*
3. *If $n$ is odd, then $f$ has neither a local maximum nor a local minimum at $a$.*

*Proof.* Let $f$ be as in the hypothesis, and without loss of generality let $f(a) = 0$, as otherwise one can replace $f$ with $f - f(a)$ without affecting the hypothesis. THen, since the first $n-1$ derivatives of $f$ at $a$ are 0, the Taylor polynomial$P_{n,a}$ of $f$ is

$$\begin{aligned} P_{n,a}(x) &= \sum_{i=0}^{n} \frac{f^{(i)}(a)}{i!}(x-a)^i \\ &= \frac{f^{(n)}(a)}{n!}(x-a)^n \end{aligned}$$

Thus, Theorem 4.1.1 states that

$$0 = \lim_{x\to a} \frac{f(x) - P_{n,a}(x)}{(x-a)^n} = \lim_{x\to a} \left[\frac{f(x)}{(x-a)^n} - \frac{f^{(n)}(a)}{n!}\right]$$

Consequently, if $x$ is sufficiently close to $a$, then $f(x)/(x-a)^n$ has the same sign as $f^{(n)}(a)/n!$. Suppose now that $n$ is even. In this case $(x-a)^n > 0$ for all $x \neq a$. Since $f(x)/(x-a)^n$ has the same sign as $f^{(n)}(a)/n!$ for $x$ sufficeintly close to $a$, it follows that $f(x)$ itself has the same sign as $f^{(n)}(a)/n!$ for $x$ sufficiently close to $a$. If $f^{(n)}(a) > 0$, this means that

$$f(x) > 0 = f(a)$$

for $x$ close to $a$. Consequently, $f$ has a local minimum at $a$. If $f^{(n)}(a) < 0$, this means that

$$f(x) < 0 = f(a)$$

for $x$ close to $a$, so $f$ has a local minimum at $a$.

Conversely, suppose that $n$ is odd. Then if $x$ is sufficiently close to $a$ $f(x)/(x-a)^n$ always has the same sign, since it has the same sign as $f^{(n)}(a)/n!$ which is constant. But $(x-a)^n > 0$ for $x > a$ and $(x-a)^n < 0$ for $x < a$. Therefore $f(x)$ has different signs for $x > a$ and $x < a$. Hence, $f$ has neither a local maximum nor a local minimum at $a$. ∎

**Definition 4.1.2.** *Two functions $f$ and $g$ are* **equal up to order $n$ at $a$** *if*

$$\lim_{x\to a} \frac{f(x)-g(x)}{(x-a)^n} = 0$$

**Theorem 4.1.3.** *Let $P$ and $Q$ be two polynomials in $(x-a)$, of degree $\leqslant n$, and suppose that $P$ and $Q$ are equal up to order $n$ at $a$. Then $P = Q$.*

*Proof.* Let $R = P - Q$. Since $R$ is a polynomial of degree $\leqslant n$, it is only necessariy to prove that if $R(x) = b_0 + \ldots + b_n(x-a)^n$ satisfies

$$\lim_{x\to a} \frac{R(x)}{(x-a)^n} = 0$$

then $R = 0$. Now the hypothesis on $R$ surely implies that

$$\lim_{x\to a} \frac{R(x)}{(x-a)^i} = 0 \quad \text{for } 0 \leqslant i \leqslant n$$

For $i = 0$ this condition reads that $\lim_{x\to a} R(x) = 0$. On the other hand $\lim_{x\to a} R(x) = b_0$. Thus $b_0 = 0$. Similarly, we find that

$$\frac{R(x)}{x-a} = b_1 + b_2(x-a) + \ldots + b_n(x-a)^{n-1}$$

and

$$\lim_{x\to a} \frac{R(x)}{x-a} = b_1$$

so $b_1 = 0$. Continuing in this way, by induction we find that

$$b_0 = b_1 = \ldots = b_n = 0$$

so $R(x) = 0$ and $P = Q$. ∎

**Corollary 4.1.4.** *Let $f$ be $n$-times differentiable at $a$, and suppose that $P$ is a polynomial in $(x-a)$ of degree $\leqslant n$, which equals $f$ up to order $n$ at $a$. Then $P = P_{n,a,f}$ (The Taylor polynomial of $f$ of degree $n$ at $a$).*

*Proof.* Since $P$ and $P_{n,a,f}$ both equal $f$ up to order $n$ at $a$, using the triangle inequality and an epsilon-delta proof it can be shown that $P$ equals $P_{n,a,f}$ up to order $n$ at $a$. Consequently, $P = P_{n,a,f}$ by the preceding Theorem. ■

**Definition 4.1.3.** *If $f$ is a function for which $P_{n,a}(x)$ exists, we define the* **remainder term** *$R_{n,a}(x)$ by*

$$R_{n,a}(x) = f(x) - P_{n,a}(x)$$

*If $f^{(n+a)}$ is continuous on $[a, x]$, then*

$$R_{n,a}(x) = \int_a^x \frac{f^{(n+1)}(t)}{n!}(x-t)^n dt$$

**Remark 4.1.2.** If $m$ and $M$ are the minimum and maximum of $f^{(n+1)}n!$ on $[a, x]$, then $R_{n,a}(x)$ satisfies

$$m \in_a^x (x-t)^n dt \leqslant R_{n,a}(x) \leqslant M \int_a^x (x-t)^n dt$$

so we can write

$$R_{n,a}(x) = \alpha \cdot \frac{(x-a)^{n+1}}{n+1}$$

for some number $\alpha$ between $m$ and $M$. Since we have assumed that $f^{(n+1)}$ is continuous, for some $t \in (a, x)$ we can also write

$$R_{n,a}(x) = \frac{f^{(n+1)}(t)}{n!}\frac{(x-a)^{n+1}}{n+1} = \frac{f^{(n+1)}(t)}{(n+1)!}(x-a)^{n+1}$$

This is known as the ***Lagrange form of the remainder***.

**Lemma 4.1.5.** *Suppose that the function $R$ is $(n+1)$-times differentiable on $[a, b]$, and*

$$R^{(k)}(a) = 0 \quad \text{for } k = 0, 1, 2, ..., n$$

*Then for any $x \in (a, b]$ we have*

$$\frac{R(x)}{(x-a)^{n+1}} = \frac{R^{(n+1)}(t)}{(n+1)!} \quad \text{for some } t \text{ in } (a, x)$$

*Proof.* For $n = 0$, this is simply 9, and we will proceed for the remaining $n$ by induction on $n$. To do this we use 10 to write

$$\frac{R(x)}{(x-a)^{n+2}} = \frac{R'(z)}{(n+2)(z-a)^{n+1}} = \frac{1}{n+2}\frac{R'(z)}{(z-a)^{n+1}} \quad \text{for some } z \text{ in } (a, x),$$

and then apply the induction hypothesis to $R'$ on the interval $[a, z]$ to get

$$\frac{R(x)}{(x-a)^{n+2}} = \frac{1}{n+2}\frac{(R')^{(n+1)}(t)}{(n+1)!} \quad \text{for some } z \text{ in } (a, x),$$

$$= \frac{R^{(n+2)}(t)}{(n+2)!}$$

as desired. ∎

### Theorem 17 (Taylor's Theorem).

*Suppose $f', \ldots, f^{(n+1)}$ are defined on $[a, x]$, and that $R_{n,a}(x)$ is defined by*

$$R_{n,a}(x) = f(x) - P_{n,a}(x)$$

*Then*

$$R_{n,a}(x) = \frac{f^{(n+1)}(t)}{(n+1)!}(x-a)^{n+1} \quad \text{for some } t \text{ in } (a, x)$$

*Proof.* The function $R_{n,a}$ satisfies the conditions of the preceding Lemma by the definition of the Taylor polynomial, so

$$\frac{R_{n,a}(x)}{(x-a)^{n+1}} = \frac{R_{n,a}^{(n+1)}(t)}{(n+1)!}$$

for some $t$ in $(a, x)$. But,

$$R_{n,a}^{(n+1)} = f^{(n+1)}$$

since $R_{n,a} - f$ is a polynomial of degree $n$. Substituting gives the Lagrange form for the remainder, as desired. ∎

**Example 4.1.3.** Applying 17 to the functions sin, cos, and exp, with $a = 0$, we obtain the following formulas:

$$\sin x = \sum_{i=0}^{n} \frac{(-1)^i x^{2i+1}}{(2i+1)!} + \frac{\sin^{(2n+2)}(t)}{(2n+2)!} x^{2n+2}$$

$$\cos x = \sum_{i=0}^{n} \frac{(-1)^i x^{2i}}{(2i)!} + \frac{\cos^{(2n+1)}(t)}{(2n+1)!} x^{2n+1}$$

$$\exp x = \sum_{i=0}^{n} \frac{x^n}{n!} + \frac{\exp t}{(n+1)!} x^{n+1}$$

(of course, we could go one power higher in the remainder terms for sin and cos)

## 4.2.0 Infinite Sequences

**Definition 4.2.1.** *An* **infinite sequence** *of real numbers is a real valued function whose domain is $\mathbb{N}$.*

**Definition 4.2.2.** *A sequence $\{a_n\}$* **converges to** $\ell \in \mathbb{R}$*, denoted by $\lim_{n\to\infty} a_n = \ell$, if for every $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that for all $n \in \mathbb{N}$, if $n \geqslant N$ then*

$$|a_n - \ell| < \epsilon$$

*A sequence $\{a_n\}$ is said to* **converge** *if such an $\ell$ exists, and to* **diverge** *otherwise.*

**Theorem 4.2.1** (Limit Laws)**.** *If $\lim_{n\to\infty} a_n$ and $\lim_{n\to\infty} b_n$ both exist, then*

$$\lim_{n\to\infty}(a_n + b_n) = \lim_{n\to\infty} a_n + \lim_{n\to\infty} b_n$$
$$\lim_{n\to\infty} a_n \cdot b_n = \lim_{n\to\infty} a_n \cdot \lim_{n\to\infty} b_n$$

*moreover, if $\lim_{n\to\infty} b_n \neq 0$, then $b_n \neq 0$ for all $n$ greater than some $N$, and*

$$\lim_{n\to\infty} a_n/b_n = \lim_{n\to\infty} a_n/\lim_{n\to\infty} b_n$$

*Proof.* (To be completed) ∎

**Theorem 4.2.2.** *Let $f$ be a function defined in an open interval containing $c$, except perhaps at $c$ itself, with*

$$\lim_{x\to c} f(x) = l$$

*Suppose that $\{a_n\}$ is a sequence such that*

1. *each $a_n$ is in the domain of $f$,*
2. *each $a_n \neq c$,*
3. $\lim_{n\to\infty} a_n = c$

*Then the sequence $\{f(a_n)\}$ satisfies*

$$\lim_{n\to\infty} f(a_n) = l$$

*Conversely, if this is true for every sequence $\{a_n\}$ satisfying the above conditions, then $\lim_{x\to c} f(x) = l$.*

*Proof.* Suppose first that $\lim_{x\to c} f(x) = l$. Then for every $\varepsilon > 0$ there is a $\delta > 0$ such that, for all $x$,

$$\text{if } 0 < |x - c| < \delta, \text{ then } |f(x) - l| < \varepsilon$$

If the sequence $\{a_n\}$ satisfies $\lim_{n\to\infty} a_n = c$, then there is a natural number $N$ such that for all $n \in \mathbb{N}$,

$$\text{if } n \geqslant N, \text{ then } 0 < |a_n - c| < \delta$$

By our choice of $\delta$ this implies that

$$|f(a_n) - l| < \varepsilon$$

showing that $\lim_{n\to\infty} f(a_n) = l$.

Suppose, conversely, that $\lim_{n\to\infty} f(a_n) = l$ for every sequence $\{a_n\}$ satisfying our three conditions. If $\lim_{x\to c} f(x) = l$ were not true, there would be some $\varepsilon > 0$ such that for all $\delta > 0$ there exists $x$ such that $0 < |x - c| < \delta$ but $|f(x) - l| \geqslant \varepsilon$. In particular, for each $n$ there would exist $x_n$ such that $0 < |x_n - c| < 1/n$, but $|f(x_n) - l| \geqslant \varepsilon$. Define a sequence $\{x_n\}$ using these $x_n$. Then $x_n$ is in the domain of $f$ for each $n$, and as $0 < |x_n - c|$ for each $n$, $x_n \neq c$ for all $n$. Moreover, for all $\varepsilon' > 0$ there exists $N \in \mathbb{N}$ such that $\varepsilon' > 1/N > 0$ (by the Archimedean Property of $\mathbb{R}$), so for all $n \geqslant N$, $0 < |x_n - c| < 1/n < \varepsilon'$. Consequently, $\lim_{n\to\infty} x_n = c$, so the sequences satisfies all of our initial conditions. However, then by assumption $\lim_{n\to\infty} f(x_n) = l$. But, by construction, for $\varepsilon$ we have that for all $n \in \mathbb{N}$, $0 < |x_n - c| < 1/n$ but $|f(x_n) - l| \geqslant \varepsilon$, so $f(x_n)$ does not converge to $l$, contradicting our hypothesis. Thus $\lim_{x\to c} f(x) = l$ must be true. ∎

**Definition 4.2.3.** *A sequence $\{a_n\}$ is* **increasing** *if $a_{n+1} > a_n$ for all $n$,* **nondecreasing** *if $a_{n+1} \geqslant a_n$ for all $n$, and* **bounded above** *if there is a number $M$ such that $a_n \leqslant M$ for all $n$. Similarly, a sequence $\{a_n\}$ is* **decreasing** *if $a_{n+1} < a_n$ for all $n$,* **nonincreasing** *if $a_{n+1} \leqslant a_n$ for all $n$, and* **bounded below** *if there is a number $m$ such that $a_n \geqslant m$ for all $n$.*

**Theorem 4.2.3.** *If $\{a_n\}$ is nondecreasing and bounded above, then $\{a_n\}$ converges.*

*Proof.* The set $A := \{a_n : n \in \mathbb{N}\}$ is, by assumption, bounded above, so $A$ has a least upper bound $\alpha \in \mathbb{R}$. We claim that $\lim_{n\to\infty} a_n = \alpha$. If $\varepsilon > 0$, then $\alpha - \varepsilon$ is not an upper bound for $A$ so there exists $a_N$ in $A$ such that $a_N > \alpha - \varepsilon$, so $\alpha - a_N < \varepsilon$. Then for all $n \geqslant N$, $a_n \geqslant a_N$ since $\{a_n\}$ is nondecreasing so

$$|\alpha - a_n| = \alpha - a_n \leqslant \alpha - a_N < \varepsilon$$

Consequently, we conclude that $\lim_{n\to\infty} a_n = \alpha$. ∎

**Definition 4.2.4.** *A* **subsequence** *of a sequence $\{a_n\}$ is a sequence*

$$a_{n_1}, a_{n_2}, a_{n_3}, \ldots$$

*where the $n_j$ are natural numbers with $n_1 < n_2 < n_3 < \ldots.$*

**Lemma 4.2.4.** *Any sequence $\{a_n\}$ contains a subsequence which is either nondecreasing or nonincreasing.*

*Proof.* Call a natural number $n$ a "peak point" of a sequence $\{a_n\}$ if $a_m < a_n$ for all $m > n$.

Case 1. The sequence has infinitely many peak points. In this case, if $n_1 < n_2 < n_3 < \ldots$ are the peak points, then $a_{n_1} > a_{n_2} > a_{n_3} > \ldots$, so $\{a_{n_j}\}$ is a nonincreasing subsequence of $\{a_n\}$.

Case 2. THe sequence has only finitely many peak points. In this case, let $n_1$ be greater than all peak points. Since $n_1$ is not a peak point, there is some $n_2 > n_1$ such that $a_{n_2} \geqslant a_{n_1}$. Since $n_2$ is not a peak point, there is some $n_3 > n_2$ such that $a_{n_3} > a_{n_2}$. Suppose there exists $k \geqslant 3$ such that for all $1 \leqslant m < k$, $a_{n_m} \leqslant a_{n_{m+1}}$ and $n_m < n_{m+1}$. Then since $n_k$ is not a peak point, there is some $n_{k+1}$ such that $n_{k+1} > n_k$ and $a_{n_{k+1}} \geqslant a_{n_k}$. Thus, by recursive definition we have constructed a nondecreasing subsequence $\{a_{n_k}\}$ of $\{a_n\}$.

∎

**Corollary 4.2.5.** *The Bolzano-Weierstrass Theorembolz Every bounded sequence has a convergent subsequence.*

**Definition 4.2.5.** *A sequence $\{a_n\}$ is a* **Cauchy sequence** *if for every $\varepsilon > 0$ there is a natural number $N$ such that, for all $m, n \in \mathbb{N}$, if $m, n \geqslant N$, then*

$$|a_n - a_m| < \varepsilon$$

*(This can be written as $\lim_{m,n\to\infty} |a_m - a_n| = 0$)*

**Theorem 4.2.6.** *A sequence $\{a_n\}$ converges if and only if it is a Cauchy sequence.*

*Proof.* First assume that $\lim_{n\to\infty} a_n = l$ for some $l \in \mathbb{R}$. Then given $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that if $n \geqslant N$, then $|a_n - l| < \varepsilon/2$. Hence, if $m, n \geqslant N$, then

$$|a_n - a_m| \leqslant |a_n - l| + |a_m - l| < \varepsilon/2 + \varepsilon/2 = \varepsilon$$

Thus, $\{a_n\}$ is Cauchy.

Conversely, suppose that $\{a_n\}$ is a Cauchy sequence. I claim that this implies $\{a_n\}$ is bounded. First, for $\varepsilon = 1 > 0$, there exists $N \in \mathbb{N}$ such that if $m, n \geqslant N$, then $|a_m - a_n| < \varepsilon = 1$. In particular, for all $n \geqslant N$, $|a_N - a_n| < 1$. Take $M = \max(a_N + 1, a_{N-1}, ..., a_1)$, and $m = \min(a_N - 1, a_{N-1}, ..., a_1)$. Then for all $k \leqslant N$ we have that $m \leqslant a_k \leqslant M$. On the other hand, for $k \geqslant N$, we have that $a_N - 1 < a_k < a_N + 1$, so $m < a_k < M$. Thus $\{a_n : n \in \mathbb{N}\}$ is bounded.

Then, by **??** $\{a_n\}$ has a convergent subsequence $\{a_{n_k}\}$. Let $\lim_{k\to\infty} a_{n_k} = l$, for some $l \in \mathbb{R}$. Then, fix $\varepsilon > 0$. It follows that there exist $K, K' \in \mathbb{N}$ such that for $m, n \geqslant K$ and $j \geqslant K'$, $|a_m - a_n| < \varepsilon/2$, while $|a_{n_j} - l| < \varepsilon/2$. Note that $n_j \geqslant j$, since the sequence $\{n_k\}$ is increasing. Then, for all $i \geqslant \max(K, K')$ we have that

$$|a_i - l| \leqslant |a_i - a_{n_i}| + |a_{n_i} - l| < \varepsilon/2 + \varepsilon/2 = \varepsilon$$

Therefore, by definition $\{a_n\}$ converges to $l$ as well. ∎

## 4.3.0 Infinite Series

**Definition 4.3.1.** *The sequence $\{a_n\}$ is* **summable** *if the sequence $\{s_n\}$ converges, where*

$$s_n = \sum_{i=1}^{n} a_i$$

*is the $n$-th* **partial sum**. *In this case, $\lim_{n\to\infty} s_n$ is denoted by*

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

*and is called the* **sum** *of the sequence $\{a_n\}$.*

**Remark 4.3.1.** If $\{a_n\}$ and $\{b_n\}$ are summable, then

$$\sum_{n=1}^{\infty}(a_n + b_n) = \sum_{i=1}^{\infty} a_n + \sum_{i=1}^{\infty} b_n$$
$$\sum_{n=1}^{\infty} c \cdot a_n = c \cdot \sum_{i=1}^{\infty} a_n$$

for all $c \in \mathbb{R}$.

## Theorem 18 (The Cauchy Criterion).

*The sequence $\{a_n\}$ is summable if and only if for all $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that for all $m \geqslant n \in \mathbb{N}$, if $m, n \geqslant N$, then*

$$\left|\sum_{i=1}^{m} a_i - \sum_{i=1}^{n} a_i\right| = \left|\sum_{i=n+1}^{m} a_i\right| < \varepsilon$$

**Remark 4.3.2.** This result is a direct conseqeunce of the fact that a sequence in $\mathbb{R}$ is Cauchy if and only if it converges applied to the sequence of partial sums for $\{a_n\}$.

## Theorem 19 (The Vanishing Condition).

*If $\{a_n\}$ is summable, then*

$$\lim_{n\to\infty} a_n = 0$$

*Proof.* If $\lim_{n\to\infty} s_n = l$, then

$$\begin{aligned}\lim_{n\to\infty} a_n &= \lim_{n\to\infty}(s_n - s_{n-1}) = \lim_{n\to\infty} s_n - \lim_{n\to\infty} s_{n-1} \\ &= l - l = 0\end{aligned}$$

∎

**Example 4.3.1.** The ***geometric series*** are of the form

$$\sum_{n=0}^{\infty} r^n = 1 + r + r^2 + r^3 + \dots$$

For $|r| \geqslant 1$, $\lim_{n\to\infty} r^n \neq 0$, so $\{r^n\}$ is not summable. On the other hand, if $|r| < 1$ the sequence is summable. First write

$$s_n = 1 + r + r^2 + ... + r^n, \quad \text{and} \quad rs_n = r + r^2 + r^3 + ... + r^{n+1}$$

It follows that

$$s_n(1 - r) = 1 - r^{n+1}$$

so

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

since $r \neq 1$. Finally, it follows that

$$\sum_{n=0}^{\infty} r^n = \lim_{n\to\infty} s_n = \lim_{n\to\infty} \frac{1-r^{n+1}}{1-r} = \frac{1}{1-r}$$

as $|r| < 1$.

**Definition 4.3.2.** *A sequence $\{a_n\}$ such that $a_n \geqslant 0$ for all $n \in \mathbb{N}$ is said to be* **nonnegative**.

## Theorem 20 (The Boundedness Criterion).

*A nonnegative sequence $\{a_n\}$ is summable if and only if the set of partial sums $s_n$ is bounded.*

*Proof.* Since $\{a_n\}$ is nonnegative, $\{s_n\}$ is nondecreasing. Hence from our previous results on monotone sequences, $\{s_n\}$ converges if and only if $\{s_n\}$ is bounded. ∎

## Theorem 21 (The Comparison Test).

*Suppose that $\{a_n\}$ and $\{b_n\}$ are sequences such that $0 \leqslant a_n \leqslant b_n$ for all $n \in \mathbb{N}$. Then if $\sum\limits_{n=1}^{\infty} b_n$ converges, so does $\sum\limits_{n=1}^{\infty} a_n$.*

*Proof.* Let $s_n$ denote the $n$-th partial sum of $\{a_n\}$, and let $t_n$ denote the $n$-th partial sum of $\{b_n\}$. Then $0 \leqslant s_n \leqslant t_n$ for all $n \in \mathbb{N}$. Now $\{t_n\}$ converges by assumption, so it is bounded. Hence, there exists $M \in \mathbb{R}$ such that $0 \leqslant s_n \leqslant t_n \leqslant M$ for all $n \in \mathbb{N}$, so $\{s_n\}$ is also bounded. Thus by 20 $\{a_n\}$ is summable, so by definition $\sum\limits_{n=1}^{\infty} a_n$ converges. ∎

## Theorem 22 (The Limit Comparison Test).

*If $a_n, b_n > 0$ for convergent sequences $\{a_n\}$ and $\{b_n\}$, and $\lim\limits_{n\to\infty} a_n/b_n = c \neq 0$, then $\sum\limits_{n=1}^{\infty} a_n$ converges if and only if $\sum\limits_{n=1}^{\infty} b_n$ converges.*

*Proof.* Suppose $\sum\limits_{n=1}^{\infty} b_n$ converges. Since $\lim\limits_{n\to\infty} a_n/b_n = c$, there is some $N$ such that

$$a_n/b_n - c \leqslant c \implies a_n \leqslant 2cb_n, \quad \text{for } n \geqslant N$$

But the sequence $2c\sum\limits_{n=N}^{\infty} b_n$ certainly converges. Then by 21 we have that $\sum\limits_{n=N}^{\infty} a_n$ converges, and this implies convergence of the whole series $\sum\limits_{n=1}^{\infty} a_n$, which only has finitely many additional terms.

Note that

$$\lim_{n\to\infty} b_n/a_n = \frac{1}{\lim\limits_{n\to\infty} a_n/b_n} = 1/c \neq 0$$

so the converse follows immediately. ∎

## Theorem 23 (The Ratio Test).

*Let $\{a_n\}$ be a positive sequence, and suppose that*

$$\lim_{n\to\infty} \frac{a_{n+1}}{a_n} = r$$

*for some $r \geqslant 0$. Then $\sum_{n=1}^{\infty} a_n$ converges if $r < 1$. On the other hand, if $r > 1$, then the terms $a_n$ are unbounded, so $\sum_{n=1}^{\infty} a_n$ diverges.*

*Proof.* Suppose first that $r < 1$. Choose any number $s$ with $r < s < 1$. The hypothesis $\lim_{n\to\infty} a_{n+1}/a_n = r < 1$ implies that there is some $N \in \mathbb{N}$ such that if $n \geqslant N$,

$$a_{n+1}/a_n - r < s - r \implies a_{n+1}/a_n < s$$

This can be written as $a_{n+1} < sa_n$. Thus,

$$\begin{aligned}
a_{N+1} &< sa_N \\
a_{N+2} &< sa_{N+1} < s^2 a_N \\
&\vdots \\
a_{N+k} &< s^k a_N
\end{aligned}$$

Since $\sum_{k=0}^{\infty} a_N s^k = a_N \sum_{k=0}^{\infty} s^k$ converges, since $|s| < 1$, 21 shows that

$$\sum_{n=N}^{\infty} a_n = \sum_{k=0}^{\infty} a_{N+k}$$

converges as $a_{N+k} < a_N s^k$ for all $k \geqslant 0$. This implies that $\sum_{n=0}^{\infty} a_n$ as a whole converges.

If $r > 1$, choose some $s \in \mathbb{R}$ such that $1 < s < r$. Then there is a number $N \in \mathbb{N}$ such that

$$r - a_{n+1}/a_n < r - s \implies s < a_{n+1}/a_n$$

for all $n \geqslant N$. This implies that

$$a_{N+k} > a_N s^k,$$

for all $k \in \mathbb{N}$, so the terms are unbounded. ∎

## Theorem 24 (The Integral Test).

*Suppose that $f$ is positive and decreasing on $[1, \infty)$, and that $f(n) = a_n$ for all $n \in \mathbb{N}$. Then $\sum_{n=1}^{\infty} a_n$ converges if and only if the limit*

$$\int_1^{\infty} f = \lim_{A\to\infty} \int_1^A f$$

*exists.*

*Proof.* The existence of $\lim\limits_{A\to\infty} \int_1^A f$ is equivalent to the convergence of the series

$$\int_1^2 f + \int_2^3 f + \int_3^4 f + \ldots$$

Since $f$ is decreasing, we have

$$f(n+1) < \int_n^{n+1} f < f(n)$$

The first half of this double inequality shows that the series $\sum\limits_{n=1}^{\infty} a_{n+1}$ may be compared to the series $\sum\limits_{n=1}^{\infty} \int_n^{n+1}$, proving that $\sum\limits_{n=1}^{\infty} a_{n+1}$ (and hence $\sum\limits_{n=1}^{\infty} a_n$) converges if $\lim\limits_{A\to\infty} \int_1^A f$ exists.

The second half of the inequality shows that the series $\sum\limits_{n=1}^{\infty} \int_n^{n+1} f$ may be compared to the series $\sum\limits_{n=1}^{\infty} a_n$, proving that $\lim\limits_{A\to\infty} \int_1^A f$ must exist if $\sum\limits_{n=1}^{\infty} a_n$ converges. ∎

**Corollary 4.3.1.** *The series $\sum\limits_{n=1}^{\infty} 1/n^p$ converges if and only if $p > 1$.*

*Proof.* If $p < 0$ then $\lim\limits_{n\to\infty} 1/n^p \neq 0$. If $p > 0$, the convergence of $\sum\limits_{n=1}^{\infty} 1/n^p$ is equivalent, by 24, to the existence of

$$\lim_{A\to\infty} \int_1^A \frac{1}{x^p} dx$$

Now, observe that

$$\int_1^A \frac{1}{x^p} dx = \begin{cases} -\frac{1}{p-1} \cdot \frac{1}{A^{p-1}} + \frac{1}{p-1}, & p \neq 1 \\ \log A, & p = 1 \end{cases}$$

This shows that $\lim\limits_{A\to\infty} \int_1^A 1/x^p dx$ exists if $p > 1$, but not if $p \leqslant 1$. Thus, $\sum\limits_{n=1}^{\infty} 1/n^p$ converges precisely for $p > 1$. ∎

**Definition 4.3.3.** *The series $\sum\limits_{n=1}^{\infty} a_n$ is* **absolutely convergent** *if the series $\sum\limits_{n=1}^{\infty} |a_n|$ is convergent. (In formal language, the sequence $\{a_n\}$ is said to be* **absolutely summable** *if the sequence $\{|a_n\}$ is summable.)*

**Theorem 4.3.2.** *Every absolutely convergent series is convergent. Moreover, a series is absolutely convergent if and only if the series formed from its positive terms and the series formed from its negative terms both converge.*

*Proof.* If $\sum\limits_{n=1}^{\infty} |a_n|$ converges, then, by 18,

$$\lim_{m,n\to\infty} \sum_{i=n+1}^{m} |a_i| = 0$$

Since
$$\left|\sum_{i=n+1}^{m} a_i\right| \leqslant \sum_{i=n+1}^{m} |a_i|$$
it follows that
$$\lim_{m,n\to\infty} \sum_{i=n+1}^{m} a_i = 0$$
which shows that $\sum_{n=1}^{\infty} a_n$ converges.

To prove the second portion of the theorem, let $\{a_n^+\}$ and $\{a_n^-\}$ be sequences defined by
$$a_n^+ = \begin{cases} a_n, & \text{if } a_n \geqslant 0 \\ 0, & \text{if } a_n \leqslant 0 \end{cases}$$
$$a_n^- = \begin{cases} a_n, & \text{if } a_n \leqslant 0 \\ 0, & \text{if } a_n \geqslant 0 \end{cases}$$
It follows that
$$\sum_{n=1}^{\infty} |a_n| = \sum_{n=1}^{\infty} [a_n^+ - a_n^-] = \sum_{n=1}^{\infty} a_n^+ - \sum_{n=1}^{\infty} a_n^-$$
so if $\sum_{n=1}^{\infty} a_n^+$ and $\sum_{n=1}^{\infty} a_n^-$ are convergent, so is $\sum_{n=1}^{\infty} |a_n|$.

Conversely, suppose $\sum_{n=1}^{\infty} |a_n|$ converges. Then by our initial argument $\sum_{n=1}^{\infty} a_n$ converges also. Therefore,
$$\sum_{n=1}^{\infty} a_n^+ = \sum_{n=1}^{\infty} \frac{1}{2}[a_n + |a_n|] = \frac{1}{2}\left(\sum_{n=1}^{\infty} a_n + \sum_{n=1}^{\infty} |a_n|\right)$$
and
$$\sum_{n=1}^{\infty} a_n^- = \sum_{n=1}^{\infty} \frac{1}{2}[a_n - |a_n|] = \frac{1}{2}\left(\sum_{n=1}^{\infty} a_n - \sum_{n=1}^{\infty} |a_n|\right)$$
both converge. ■

**Remark 4.3.3.** A consequent of this result is that every convergent series with positive terms can be used to obtain infinitely many other convergent series simply by putting minus sings at random.

**Definition 4.3.4.** *A convergent series which is not absolutely convergent is said to be* **conditionally convergent**.

## Theorem 25 (Leibniz's Theorem).

*Suppose that $\{a_n\}$ is a non-decreasing non-negative sequence such that*
$$\lim_{n\to\infty} a_n = 0$$
*Then the series*
$$\sum_{n=1}^{\infty} (-1)^{n+1} a_n = a_1 - a_2 + a_3 - a_4 + \ldots$$
*converges.*

*Proof.* First, observe that

1. $s_2 \leqslant s_4 \leqslant s_6 \leqslant \ldots$
2. $s_1 \geqslant s_3 \geqslant s_5 \geqslant \ldots$
3. $s_k \leqslant s_l$ if $k$ is even and $l$ is odd.

Indeed, note that for all $n$, $a_{2n+1} \geqslant a_{2n+2}$, so $a_{2n+1} - a_{2n+2} \geqslant 0$, so

$$s_{2n+2} = s_{2n} + a_{2n+1} - a_{2n+2} \geqslant s_{2n}$$

and similarly as $a_{2n+2} \geqslant a_{2n+3}$, we have

$$s_{2n+3} = s_{2n+1} - a_{2n+2} + a_{2n+3} \leqslant s_{2n+1}$$

To prove the third inequality, first notice that

$$s_{2n} = s_{2n-1} - a_{2n} \leqslant s_{2n-1}$$

since $a_{2n} \geqslant 0$. Now, if $k$ is even and $l$ is odd, choose $n$ such that $k \leqslant 2n$ and $l \leqslant 2n-1$. Then

$$s_k \leqslant s_{2n} \leqslant s_{2n-1} \leqslant s_l$$

which proves the third inequality.

Now, the sequence $\{s_{2n}\}$ converges, because it is nondecreasing and is bounded above (by $s_l$ for any odd $l$). Let $\alpha = \sup\{s_{2n}\} = \lim\limits_{n\to\infty} s_{2n}$. Similarly, let $\beta = \inf\{s_{2n+1}\} = \lim\limits_{n\to\infty} s_{2n+1}$. It follows from our third inequality that $\alpha \leqslant \beta$; since

$$s_{2n+1} - s_{2n} = a_{2n+1} \quad \text{and} \quad \lim_{n\to\infty} a_n = 0$$

it is actually the case that $\alpha = \beta$. This proves that $\alpha = \beta = \lim\limits_{n\to\infty} s_n$. ■

**Definition 4.3.5.** *A sequence $\{a_n\}$ is a* **rearrangement** *of a sequence $\{a_n\}$ if each $b_n = a_{f(n)}$ where $f$ is a certain permutation on the natural numbers.*

**Theorem 4.3.3.** *If $\sum\limits_{n=1}^{\infty} a_n$ converges, but does not converge absolutely; then for any number $\alpha$ there is a rearrangement $\{b_n\}$ of $\{a_n\}$ such that $\sum\limits_{n=1}^{\infty} b_n = \alpha$.*

*Proof.* Let $\sum\limits_{n=1}^{\infty} p_n$ denote the series formed from the positive terms of $\{a_n\}$ and let $\sum\limits_{n=1}^{\infty} q_n$ denote the series of negative terms. It follows from Theorem 4.3.2 that at least one of these series does not converge. As a matter of fact, both must fail to converge, for if one had bounded partial sums, and the other had unbounded partial sums, then the original series $\sum\limits_{n=1}^{\infty} a_n$ would also have unbounded partial sums, contradicting the assumption that it converges.

Let $\alpha$ be any number. Assume, for simplicity, that $\alpha > 0$. Since the series $\sum_{n=1}^{\infty} p_n$ there is a number $N$ such that

$$\sum_{n=1}^{N} p_n > \alpha$$

We will choose $N_1$ to be the smallest $N$ with this property. This means that

$$\sum_{n=1}^{N_1-1} p_n \leqslant \alpha \quad \text{and} \quad \sum_{n=1}^{N_1} p_n > \alpha$$

Then if $S_1 = \sum_{n=1}^{N_1} p_n$, we have $S_1 - \alpha \leqslant p_{N_1}$. Next, choose the smallest integer $M_1$ for which

$$T_1 = S_1 + \sum_{n=1}^{M_1} q_n < \alpha$$

As before, we have $\alpha - T_1 \leqslant -q_{M_1}$. We continue this process indefinitely. The sequence

$$p_1, ..., p_{N_1}, q_1, ..., q_{M_1}, q_{N_1+1}, ..., p_{N_2}, ...$$

is a rearrangement of $\{a_n\}$. The partial sums of this rearrangement increase to $S_1$, then decrease to $T_1$, then increase to $S_2$, etc. To complete the proof we note that $|S_k - \alpha|$ and $|T_k - \alpha|$ are less that or equal to $p_{N_k}$ or $-q_{M_k}$, respectively, and that these terms, being numbers of the original sequence $\{a_n\}$, must decrease to 0, since $\sum_{n=1}^{\infty} a_n$ converges. ■

**Theorem 4.3.4.** *If $\sum_{n=1}^{\infty} a_n$ converges absolutely, and $\{b_n\}$ is any rearrangement of $\{a_n\}$, then $\sum_{n=1}^{\infty} b_n$ also converges (absolutely), and in particular*

$$\sum_{n=1}^{\infty} a_n = \sum_{n=1}^{\infty} b_n$$

*Proof.* Denote the partial sums of $\{a_n\}$ by $s_n$, and the partial sums of $\{b_n\}$ by $t_n$. Suppose that $\varepsilon > 0$. Since $\sum_{n=1}^{\infty} a_n$ converges, there is some $N$ such that

$$\left|\sum_{n=1}^{\infty} a_n - s_N\right| < \varepsilon$$

Moreover, since $\sum_{n=1}^{\infty} |a_n|$ converges, we can also choose $N'$ so that

$$\sum_{n=1}^{\infty} |a_n| - (|a_1| + \ldots + |a_{N'}|) < \varepsilon$$

so that
$$\sum_{n=N+1}^{\infty} |a_n| < \varepsilon$$
Choose $M$ such that each $a_1, ..., a_N$ appear among $b_1, ..., b_M$. Then whenever $m > M$, the difference $t_m - s)N$ is the sum of certain $a_i$, where $i > N$. Consequently,
$$|t_m - s_N| \leqslant \sum_{n=N+1}^{\infty} |a_n| < \varepsilon$$
Thus, if $m >$, then
$$\begin{aligned}\left|\sum_{n=1}^{\infty} a_n - t_m\right| &= \left|\sum_{n=1}^{\infty} a_n - s_N - (t_m - s_N)\right| \\ &\leqslant \left|\sum_{n=1}^{\infty} a_n - s_N\right| + |(t_m - s_N)| \\ &< \varepsilon + \varepsilon\end{aligned}$$
Since this is true for every $\varepsilon > 0$, the series $\sum_{n=1}^{\infty} b_n$ converges to $\sum_{n=1}^{\infty} a_n$.

To show that $\sum_{n=1}^{\infty} b_n$ converges absolutely, note that $\{|b_n|\}$ is a rearrangement of $\{|a_n|\}$; since $\sum_{n=1}^{\infty} |a_n|$ converges absolutely, $\sum_{n=1}^{\infty} |b_n|$ converges by the first part of the theorem. ■

**Theorem 4.3.5.** *If $\sum_{n=1}^{\infty} a_n$ and $\sum_{n=1}^{\infty} b_n$ converge absolutely, and $\{c_n\}$ is any sequence containing the products $a_i b_j$ for each pair $(i, j) \in \mathbb{N} \times \mathbb{N}$, then*
$$\sum_{n=1}^{\infty} c_n = \sum_{n=1}^{\infty} a_n \cdot \sum_{n=1}^{\infty} b_n$$

*Proof.* Notice first that the sequence
$$p_L = \sum_{i=1}^{L} |a_i| \cdot \sum_{j=1}^{L} |b_j|$$
converges since $\{a_n\}$ and $\{b_n\}$ are absolutely convergent, and since the limit of a product is the product of the limits. So $\{p_L\}$ is a Cauchy sequence, which means that for any $\varepsilon > 0$, if there exists $N$ such that for all $L, L' \geqslant N$,
$$\left|\sum_{i=1}^{L'} |a_i| \cdot \sum_{j=1}^{L'} |b_j| - \sum_{i=1}^{L} |a_i| \cdot \sum_{j=1}^{L} |b_j|\right| < \varepsilon/2$$
It follows that
$$\sum_{i \text{ or } j>L} |a_i| \cdot |b_i| \leqslant \varepsilon/2 < \varepsilon \tag{1}$$

Now suppose that $N$ is such that the terms $c_n$ for $n \leqslant N$ include all $a_i b_j$ for $i, j \leqslant L$. Then the difference

$$\sum_{n=1}^{N} c_n - \sum_{i=1}^{L} a_i \cdot \sum_{j=1}^{L} b_j$$

consists of terms $a_i b_j$ with $i > L$ of $j > L$, so

$$\left| \sum_{n=1}^{N} c_n - \sum_{i=1}^{L} a_i \cdot \sum_{j=1}^{L} b_j \right| \leqslant \sum_{i \text{ or } j > L} |a_i| \cdot |b_j| < \varepsilon \tag{2}$$

But since the limit of a product is the product of the limits we also have

$$\left| \sum_{i=1}^{\infty} a_i \cdot \sum_{j=1}^{\infty} b_j - \sum_{i=1}^{L} a_i \cdot \sum_{j=1}^{L} b_j \right| < \varepsilon \tag{3}$$

for sufficiently large $L$. Consequently, if we choose $L$ and then $N$ large enough, we will have

$$\begin{aligned} \left| \sum_{i=1}^{\infty} a_i \cdot \sum_{j=1}^{\infty} b_j - \sum_{i=1}^{N} c_i \right| \leqslant \left| \sum_{i=1}^{\infty} a_i \cdot \sum_{j=1}^{\infty} b_j - \sum_{i=1}^{L} a_i \cdot \sum_{j=1}^{L} b_j \right| & \\ + \left| \sum_{n=1}^{N} c_n - \sum_{i=1}^{L} a_i \cdot \sum_{j=1}^{L} b_j \right| & \\ < 2\varepsilon & \end{aligned}$$

which proves the theorem. ■

# 4.4.0 Uniform Convergence and Power Series

**Remark 4.4.1.** We are now interested in the study of series of functions, or in other words functions of the form

$$f(x) = f_1(x) + f_2(x) + f_3(x) + \ldots$$

In such a situation $\{f_n\}$ will be some sequence of functions; for each $x$ we obtain a sequence of numbers $\{f_n(x)\}$, and $f(x)$ is the sum of this sequence. Recall that each sum $f_1(x) + f_2(x) + f_3(x) + \ldots$ is, by definition, the limit of the sequence $f_1(x), f_1(x) + f_2(x), f_1(x) + f_2(x) + f_3(x), \ldots$. If we define a new sequence of functions $\{s_n\}$ by

$$s_n = f_1 + \ldots + f_n$$

then we can express this fact more succinctly by writing

$$f(x) = \lim_{n\to\infty} s_n(x)$$

for some $x \in \mathbb{R}$.

**Remark 4.4.2.** First let us consider functions of the form

$$f(x) = \lim_{n\to\infty} f_n(x)$$

All this form may seem simple, it is very important to note that ***nothing one would hope to be true actually is***. Instead we have a flurry of lovely counter-examples.

**Example 4.4.1** (Counter-Example 1)**.** Even if each $f_n$ is continuous, the function $f$ may not be! Indeed, consider the sequence of functions

$$f_n(x) = \begin{cases} x^n, & 0 \leqslant x \leqslant 1 \\ 1, & x \geqslant 1 \end{cases}$$

These functions are all continuous, but the function $f(x) = \lim_{n\to\infty} f_n(x)$ is not continuous; in fact,

$$\lim_{n\to\infty} f_n(x) = \begin{cases} 0, & 0 \leqslant x < 1 \\ 1, & x \geqslant 1 \end{cases}$$

Another example of this phenomenon is illustrated by the family of functions

$$f_n(x) = \begin{cases} -1, & x \leqslant -\frac{1}{n} \\ nx, & -\frac{1}{n} \leqslant x \leqslant \frac{1}{n} \\ 1, & \frac{1}{n} \leqslant x \end{cases}$$

In this case, if $x < 0$ $f_n(x)$ is eventually $-1$, and if $x > 0$, then $f_n(x)$ is eventually 1, while $f_n(0) = 0$ for all $n$. Thus,

$$\lim_{n\to\infty} f_n(x) = \begin{cases} -1, & x < 0 \\ 0, & x = 0 \\ 1, & x > 0 \end{cases}$$

so once again $f(x) = \lim_{n\to\infty} f_n(x)$ is not continuous.

**Example 4.4.2** (Counter-Example 2)**.** It is even possible to produce a sequence of differentiable functions $\{f_n\}$ for which the function $f(x) = \lim_{n\to\infty} f_n(x)$ is not continuous. One such sequence is

$$f_n(x) = \begin{cases} -1, & x \leqslant -\frac{1}{n} \\ \sin\left(\frac{n\pi x}{2}\right), & -\frac{1}{n} \leqslant x \leqslant \frac{1}{n} \\ 1, & \frac{1}{n} \leqslant x \end{cases}$$

These functions are differentiable, but we still have

$$\lim_{n\to\infty} f_n(x) = \begin{cases} -1, & x < 0 \\ 0, & x = 0 \\ 1, & x > 0 \end{cases}$$

**Definition 4.4.1.** *If $f$ is a function defined on some set $A$, and a sequence of functions $\{f_n\}$, all defined on the same set $A$, are such that only*

$$f(x) = \lim_{n\to\infty} f_n(x)$$

*for all $x \in A$. Precisely, $\{f_n\}$ is said to* **converge pointwise to $f$ on $A$** *if for all $\varepsilon > 0$, and for all $x \in A$, there is some $N$ such that if $n \geqslant N$, then $|f(x) - f_n(x)| < \varepsilon$.*

**Definition 4.4.2 (Pointwise Convergence (alt.)).** *Suppose $S \subseteq \mathbb{R}$ and $f_n : S \to \mathbb{R}$ is a real-valued function for each $n \in \mathbb{N}$. We say that the sequence of functions $\{f_n\}$* **converges pointwise on $S$** *to $f : S \to \mathbb{R}$ if for every $x \in S$ and every $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that if $n \geqslant N$, $|f_n(x) - f(x)| < \varepsilon$.*

**Example 4.4.3.** Take $f_n(x) = x^n$ on $S = [0, 1]$. If $0 \leqslant x < 1$ notice that $\lim_{n\to\infty} x^n = 0$ (geometric sequence). If $x = 1$, then $f_n(1) = 1^n = 1$, which converges to 1 as $n$ goes to infinity. Thus, $f)n$ converges pointwise to

$$f(x) = \begin{cases} 0 & 0 \leqslant x < 1 \\ 1 & x = 1 \end{cases}$$

Notice each $f_n$ is continuous on $[0, 1]$, but $f$ is not.

This example answers the following question in the negative:

**Question 4.4.3.** Suppose $f_n$ converges pointwise to $f$ on $S \subseteq \mathbb{R}$. If $a \in S'$ (an accumulation/limit point for $S$), $\lim_{x\to a} f(x)$ exists and $\lim_{x\to a} f_n(x)$ exists for all $n$, is it true that

$$\lim_{x\to a} \lim_{n\to\infty} f_n(x) = \lim_{n\to\infty} \lim_{x\to a} f_n(x)?$$

**Example 4.4.4.** Consider the sequence $g_n(x) = \frac{1}{1+x^n}$ on $S = (-\infty, -1) \cup (-1, \infty)$. As $n$ goes to infinity we have that the sequence converges pointwise to 1 for $|x| < 1$, 1/2 for $x = 1$, and 0 for $|x| > 1$.

**Example 4.4.5.** Let $S = [0, \infty)$ and define

$$f_n(x) = \begin{cases} n^2 x & 0 \leqslant x \leqslant \frac{1}{n} \\ -n^2\left(x - \frac{2}{n}\right) & \frac{1}{n} < x \leqslant \frac{2}{n} \\ 0 & x > \frac{2}{n} \end{cases}$$

We claim that $\lim_{n\to\infty} f_n(x) = 0$ for all $x \geqslant 0$. When $x = 0$, $f_n(0) = 0$ which converges to 0. If $0 < x$: By the Archimedean property there exists $N \in \mathbb{N}$ such that $\frac{2}{N} < x$. Then $f_N(x) = 0$ and $f_n(x) = 0$ for all $n \geqslant N$. Thus, $\lim_{n\to\infty} f_n(x) = 0$, as claimed. This argument can be intuitively realized by noting that for $n$ large enought, the tent is always to the left of any $0 < x$.

We note that this gives an example of an unbounded sequence $\{f_n\}$ converging pointwise to a bounded function.

**Example 4.4.6.** Take $S = \mathbb{R}$ and define $f_n(x) = \frac{\sin(nx)}{\sqrt{n}}$, waves of declining amplitude but increasing frequency. Note that $0 \leqslant |f_n(x)| = \left|\frac{\sin(nx)}{\sqrt{n}}\right| \leqslant \frac{1}{\sqrt{n}}$ which goes to 0 as $n$ goes to infinity, so the sequence converges pointwise to 0. Notice $f_n'(x) = \frac{n}{\sqrt{n}}\cos(nx) = \sqrt{n}\cos(nx)$, which has no limit for any $x \in \mathbb{R}$.

This is an example of pointwise convergence where the derivatives do not converge pointwise. Moreover, as we will see, the original sequence actually converges uniformly (the bound on the terms does not depend on $x$), which suggests we need a stronger requirement for convergence of derivatives.

**Definition 4.4.3.** *Let $\{f_n\}$ be a sequence of functions defined on $A$, and let $f$ be a function which is also defined on $A$. Then $f$ is called the* **uniform limit of $\{f_n\}$ on $A$** *if for every $\varepsilon > 0$ there is some $N$ such that for all $x \in A$,*

$$\text{if } n > N, \text{ then } |f(x) - f_n(x)| < \varepsilon$$

*We also say that $\{f_n\}$* **converges uniformly to $f$ on $A$**, *or that $f_n$* **approaches $f$ uniformly on $A$**.

**Definition 4.4.4 (Uniform Convergence (alt.)).** *Suppose $S \subseteq \mathbb{R}$ and $f_n : S \to \mathbb{R}$ are real-valued functions for each $n \in \mathbb{N}$. We say that $f_n$* **converges uniformly on $S$** *to $f : S \to \mathbb{R}$ if for all $\varepsilon > 0$, there is an $N \in \mathbb{N}$ (depends only on $\varepsilon$) such that $n \geqslant N$ implies*

$$|f_n(x) - f(x)| < \varepsilon$$

*for all $x \in S$.*

We will use the notation $f_n \to_u f$ sometimes to denote uniform convergence. Intuitively uniform convergence can be understood by saying that given any band or tube containing $f$ on $S$, there is a point past which the tail functions of the sequence reside entirely in this band.

**Remark 4.4.4.** Note that uniform convergence implies pointwise convergence, but the converse is not true.

**Definition 4.4.5 (Uniform Norm).** *Suppose $S \subseteq \mathbb{R}$ and $f$ is a function on $S$. The* **uniform norm** *of $f$ on $S$ is given by*

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

We note that this may not be finite. When the context is clear we will write $||f||_\infty$.

**Remark 4.4.5.** Suppose $f : S \to \mathbb{R}$ is a function:

(i) $f$ is bounded on $S$ if and only if $||f||_S < \infty$

(ii) If $f$ is continous on $S$ and $S$ is compact, then $||f||_S = \max_{s \in S} |f(x)|$

**Proposition 4.4.1.** *A sequence of functions $f_n$ converges uniformly to $f$ on $S$ if and only if $||f_n - f||_S \to 0$.*

*Proof.* Let $f_n$ be a sequence of functions, each defined on $S$, and let $f$ be another function on $S$.

First, let us suppose that the $f_n$ converge uniformly to $f$ on $S$. Fix $\varepsilon > 0$. Then there exists $N \in \mathbb{N}$ such that for $n \geqslant N$, $|f_n(x) - f(x)| < \varepsilon/2$ for all $x \in S$. This implies that $\varepsilon/2$ is an upper bound of all $|f_n(x) - f(x)|$, so by definition $||f_n - f||_S \leqslant \varepsilon/2 < \varepsilon$ for all $n \geqslant N$. Hence, as $||f_n - f||_S = |||f_n - f||_S - 0|$, we have that the sequence $||f_n - f||_S$ converges to 0.

Conversely, suppose that $||f_n - f||_S$ converges to 0, and let $\varepsilon > 0$. Then, there exists $N \in \mathbb{N}$ such that for all $n \geqslant N$, $||f_n - f||_S < \varepsilon$. It follows that for all $x \in S$, $|f_n(x) - f(x)| \leqslant ||f_n - f||_S < \varepsilon$ for $n \geqslant N$, so we find that $f_n$ converges uniformly to $f$ on $S$ by definition. ■

**Example 4.4.7.** Let $S = [0, 1]$ and $f_n(x) = x^n$. $f_n$ does not converge uniformly to any function. Indeed if $f_n$ converges uniformly to anything on $[0, 1]$, it must be the pointwise limit $f(x) =$

$0, 0 \leqslant x < 1$ and $f(x) = 1, x = 1$, since uniform implies pointwise convergence (and limits are unique in Hausdorff spaces). This implies $\sup_{x\in[0,1]} |f_n(x) - f(x)|$ goes to zero as $n$ goes to infinity. But if $x = 1 - \frac{1}{n}$, then $|f_n(1-1/n) - f(1-1/n)| = (1-1/n)^n$, which converges to $e^{-1} = \frac{1}{e}$, and not zero. Further, this must be less than what the limit of the supremums converges to, so the supremums cannot converge to 0, as that would result in a contradiction. Hence, the sequence does not converge uniformly.

The tent function is another example of pointwise but not uniform convergence, since $||f_n - 0||_\infty = n$, which does not converge to 0. (CONTINUE HAM HERE)

**Theorem 4.4.2.** *Suppose that $\{f_n\}$ is a sequence of functions which are integrable on $[a,b]$, and that $\{f_n\}$ converges uniformly on $[a,b]$ to a function $f$ which is also integrable on $[a,b]$. Then*

$$\int_a^b f = \lim_{n\to\infty} \int_a^b f_n$$

*Proof.* Let $\varepsilon > 0$. Then since $\{f_n\}$ converges uniformly to $f$, there exists $N \in \mathbb{N}$ such that for all $x \in [a,b]$, if $n \geqslant N$

$$|f(x) f_n(x)| < \varepsilon$$

Then for all $n \geqslant N$, we have that

$$\begin{aligned} \left| \int_a^b f(x)dx - \int_a^b f_n(x)dx \right| &= \left| \int_a^b [f(x) - f_n(x)]dx \right| \\ &\leqslant \int_a^b |f(x) - f_n(x)|dx \\ &\leqslant \int_a^b \varepsilon dx \\ &= \varepsilon(b-a) \end{aligned}$$

Since this is true for all $\varepsilon > 0$, it follows that

$$\int_a^b f = \lim_{n\to\infty} \int_a^b f_n$$

■

**Theorem 4.4.3.** *Suppose that $\{f_n\}$ is a sequence of functions which are continuous on $[a,b]$, and that $\{f_n\}$ converges uniformly on $[a,b]$ to $f$. Then $f$ is also continuous on $[a,b]$.*

*Proof.* For each $x \in [a,b]$ we must prove that $f$ is continuous at $x$. We first deal with $x \in (a,b)$. Fix $\varepsilon > 0$. Since $\{f_n\}$ converges uniformly to $f$ on $[a,b]$, there is some $N$ such that for all $n \geqslant N$ and all $y \in [a,b]$,

$$|f(y) - f_n(y)| < \varepsilon/3$$

In particular, for all $h$ such that $x + h \in [a,b]$, we have

$$|f(x) - f_n(x)| < \varepsilon/3,$$

$$|f(x+h) - f_n(x+h)| < \varepsilon/3$$

Now $f_n$ is continuous, so there is some $\delta > 0$ such that for $|h| < \delta$ we have

$$|f_n(x) - f_n(x+h)| < \varepsilon/3$$

Thus, if $|h| < \delta$, then

$$\begin{aligned}
|f(x+h) - f(x)| &= |f(x+h) - f_n(x+h) + f_n(x+h) - f_n(x) + f_n(x) - f(x)| \\
&\leqslant |f(x+h) - f_n(x+h)| + |f_n(x+h) - f_n(x)| + |f_n(x) - f(x)| \\
&< \varepsilon/3 + \varepsilon/3 + \varepsilon/3 \\
&= \varepsilon
\end{aligned}$$

This proves that $f$ is continuous at $x$. ■

**Remark 4.4.6.** Allow this last two theorems are great successes, differentiability sadly fails. Even if each $f_n$ is differentiable and $\{f_n\}$ converges uniformly to $f$, it need not be the case that $f$ is differentiable. Moreover, even if $f$ is itself differentiable, it need not be the case that

$$f'(x) = \lim_{n\to\infty} f_n'(x)$$

**Example 4.4.8** (Counter Example 3)**.** Consider the family of functions

$$f_n(x) = \frac{1}{n}\sin(n^2 x)$$

then $\{f_n\}$ converges uniformly to the function $f(x) = 0$, but

$$f_n'(x) = n\cos(n^2 x)$$

and $\lim\limits_{n\to\infty} n\cos(n^2 x)$ does not even always exist (for example if $x = 0$).

**Theorem 4.4.4.** *Suppose that $\{f_n\}$ is a sequence of functions which are differentiable on $[a,b]$, with integrable derivatives $f_n'$, and that $\{f_n\}$ converges (pointwise) to $f$. Suppose, moreover, that $\{f_n'\}$ converges uniformly on $[a,b]$ to some continuous function $g$. Then $f$ is differentiable and*

$$f'(x) = \lim_{n\to\infty} f_n'(x)$$

*Proof.* Applying Theorem 1 to the interval $[a,x]$, we see that for each $x$ we have

$$\begin{aligned}
\int_a^x g &= \lim_{n\to\infty} \int_a^x f_n' \\
&= \lim_{n\to\infty} [f_n(x) - f_n(a)] && \text{(by \textbf{??})} \\
&= f(x) - f(a)
\end{aligned}$$

Since $g$ is continuous, it follows that $f'(x) = g(x) = \lim\limits_{n\to\infty} f_n'(x)$ for all $x$ in the interval $[a,b]$, by **??**. ■

**Definition 4.4.6.** *The series $\sum_{n=1}^{\infty} f_n$* **converges uniformly** *(more formally, the sequence $\{f_n\}$ is* **uniformly summable***)* **to** $f$ **on** $A$*, if the sequence*

$$f_1, f_1 + f_2, f_1 + f_2 + f_3, \ldots$$

*converges uniformly to $f$ on $A$.*

**Corollary 4.4.5.** *Let $\sum_{n=1}^{\infty} f_n$ converge uniformly to $f$ on $[a,b]$.*

1. *If each $f_n$ is continuous on $[a,b]$, then $f$ is continuous on $[a,b]$.*
2. *If $f$ and each $f_n$ is integrable on $[a,b]$, then*

$$\int_a^b f = \sum_{n=1}^{\infty} \int_a^b f_n$$

*Moreover, if $\sum_{n=1}^{\infty} f_n$ converges (pointwise) to $f$ on $[a,b]$, each $f_n$ has an integrable derivative $f_n'$ and $\sum_{n=1}^{\infty} f_n'$ converges uniformly on $[a,b]$ to some continuous function, then*

3. *$f'(x) = \sum_{n=1}^{\infty} f_n'(x)$ for all $x \in [a,b]$*

*Proof.* Let $\{s_n\}$ be the sequence of partial sums of the $\{f_n\}$. Then since each $f_n$ is continuous, so is each $s_n$. Then as $\{s_n\}$ converges uniformly to $f$ we have by a previous theorem that $f$ is also continuous on $[a,b]$. Next, since each $f_n$ is integrable on $[a,b]$, so is each $s_n$. Then as $\{s_n\}$ converges uniformly to $f$ we have that

$$\begin{aligned}\int_a^b f &= \lim_{n\to\infty} \int_a^b s_n \\ &= \lim_{n\to\infty} t_n \\ &= \sum_{n=1}^{\infty} \int_a^b f_n\end{aligned}$$

where $\{t_n\}$ is the sequence such that

$$t_n = \sum_{i=1}^{n} \int_a^b f_n = \int_a^b s_n$$

Finally, suppose $s_n$ converges (pointwise) to $f$ on $[a,b]$, and each $f_n$ has an integrable derivative $f_n'$. Then each $s_n$ has an integrable derivative $s_n'$ on $[a,b]$ by the linearity of the derivative and integral operators. Moreover, suppose $s_n'$ converges uniformly on $[a,b]$ to some continuous function $g$. Then it follows that for all $x \in [a,b]$

$$f'(x) = \lim_{n\to\infty} s_n'(x) = \sum_{n=1}^{\infty} f_n'(x)$$

∎

## Theorem 26 (The Weierstrass M-Test).

*Let $\{f_n\}$ be a sequence of functions defined on $A$, and suppose that $\{M_n\}$ is a sequence of numbers such that*

$$|f_n(x)| \leqslant M_n, \forall x \in A$$

*Suppose moreover that $\sum_{n=1}^{\infty} M_n$ converges. Then for each $x$ in $A$ the series $\sum_{n=1}^{\infty} f_n(x)$ converges absolutely, and $\sum_{n=1}^{\infty} f_n$ converges uniformly on $A$ to the function*

$$f(x) = \sum_{n=1}^{\infty} f_n(x)$$

*Proof.* For each $x \in A$, the series $\sum_{n=1}^{\infty} |f_n(x)|$ converges by 21; consequently $\sum_{n=1}^{\infty} f_n(x)$ converges absolutely. Moreover, for all $x \in A$ we have

$$\begin{aligned}\left|f(x) - \sum_{i=1}^{N} f(x)\right| &= \left|\sum_{n=N+1}^{\infty} f_n(x)\right| \\ &\leqslant \sum_{n=N+1}^{\infty} |f_n(x)| \\ &\leqslant \sum_{n=N+1}^{\infty} M_n\end{aligned}$$

Since $\sum_{n=1}^{\infty} M_n$ converges, the number $\sum_{n=N+1}^{\infty} M_n$ can be made as small as desired (by 18), by choosing $N$ sufficiently large. ∎

**Definition 4.4.7.** *An infinite sum of functions of the form*

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

*is called a **power series centered at** $a$. One especially important family of power series are those of the form*

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

*where $f$ is some infinitely differentiable function at $a$; this series is called the **Taylor series for $f$ at** $a$.*

**Remark 4.4.7.** Given a function $f$ infinitely differentiable at $a$, we have for $x \in \mathbb{R}$ that

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

if and only if the remainder terms satisfy $\lim_{n\to\infty} R_{n,a}(x) = 0$.

**Theorem 4.4.6.** *Suppose that the series*

$$f(x_0) = \sum_{n=0}^{\infty} a_n x_0^n$$

*converges, and let $a$ be any number with $0 < a < |x_0|$. Then on $[-a, a]$ the series*

$$f(x) = \sum_{n=0}^{\infty} a_n x^n$$

*converges uniformly (and absolutely). Moreover, the same is true for the series*

$$g(x) = \sum_{n=1}^{\infty} n a_n x^{n-1}$$

*Finally, $f$ is differentiable and*

$$f'(x) = \sum_{n=1}^{\infty} n a_n x^{n-1}$$

*for all $x$ with $|x| < |x_0|$.*

*Proof.* First, since $\sum\limits_{n=0}^{\infty} a_n x_0^n$ converges, $\lim\limits_{n\to\infty} a_n x_0^n = 0$. Hence, the sequence $\{a_n x_0^n\}$ is surely bounded: there is some number $M$ such that

$$|a_n x_0|^n = |a_n| \cdot |x_0|^n \leqslant M$$

for all $n$. Now if $x$ is in $[-a, a]$, then $|x| \leqslant |a|$, so

$$\begin{aligned} |a_n x^n| &= |a_n| \cdot |x|^n \\ &\leqslant |a_n| \cdot |a|^n \\ &= |a_n| \cdot |x_0|^n \cdot \left|\frac{a}{x_0}\right|^n \\ &\leqslant M \left|\frac{a}{x_0}\right|^n \end{aligned}$$

But $|a/x_0| < 1$, so the (geometric) series

$$\sum_{n=0}^{\infty} M \left|\frac{a}{x_0}\right|^n = M \sum_{n=0}^{\infty} \left|\frac{a}{x_0}\right|^n$$

converges. Choosing $M \cdot |a/x_0|^n$ as the number $M_n$ in 26, it follows that $\sum\limits_{n=0}^{\infty} a_n x^n$ converges uniformly on $[-a, a]$.

To prove the same assertion for $g(x) = \sum\limits_{n=1}^{\infty} n a_n x^{n-1}$ notice that

$$|n a_n x^{n-1}| = n|a_n| \cdot |x^{n-1}|$$

$$
\begin{aligned}
&\leqslant n|a_n| \cdot |a^{n-1}| \\
&= \frac{|a_n|}{|a|} \cdot |x_0|^n n \left|\frac{a}{x_0}\right|^n \\
&\leqslant \frac{M}{|a|} n \left|\frac{a}{x_0}\right|^n
\end{aligned}
$$

Since $|a/x_0| < 1$, the series

$$\sum_{n=1}^{\infty} \frac{M}{|a|} n \left|\frac{a}{x_0}\right|^n = \frac{M}{|a|} \sum_{n=1}^{\infty} n \left|\frac{a}{x_0}\right|^n$$

converges (by an application of the Ratio Tes). Another appeal to 26 proves that $\sum\limits_{n=1}^{\infty} na_n x^{n-1}$ converges uniformly on $[-a, a]$.

Finally, our corollary proves, first that $g$ is continuous, and then that

$$f'(x) = g(x) = \sum_{n=1}^{\infty} na_n x^{n-1}$$

for all $x \in [-a, a]$. Since we could have chosen any $a$ with $0 < a < |x_0|$, this result holds for all $x$ with $|x| < |x_0|$. ∎

## Complex Power Series

We now investigate power series in the complex numbers. First, recall that if $r \in \mathbb{C}$, $|r| < 1$, then we have the geometric series

$$\sum_{k=0}^{\infty} r^k = \frac{1}{1-r}$$

**Definition 4.4.8.** *A* **power series** *(in $\mathbb{C}$) is a series of functions of the form*

$$f(z) = \sum_{k=0}^{\infty} a_k z^k, \quad a_k \in \mathbb{C}, z \in \mathbb{C}$$

Note by definition $f(z) = \lim\limits_{n\to\infty} \sum_{k=0}^{n} a_k z^k$.

**Proposition 4.4.7.** *If $f(z) = \sum_{k=0}^{\infty} a_k z^k$ is defined for some $z_0 \in \mathbb{C}, z_0 \neq 0$, then $f(z)$ converges absolutely and uniformly for all $|z| < |z_0|$.*

*Proof.* Note that if $|z| < |z_0|$, $|a_k z^k| < |a_k||z_0|^k$ for all $k \geqslant 0$. Further, as $\sum_{k=0}^{n} a_k z_0^k$ is convergent, it is Cauchy so $|a_k z^k| \to 0$, and in particular $|a_k z_0^k| \leqslant C$ for all $k \geqslant 0$ for some $C \in \mathbb{R}$. Then $0 \leqslant \frac{|z|}{|z_0|} = r < 1$. So

$$\left|\sum_{k=0}^{n} a_k z^k\right| \leqslant \sum_{k=0}^{n} |a_k z_0^k| r^k \leqslant C \sum_{k=0}^{n} r^k \leqslant \frac{C}{1-r}$$

Thus $\sum_{k=0}^{n}|a_kz^k|$ is a bounded increasing sequence, and so converges. In particular, letting $M_k = Cr^k = C\frac{|z_1|^k}{|z_0|^k}$, we have that for all $|z| < |z_1| < |z_0|$, $|a_kz^k| \leqslant M_k$ and $\sum_{k=0}^{\infty} M_k < \infty$, so $\sum_{k=0}^{n} a_kz^k$ is uniformly convergent to $f(z)$ on $|z| < |z_1|$ for all $|z_1| < |z_0|$, so in particular it is uniformly convergent on $|z| < |z_0|$. ■

From this theorem we have a few cases for $f(z)$:

- $f(z)$ converges for all $z \in \mathbb{C}$
- There exists $R > 0$ such that the series converges absolutely for $|z| < R$, but diverges for $|z| > R$.
- The series does not converge for any $z \neq 0$ ($R = 0$)

**Proposition 4.4.8.** *If $f(z) = \sum_{k=0}^{\infty} a_kz^k$ converges in $D_R = \{z \in \mathbb{C} : |z| < R\}$, then it converges uniformly in any disk $D_S$ with $0 < S < R$. In particular, this yields $f(z)$ is continuous on $D_R$.*

*Proof.* Let $0 < S < R$ and pick $T$ such that $0 < S < T < R$. Then the series converges for any $|z| = T$. In particular, there exists $C > 0$ such that $|a_kT^k| < C$ for all $k \geqslant 0$. If $z \in D_S$, then $|a_kz^k| \leqslant |a_kT^k|\frac{|z|^k}{|T|^k} \leqslant Cr^k$, where $0 \leqslant r = \frac{|z|}{|T|} < 1$. As $|r| < 1$, $C\sum_{k=0}^{n} r^k$ converges, and so is Cauchy. Thus, for $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that if $m > n \geqslant N$,

$$\sum_{k=n}^{m} ||a_kz^k||_{D_S} \leqslant \sum_{k=n}^{m} Cr^k < \varepsilon$$

so $\sum_{k=0}^{n} a_kz^k$ is uniformly Cauchy on $D_S$. So it converges uniformly to $f(z)$ in $D_S$ for all $S < R$. So $f(z)$ is continuous in $D_S$ for all $S < R$, and in particular $f$ is continuous on $D_R$. ■

**Definition 4.4.9.** *For $f(z) = \sum_{k=0}^{\infty}(z - z_0)^k$, let the* **<u>radius of convergence</u>** *$R$ be defined by*

$$\frac{1}{R} = \limsup_{n\to\infty} |a_n|^{1/n}$$

*If the right hand side is $0$, $R = \infty$, and if the right hand side is $\infty$, $R = 0$.*

**Proposition 4.4.9.** *The series $f(z) = \sum_{k=0}^{\infty} a_k(z - z_0)^k$ converges absolutely for $|z - z_0| < R$, and diverges when $|z - z_0| > R$. Then when $R > 0$, $f$ is a continuous function $f : D_R(z_0) \to \mathbb{C}$.*

*Proof.* If $R' < R$, and $R' \neq 0$, $\frac{1}{R} < \frac{1}{R'}$. Note $\frac{1}{R} = \limsup_{n\to\infty} |a_n|^{1/n}$, and let $\varepsilon = \left(\frac{1}{R'} - \frac{1}{R}\right)$. Then there exists $N \in \mathbb{N}$ such that for $n \geqslant N$, $\sup_{k\geqslant n} |a_k|^{1/k} < \frac{1}{R} + \varepsilon$, so $\sup_{k\geqslant n} |a_k|^{1/k} < \frac{1}{R'}$, so $|a_n|^{1/n}R' < 1$ and $|a_n|R'^n < 1$ for all $n \geqslant N$. If $|z - z_0| < R'$, then $\frac{|z-z_0|}{R'} = r < 1$, so $|a_n(z - z_0)^n| = |a_n|R'^n r^n < r^n$ for all $n \geqslant N$. Then $\sum_{n=m}^{'infty} ||a_n(z - z_0)||_{D_{R'}} \leqslant \sum_{n=m}^{\infty} r^n \to 0$ for all $m \geqslant N$. So we have absolute uniform convergence in $D_{R'}(z_0)$, for all $0 < R' < R$.

Divergence: Let $R'' > R$, so $\frac{1}{R''} < \frac{1}{R} = \limsup_{n\to\infty} |a_n|^{1/n}$. Then for all $n$, $\sup_{k\geqslant n} |a_n|^{1/n} > \frac{1}{R''}$, and in particular there are infinitely many $k \geqslant n$, for all $n$, such that $|a_k|^{1/k} > \frac{1}{R''}$, so $|a_k|R''^k > 1$. If $|z - z_0| \geqslant R''$, then $|a_n(z - z_0)^n| \geqslant |a_n|R''^n > 1$ for infinitely many $n \geqslant 0$. Thus, $|a_n(z - z_0)^n|$ does not converge to 0, so the series diverges. ■

**Definition 4.4.10.** *A function defined by $\sum_{k=0}^{\infty} a_k z^k$ with radius of convergence $R > 0$ is said to be* **analytic** *on $D_R$.*

**Definition 4.4.11.** *In $\mathbb{R}^n$ with $x = (x_1, ..., x_n)$, for any multi-index $\alpha = (\alpha_1, ..., \alpha_n)$ with $\alpha_j \in \mathbb{N}_0$, we define*

$$x^\alpha = x_1^{\alpha_1} \cdots x_n^{\alpha_n}$$

*and we define a power series in $\mathbb{R}^n$ by*

$$f(x) = \sum_{|\alpha| \geqslant 0} a_\alpha x^\alpha$$

## Products of Series

We investigate when we can distribute multiplication of two series:

**Proposition 4.4.10.** *If $A = \sum_{k=0}^{\infty} a_k$ and $B = \sum_{k=0}^{\infty} b_k$ are absolutely convergent, then*

$$AB = \sum_{k=0}^{\infty} c_k, \quad c_k = \sum_{j=0}^{k} a_{k-j} b_j$$

*Proof.* Let $A_k = \sum_{n=0}^{k} a_n$ and $B_k = \sum_{n=0}^{k} b_n$. Then

$$\begin{aligned} A_k B_k &= \sum_{n=0}^{k} \sum_{m=0}^{k} a_n b_m \\ &= \sum_{l=0}^{k} \sum_{j=0}^{l} a_j b_{l-j} + R_l = \sum_{l=0}^{k} c_l + R_k \end{aligned}$$

where $R_k = \sum_{(n,m)\in\sigma(k)} a_n b_m$ where $\sigma(k) := \{(n,m) \in \mathbb{N}_0 \times \mathbb{N}_0 : n, m \leqslant k, n+m > k\}$. Then

$$\begin{aligned} |R_k| &\leqslant \sum_{(n,m)\in\sigma(k)} |a_n b_m| \\ &\leqslant \sum_{k \geqslant n \geqslant k/2} \sum_{m \leqslant k} |a_n||b_m| + \sum_{n \leqslant k} \sum_{k/2 \leqslant m \leqslant k} |a_n||b_m| \\ &\leqslant \sum_{k/2 \leqslant n \leqslant k} \sum_{m=0}^{\infty} |a_n||b_m| + \sum_{k/2 \leqslant m \leqslant k} \sum_{n=0}^{\infty} |b_m||a_n| \\ &\leqslant A \sum_{k/2 \leqslant m} |b_m| + B \sum_{k/2 \leqslant n} |a_n| \end{aligned}$$

where both sums on the right go to zero as $k$ goes to infinity. So for all $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that $k \geqslant N$ implies $|R_k| < \varepsilon$. Then

$$A_k B_k = \sum_{l=0}^{k} c_l + R_k$$

so

$$\lim_{k\to\infty}\sum_{l=0}^{k} c_l = \lim_{k\to\infty}(A_kB_k - R_k) = AB + 0 + AB$$

and we conclude

$$AB = \sum_{l=0}^{\infty} c_l$$

as claimed. ∎

**Corollary 4.4.11.** *If the series*

$$f(z) = \sum_{n=0}^{\infty} a_n(z-z_0)^n, \quad g(z) = \sum_{n=0}^{\infty} b_n(z-z_0)^n$$

*converge for some* $|z-z_0| < R$*, then*

$$f(z)g(z) = \sum_{n=0}^{\infty} c_n(z-z_0)^n$$

*converges for* $|z-z_0| < R$ *with*

$$c_n = \sum_{j=0}^{n} a_j b_{n-j}$$

It follows that the set of analytic functions on some disk is an algebra over $\mathbb{C}$.

**Proposition 4.4.12.** *If* $a_{jk} \in \mathbb{C}$ *and* $\sum_{j,k=0}^{\infty} |a_{j,k}| < \infty$*, then for each* $k$*,* $\alpha_k = \sum_{j=0}^{\infty} a_{j,k}$ *and for each* $j$ $\beta_j = \sum_{k=0}^{\infty} a_{j,k}$ *are absolutely convergent, where*

$$\lim_{n\to\infty}\sum_{j=0}^{n}\sum_{k=0}^{n} |a_{j,k}| =: \sum_{j,k=0}^{\infty} |a_{j,k}|$$

*Then*

$$\sum_{j=0}^{\infty} \beta_j = \sum_{k=0}^{\infty} \alpha_k = \sum_{j,k=0}^{\infty} a_{j,k}$$

*or equivalently*

$$\sum_{j=0}^{\infty}\left(\sum_{k=0}^{\infty} a_{j,k}\right) = \sum_{k=0}^{\infty}\left(\sum_{j=0}^{\infty} a_{j,k}\right) = \sum_{j,k=0}^{\infty} a_{j,k}$$

*Proof.* For all $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that

$$\sum_{j,k\geqslant n} |a_{j,k}| < \varepsilon$$

Thus for all $j \in \mathbb{N}$ and all $k \in \mathbb{N}$,

$$\sum_{k=0}^{\infty} |a_{j,k}| < \sum_{j,k} |a_{j,k}| < \infty$$

and

$$\sum_{j=0}^{\infty} |a_{j,k}| < \sum_{j,k} |a_{j,k}| < \infty$$

Then for all $M, K \geqslant N$,

$$\left| \sum_{k=0}^{M} \sum_{j=0}^{K} a_{j,k} - \sum_{j,k}^{N} a_{j,k} \right| < \sum_{j,k \geqslant N} |a_{j,k}| < \varepsilon$$

In particular, taking the limit as $K$ goes to infinity,

$$\left| \sum_{k=0}^{M} \sum_{j=0}^{\infty} a_{j,k} - \sum_{j,k}^{N} a_{j,k} \right| \leqslant \sum_{j,k \geqslant N} |a_{j,k}| < \varepsilon$$

and

$$\left| \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} a_{j,k} - \sum_{j,k}^{N} a_{j,k} \right| \leqslant \sum_{j,k \geqslant N} |a_{j,k}| < \varepsilon$$

Reversing the sums before taking the limits, we obtain the other direction, so

$$\sum_{j=0}^{\infty} \left( \sum_{k=0}^{\infty} a_{j,k} \right) = \sum_{j,k=0}^{\infty} a_{j,k} = \sum_{k=0}^{\infty} \left( \sum_{j=0}^{\infty} a_{j,k} \right)$$

as desired. ■

# Part II

# Higher-Dimesional Analysis

# Chapter 5

# Metric Spaces

## 5.1.0 Euclidean Spaces

We begin our study with one of the most well studied metric spaces, Euclidean $n$-space.

**Definition 5.1.1.** ***Euclidean*** *$n$-space is defined as the product*

$$\mathbb{R}^n := \mathbb{R} \times \cdots \times \mathbb{R} = \{(x_1, ..., x_n) : x_1, ..., x_n \in \mathbb{R}\}$$

*We define $+ : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ component-wise, and we define a module action $\cdot : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$ also component-wise, turning $\mathbb{R}^n$ into an $\mathbb{R}$-linear space.*

Euclidean space is a special type of vector space which satisfies certain additional structures:

**Definition 5.1.2.** *A* ***real-inner product*** *on a real vector space $V$ is a map $\langle , \rangle : V \times V \to \mathbb{R}$ such that*

1. *$\langle v, v\rangle \geqslant 0$ for all $v \in V$, and $\langle v, v\rangle = 0$ if and only if $v = 0_V$*
2. *$\langle v, w\rangle = \langle w, v\rangle$ for all $v, w \in V$*
3. *$\langle av + bu, w\rangle = a\langle v, w\rangle + b\langle u, w\rangle$ for all $v, u, w \in V$ and all $a, b \in \mathbb{R}$.*

**Definition 5.1.3.** *A* ***norm*** *on a real vector space $V$ is a map $|| \cdot || : V \to \mathbb{R}$ such that*

1. *$||v|| \geqslant 0$ for all $v \in V$, and $||v|| = 0$ if and only if $v = 0_V$*
2. *$||av|| = |a| \cdot ||v||$ for all $v \in V$ and all $a \in \mathbb{R}$*
3. *$||v + w|| \leqslant ||v|| + ||w||$ for all $v, w \in V$ (triangle inequality)*

*Then $(V, || \cdot ||)$ is called a* ***normed linear space****.*

**Definition 5.1.4.** *A* **metric** *on a set $X$ is a map $d : X \times X \to \mathbb{R}$ such that*

1. $d(x, y) \geqslant 0$ *for all* $x, y \in X$, *and* $d(x, y) = 0$ *if and only if* $x = y$
2. $d(x, y) = d(y, x)$ *for all* $x, y \in X$
3. $d(x, z) \leqslant d(x, y) + d(y, z)$ *for all* $x, y, z \in X$ *(triangle inequality)*

*In this case we call $(X, d)$ a* **metric space**.

**Proposition 5.1.1.** *If $\langle , \rangle : V \times V \to \mathbb{R}$ is an inner product, $|| \cdot || : V \to \mathbb{R}$ defined by $|| \cdot || = \sqrt{\langle \cdot, \cdot \rangle}$ is a norm, and $d : V \times V \to \mathbb{R}$ defined by $d(a, b) = ||a - b||$ is a metric.*

We now define the inner product on Euclidean $n$-space:

**Definition 5.1.5.** *The Euclidean inner product on $\mathbb{R}^n$ is given by*

$$\langle x, y \rangle = x \cdot y = \sum_{i=1}^{n} x_i y_i$$

*for all $x = (x_1, ..., x_n), y = (y_1, ..., y_n) \in \mathbb{R}^n$.*

**Proposition 5.1.2** (Cauchy-Schwarz Inequality)**.** *For any inner product $\langle , \rangle : V \times V \to \mathbb{R}$*

$$|\langle v, w \rangle| \leqslant |\langle v, v \rangle||\langle w, w \rangle|$$

*Proof.* Let $t > 0$, and consider $0 \leqslant \langle tx - t^{-1}y, tx - t^{-1}y \rangle$, so $0 \leqslant t^2||x||^2 - 2\langle x, y \rangle + t^{-2}||y||^2$. Without loss of generality suppose $x, y \neq 0$. Then let $t^2 = \frac{||y||}{||x||}$, so we have $2\langle x, y \rangle \leqslant 2||y|| \cdot ||x||$, so we have our desired result. Replacing $x$ with $-x$ we obtain $-\langle x, y \rangle \leqslant ||y|| \cdot ||x||$, so $|\langle x, y \rangle| \leqslant ||x|| \cdot ||y||$. ∎

The triangle inequality follows from this Cauchy-Schwarz inequality. Thus, $\mathbb{R}^n$ is a metric space, and in particular it is a normed linear space.

## Sequences and Convergence

We can now use the metric on $\mathbb{R}^n$ to define notions of convergence for sequences.

**Definition 5.1.6.** *If $(\vec{x}_j)_{j=1}^{\infty} \subset \mathbb{R}^n$, then $\vec{x}_j \to \vec{x} \in \mathbb{R}^n$ if and only if $||\vec{x}_j - \vec{x}||$ converges to $0$ in $\mathbb{R}$. Further, we say such a sequence is* **Cauchy** *if and only if for all $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that for $j, k \geqslant N$, $||\vec{x}_j - \vec{x}_k|| < \varepsilon$.*

We note that we can consider $\mathbb{C} \cong \mathbb{R}^2$ as normed linear spaces over $\mathbb{R}$.

**Proposition 5.1.3.** *$\mathbb{R}^n$ is a complete metric space.*

*Proof.* By completeness of $\mathbb{R}$, and the fact that $(x_j)_{j=1}^{\infty} = (\langle x_{j,1}, ..., x_{j,n}\rangle)_{j=1}^{\infty}$ is Cauchy if and only if the $(x_{j,i})_{j=1}^{\infty}$ are Cauchy for all $i$. ■

## Topological Properties of Euclidean Space

First we define the notions of open and closed sets in $\mathbb{R}^n$:

**Definition 5.1.7.** *We say that $S \subseteq \mathbb{R}^n$ is **closed** if and only if for all sequences $(x_j)_{j=1}^{\infty} \subseteq S$ such that $x_j$ converges to $x \in \mathbb{R}^n$, $x \in S$.*

**Definition 5.1.8.** *We say that a set $U \subseteq \mathbb{R}^n$ is **open** if and only if $\mathbb{R}^n \backslash U = U^C$ is closed. This holds if and only if for all $x \in U$, there exists $\varepsilon > 0$ such that $B_\varepsilon(x) \subseteq U$, where*

$$B_\varepsilon(x) := \{y \in \mathbb{R}^n : d(x, y) < \varepsilon\}$$

*is the $\varepsilon$-ball centered at $x \in \mathbb{R}^n$.*

**Definition 5.1.9.** *A set $K \subseteq \mathbb{R}^n$ is **sequentially compact** if and only if for all $(p_j)_{j=1}^{\infty} \subseteq K$, there exists a subsequence $(p_{j_k})_{k=1}^{\infty} \subseteq K$ and a $p \in K$ such that $p_{j_k}$ converges to $p$.*

That is every sequence in a sequentially compact set has a convergent subsequence.

**Definition 5.1.10.** *A set $S \subseteq T \subseteq \mathbb{R}^n$ is said to be **dense** in $T$ if $\overline{S} \supseteq T$, where $\overline{S}$ is the **closure** of $S$:*

$$\overline{S} := \bigcap \{C \subseteq \mathbb{R}^n : S \subseteq C \text{ and } C \text{ is closed}\}$$

Note that the closure of a set $S$ is the smallest closed set containing $S$.

**Definition 5.1.11.** *A topological space $(X, \tau)$ is said to be **separable** if it has a **countable dense subset**.*

We note that $\mathbb{R}^n$ is separable with countable dense subset $\mathbb{Q}^n$. Indeed, for all $U \subseteq \mathbb{R}^n$, $U$ is open if and only if for all $p \in U$, there exists $q \in \mathbb{Q}^n$ and $r \in \mathbb{Q}^+$ such that $p \in B_r(q) \subseteq U$. Hence,

$$\mathcal{B} := \{B_r(q) : r \in \mathbb{Q}^+, q \in \mathbb{Q}^n\}$$

is a countable base for the topology on $\mathbb{R}^n$, so $\mathbb{R}^n$ is a ***second countable space***. We have the following useful results about compactness in Euclidean spaces:

**Corollary 5.1.4.** *If $K \subseteq \mathbb{R}^n$ is sequentially compact then it is topologically compact.*

To prove this corollary we first prove some intermediate results.

**Proposition 5.1.5.** *If $K \subseteq \mathbb{R}^n$ is sequentially compact, and $X_1 \supseteq X_2 \supseteq X_3 \supseteq \ldots$ is a chain of nonempty closed subsets of $K$, then $\bigcap_{j\geqslant 1} X_j \neq \varnothing$.*

*Proof.* Let $x_j \in X_j \subseteq K$ for each $j$. Then $(x_j)_{j=1}^{\infty} \subseteq K$, so as $K$ is compact we have a subsequence $(x_{j_k})_{k=1}^{\infty} \subseteq K$ such that $x_{j_k} \to x \in K$. But, for each $m \in \mathbb{N}$, $\{x_{j_k} : k \geqslant m\} \subseteq X_m$, so as $X_m$ is closed it follows that $x \in X_m$. Thus, $x \in \bigcap_{j\geqslant 1} X_j$, so $\bigcap_{j\geqslant 1} X_j \neq \varnothing$. ■

**Proposition 5.1.6.** *If $K \subseteq \mathbb{R}^n$ is compact and $U_1 \subseteq U_2 \subseteq \ldots$ is a chain of open sets which cover $K$, $K \subseteq \bigcup_{j\geqslant 1} U_j$, then there exists $M \in \mathbb{N}$ such that $K \subseteq U_M$.*

*Proof.* Consider $X_m = K \backslash U_m$, which gives a chain of open sets. We have that $\bigcap_{j\geqslant 1} X_j = \varnothing$, so by the contrapositive of the previous proposition there exists $M \in \mathbb{N}$ such that $X_M = \varnothing$. Then $K \subseteq U_M$, as desired. ■

Now we proceed to the proof of the corollary:

*Proof.* Let $\{U_\alpha\}_{\alpha\in J}$ be an open cover of $K$, which is sequentially compact in $\mathbb{R}^n$. For each $\alpha$ we have $U_\alpha = \bigcup_{j\geqslant 1} B_{r_{\alpha_j}}(q_{\alpha_j})$ is the union of a countable number of open sets in our countable base. Then, $K \subseteq \bigcup_{\alpha\in J} \bigcup_{j\geqslant 1} B_{r_{\alpha_j}}(q_{\alpha_j})$, which is a union over a countable collection of open sets since $\mathcal{B}$ is countable. We claim any countable cover has a finite subcover for $K$. Indeed, consider $U_m = \bigcup_{j=1}^{m} B_j$ for a countable cover $\bigcup_{j\geqslant 1} B_j$. Then by the previous proposition there exists $M \in \mathbb{N}$ such that $K \subseteq \bigcup_{j=1}^{M} B_j$. Thus, there exist $\alpha_1, \ldots, \alpha_N \in J$, and $M_j$ such that $K \subseteq \bigcup_{j=1}^{N} \bigcup_{k=1}^{M_j} B_{r_{\alpha_{j,k}}}(q_{\alpha_{j,k}}) \subseteq \bigcup_{j=1}^{N} U_{\alpha_j}$, so $K$ it topologically compact. ■

## 5.2.0 Metric Space Properties

Throughout this section let $(X, d)$ be a metric space. Then we can define convergence of sequences as follows:

**Definition 5.2.1.** *A sequence $(x_j)_{j=1}^{\infty} \subseteq X$ converges to $x \in X$ if and only if $d(x_j, x)$ converges to $0$ in $\mathbb{R}$.*

**Definition 5.2.2.** *A sequence $(x_j)_{j=1}^{\infty} \subseteq X$ is said to be Cauchy if and only if $d(x_j, x_k)$ converges to $0$ in $\mathbb{R}$ as $k$ and $j$ go to infinity.*

**Definition 5.2.3.** *$X$ is said to be a* **complete metric space**, *or a* **Fréchet space**, *if and only if for all Cauchy sequences $(x_j)_{j=1}^{\infty} \subseteq X$, there exists $x \in X$ such that $x_j$ converges to $x$.*

### Completion of a Metric Space

As in the case of $\mathbb{Q}$ to $\mathbb{R}$, we can form the completion of a general metric space.

**Definition 5.2.4.** *If $(X, d)$ is not complete, we can define its completion $(\hat{X}, \hat{d})$ by taking $\hat{X} = \{[(x_j)_{j=1}^{\infty}] : (x_j)_{j=1}^{\infty} \subseteq X$ is Cauchy$\}$, where $[(x_j)]$ is the the equivalence class defined by $(x_j) \sim (x'_j)$ if and only if $d(x_j, x'_j) \to 0$ in $\mathbb{R}$. For $\xi = [(x_j)]$ and $\eta = [(y_j)]$, we define*

$$\hat{d}(\xi, \eta) := \lim_{j\to\infty} d(x_j, y_j)$$

It remains to show that $\hat{d}$ is a well defined metric. Note by the triangle inequality $d(x, y) \leqslant d(x, z) + d(z, y)$, so $d(x, y) - d(x, z) \leqslant d(y, z)$ and using $d(x, z) \leqslant d(x, y) + d(y, z)$, $d(x, z) - d(x, y) \leqslant d(y, z)$. Thus

$$|d(x, z) - d(x, y)| \leqslant d(y, z)$$

Then suppose $(x_j) \sim (x'_j)$ and $(y_j) \sim (y'_j)$. It follows that

$$\begin{aligned}
|d(x_j, y_j) - d(x'_j, y'_j)| &= |d(x_j, y_j) - d(x_j, y'_j) + d(x_j, y'_j) - d(x'_j, y'_j)| \\
&\leqslant |d(x_j, y_j) - d(x_j, y'_j)| + |d(x_j, y'_j) - d(x'_j, y'_j)| \\
&\leqslant d(y_j, y'_j) + d(x_j, x'_j)
\end{aligned}$$

which goes to 0 as $(x_j) \sim (x'_j)$ and $(y_j) \sim (y'_j)$. This proves that if the limit exists, then it is independent of representative. Now, as $\mathbb{R}$ is complete, to show $\hat{d}(\xi, \eta) = \lim_{j\to\infty} d(x_j, y_j)$ exists it is sufficient to show $d(x_j, y_j)$ is Cauchy. Then

$$\begin{aligned}
|d(x_j, y_j) - d(x_k, y_k)| &= |d(x_j, y_j) - d(x_j, y_k) + d(x_j, y_k) - d(x_k, y_k)| \\
&\leqslant |d(x_j, y_j) - d(x_j, y_k)| + |d(x_j, y_k) - d(x_k, y_k)| \\
&\leqslant d(y_j, y_k) + d(x_j, x_k)
\end{aligned}$$

which goes to 0 as $j$ and $k$ go to $\infty$ as $(y_j)$ and $(x_j)$ are Cauchy, so $d(x_j, y_j)$ is Cauchy. Thus $\hat{d}$ is well defined. Further, $\hat{d}(\xi, \eta) = \lim_{j\to\infty} d(x_j, y_j) \geqslant 0$, $\hat{d}(\xi, \eta) = 0$ if and only if $d(x_j, y_j) \to 0$, which holds if and only if $\xi = \eta$ by definition of $\sim$,

$$\hat{d}(\xi, \eta) = \lim_{j\to\infty} d(x_j, y_j) = \lim_{j\to\infty} d(y_j, x_j) = \hat{d}(\eta, \xi)$$

and

$$\hat{d}(\xi, \eta) = \lim_{j\to\infty} d(x_j, y_j) \leqslant \lim_{j\to\infty} (d(x_j, z_j) + d(z_j, y_j) = \hat{d}(\xi, \mu) + \hat{d}(\mu, \eta)$$

for any $\mu = [(z_j)] \in \hat{X}$. Thus $(\hat{X}, \hat{d})$ is indeed a metric space.

**Example 5.2.1.** Consider $\mathbb{Q}[x]$, the set of polynomials over $\mathbb{Q}$, and define $d(p, q) = \max_{x\in[0,1]} |p(x) - q(x)|$. The max exists as $[0, 1]$ is a compact set in $\mathbb{R}$ and all polynomials are continuous. Then $d(p, q) = 0$ if and only if $p = q$, as polynomials of degree $> 0$ have only a finite number of roots, $d(p, q) = d(q, p)$, and $d(p, q) \leqslant d(p, r) + d(r, q)$ using the triangle inequality for $|\cdot|$ on $\mathbb{R}$. Note this is a countable metric space. But, this is not complete as $\mathbb{Q}$ is not complete. Then $(\hat{X}, \hat{d}) = \{[p_j] : (p_j)$ is Cauchy in $(\mathbb{Q}[x], d)\}$. We will see that $\hat{X} = C([0, 1])$, all continuous functions in $[0, 1]$. This says if $f$ is continuous in $[0, 1]$, for all $\varepsilon > 0$ there exists $p \in \mathbb{Q}[x]$ such that $\hat{d}(p, f) < \varepsilon$. Suppose now we define $X = \mathbb{Q}[x]$ with the distance

$$d_1(p, q) = \max_{x\in[0,1]} |p(x) - q(x)| + \max_{x\in[0,1]} |p'(x) - q'(x)|$$

Note $X \subseteq \mathbb{C}^{\infty}([0, 1])$. Upon completion we obtain $(\hat{X}, \hat{d}_1) = C^1([0, 1])$, the space of all continuous functions with continuous first derivative.

We claim that $\hat{X}$ is indeed a complete metric space:

**Lemma 5.2.1.** *$X$ is dense in $\hat{X}$.*

and

**Proposition 5.2.2.** *$(\hat{X}, \hat{d})$ is complete.*

which follow similarly to the case of $\mathbb{R}$.

## 5.3.0 Compactness for Metric Spaces

Recall that $(X, d)$ denotes an arbitrary metric space.

**Definition 5.3.1.** *$X$ is **sequentially compact** if every sequence $(x_j) \subseteq X$ has a convergent subsequence.*

**Definition 5.3.2.** *$X$ is **limit point compact** if every infinite subset of $X$ has an accumulation point in $X$.*

These are equivalent for metric spaces by the axiom of choice.

**Definition 5.3.3.** *$X$ is **totally bounded** if for every $\varepsilon > 0$ there exists a finite set $\{x_1, ..., x_N\} \subseteq X$ such that*

$$X = \cup_{j=1}^{N} B_\varepsilon(x_j)$$

**Proposition 5.3.1.** *If $X$ is a sequentially compact metric space, then $X$ is totally bounded.*

*Proof.* Let $\varepsilon > 0$. If $X$ is empty, there is nothing to prove, so suppose $X \neq \emptyset$. Let $x_1 \in X$. If $X = B_\varepsilon(x_1)$ we're done. Otherwise, choose $x_2 \in X \backslash B_\varepsilon(x_1)$. If $X = B_\varepsilon(x_1) \cup B_\varepsilon(x_2)$, we're done. Then either this process terminates at a finite step, in which case we're done, or we obtain a sequence $x_1, x_2, ...$ such that $d(x_j, x_i) \geqslant \varepsilon$ for all $i \neq j$. But then $(x_j)$ has no convergent subsequence, contradicting the assumption that $X$ is compact. Thus, this process must terminate at a finite step, so there exists $N \in \mathbb{N}$ such that

$$X = \bigcup_{j=1}^{N} B_\varepsilon(x_j)$$

■

**Corollary 5.3.2.** *If $X$ is a sequentially compact metric space, then $X$ has a countable dense subset, which is to say $X$ is separable.*

*Proof.* By Proposition 5.3.1 $X$ is totally bounded. Let $S_n = \{x_{n1}, ..., x_{nm_n}\}$ such that

$$X = \bigcup_{j=1}^{m_n} B_{2^{-n}}(x_{nj}), \ C := \bigcup_{n=1}^{\infty} S_n$$

Then $C$ is countable being the countable union of finite sets. Let $x \in X$ and $\varepsilon > 0$. Let $n \in \mathbb{N}$ such that $2^{-n} < \varepsilon$. Then $X = \bigcup_{j=1}^{m_n} B_{2^{-n}}(x_{nj})$ so $x \in B_{2^{-n}}(x_{nj})$ for some $1 \leqslant j \leqslant m_n$. Thus $d(x, x_{nj}) < 2^{-n} < \varepsilon$, so $x_{nj} \in B_\varepsilon(x)$ and $B_\varepsilon(x) \cap C \neq \varnothing$. Thus $C$ is dense in $X$. ∎

This implies a relation between sequentially compact metric spaces and its size. In particular, metric plus compact implies separable.

**Proposition 5.3.3.** *If $X$ is a sequentially compact metric space and $K_1 \supseteq K_2 \supseteq ...$ is a chain of non-empty closed subsets, then $\bigcap_{j=1}^{\infty} K_j \neq \varnothing$.*

The proof of this result is identical to that of the case for Euclidean Spaces, and the same holds for its corollary:

**Corollary 5.3.4.** *If $X$ is a sequentially compact metric space and $U_1 \subseteq U_2 \subseteq ...$ is a chain of open sets such that $X = \bigcup_{j=1}^{\infty} U_j$, then there exists $M \in \mathbb{N}$ such that $X = U_M$.*

**Proposition 5.3.5.** *If $X$ is a sequentially compact metric space, then $X$ is topologically compact.*

*Proof.* Let $\{U_\alpha\}_{\alpha \in J}$ be an open cover of $X$. By Proposition **??** $X$ is separable, so we have a countable dense subset $C$. Let $\mathcal{R} = \{B_q(x) : x \in C, q \in \mathbb{Q}^+\}$. Then $\mathcal{R}$ is countable. Further, if $U \subseteq X$ is open, for all $p \in U$ there exists $\varepsilon > 0$ such that $B_\varepsilon(p) \subseteq U$. Then as $C$ is dense, $C \cap B_{\varepsilon/3}(p) \neq \varnothing$, so there exists $c \in C$ such that $c \in B_{\varepsilon/3}(p)$. Then as $\mathbb{Q}$ is dense in $\mathbb{R}$ there exists $q \in \mathbb{Q}$ such that $\varepsilon/3 < q < 2\varepsilon/3$. It follows that $p \in B_q(c) \subseteq B_\varepsilon(p) \subseteq U$. Thus $U = \bigcup\{B \in \mathcal{R} : B \subseteq U\}$, so every open set can be written as a countable union of open sets in $\mathcal{R}$. Then $X = \bigcup_{\alpha \in J} U_\alpha = U_{\alpha \in J} \bigcup_{j \geqslant 1} B_{q_{\alpha,j}}(c_{\alpha,j})$. Suppose $\{B_1, B_2, ...\}$ is a countable cover of $X$. Define $U_m = B_1 \cup ... \cup B_m$. Then by Corollary 5.3.4 there exists $M \in \mathbb{N}$ such that $X = \bigcup_{j=1}^{M} B_j$. Thus as $\bigcup_{\alpha \in J} \bigcup_{j \geqslant 1} B_{q_{\alpha,j}}(c_{\alpha,j})$ is countable, there exists $M \in \mathbb{N}$ such that $X = B_{q_{\alpha_1,i_1}}(c_{\alpha_1,i_1}) \cup ... \cup B_{q_{\alpha_M,i_M}}(c_{\alpha_M,i_M})$, so $X = U_{\alpha_1} \cup ... \cup U_{\alpha_M}$ is a finite subcover. ∎

# Theorem 27 (Heine-Borel Property).

*If $X$ is a metric space, then $X$ is sequentially compact if and only if it is compact in terms of open covers.*

*Proof.* Proposition 5.3.5 is the forward implication, so suppose $X$ is topologically compact. We show the equivalence with limit point compactness. We argue by contrapositive and suppose $S \subseteq X$ has no accumulation points. Then in particular $S$ is closed, as $\overline{S} = S \cup S' = S$, since $S' = \varnothing$. Then $S_x = S \backslash \{x\}$ for all $x \in S$ is also closed, also having no accumulation points. Then $U_x \in X \backslash S_x$ is a cover for $S$, with $U_x \cap S = \{x\}$. But then $\{U_x\}_{x \in S} \cup \{X \backslash S\}$ is an open cover for $X$. As $X$ is open cover compact, we have $x_1, ..., x_n$ such that $U_{x_1}, ..., U_{x_n}, X \backslash S$ covers

$X$. But then $U_{x_1}, ..., U_{x_n}$ covers $S$, so $S = \bigcup_{i=1}^n S \cap U_{x_i} = \{x_1, ..., x_n\}$ so $S$ is finite. Thus, $X$ is limit point compact, completing the proof. ■

**Definition 5.3.4.** *Let $S \subseteq X$ be a subset of a metric space $(X, d)$. We say that $p \in X$ is an* **accumulation point** *of $S$ if for every $\varepsilon > 0$ there exists $q \neq p$ with $q \in S$ such that $q \in B_\varepsilon(p)$. That is $B_\varepsilon(p)^* \cap S \neq \emptyset$ for all $\varepsilon > 0$.*

Note that a metric space which is compact is totally bounded and complete. Now we have the converse:

**Proposition 5.3.6.** *If $X$ is a complete metric space which is totally bounded, then $X$ is compact.*

*Proof.* Let $S \subseteq X$ be infinite. Because $X$ is totally bounded, there exist $x_1, ..., x_N \in X$ such that $X \subseteq \bigcup_{j=1}^N B_{1/2}(x_j)$. Since $S \subseteq X$ is infinite, by the pigeon hole principle there exists $x_j =: x^1$ such that $B_{1/2}(x^1) \cap S$ is infinite. Then there exists $\{x_{2,1}, ..., x_{2,N_2}\} \subseteq X$ such that $X \subseteq \bigcup_{j=1}^{N_2} B_{1/2^2}(x_{2,j})$ and again there exists $x_{2j} =: x^2$ such that $B_{1/2^2}(x^2) \cap (B_{1/2}(x^1) \cap S)$ is infinite. Continuing in this way there exists $x^j \in X$ such that

$$B_{1/2^j}(x^j) \cap ... \cap B_{1/2}(x^1) \cap S$$

is infinite. Let $X_j$ be the closure of this $j$th set, so we have a decreasing chain $X_1 \supseteq X_2 \supseteq ...$ of non-empty closed sets, such that $X_j \cap S$ is infinite for all $j$. Pick $z_1 \in X_1 \cap S$, and $z_{j+1} \in X_{j+1} \cap S \backslash \{z_1, ..., z_j\}$, which is possible using the axiom of choice as each set is infinite. Then $(z_j) \subseteq X_1 \cap S$ is a Cauchy sequence. By completeness there exists $z \in X$ such that $z_j$ converges to $z$. But $(z_j) \subseteq S$, and the $z_j$ are distinct, so $z \in X$ is an accumulation point of $S$. ■

**Proposition 5.3.7.** *If $X$ is a compact metric space, then diam$(X) < \infty$, where*

$$\text{diam}(X) = \sup\{d(x, y) : x, y \in X\}$$

*Proof.* As $X$ is compact there exist $x_1, ..., x_N \in X$ such that $X \subseteq \bigcup_{j=1}^N B_1(x_j)$. Then, let $M = \max\{d(x_i, x_j) : 1 \leqslant i, j \leqslant N\}$. Now, let $x, y \in X$. Then there exist $i, j$ such that $x \in B_1(x_i)$ and $y \in B_1(x_j)$. It follows that

$$d(x, y) \leqslant d(x, x_i) + d(x_i, x_j) + d(x_j, y) < 1 + M + 1 = M + 2$$

Thus, we have that diam$(X) \leqslant M + 2 < \infty$, as desired. ■

## 5.4.0 Product Spaces

We now define finite and countable products of metric spaces, and their associated properties.

**Definition 5.4.1.** *If $(X_1, d_1), ..., (X_N, d_N)$ are metric spaces, we define the product metric space*

$$X := X_1 \times \cdots \times X_N = \prod_{j=1}^N X_j$$

*and define a metric $d$ in $X$ for $x = (x_1, ..., x_N), y = (y_1, ..., y_N)$ by*

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

Equivalently, we could define

$$\delta(x, y) = \sqrt{\sum_{j=1}^{N} d_j(x_j, y_j)^2}$$

where the equivalence is in the sense that they define the same topology on the product. In particular, we characterize two metrics being equivalent as follows:

**Definition 5.4.2.** *If $X$ is a set with metrics $d_1$ and $d_2$, then $d_1$ and $d_2$ are said to be* **equivalent** *if there exists $0 < C_0 \leqslant C_1 < \infty$ such that*

$$C_0 d_1(x, y) \leqslant d_2(x, y) \leqslant C_1 d_1(x, y)$$

*for all $x, y \in X$.*

**Definition 5.4.3.** *For $p \in \mathbb{N}$, the $\ell_p$ norm on $\mathbb{R}^n$ is*

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

*which gives the metric*

$$d_p(\vec{x}, \vec{y}) = ||\vec{x} - \vec{y}||_p$$

*For $p = \infty$, define*

$$||\vec{x}||_\infty = \max_{1 \leqslant j \leqslant n} |x_j|$$

*and*

$$d_\infty(\vec{x}, \vec{y}) = \max_{1 \leqslant j \leqslant n} |x_j - y_j|$$

It is an important, but non-trivial result, that $d_p$ and $d_q$ are equivalent for all $0 < p, q \leqslant \infty$. Now we define countable products:

**Definition 5.4.4.** *Let $(X_1, d_1), (X_2, d_2), ...$ be a countable collection of metric spaces. Define*

$$X = \prod_{j=1}^{\infty} X_j$$

*where $x \in X$ is a sequence $x = (x_j)$, $x_j \in X_j$. We define the metric by*

$$d(x, y) = \sum_{j=1}^{\infty} 2^{-j} \frac{d_j(x_j, y_j)}{1 + d_j(x_j, y_j)}$$

We have that compactness of product spaces occurs if and only if we have compactness of the individual component spaces.

**Proposition 5.4.1.** *The product $X = \prod_{j=1}^{N} X_j$ is compact if and only if $X_j$ is compact for all $j$.*

*Proof.* First, suppose the product is compact and let $(x_{n,j})_{n=1}^{\infty} \subseteq X_j$. Let $x_i \in X_i$ for $i \neq j$. Then $(x_1, ..., x_{n,j}, ..., x_N) \subseteq X$ has a convergent subsequence $(x_1, ..., x_{n_k,j}, ..., x_N)$ converging to $(x_1, ..., x_j, ..., x_N)$ in $X$ since $X$ is compact. Then for all $\varepsilon > 0$, there exists $M \in \mathbb{N}$ such that for $k \geqslant M$, $d_j(x_{n_k,j}, x_j) = d((x_1, ..., x_{n_k,j}, ..., x_N), (x_1, ..., x_j, ..., x_N)) < \varepsilon$, so $(x_{n_k,j})$ is a convergent subsequence of $(x_j)$. Thus $X_j$ is compact for all $j$, as desired. Now suppose that $X_j$ is compact for each $j$, and $((x_{n,1}, ..., x_{n,N})) \subseteq X$. As $X_1$ is compact there exists a subsequence $(x_{s_1(n),1})$ which converges to some $x_1 \in X_1$. Then $(x_{s_1(n),2}) \subseteq X_2$ has a convergent subsequence $(x_{s_2(n),2})$ since $X_2$ is compact, which converges to $x_2 \in X$. Proceeding in this way we arrive at $((x_{s_N(n),1}, ..., x_{s_N(n),N})) \subseteq ((x_{n,1}, ..., x_{n,N}))$, where $(x_{s_N(n),j}) \subseteq (x_{s_j(n),j})$ which converges to $x_j \in X_j$, and hence $(x_{s_N(n),1}, ..., x_{s_N(n),N})$ converges to $(x_1, ..., x_N)$. Thus $X$ is compact, as desired. ∎

**Proposition 5.4.2.** *The product $X = \prod_{j=1}^{\infty} X_j$ is compact if and only if $X_j$ is compact for all $j$.*

*Proof.* The forward implication follows analogously to the proof of the previous proposition. Now, suppose each $X_j$ is compact, and let $((x_{\nu,1}, x_{\nu,2}, ...)) \subseteq X$ be an arbitrary sequence. Then, as $X_1$ is compact, we have a subsequence $x_{s_1(\nu),1}$ which converges to some $x_1 \in X_1$. Then, we can take a subsequence $x_{s_2(\nu),2}$ of $x_{s_1(\nu),2}$ which converges to some $x_2 \in X_2$ since $X_2$ is compact. Let $(x_\nu^j) = ((x_{s_j(\nu),1}, x_{s_j(\nu),2}, ...))$. Then we have a decreasing sequence of subsequences $(x_\nu) \supseteq (x_\nu^1) \supseteq (x_\nu^2) \supseteq ...$, such that $x_{s_j(\nu),i}$ converges to $x_i \in X_i$ for all $1 \leqslant j \leqslant i$. Then, let $(\xi_\nu)$ be the subsequence defined by $\xi_\nu = x_\nu^\nu$, the diagonal. Then for all $j \in \mathbb{N}$, $(\xi_\nu)_{\nu=j}^{\infty} \subseteq (x_\nu^j)_{\nu=1}^{\infty}$, so for each $j$ $x_{s_\nu(\nu),j}$ converges to $x_j \in X_j$. Thus, $\xi_\nu$ converges to $(x_1, x_2, ....) \in X$, and hence $X$ is compact. ∎

## 5.5.0 Baire Category Theorem

We now construct a result on the size of complete metric spaces, known as the Baire's Category Theorem. First we define the notion of a category for a topological space:

**Definition 5.5.1.** *A topological space $(X, \tau)$ is said to be of the* **first category** *if and only if $X$ can be written as a countable union of nowhere dense sets.*

**Definition 5.5.2.** *A subset $S \subseteq X$ is* **nowhere dense** *if and only if $\overline{S}$ does not contain any non-empty open subsets, which occurs in a metric space if and only if $\overline{S}$ does not contain any open ball $B_r(x)$ for $r > 0$ and $x \in X$.*

**Definition 5.5.3.** *A topological space $X$ is said to be of the* **second category** *if and only if $X$ is not of the first category.*

### Theorem 28 (Baire's Category Theorem).

*If $X$ is a complete metric space, then $X$ is of second category.*

*Proof.* Let $S_k \subseteq X$ be a sequence of nowhere dense sets. We claim $X\backslash\bigcup_{k\geq 1} S_k \neq \emptyset$. Let $T_k = \overline{\bigcup_{j=1}^k S_j} = \bigcup_{j=1}^k \overline{S}_j$, so $T_k$ is closed and nowhere dense. Further, $T_1 \subseteq T_2 \subseteq ....$ Let $U_k = X\backslash T_k$, so $U_k$ is open and dense. Indeed, if $x \in X$, and $N_x \subseteq N(x)$, an open neighborhood, $N_x \not\subseteq T_k$, so $N_x \cap U_k \neq \emptyset$. Thus $\overline{U_k} = X$, so $U_k$ is dense. Then we have $U_1 \supseteq U_2 \supseteq ....$ We claim that the intersection is non-empty. Let $p_1 \in U_1$, so there exists $\varepsilon_1 > 0$ such that $\overline{B_\varepsilon(p_1)} \subseteq U_1$. By density of $U_2$, there exists $p_2 \in B_{\varepsilon/2} \cap U_2$. But $U_2$ is open so there exists $\varepsilon_2 < \varepsilon/2$ such that $\overline{B_{\varepsilon_2}(p_2)} \subseteq B_\varepsilon(p_1) \cap U_2$. Inductively, take $p_{k+1} \in B_{\varepsilon_k}(p_k) \cap U_{k+1}$, and $\varepsilon_{k+1} < \varepsilon_k/2$ such that $\overline{B_{\varepsilon_{k+1}}(p_{k+1})} \subseteq B_{\varepsilon_k}(p_k) \cap U_{k+1}$. Then $\varepsilon_k < \varepsilon/2^{k-1}$. Note $(p_k)$ is a Cauchy sequence, because $p_l \in \overline{B_{\varepsilon_l}(p_l)} \subseteq B_{\varepsilon_k}(p_k)$ for all $l > k$, where $\varepsilon_k < \varepsilon/2^{k-1}$. By completeness of $X$, there exists $p \in X$ such that $p_l \to p$. Then $p \in \overline{B_{\varepsilon_k}(p_k)}$ for all $k$ as they are closed, so in particular $p \in U_k$ for all $k$. Therefore, $p \in \bigcap_{i=1}^\infty U_i$, so $\bigcap_{i=1}^\infty U_i \neq \emptyset$, as claimed. ■

## 5.6.0 continuous Functions on Metric Spaces

We now explore the homomorphisms of the category of metric spaces: continuous functions.

**Definition 5.6.1.** *A function $f : X \to Y$ for metric spaces $(X, d_X), (Y, d_Y)$, is continuous at $x \in X$ if whenever $x_j \to x$ in $X$, then $f(x_j) \to f(x)$ in $Y$. We say $f$ is continuous on $X$ if and only if $f$ is continuous at $x \in X$ for all $x$.*

**Proposition 5.6.1.** *A function $f : X \to Y$ is continuous if and only if $U \subseteq Y$ open implies $f^{-1}(U) \subseteq X$ is open, where $f^{-1}(U) = \{x \in X : f(x) \in U\}$.*

This is a common result which can be shown using the axiom of choice.

**Proposition 5.6.2.** *If $f : X \to Y$ is continuous and $K \subseteq X$ is compact in $X$, then $f(K)$ is compact in $Y$.*

*Proof.* Let $(y_j) \subseteq f(K)$. Then there exists $(x_j) \subseteq K$ such that $f(x_j) = y_j$ for all $j$. But, $K$ is compact so there exists a convergent subsequence $(x_{j_\nu})$ which converges to some $x \in K$. As $f$ is continuous, $f(x_{j_\nu}) \to f(x)$. But then $y_{j_\nu} = f(x_{j_\nu}) \to f(x) \in f(K)$, so $(y_j)$ has a convergent subsequence and $f(K)$ is compact. ■

**Proposition 5.6.3.** *If $X$ is a compact metric space and $f : X \to \mathbb{R}$ is continous, then $f$ assumes a max and min value in $X$.*

This follows from the characterization of compact sets in $\mathbb{R}$, which are precisely the closed and bounded sets.

**Definition 5.6.2.** *For $f : X \to \mathbb{R}$, we define*

$$\sup_X f = \begin{cases} \sup_{x\in X} f(x) & \text{if } f(X) \text{ is bounded from above} \\ \infty & \text{if not bounded above} \end{cases}$$

*and*

$$\inf_X f = \begin{cases} \inf_{x \in X} f(x) & \text{if } f(X) \text{ is bounded from below} \\ -\infty & \text{if not bounded below} \end{cases}$$

Using this convention, we define a notion of a limit which always exists in the extended reals, even if the actual limit does not exist.

**Definition 5.6.3.** *For any sequence $(x_n) \subseteq \mathbb{R}$, we define*

$$\limsup_{n\to\infty} x_n := \lim_{n\to\infty} \left( \sup_{k \geqslant n} x_k \right)$$

*and*

$$\liminf_{n\to\infty} x_n := \lim_{n\to\infty} \left( \inf_{k \geqslant n} x_k \right)$$

*Note $\sup_{k \geqslant n} x_k$ is a* **decreasing sequence** *and $\inf_{k \geqslant n} x_k$ is an* **increasing sequence**.

So $\limsup$ is the limit of a monotone decreasing sequence and $\liminf$ is the limit of a monotone increasing sequence. Further we have that $(x_n) \subseteq \mathbb{R}$ is convergent if and only if $\limsup x_n = \liminf x_n$.

## Uniform Continuity

We now define a more powerful notin of continuity of functions on metric spaces.

**Definition 5.6.4.** *A function $f : X \to Y$ is said to be* **uniformly continuous** *if for all $\varepsilon > 0$ there exists $\delta > 0$ such that*

$$d_X(x, y) < \delta \implies d_Y(f(x), f(y)) < \varepsilon$$

*for all $x, y \in X$.*

**Example 5.6.1.** An example of a continuous function which is not uniformly continuous is the ***topoligists sine curve***, $\sin(1/x)$ (Image here)

**Example 5.6.2.** $f(x) = x^2$ is continuous. For any $\varepsilon > 0$, observe $|x^2 - y^2| < \varepsilon \implies |x - y||x + y| < \varepsilon$. If $|x - y| < \delta$, $|x - y||x + y| \leqslant (2|x| + 1)\delta$ if $\delta = \min\left\{\frac{\varepsilon}{2|x|+1}, 1\right\}$. But his depends on $x$. This does work for $x'$ such that $|x'| \leqslant |x|$, but for bigger $x'$ we would need a smaller $\delta$, so $f$ is not uniformly continuous.

**Proposition 5.6.4.** *If $X$ is compact and $f : X \to Y$ is continuous, then $f$ is uniformly continuous.*

*Proof.* Let $\varepsilon > 0$. Let $f(X) = \bigcup_{x \in X} B_{\varepsilon/2}(f(x))$ cover the image. For each $x$ there exist $\delta_x > 0$ such that $f(B_{\delta_x}(x)) \subseteq B_{\varepsilon/2}(f(x))$. Then $X \subseteq \bigcup_{x \in X} B_{\delta_x/2}(x)$ is an open cover, so there exists

$x_1, ..., x_N \in X$ such that $X \subseteq \bigcup_{i=1}^N B_{\delta_{x_i}/2}(x_i)$ since $X$ is compact. Let $\delta = \min_{1 \leqslant i \leqslant N}(\delta_{x_i}/2)$. Then let $x, y \in X$ such that $d_X(x, y) < \delta$. Since the $B_{\delta_{x_i}/2}(x_i)$ cover $X$, there exists $1 \leqslant i \leqslant N$ such that $x \in B_{\delta_{x_i}/2}(x_i)$. Then

$$d_X(x_i, y) \leqslant d_X(x_i, x) + d_X(x, y) < \delta_{x_i}/2 + \delta \leqslant \delta_{x_i}$$

so $x, y \in B_{\delta_{x_i}}(x_i)$. It follows that $f(x), f(y) \in B_{\varepsilon/2}(f(x_i))$ so

$$d_Y(f(x), f(y)) \leqslant d_Y(f(x), f(x_i)) + d_Y(f(x_i), f(y) < \varepsilon/2 + \varepsilon/2 = \varepsilon$$

completing the proof. ■

Now we come to the notion of an isomorphism in the category of metric spaces:

**Definition 5.6.5.** *A function $f : X \to Y$ is said to be a* **homeomorphism** *if it is continuous, bijective, and $f^{-1}$ is continuous.*

**Proposition 5.6.5.** *If $X$ is a compact metric space, then $f : X \to Y$ being continuous and bijective implies $f^{-1}$ is continuous.*

*Proof.* Let $g = f^{-1} : Y \to X$. Note $g$ is continuous in $Y$ if and only if $g^{-1}(V) = f(V)$ is closed in $Y$ for all $V$ closed in $X$. Note $V$ is compact in $X$ if it is closed, so $f(V)$ is compact. Then as $Y$ is a metric space $V$ being compact implies it is closed. Thus $g$ is continuous, as desired. ■

## Sequences and Series of Functions

Next we consider convergence of sequences of functions:

**Definition 5.6.6.** *Let $f_j : X \to Y$, $j \in \mathbb{N}$, be a sequence of functions. If $f : X \to Y$, and $f_j(x) \to f(x)$ for all $x \in X$, we say that $f_j \to f$* **pointwise** *on $X$.*

**Definition 5.6.7.** *A sequence $f_j : X \to Y$ converges* **uniformly** *to $f : X \to Y$ if and only if for all $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that if $j \geqslant N$,*

$$d_Y(f_j(x), f(x) < \varepsilon, \forall x \in X$$

*or equivalently*

$$\sup_{x \in X} d_Y(f_j(x), f(x)) \leqslant \varepsilon$$

**Proposition 5.6.6.** *If $f_j : X \to Y$ are continuous and converge uniformly to $f : X \to Y$, then $f$ is continuous.*

*Proof.* Let $\varepsilon > 0$ and $x \in X$. Then there exists $N \in \mathbb{N}$ such that for $j \geqslant N$, $d_Y(f_j(y), f(y)) < \varepsilon/3$ for all $y \in X$. As $f_N$ is continuous, there exists $\delta > 0$ such that $f_N(B_\delta(x)) \subseteq B_{\varepsilon/3}(f_N(x))$. Then for $d_X(x, y) < \delta$,

$$d_Y(f(x), f(y)) \leqslant d_Y(f(x), f_N(x)) + d_Y(f_N(x), f_N(y)) + d_Y(f_N(y), f_N(x) < \varepsilon$$

so $f$ is continuous. ■

**Example 5.6.3.** If $f_j : [0,1] \to [0,1]$ is defined by $f_j(x) = x^j$, all of which are continuous, then $f_j(x)$ converge pointwise to $f(x) = \begin{cases} 0 & x < 1 \\ 1 & x = 1 \end{cases}$, which is discontinuous, and hence the convergence can't be uniform.

**Definition 5.6.8.** *A sequence of functions $f_j : X \to Y$ is said to be* **uniformly Cauchy** *if for all $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that for $j, k \geqslant N$,*

$$\sup_{x\in X} d_Y(f_j(x), f_k(x)) \leqslant \varepsilon$$

*or equivalently* $\lim_{j,k\to\infty} \sup_{x\in X} d_Y(f_j(x), f_k(x)) = 0$.

**Proposition 5.6.7.** *If $Y$ is a complete metric space and $f_j : X \to Y$ is uniformly Cauchy, then there exists $f : X \to Y$ such that $f_j$ converge uniformly to $f$.*

*Proof.* As $f_j$ is uniformly Cauchy, for each $x \in X$, $(f_j(x)) \subseteq Y$ is Cauchy. As $Y$ is complete there exists $y_x \in Y$ such that $f_j(x) \to y_x$. Then define $f : X \to Y$ by $f(x) = \lim_{j\to\infty} f_j(x)$. Fix $\varepsilon > 0$. As $f_j$ is uniformly Cauchy there exists $N \in \mathbb{N}$ such that $j \geqslant N$ and $k \geqslant 0$ implies $d_Y(f_j(x), f_{j+k}(x)) < \varepsilon$. Taking the limit as $k$ goes to infinity we have $d_Y(f_j(x), f(x)) \leqslant \varepsilon$ for all $x \in X$. Thus, $f_j$ converges uniformly to $f$. ∎

Now we consider basic properties of series:

**Definition 5.6.9.** *We say a sequence $f_j : X \to \mathbb{R}^n$ is* **pointwise summable** *if and only if the sequence*

$$s_n(x) = \sum_{j=0}^{n} f_j(x)$$

*is* **pointwise convergent**.

**Definition 5.6.10.** *$f_j : X \to \mathbb{R}^n$ is* **uniformly summable** *if and only if $s_n(x) = \sum_{j=0}^{n} f_j(x)$ is* **uniformly convergent**.

# Theorem 29 (Weierstrass M-Test (General)).

*Let $f_j : X \to \mathbb{R}^n$ be a sequence of functions. Assume there exist $M_k \in \mathbb{R}$ such that $\sup_{x\in X} ||f_k(x)|| \leqslant M_k$ and*

$$\sum_{k=0}^{\infty} M_k < \infty$$

*Then the series $\sum_{k=0}^{n} f_k(x)$ converges uniformly on $X$ to a limit $s(x)$.*

*Proof.* Suppose the hypotheses of the theorem. As $\sum_{k=0}^{n} M_k$ is Cauchy, for $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that for $j, k \geqslant N$,

$$\left| \sum_{n=j+1}^{k} f_n(x) \right| \leqslant \sum_{n=j+1}^{k} |f_n(x)| \leqslant \sum_{n=j+1}^{k} M_n < \varepsilon$$

for all $x \in X$, so $\sum_{n=0}^{k} f_n(x)$ is uniformly Cauchy, and hence uniformly convergent since $\mathbb{R}^n$ is complete. ■

## 5.7.0 Connected Sets

We now define a notion for what it means for a space to be connected, or not disconnected.

**Definition 5.7.1.** *A topological space $(X, \tau)$ is* **not connected** *or* **disconnected** *if there exists $U, V \in \tau$ such that $U, V \neq \varnothing$, $U \cap V = \varnothing$, and*

$$X = U \sqcup V$$

*If no such $U$ and $V$ exist, then $X$ is said to be* **connected**.

Equivalently a space $X$ is connected if and only if the only clopen sets are $X$ and $\varnothing$.

**Proposition 5.7.1.** *All intervals in $\mathbb{R}$ are connected.*

*Proof.* Suppose $I$ is an interval in $\mathbb{R}$. That is for all $a, b \in I$, with $a < b$, $[a, b] \subseteq I$. Suppose we have a separation $A, B \subseteq \mathbb{R}$ open such that $I = (I \cap A) \cup (I \cap B)$, with $I \cap A, I \cap B \neq \varnothing$, and $(I \cap A) \cap (I \cap B) = \varnothing$. Let $A' = I \cap A$ and $B' = I \cap B$. Let $a \in A' \subseteq I$ and $b \in B' \subseteq I$ so $a \neq b$. Without loss of generality suppose $a < b$. Then $[a, b] \subseteq I$ and covered by $A', B'$. Then let $s = \sup A' \cap [a, b]$. Then $s \leqslant b$. We proceed by cases:

- Suppose $s \in A'$. Then as $A$ is open, there exists $\varepsilon > 0$ such that $(s - \varepsilon, s + \varepsilon) \subseteq A$. Let $\varepsilon' = \min\{\varepsilon, |s - b|\} > 0$. Then $[s, s + \varepsilon') \subseteq A' \cap [a, b]$. In particular $s + \varepsilon'/2 \in A' \cap [a, b]$, contradicting the fact that $s = \sup A' \cap [a, b]$.
- Suppose $s \in B' \cap [a, b]$. As $B$ is open, there exists $\varepsilon > 0$ such that $B_\varepsilon(s) \subseteq B$. Then let $\varepsilon' = \min\{\varepsilon, |s - a|\} > 0$, so $(s - \varepsilon', b] \subseteq B' \cap [a, b]$ and $s - \varepsilon'/2 \in B' \cap [a, b]$, so $t \leqslant s - \varepsilon/2$ for all $t \in A' \cap [a, b]$, contradicting the fact $s = \sup A' \cap [a, b]$.

■

**Definition 5.7.2.** *If $(X, \tau)$ is a topological space and $A \subseteq X$, the subspace topology on $A$ is defined by*

$$\tau_A := \{U \cap A \subseteq A : U \in \tau\}$$

*Note if $\iota : A \hookrightarrow X$ is the inclusion, $\tau_A$ is the coarsest topology/weakest topology making $\iota$ continuous*

$$U \subseteq A \text{ open} \iff \exists V \in \tau; \iota^{-1}(V) = V \cap A = U$$

**Definition 5.7.3.** *A topological space $(X, \tau)$ is* **path connected** *if and only if for all $p, q \in X$, there exists a continuous map*

$$\gamma : [0, 1] \to X,\ \gamma(0) = p,\ \gamma(1) = q$$

**Proposition 5.7.2.** *Path connected implies connected.*

*Proof.* Suppose $X$ is path connected. Towards a contradiction suppose $X$ has a separation $A, B$. Let $a \in A$, $b \in B$, and $\gamma : [0,1] \to X$ with $\gamma(0) = a, \gamma(1) = b$. Then $\gamma^{-1}(A) \subseteq [0,1]$, $\gamma^{-1}(B) \subseteq [0,1]$ are open, disjoint, and cover $[0,1]$, contradicting the fact that $[0,1]$ is connected. Thus $X$ must be connected. ■

**Example 5.7.1.** The metric space

$$X = \{(0,y) \in \mathbb{R}^2 : y \in [-1,1]\} \cup \{(x, \sin 1/x) \in \mathbb{R}^2 : x \in (0,1]\}$$

with metric $d_X = d_{\mathbb{R}^2}$ is compact and connected, but not path connected.

## Theorem 30 (Intermediate Value Theorem).

*Suppose $(X,\tau)$ is a connected space and $f : X \to \mathbb{R}$ is continuous. Suppose $p, q \in X$ such that $f(p) = a < b = f(q)$. Then for all $c \in (a,b)$ there exists $z \in X$ such that $f(z) = c$.*

*Proof.* Let $A = f^{-1}((-\infty, c))$ and $B = f^{-1}((c, \infty))$, so $A$ and $B$ are open, non-empty, and disjoint. Thus, as $X$ is connected, $X \neq A \cup B$ so there must exist $t \in f^{-1}(\{c\})$, so in particular $f(t) = c$. ■

# Chapter 6

# Higher-Dimensional Differentiation

# Chapter 7

# Higher-Dimensional Integration

# Chapter 8

# Manifolds

# Chapter 9

# Differential Forms

# Chapter 10

# Integration On Chains

# Chapter 11

# Integration on Manifolds

# Part III

# Function Spaces

# Chapter 12

# Normed Spaces

# Chapter 13

# Hilbert Spaces

# Chapter 14

# Banach Spaces

# Chapter 15

# Differentiation and Integration on Functions

# Chapter 16

# Banach Algebras

# Part IV

# Measure Theory

# Chapter 17

# Basic Measure Spaces

# Chapter 18

# $L^p$ Spaces

# Chapter 19

# Radon Measures

# Appendices

# .1.0 Multivariate Calculus - with Applications

## Vector Functions and Derivatives

**Definition .1.1.** *A* **vector function** *of one variable is a function $\vec{f} : J \subseteq \mathbb{R} \to \mathbb{R}^n$ defined by $t \mapsto \vec{f}(t)$, where $\vec{f}(t) \in \mathbb{R}^n$ is unique.*

**Definition .1.2.** *Let $\vec{f} : \mathbb{R} \to \mathbb{R}^n$ be a vector valued function. Then we define the derivative of $\vec{f}$ at $t$ by*

$$\frac{d\vec{f}(t)}{dt} = \lim_{\Delta t \to 0} \frac{\vec{f}(t + \Delta t) - \vec{f}(t)}{\Delta t} \tag{.1.1}$$

**Remark .1.1** (Properties)**.** Let $\vec{f} : J \subseteq \mathbb{R} \to \mathbb{R}^n$ and $\vec{g} : I \subseteq \mathbb{R} \to \mathbb{R}^n$ be vector functions such that

$$\vec{f} = \langle f_1, ..., f_n \rangle \text{ and } \vec{g} = \langle g_1, ..., g_n \rangle \tag{.1.2}$$

The for all $t \in J \cap I$ and all $\lambda : D_\lambda \subseteq \mathbb{R} \to \mathbb{R}$ we have

1. $(\vec{f} + \vec{g})(t) := \vec{f}(t) + \vec{g}(t)$
2. $(\lambda \vec{f})(t) := \lambda(t)\vec{f}(t)$
3. $(\vec{f} \cdot \vec{g})(t) = \vec{f}(t) \cdot \vec{g}(t)$
4. For $n = 3$, $(\vec{f} \times \vec{g})(t) = \vec{f}(t) \times \vec{g}(t)$

**Recall .1.2.** Given $A \in \mathbb{R}^{n \times n}$, the lagrange expansion is

$$\det(A) = \sum_{i=1}^{n} (-1)^{i+j} \det(A_{ij}) \text{ (along } j\text{-th column)} \tag{.1.3}$$

$$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} \det(A_{ij}) \text{ (along } i\text{-th row)} \tag{.1.4}$$

where $A_{ij}$ is the minor matrix of $A$ obtained by deleting the ith ro and jth column of $A$.

**Definition .1.3.** *Suppose $\vec{f}(t) = \langle f_1(t), ..., f_n(t) \rangle$ and $\vec{L} = \langle L_1, ..., L_n \rangle$ Then*

$$\lim_{t \to t_0} \vec{f}(t) = \vec{L} \implies \lim_{t \to t_0} f_i(t) = L_i, \forall 1 \leqslant i \leqslant n \tag{.1.5}$$

**Remark .1.3.** The right hand limit $\lim_{t \to t_0^+} \vec{f}(t)$ and the left hand limit $\lim_{t \to t_0^-} \vec{f}(t)$ are defined in the same way.

**Remark .1.4** (Limit Rules)**.** If $\lim_{t \to t_0}$ for $\vec{f}(t)$, $\vec{g}(t)$, and $\lambda(t)$ exist and $k \in \mathbb{R}$, then

1. $\lim_{t \to t_0} (\vec{f}(t) + \vec{g}(t)) = \lim_{t \to t_0} \vec{f}(t) + \lim_{t \to t_0} \vec{g}(t)$

2. $\lim_{t \to t_0} k\vec{f}(t) = k \lim_{t \to t_0} \vec{f}(t)$

3. $\lim_{t \to t_0} \lambda(t)\vec{f}(t) = (\lim_{t \to t_0} \lambda(t))(\lim_{t \to t_0} \vec{f}(t))$

4. $\lim_{t \to t_0} \vec{f}(t) \cdot \vec{g}(t) = (\lim_{t \to t_0} \vec{f}(t)) \cdot (\lim_{t \to t_0} \vec{g}(t))$

5. $\lim_{t \to t_0} \vec{f}(t) \times \vec{g}(t) = (\lim_{t \to t_0} \vec{f}(t)) \times (\lim_{t \to t_0} \vec{g}(t))$, for $n = 3$.

**Definition .1.4 (Continuity).** *A vector function $\vec{f}(t) = \langle f_1(t), ..., f_n(t) \rangle$ is said to be continuous at $t = t_0$ if*

$$\lim_{t \to t_0} \vec{f}(t) = \vec{f}(t_0) \tag{.1.6}$$

*In other words, each component function is continuous at $t = t_0$.*

**Definition .1.5 (Differentiability).** *A vector function $\vec{f}(t) = \langle f_1(t), ..., f_n(t) \rangle$ is said to be differentiable at $t = t_0$ if $\vec{f}(t)$ is defined at and around $t$ and*

$$\lim_{t \to t_0} \frac{\vec{f}(t) - \vec{f}(t_0)}{t - t_0} \tag{.1.7}$$

*exists. We call this limit the derivative of $\vec{f}(t)$ at $t = t_0$ and is denoted by $\vec{f'}(t_0)$ if it exists.*

**Theorem .1.1.** *We say that the vector function is differentiable at $t = t_0$ if and only if its component functions are differentiable at $t = t_0$, and*

$$\vec{f'}(t_0) = \langle f_1'(t_0), ..., f_n'(t_0) \rangle \tag{.1.8}$$

**Remark .1.5** (Differentiation Rules)**.** Let $\vec{f}(t), \vec{g}(t)$ and $\lambda(t)$ be differentiable and $k \in \mathbb{R}$. Then

1. $(\vec{f} + \vec{g})'(t) = \vec{f'}(t) + \vec{g'}(t)$
2. $(k\vec{f})'(t) = k\vec{f'}(t)$
3. $(\lambda\vec{f})'(t) = \lambda'(t)\vec{f}(t) + \lambda(t)\vec{f'}(t)$
4. $(\vec{f} \cdot \vec{g})'(t) = \vec{f'}(t) \cdot \vec{g}(t) + \vec{f}(t) \cdot \vec{g'}(t)$
5. $(\vec{f} \times \vec{g})'(t) = \vec{f'}(t) \times \vec{g}(t) + \vec{f}(t) \times \vec{g'}(t)$ for $n = 3$
6. $(\vec{f}(\lambda(t)))' = \vec{f'}(\lambda(t))\lambda'(t)$

**Definition .1.6.** *Let $\vec{f}(t) = \langle f_1(t), ..., f_n(t) \rangle$ be a vector function defined on a closed interval $[a, b]$. We say that $\vec{f}(t)$ is* **integrable** *on $[a, b]$ if each $f_i(t)$ is integrable on $[a, b]$. When that is the case we define*

$$\int_a^b \vec{f}(t)dt := \left\langle \int_a^b f_1(t)dt, ..., \int_a^b f_n(t)dt \right\rangle \tag{.1.9}$$

*the definite integral of $\vec{f}(t)$ on $[a, b]$*

**Remark .1.6** (Properties)**.** Let $\vec{f}(t), \vec{g}(t)$ be integrable on $[a,b]$ and $k \in \mathbb{R}$. Then

1. $\int\limits_a^b k\vec{f}(t)dt = k\int\limits_a^b \vec{f}(t)dt$
2. $\int\limits_a^b (\vec{f}(t) + \vec{g}(t))dt = \int\limits_a^b \vec{f}(t)dt + \int\limits_a^b \vec{g}(t)dt$
3. $\int\limits_a^b \vec{f}(t)dt = \int\limits_a^c \vec{f}(t)dt + \int\limits_c^b \vec{f}(t)dt,\ a \leqslant c \leqslant b.$
4. $\left|\int\limits_a^b \vec{f}(t)dt\right| \leqslant \int\limits_a^b \left|\vec{f}(t)\right| dt$ (***The triangle inequality***)

## Parametric Curves and Paths

**Definition .1.7.** *Let $\vec{f} : J \subseteq \mathbb{R} \to \mathbb{R}^n$ given by $t \mapsto \langle f_1(t), ..., f_n(t)\rangle$, and $f_j : J \subseteq \mathbb{R} \to \mathbb{R}$ be the $j$th component function of $\vec{f}$. We define the image*

$$\vec{f}(J) = \mathscr{C} \tag{.1.10}$$

*and call $\mathscr{C}$ a* **curve parametrized** *by $\vec{f}$. If $J = [a,b]$ for $a \leqslant b \in \mathbb{R}$ and $\vec{f}(a) = \vec{f}(b)$, then $\mathscr{C}$ is a* **closed curve***. If there exists a parametrization $\vec{g} : I \subseteq \mathbb{R} \to \mathbb{R}^n$ of $\mathscr{C}$ such that $\vec{g}$ is injective (except maybe at end points), then $\mathscr{C}$ is said to be a* **non-self intersecting** *curve. If such a curve is closed it is called a* **simple closed curve***.*

**Remark .1.7.** The pair $(\vec{f}(t), J)$ is called a ***parametrization*** of the curve $\mathscr{C}$. The triple $(\vec{f}(t), J, \mathscr{C})$ is called a ***path with curve $\mathscr{C}$***.

**Definition .1.8.** *If $\vec{f} : J \subseteq \mathbb{R} \to \mathbb{R}^n$ is a one-to-one function, then the image $\vec{f}(J) = \mathscr{C}$ is an* **oriented curve** *and $\vec{f}$ is a* **consistently oriented path** *which covers $\mathscr{C}$.*

**Remark .1.8** (Ellipse)**.** The parametrization of an ellipse with equation $\frac{(x-x_0)^2}{a^2} + \frac{(y-y_0)^2}{b^2} = 1$, $a, b > 0$ are constants, is $\vec{f}(t) = \langle x_0 + a\cos(t), y_0 + b\sin(t)\rangle$, for $t \in [0, 2\pi)$.

**Remark .1.9** (Line)**.** Let $A$ and $B$ be points in $\mathbb{R}^n$. The parametrization of the line segment connecting $A$ to $B$ is

$$(\vec{f}(t) = \overline{OA} + t(\overline{OB} - \overline{OA}), [0,1]) \tag{.1.11}$$

**Definition .1.9.** *The tangent vector to a curve $\mathscr{C}$ with parametrization $(\vec{f}, I)$ exists at a point $t = t_0$ if $\vec{f}$ is $\vec{f}'(t)$ exists.*

**Definition .1.10.** *Let $\mathscr{C}$ be a parametric curve with parametrization $(\vec{f}(t), [a,b])$. If $\vec{f}(t)$ is differentiable at $t = t_0$, then $\vec{f}(t_0)$ is called a* **tangent vector** *to $\mathscr{C}$ at $P_0 = tip(\vec{f}(t_0))$, provided $\vec{f}'(t_0) \neq \vec{0}$. If $\vec{f}(t)$ is differentiable at every $t \in (a,b)$ and $\vec{f}'(t) \neq \vec{0}$, we say that the parametric curve $\mathscr{C}$ is a* **smooth parametric curve***.*

**Definition .1.11.** *Let $\mathscr{C}$ be a parametric curve and let $(\vec{f}(t), I)$ be a parametrization. If $P_0 = tip(\vec{f}(t_0))$, for $t_0 \in I$, such that $\vec{f}'(t_0) \neq \vec{0}$. Then*

$$\vec{T}(t_0) = \frac{1}{|\vec{f}'(t_0)|}\vec{f}'(t_0) \tag{.1.12}$$

*is called the* **unit tangent vector** *associated with the parametrization $(\vec{f}(t), I)$. $\vec{T}(t_0)$ always forces the direction in which $\vec{f}(t)$ traces $\mathscr{C}$. The vector*

$$\vec{N}(t_0) = \frac{1}{|\vec{T}'(t_0)|}\vec{T}'(t_0) \tag{.1.13}$$

*is perpendicular to $\vec{T}(t_0)$ and is called the* **unit principal normal** *to $\mathscr{C}$ at $P_0$.*

**Definition .1.12.** *A curve $\mathscr{C}$ is called* **piecewise smooth** *if it consists of a finite number of smooth parametric curves $\mathscr{C}_1, ..., \mathscr{C}_k$, where the endpoint of $\mathscr{C}_i$ is the starting point of $\mathscr{C}_{i+1}$ for $i = 1, 2, ..., k-1$.*

**Definition .1.13.** *Let $\mathscr{C}$ be a bounded continuous curve specified by a parametrization $\vec{f} : [a, b] \to \mathbb{R}^n$. We consider partitions of $[a, b]$ into $n$-subintervals by*

$$a = t_0 < t_1 < ... < t_n = b \tag{.1.14}$$

*So the points $\vec{f}(t_i)$ subdivide $\mathscr{C}$, and using the* **chord length** *$|\vec{f}(t_i) - \vec{f}(t_{i-1})|$ we define the sequence of lengths approximating $\mathscr{C}$ by*

$$s_n = \sum_{i=1}^{n} |\vec{f}(t_i) - \vec{f}(t_{i-1})| \tag{.1.15}$$

*We say $\mathscr{C}$ is* **rectifiable** *if there exists $K \in \mathbb{R}$ such that $s_n \leqslant K$ for all $n \in \mathbb{N}$ and all choices of points. From the completeness axiom of $\mathbb{R}$ there exists a least such $K$. This $K$ we define as the* **length** *of $\mathscr{C}$ and we denote it by $s$. Let $\Delta t_i = t_i - t_{i-1}$ and $\Delta\vec{f}_i = \vec{f}(t_i) - \vec{f}(t_{i-1})$ so*

$$s_n = \sum_{i=1}^{n} \left|\frac{\Delta\vec{f}_i}{\Delta t_i}\right| \Delta t_i \tag{.1.16}$$

*If $\vec{f}(t)$ has a continuous derivative, then*

$$s = \lim_{\substack{n\to\infty \\ sup\Delta t_i \to 0}} s_n = \int_a^b \left|\frac{d\vec{f}}{dt}\right| dt \tag{.1.17}$$

*is the* **arclength**.

**Definition .1.14.** *Let $\vec{f} : [a, b] \to \mathbb{R}^n$ be a smooth parammetrization of a curve $\mathscr{C}$. Then the* **arclength function** *of $\vec{f}$ is a function $s : [a, b] \to \mathbb{R}$ where*

$$s(t) := \int_a^b \left|\frac{d\vec{f}}{dt}\right| dt \tag{.1.18}$$

*and the* **arclength element** *for* $\mathscr{C}$ *is given by*

$$ds := \left|\frac{d}{dt}\vec{f}\right| dt \tag{.1.19}$$

**Definition .1.15.** *If* $\vec{f} : J \subseteq \mathbb{R} \to \mathbb{R}^n$ *parametrizes a curve* $\mathscr{C}$ *with the parameter being the arclenght along the curve relative to some inital point, then we call this an* **arclength** *or* **intrinsic parametrization***. Such a parametrization traces* $\mathscr{C}$ *at unit speed*

$$\left|\frac{d\vec{f}(s)}{dt}\right| = 1 \tag{.1.20}$$

## Curvature and the Frenet Frame

The Frenet Frame forms a right handed orthogonal basis at any point along a smooth curve. The curvature measures the rate at which a curve is turning (away from its tangent line) at any point. The Torsion measures the rate at which the curve is twisting (out of the plane in which it is turning) at any point.

**Definition .1.16.** *Let* $\gamma : I \subseteq \mathbb{R} \to \mathbb{R}^n$ *denote a smooth parametric curve. Then* $d\gamma/dt$ *gives a tangent vector to the curve at any point, and it points in the direction of the orientation of the curve. We assume that this derivative is nowhere zero so we can define a* **unit tangent vector***,* $\hat{T}(t)$*, at* $\gamma(t)$ *by*

$$\hat{T}(t) := \frac{d\gamma(t)/dt}{||d\gamma(t)/dt||}$$

Note that when the parameterization is in terms of arclength the denominator is 1. If $\gamma(t)$ has a continuous nonvanishing derivative, then $\hat{T}(t)$ is continuous and the angle it makes with any fixed vector is also continuous.

Having unit length the tangent vector field satisfies $\hat{T}(t) \cdot \hat{T}(t) = 1$, so $2\hat{T}(t) \cdot \frac{d\hat{T}(t)}{dt} = 0$. Thus, the derivative of the tangent is orthogonal to the tangent at any point on the curve.

**Definition .1.17.** *The* **curvature** *of a smooth curve* $\gamma : I \subseteq \mathbb{R} \to \mathbb{R}^n$ *parameterized in terms of arclength is defined to be*

$$\kappa(s) = \left|\frac{d\hat{T}(s)}{ds}\right|$$

*The* **radius of curvature***, denoted* $\rho$*, is the reciprocal of the curvature:*

$$\rho(s) = \frac{1}{\kappa(s)}$$

The curvature of $\gamma$ measures the rate of turning of the tangent line to the curve at any point.

**Definition .1.18.** *If the curvature of a smooth curve $\gamma$ is nonzero, we define the* **unit principal normal** *to $\gamma$ in terms of an arclength parameterization to be*

$$\hat{N}(s) = \frac{d\hat{T}(s)/ds}{||d\hat{T}(s)/ds||}$$

Note that the principal normal is perpendicular to the curve at any point, and points in the direction that the unit tangent, and consequently the curve, is turning.

**Theorem .1.2.** *Let $\kappa > 0$ on an interval containing $s$ and let $\delta\theta$ be the angle between $\hat{T}(s+\delta s)$ and $\hat{T}(s)$, the unit tangent vectors at neighboring points on the curve. Then*

$$\kappa(s) = \lim_{\delta s \to 0} \left| \frac{\delta s}{\delta s} \right|$$

The plane passing through $\gamma(s)$ and spanned by $\hat{T}(s)$ and $\hat{N}(s)$ is called the ***osculating plane*** to the curve at $\gamma(s)$.

Assuming $\kappa(s) \neq 0$, let $\gamma_c(s) = \gamma(s) + \rho(s)\hat{N}(s)$. Then $\gamma_c(s)$ is called the ***center of curvature*** to $\gamma$ at $\gamma(s)$, and the circle in the osculating plane with center $\gamma_c(s)$ and radius $\rho(s)$ is the best approximation of the curve near $\gamma(s)$, and is called the ***osculating circle***.

**Definition .1.19.** *At any point on $\gamma(s)$ where $\hat{T}(s), \hat{N}(s)$ are both defined, we define the* **unit binormal** *field to be*

$$\hat{B}(s) = \hat{T}(s) \times \hat{N}(s)$$

The three vectors $\{\hat{T}(s), \hat{N}(s), \hat{B}(s)\}$ correspond to a right-handed coordinate system, the ***Frenet Frame*** for $\gamma$ at any point $\gamma(s)$. Note that

$$\frac{d\hat{B}}{ds} = \hat{T} \times \frac{d\hat{N}}{ds}$$

and as with any unit normal field, $\hat{B}$'s derivative is orthogonal to itself, so we have that $\frac{d\hat{B}}{ds}$ lies on the line spanned by $\hat{N}$.

**Definition .1.20.** *On any interval where $\kappa(s) \neq 0$ there exists a function $\tau(s)$ such that*

$$\frac{d\hat{B}}{ds} = -\tau(s)\hat{N}(s)$$

*The number $\tau(s)$ is called the* **torsion** *of $\gamma$ at $\gamma(s)$.*

The torsion measures the degree of twisting that the curve exhibits near a point, that is, the extent to which the curve fails to be planar.

Let $\delta\psi$ be the angle between $\hat{B}(s)$ and $\hat{B}(s+\delta s)$. Then

$$|\tau(s)| = \lim_{\delta s \to 0} \left| \frac{\delta\psi}{\delta s} \right|$$

**Proposition .1.3.** *The* **Frenet-Serret Formulas** *for our Frenet Frame are given by*

$$\frac{d\hat{T}}{ds} = \kappa(s)\hat{N}(s)$$
$$\frac{d\hat{N}}{ds} = -\kappa(s)\hat{T}(s) + \tau(s)\hat{B}(s)$$
$$\frac{d\hat{B}}{ds} = -\tau(s)\hat{N}(s)$$

*These can be summarized by the matrix equation*

$$\frac{d}{ds}\begin{bmatrix}\hat{T}(s)\\ \hat{N}(s)\\ \hat{B}(s)\end{bmatrix} = \begin{bmatrix}0 & \kappa(s) & 0\\ -\kappa(s) & 0 & \tau(s)\\ 0 & -\tau(s) & 0\end{bmatrix}\begin{bmatrix}\hat{T}(s)\\ \hat{N}(s)\\ \hat{B}(s)\end{bmatrix}$$

**Theorem .1.4.** *Let $\gamma_1$ and $\gamma_2$ be two curves, both of which have the same nonvanishing curvature function $\kappa(s)$ and the same torsion function $\tau(s)$. Then the curves are congruent. That is, there exists a rigid motion so as to coincide one curve exactly with the other.*

## Functions of Several Variables

**Definition .1.21.** *A function $f : \mathscr{D}(f) \subseteq \mathbb{R}^n \to \mathbb{R}$, where $\mathscr{D}(f)$ is the* **domain** *of $f$, is called a* **scalar field on $\mathscr{D}(f)$**. *The image of $f$ is*

$$\mathrm{Im}(f) := \{x \in \mathbb{R} : \exists \vec{v} \in \mathscr{D}(f), f(\vec{v}) = x\} \tag{.1.21}$$

*The natural domain of $f$ is the largest subset of $\mathbb{R}^n$ such that $f$ is well-defined.*

**Definition .1.22.** *The* **graph** *of a function $f : A \to B$ is the set*

$$\Gamma(f) := \{(x, f(x)) : x \in A\} \tag{.1.22}$$

*For*

$$f : \prod_{i=1}^{n} X_i \to B \tag{.1.23}$$

*we have the graph*

$$\Gamma(f) := \{((x_1, ..., x_n), f(x_1, ..., x_n)) : (x_1, ..., x_n) \in \prod_{i=1}^{n} X_i\} \tag{.1.24}$$

**Remark .1.10.** The grap of a function $f : \mathbb{R}^n \to \mathbb{R}$ can be considered as a surface in $\mathbb{R}^{n+1}$ by a natural embedding.

**Definition .1.23.** *Given a function $f : \mathbb{R}^n \to \mathbb{R}$, a* **k-level surface** *of $f$ is a set*

$$S_k := \{\vec{x} \in \mathbb{R}^n : f(\vec{x}) = k\} \tag{.1.25}$$

*where $k$ is a fixed constant.*

**Definition .1.24.** *Let $\vec{x}_0 \in \mathbb{R}^n$ and $r > 0$ a real number. Then the* **open ball of radius** $r$ *centered at $\vec{x}_0$ is defined as*

$$B_r(\vec{x}_0) := \{\vec{x} \in \mathbb{R}^n : ||\vec{x} - \vec{x}_0|| < r\} \qquad (.1.26)$$

*the* **closed ball** *is defined by*

$$\overline{B}_r(\vec{x}_0) := \{\vec{x} \in \mathbb{R}^n : ||\vec{x} - \vec{x}_0|| \leqslant r\} \qquad (.1.27)$$

**Definition .1.25.** *A* **neighborhood** *of a point $\vec{x}_0 \in \mathbb{R}^n$ is any set $U \subseteq \mathbb{R}^n$ such that there exists $r > 0$ so that $B_r(\vec{x}_0) \subseteq U$.*

**Definition .1.26.** *Let $E \subseteq \mathbb{R}^n$, where we equip $\mathbb{R}^n$ with $\mathcal{T}_{st}$. We say $E$ is* **open** *if $E \in \mathcal{T}_{st}$. Equivalently, $E$ is* **open** *if for all $\vec{x} \in E$, $E$ is a* **neighborhood** *of $\vec{x}$. We sau $E$ is* **closed** *if its* **complement** *$E^C = \mathbb{R}^n \backslash E$ is* **open**.

**Definition .1.27.** *A point $\vec{x}_0 \in \mathbb{R}^n$ is called a* **boundary point** *of $E$ if for any $r > 0$, $B_r(\vec{x}_0) \cap E \neq \varnothing$ and $B_r(\vec{x}_0) \cap E^C \neq \varnothing$.*

**Definition .1.28.** *The set of all* **boundary points** *of a set $E \subseteq \mathbb{R}^n$ is called the* **boundary** *of $E$.*

**Definition .1.29.** *Let $E \subseteq \mathbb{R}^n$. We say that $E$ is* **bounded** *if there exists $R > 0$ such that $||\vec{x}|| \leqslant R$ for all $\vec{x} \in E$. $E$ is* **unbounded** *if for all $R > 0$ there exists $\vec{x}_0 \in E$ such that $||\vec{x}_0|| > R$.*

**Definition .1.30.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function. We say $\lim_{\vec{x} \to \vec{x}_0} f(\vec{x}) = L$, provided that*

1. *Every punctered neighborhood $B_r^*(\vec{x}_0)$ of $\vec{x}_0$ intersects $\mathscr{D}(f)$*

$$B_r^*(\vec{x}_0) \cap \mathscr{D}(f) \neq \varnothing \qquad (.1.28)$$

   *that is, $\vec{x}_0$ is a limit point of $\mathscr{D}(f)$.*

2. *For all $\varepsilon > 0$, there exists $\delta > 0$ such that $f(\vec{x}) \in B_\varepsilon(L)$ whenever $\vec{x} \in B_\delta^*(\vec{x}_0) \cap \mathscr{D}(f)$. That is*

$$f(\mathscr{D}(f) \cap B_\delta^*(\vec{x}_0)) \subseteq B_\varepsilon(L) \qquad (.1.29)$$

**Remark .1.11.** As $\mathbb{R}^n$ and $\mathbb{R}$ are metric spaces, they are Hausdorff, so if the limit exists it is unique.

**Remark .1.12** (Limit Properties)**.** Let $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}$ and $\vec{x}_0 \in \mathbb{R}^n$ such that $\lim_{\vec{x} \to \vec{x}_0} f(\vec{x}_0) = L$ and $\lim_{\vec{x} \to \vec{x}_0} g(\vec{x}) = M$. Then if $\vec{x}_0$ is not an isolated point of $\mathscr{D}(f) \cap \mathscr{D}(g)$, then

1. $\lim_{\vec{x} \to \vec{x}_0}(f(\vec{x}) \pm g(\vec{x})) = L \pm M$
2. $\lim_{\vec{x} \to \vec{x}_0} f(\vec{x})g(\vec{x}) = LM$
3. $\lim_{\vec{x} \to \vec{x}_0} \frac{f(\vec{x})}{g(\vec{x})} = \frac{L}{M}$ if $M \neq 0$
4. If $F : \mathbb{R} \to X$ is continuous at $L$, then $\lim_{\vec{x} \to \vec{x}_0} F(f(\vec{x})) = F(L)$.

**Definition .1.31.** *We say a function $f : \mathbb{R}^n \to \mathbb{R}$ is* **continuous at a point** $\vec{x}_0 \in \mathbb{R}^n$ *if*

$$\lim_{\vec{x}\to\vec{x}_0} f(\vec{x}) = f(\vec{x}_0) \tag{.1.30}$$

**Remark .1.13.** If setting $x = x_0$ and $y = y_0$ in the expression for $f(x,y)$ does not evalutate to a real number, then we can try using polar coordinates: $x = x_0 + r\cos(\theta)$ and $y = y_0 + r\sin(\theta)$. Recall $r = \sqrt{(x-x_0)^2 + (y-y_0)^2}, 0 \leqslant \theta < 2\pi$. As a result, $(x,y) \to (x_0,y_0)$ is equivalent to $r \to 0$, so

$$\lim_{(x,y)\to(x_0,y_0)} f(x,y) = \lim_{r\to 0} f(x_0 + r\cos(\theta), y_0 + r\sin(\theta)) \tag{.1.31}$$

**Theorem .1.5** (Squeeze Theorem)**.** *Let $f(x,y)$, $g(x,y)$ and $h(x,y)$ be defined in a neighborhood $U$ of $(x_0,y_0)$, except maybe at $(x_0,y_0)$, and such that*

$$g(x,y) \leqslant f(x,y) \leqslant h(x,y), \forall (x,y) \in U\backslash\{(x_0,y_0)\} \tag{.1.32}$$

*If*

$$\lim_{(x,y)\to(x_0,y_0)} g(x) = L, \quad \lim_{(x,y)\to(x_0,y_0)} h(x) = L \tag{.1.33}$$

*Then* $\lim_{(x,y)\to(x_0,y_0)} f(x) = L$.

**Theorem .1.6.** *If one can find two continuous parametric curves $\mathscr{C}_1$ and $\mathscr{C}_2$ that pass through the point $(x_0,y_0)$ such that*

$$\lim_{\substack{(x,y)\to(x_0,y_0)\\(x,y)\in\mathscr{C}_1}} f(x,y) = L_1, \quad \lim_{\substack{(x,y)\to(x_0,y_0)\\(x,y)\in\mathscr{C}_2}} f(x,y) = L_2, \text{ and } L_1 \neq L_2 \tag{.1.34}$$

*then* $\lim_{(x,y)\to(x_0,y_0)} f(x,y)$ *does not exist.*

**Definition .1.32.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function.*

1. *We say that $f$ is* **continuous** *at $\vec{x}_0 \in \mathbb{R}^n$ if*
   (a) *There exists a neighborhood $U$ of $\vec{x}_0$ such that $f(U)$ is defined*
   (b) $\lim_{\vec{x}\to\vec{x}_0} f(\vec{x}) = f(\vec{x}_0)$
2. *We say that $f$ is continuous on a region $D$ if it s continuous at every point $\vec{x}$ in the region.*

**Remark .1.14** (Constructing Continuous Functions)**.** Let $f, g : \mathbb{R}^n \to \mathbb{R}$ be continuous at $\vec{x}_0 \in \mathbb{R}^n$, and if $\lambda \in \mathbb{R}$, then

1. $f \pm g$, $f \cdot g$, and $\lambda f$ are continuous at $\vec{x}_0$
2. $f/g$ is continuous at $\vec{x}_0$ provided $g(\vec{x}_0) \neq 0$.

Suppose $u(t)$ is continuous at $t_0 = f(\vec{x}_0)$. Then $u(f(\vec{x}))$ is continuous at $\vec{x}_0$.

## Theorem 31 (Extreme Value Theorem).

*If $f : \mathbb{R}^n \to \mathbb{R}$ is continuous on a closed and bounded region $D \subseteq \mathbb{R}^n$, then there exist $\vec{x}_m, \vec{x}_M \in D$ such that*

$$f(\vec{x}_m) \leqslant f(\vec{x}) \leqslant f(\vec{x}_M), \forall \vec{x} \in D \tag{.1.35}$$

*$m = f(\vec{x}_m)$ is called the **absolute minimum** of $f$ on $D$, while $M = f(\vec{x}_M)$ is called the **absolute maximum** of $f$ on $D$.*

### Partial Derivatives

**Definition .1.33.** *The **first partial derivatives** of a function $f : \mathbb{R}^n \to \mathbb{R}$ with respect to the variable $x_i$, $1 \leqslant i \leqslant n$, is the function*

$$f_i(x_1, ..., x_n) = \lim_{h \to 0} \frac{f(x_1, ..., x_i + h, ..., x_n) - f(x_1, ..., x_i, ..., x_n)}{h} \tag{.1.36}$$

*provided the limit exists and $f$ is defined in a neighbrohood of $(x_1, ..., x_n)$.*

**Notation .1.15.** We often write

$$\frac{\partial}{\partial x_i} f(x_1, ..., x_n) = f_i(x_1, ..., x_n) = D_i f(x_1, ..., x_n) \tag{.1.37}$$

and

$$\left(\frac{\partial}{\partial x_i} f(\vec{x})\right)\bigg|_{\vec{x}_0} = f_i(\vec{x}_0) = D_i f(\vec{x}_0) \tag{.1.38}$$

**Remark .1.16.** If a function $f : J \subseteq \mathbb{R}^n \to \mathbb{R}$ has first partial derivatives at $\vec{x}_0$ in a region $D \subseteq \mathbb{R}^n$, then this defines $n$ new functions

$$\frac{\partial f}{\partial x_i}\bigg| : \mathbb{R} \to \mathbb{R} \tag{.1.39}$$

where we differentiate $f$ with respect to $x_i$.

**Definition .1.34.** *Given a function $f : \mathbb{R}^n \to \mathbb{R}$, the **gradient** of $f$ is the vector function*

$$\nabla f = grad(f) : \mathbb{R}^n \to \mathbb{R}^n \tag{.1.40}$$

*such that*

$$\nabla f(x_1, ..., x_n) = \left\langle \frac{\partial}{\partial x_1} f, ..., \frac{\partial}{\partial x_n} f \right\rangle \tag{.1.41}$$

*where $\nabla$ is the **del operator***

$$\nabla = \left[\frac{\partial}{\partial x_1}, ..., \frac{\partial}{\partial x_n}\right]^T \tag{.1.42}$$

**Definition .1.35.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be defined in a neighborhood of a point $\vec{x}_0$ such that its first partial derivatives exist at $\vec{x}_0$. Then by definition, the **linear approximation** of $f$ at $\vec{x}_0$ is the polynomial of degree 1, $L(\vec{x})$, that matches $f$ at $\vec{x}_0$ and matches its partials at $\vec{x}_0$. In particular, we have that*

$$L(\vec{x}, \vec{x}_0) = f(\vec{x}_0) + \sum_{i=1}^{n} \frac{\partial}{\partial x_i} f(\vec{x}_0)(x_i - x_{i,0}) \tag{.1.43}$$

**Definition .1.36.** *Let $f : D \subseteq \mathbb{R}^n \to \mathbb{R}$ be a function defined around $\vec{x}_0 \in \mathbb{R}^n$ with first partials also defined. Let $L(\vec{x})$ be its linear approximation at $\vec{x}_0$. We say that $f$ is* **differentiable** *at $\vec{x}_0$ if the limit*

$$\lim_{\vec{x}\to\vec{x}_0} \frac{f(\vec{x}) - L(\vec{x})}{||\vec{x} - \vec{x}_0||} = 0 \tag{.1.44}$$

**Definition .1.37.** *Let $f : D \subseteq \mathbb{R}^n \to \mathbb{R}^m$ be a multivariate function defined in a neighborhood of $\vec{x}_0 \in \mathbb{R}^n$ with first partial derivatives also defined. Then the* **Jacobian matrix** *of $f$ is defined to be*

$$Df(\vec{x}) = \begin{bmatrix} \partial_1 f_1(\vec{x}) & \partial_2 f_1(\vec{x}) & \dots & \partial_n f_1(\vec{x}) \\ \partial_1 f_2(\vec{x}) & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \partial_1 f_m(\vec{x}) & \dots & \dots & \partial_n f_m(\vec{x}) \end{bmatrix} \tag{.1.45}$$

*Then we say that $f$ is differentiable at $\vec{x}_0$ if*

$$\lim_{\vec{x}\to\vec{x}_0} \frac{||f(\vec{x}) - f(\vec{x}_0) - Df(\vec{x} - \vec{x}_0)||}{||\vec{x} - \vec{x}_0||} = 0 \tag{.1.46}$$

*If all the first partial derivatives of $f$ are continuous in a neighborhood of $\vec{x}_0$ then this holds.*

**Remark .1.17** (Properties)**.** For $f : D \subseteq \mathbb{R}^n \to \mathbb{R}^m$:

1. If $f$ is differentiable at $\vec{x}_0$, then $f$ is continuous at $\vec{x}_0$
2. If $f$ and $g$ are differentiable at $\vec{x}_0$, then $f \pm g, kf, fg$ $(m = 1)$ are differentiable at $\vec{x}_0$.
3. If the partials of $f$ are continuous in a neighborhood of $\vec{x}_0$, then $f$ is differentiable at $\vec{x}_0$. The converse is not true in general.

**Definition .1.38.** *Consider $f : D \subseteq \mathbb{R}^n \to \mathbb{R}$, we have*

$$\partial_i f(\vec{x}_0) = \lim_{t\to 0} \frac{f(\vec{x}_0 + t\hat{e}_i) - f(\vec{x}_0)}{t} \tag{.1.47}$$

*so we can generalize this to define the* **directional derivative** *in the direction of $\hat{u}$:*

$$\partial_{\hat{u}} f(\vec{x}_0) = \lim_{t\to 0} \frac{f(\vec{x}_0 + t\hat{u}) - f(\vec{x}_0)}{t} \tag{.1.48}$$

**Theorem .1.7.** *If $f : D \subseteq \mathbb{R}^n \to \mathbb{R}$ is differentiable at $\vec{x}_0$, then the directional derivative of $f$ at $\vec{x}_0$ exists in the direction of $\hat{u}$, and is equal to*

$$\partial_{\hat{u}} f(\vec{x}_0) = \nabla f(\vec{x}_0) \cdot \hat{u} \tag{.1.49}$$

**Theorem .1.8.** *Notice if $f : D \subseteq \mathbb{R}^n \to \mathbb{R}$ is differentiable at $\vec{x}_0$, and $\hat{u}$ is a unit vector, we have that*

$$\partial_{\hat{u}} f(\vec{x}_0) = \nabla f(\vec{x}_0) \cdot \hat{u} = |\nabla f(\vec{x}_0)| \cos(\theta) \tag{.1.50}$$

*with $\theta$ being the angle between $\nabla f(\vec{x}_0)$ and $\hat{u}$. As a result:*

1. *The largest value of $\partial_{\hat{u}} f(\vec{x}_0)$ is equal to $|\nabla f(\vec{x}_0)|$, and occurs when $\hat{u}$ is in the same direction as the gradient*
2. *The smallest value of $\partial_{\hat{u}} f(\vec{x}_0)$ is equal to $-|\nabla f(\vec{x}_0)$ when $\hat{u}$ is in the same direction as $-\nabla f(\vec{x}_0)$*
3. *When $\hat{u}$ is perpendicular to $\nabla f(\vec{x}_0)$, the directional derivative is zero.*

**Theorem .1.9** (Chain Rule V1)**.** *Let $f : D_f \subseteq \mathbb{R}^n \to \mathbb{R}$ and $g : D_g \subseteq \mathbb{R} \to \mathbb{R}$ such that $g(t) = f(\vec{x}(t))$. Then*

$$\frac{dg}{dt} = \nabla f(\vec{x}) \cdot \frac{d}{dt}\langle x_1, ..., x_n \rangle = \sum_{i=1}^{n} \partial_i f \frac{dx_i}{dt} \tag{.1.51}$$

**Theorem .1.10** (Chain Rule V2)**.** *Let $f : D_f \subseteq \mathbb{R}^n \to \mathbb{R}$ and $g : D_g \subseteq \mathbb{R}^m \to \mathbb{R}$ such that $g(\vec{x}) = f(y_1(\vec{x}), ..., y_n(\vec{x}))$. Then*

$$\partial_i g(\vec{x}) = \nabla f(y_1(\vec{x}), ..., y_n(\vec{x})) \cdot \frac{\partial}{\partial x_i}\vec{y} \tag{.1.52}$$

*for $\vec{y} = \langle y_1, ..., y_n \rangle$.*

**Theorem .1.11** (Clairout's Theorem)**.** *Suppose $f : \mathbb{R}^n \to \mathbb{R}$ has continous first and second partials on an open ball $B_r$. Then $f_{ij}(\vec{x}) = f_{ji}(\vec{x})$ for all $\vec{x} \in D$.*

## Implicit Differentiation

**Theorem .1.12** (Implicit Function Theorem (Two variables))**.** *Consider $F(x, y) = 0$. Let $(x_0, y_0) \in \mathbb{R}^2$ such that $F(x_0, y_0) = 0$, and suppose $F$'s first partials are continuous in a neighborhood of $(x_0, y_0)$. Then*

1. *If $F_y(x_0, y_0) \neq 0$, then $F(x, y) = 0$ uniquely defines $y$ as a continuously differentiable function of $x$ in a neighborhood of $x_0$, and we have that*

$$\frac{dy}{dx} = -\frac{\partial_x F(x, y)}{\partial_y F(x, y)} \tag{.1.53}$$

2. *Similarly for $F_x(x_0, y_0) \neq 0$.*

**Theorem .1.13** (Implicit Function Theorem (n variables))**.** *Consider $F(\vec{x}) = 0(\star)$, $\vec{x} = (x_1, ..., x_n)$. Let $\vec{a}$ satisfy $F(\vec{a}) = 0$, and suppose $F(\vec{x})$ has continuous first partial derivatives at and in a neighborhood of $\vec{a}$. Let $\beta$ be one of the variables $\{x_1, ..., x_n\}$, and let $\vec{\alpha}$ be the rest. If $\partial_\beta F(\vec{a}) \neq 0$, then the equation $(\star)$ uniquely defines the variable $\beta$ as a continuously differentiable function of $\vec{\alpha}$, and for $x_j \neq \beta$, we have*

$$\partial_{x_j} \beta(\vec{\alpha}) = -\frac{\partial_{x_j} F(\vec{a})}{\partial_\beta F(\vec{a})} \tag{.1.54}$$

## Theorem 32 (Implicit Function Theorem (General)).

*Consider a system of $n$ equations in $n+m$ variables*

$$\begin{cases} F_{(1)}(x_1, ..., x_m, y_1, ..., y_n) = 0 \\ \vdots \\ F_{(n)}(x_1, ..., x_m, y_1, ..., y_n) = 0 \end{cases} \tag{.1.55}$$

*and a point $P_0$ which satisfies the system. Suppose each $F_{(i)}$ is differentiable near $P_0$, so they have continuous first partial derivatives. Finally, suppose*

$$\left.\frac{\partial(F_{(1)}, ..., F_{(n)})}{\partial(y_1, ..., y_n)}\right|_{P_0} \neq 0 \tag{.1.56}$$

*Then the system defines $y_1, ..., y_n$ uniquely as continuously differentiable functions of $x_1, ..., x_m$ in some neighborhood of $P_0$. Moreover,*

$$\partial_{x_j} y_i = -\frac{\frac{\partial(F_{(1)},...,F_{(n)})}{\partial(y_1,...,x_j,...,y_n)}}{\frac{\partial(F_{(1)},...,F_{(n)})}{\partial(y_1,...,y_i,...,y_n)}} \tag{.1.57}$$

*This formula is a consequence of Cramer's Rule applied to the n linear equations in $n$ unknowns which is the system differentiated with respect to $x_j$.*

### Differentials

**Definition .1.39.** *Let $f : D \subseteq \mathbb{R}^n \to \mathbb{R}$ be a function defined at and around a point $\vec{a}$. Given $\Delta\vec{x}$, of small magnitude,*

$$\Delta f_{\vec{a}}(\Delta\vec{x}) = f(\vec{a} + \Delta\vec{x}) - f(\vec{a}) \tag{.1.58}$$

*represents the change in the value of the function associated with the change $\Delta\vec{x}$ in $\vec{x}$ at $\vec{a}$. Then, we approximate this change with the* **differential** *at $\vec{a}$ defined by*

$$df_{\vec{a}}(\Delta\vec{x}) = \nabla f(\vec{a}) \cdot \Delta\vec{x} \tag{.1.59}$$

*If $\Delta\vec{x}$ is sufficiently small, these changes are approximately equal.*

### Taylor Polynomials

## Theorem 33 (Taylor's Theorem (One Variable)).

*Let $f(x)$ be a function wit $n+1$ continuous derivatives in the open interval $(a, b)$. Let*

$$T_n(x) := \sum_{i=0}^{n} \frac{f^{(i)}(c)(x-c)^i}{i!} \tag{.1.60}$$

*be the* **degree n Taylor polynomial** *of $f(x)$ centered at $x = c \in (a, b)$. Then, for any $x \in (a, b)$, there exists a number $\theta$ between $c$ and $x$ such that*

$$f(x) = T_n(x) + \frac{f^{(n+1)}(\theta)}{(n+1)!}(x-c)^{n+1} \tag{.1.61}$$

**Definition .1.40 (Two Variable Taylor Polynomial).** *Let $f(x,y)$ be a smooth function (continuous partial derivatives up to whatever degree needed) in a open set $D \subset \mathbb{R}^2$. The* **degree $n$ Taylor polynomial** *of $f(x,y)$ at a point $(a,b) \in D$, is the polynomial $T_n(x,y)$ of degree $n$ that equals $f(x,y)$ and its first $n$ partial derivatives at $(a,b)$. It can be written as*

$$T_n(x,y) := \sum_{i=0}^{n} \frac{[(x-a)\partial_x + (y-b)\partial_y]^{(i)} f(a,b)}{i!} \tag{.1.62}$$

*where $[(x-a)\partial_x + (y-b)\partial_y]^{(i)}$ is to be expanded as an algebraic expression and the products of $\partial_x$ and $\partial_y$ correspond to composition of operators.*

## Theorem 34 (Taylor's Theorem (Two variables)).

*Let $f(x,y)$ be a function with continuous partial derivatives up to $(n+1)$ in some neighborhood $D$ of $(a,b) \in \mathbb{R}^2$. Let $T_n(x,y)$ be the degree $n$ Taylor polynomial of $f(x,y)$ at $(a,b)$. Then for any $(x,y) \in D$, there exists $(\alpha,\beta) \in D$ such that*

$$f(x,y) = T_n(x,y) + \frac{[(x-a)\partial_x + (y-b)\partial_y]^{(n+1)} f(\alpha,\beta)}{(n+1)!} \tag{.1.63}$$

*This is called the* **Taylor formula/expansion of order** *$n$ of $f$ at $(a,b)$.*

$$R_n(x,y) := \frac{[(x-a)\partial_x + (y-b)\partial_y]^{(n+1)} f(\alpha,\beta)}{(n+1)!} = f(x,y) - T_n(x,y) \tag{.1.64}$$

*is called the* **remainder** *of the expansion.*

**Remark .1.18.** If all partial derivatives of order $(n+1)$ are bounded by some constant $M > 0$, then

$$|f(x,y) - T_n(x,y)| \leqslant \frac{M}{(n+1)!} \left[ \sum_{j=0}^{n+1} \binom{n+1}{j} |x-a|^{n+1-j} |y-b|^{j} \right] \tag{.1.65}$$

## Theorem 35 (Taylor's Theorem (General)).

*Let $f : D \subset \mathbb{R}^n \to \mathbb{R}$ be a function with continuous partial derivatives of order up to $m+1$ in a neighborhood $D$ of $\vec{a} \in \mathbb{R}^n$. Then for all $\vec{x} \in D$ there exists $\vec{theta} \in D$ such that*

$$f(\vec{x}) = T_m(\vec{x}) + R_m(\vec{x}, \vec{\theta}) \tag{.1.66}$$

*where*

$$T_m(\vec{x}) := \sum_{k=0}^{m} \frac{[(\vec{x}-\vec{a}) \cdot \nabla]^{(k)} f(\vec{a})}{k!} \tag{.1.67}$$

*is the degree $m$ Taylor polynomial of $f$ at $\vec{a}$, and*

$$R_m(\vec{x}, \vec{\theta}) := \frac{[(\vec{x}-\vec{a}) \cdot \nabla]^{(m+1)} f(\vec{\theta})}{(m+1)!} \tag{.1.68}$$

*is the remainder. If all partial derivatives of $f$ are continuous and there exists $r \in \mathbb{R}^+$ such that whenever $||\vec{x} - \vec{a}|| < r$ we have for all $t \in [0,1]$*

$$\lim_{m \to \infty} R_m(\vec{x}, \vec{a} + t(\vec{x} - \vec{a})) = 0 \tag{.1.69}$$

*Then we can represent $f(\vec{x})$ as the* **Taylor series**

$$f(\vec{x}) = \sum_{n=0}^{\infty} \frac{\left[(\vec{x}-\vec{a})\cdot\nabla\right]^{(n)} f(\vec{a})}{n!} \tag{.1.70}$$

## Local Extrema of Multivariate Functions

**Definition .1.41.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function of $n$ variables defined in a neighborhood of $\vec{x}_0 \in \mathbb{R}^n$:*

1. *We say $f$ has a* **local maximum** *at $\vec{x}_0$ if there exists a neighborhood $D$ of $\vec{x}_0$ for which $f$ is defined and*
$$f(\vec{x}) \leqslant f(\vec{x}_0), \forall \vec{x} \in D \tag{.1.71}$$
2. *We say $f$ has a* **local minimum** *at $\vec{x}_0$ if there exists a neighborhood $D$ of $\vec{x}_0$ for which $f$ is defined and*
$$f(\vec{x}) \geqslant f(\vec{x}_0), \forall \vec{x} \in D \tag{.1.72}$$

↳ *If $f$ has a local maximum or minimum at $\vec{x}_0$, we say $f$ has a* **local extremum** *at $\vec{x}_0$.*

**Theorem .1.14** (Fermat). *If $f : \mathbb{R}^n \to \mathbb{R}$ has a local extremum at $\vec{x}_0$, then one of the following must hold:*

1. *$\nabla f(\vec{x}_0) = \vec{0}$ when all first partials of $f$ exist at $\vec{x}_0$*
2. *At least one of the first partials of $f$ are not defined at $\vec{x}_0$*

**Definition .1.42.** *Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is defined in a neighborhood of $\vec{x}_0$. If one of the conditions of Fermat's Theorem is satisfied by $\vec{x}_0$, we say $\vec{x}_0$ is a* **critical point** *of $f$.*

**Remark .1.19.** To determine if $f$ has a local max or min at a critical point $\vec{x}_0$, study the sign of

$$f(\vec{x}_0 + \vec{h}) - f(\vec{x}_0) \tag{.1.73}$$

for small $|\vec{h}|$. If it is always positive, $f$ has a local minimum, if it is always negative $f$ has a local maximum, and if it changes sign, $f$ does not have a local extremum and in this case we say $f$ has a ***saddle point*** at $\vec{x}_0$.

# Theorem 36 (Second Derivative Test).

*Suppose $f(x,y)$ has continuous second partial derivatives in a neighborhood of a critical point $(x_0,y_0)$. Define the* **Hessian** *matrix of $f$ at $(x_0,y_0)$ to be*

$$H_f(x_0,y_0) := \begin{bmatrix} f_{xx}(x_0,y_0) & f_{xy}(x_0,y_0) \\ f_{yx}(x_0,y_0) & f_{yy}(x_0,y_0) \end{bmatrix} \tag{.1.74}$$

*Let $\delta_1 = f_{xx}(x_0,y_0)$ and $\delta_2 = \det(H_f(x_0,y_0))$. Then*

1. *If $\delta_1 > 0$ and $\delta_2 > 0$, then $f$ has a local minimum at $(x_0, y_0)$*
2. *If $\delta_1 < 0$ and $\delta_2 > 0$, then $f$ has a local maximum at $(x_0, y_0)$*
3. *If $\delta_2 \neq 0$ but neither case 1 nor case 2 hold, then $f$ has a saddle point at $(x_0, y_0)$.*
4. *If $\delta_2 = 0$ the test is inconclusive.*

## Theorem 37 (Second Derivative Test (general)).

*Suppose $f : D \subset \mathbb{R}^n \to \mathbb{R}$ has continuous second partial derivatives in a neighborhood of a critical point $\vec{x}_0 \in D$. Define the* **Hessian** *matrix of $f$ at $\vec{x}_0$ to be*

$$H_f(\vec{x}_0) := \begin{bmatrix} f_{11}(\vec{x}_0) & f_{12}(\vec{x}_0) & \dots & f_{1n}(\vec{x}_0) \\ f_{21}(\vec{x}_0) & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ f_{n1}(\vec{x}_0) & \dots & \dots & f_{nn}(\vec{x}_0) \end{bmatrix} \tag{.1.75}$$

*Denote the the $k$th principal minor of $H_f(\vec{x}_0)$ by*

$$\delta_k := \begin{vmatrix} f_{11}(\vec{x}_0) & f_{12}(\vec{x}_0) & \dots & f_{1k}(\vec{x}_0) \\ f_{21}(\vec{x}_0) & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ f_{k1}(\vec{x}_0) & \dots & \dots & f_{kk}(\vec{x}_0) \end{vmatrix} \tag{.1.76}$$

*. Then*

1. *If for all $i \in \{1, 2, ..., n\}$, $\delta_i > 0$, then $f$ has a local minimum at $\vec{x}_0$*
2. *If for all $i \in \{1, 2, ..., n\}$, $\delta_{2i-1} < 0$ and $\delta_{2i} > 0$, then $f$ has a local maximum at $\vec{x}_0$*
3. *If $\delta_n = \det(H_f(\vec{x}_0)) \neq 0$ but neither case 1 nor case 2 hold, then $f$ has a saddle point at $\vec{x}_0$.*
4. *If $\delta_n = \det(H_f(\vec{x}_0)) = 0$ the test is inconclusive.*

### Vector Fields

**Definition .1.43.** *A* **vector field** *is a vector function $\vec{F} : D \subset \mathbb{R}^n \to \mathbb{R}^n$. In the case of three variables we write*

$$\vec{F}(x, y, z) = \langle P(x, y, z), Q(x, y, z), R(x, y, z) \rangle \tag{.1.77}$$

**Remark .1.20.** A vector field $\vec{F} : D \subset \mathbb{R}^n \to \mathbb{R}^n$ is said to be of class $C^k$ for $k \in \mathbb{Z}^+$ in $D$ if the first $k$ partial derivatives of the component functions of $\vec{F}$ are continuous in $D$.

**Definition .1.44 (Conservative Fields).** *A vector field $\vec{F} : D \subset \mathbb{R}^n \to \mathbb{R}^n$ is called* **conservative** *in a region $E \subseteq D$ if there exists a scalar function $f : D_f \subset \mathbb{R}^n \to \mathbb{R}$ such that*

$$\vec{F}(\vec{x}) = \nabla f(\vec{x}), \forall \vec{x} \in E \tag{.1.78}$$

*where $f$ is called a* **potential function** *of the vector field $\vec{F}$.*

**Definition .1.45.** *Let $\vec{F}: D \subset \mathbb{R}^n \to \mathbb{R}^n$ be a differentiable vector field. The* **divergence** *of $\vec{F}$ is the scalar field*

$$\nabla \cdot \vec{F} = \sum_{i=1}^{n} \partial_i F_i \tag{.1.79}$$

*where $F_i$ are the component function of $\vec{F}$.*

**Definition .1.46.** *Let $\vec{F}: D \subset \mathbb{R}^3 \to \mathbb{R}^3$ be a differentiable vector field. The* **curl** *of $\vec{F}$ is the vector field*

$$\nabla \times \vec{F} = \begin{vmatrix} \hat{i} & \hat{j} & \hat{k} \\ \partial_x & \partial_y & \partial_z \\ P & Q & R \end{vmatrix} \tag{.1.80}$$

*where $F_i$ are the component function of $\vec{F}$.*

**Proposition .1.15** (Properties of Divergence)**.** *If $\vec{F}: D \subset \mathbb{R}^n \to \mathbb{R}^n$ and $\vec{G}: D \subset \mathbb{R}^n \to \mathbb{R}^n$ vector fields and $f: D \subset \mathbb{R}^n \to \mathbb{R}$ is a scalar field, and $C_1, C_2 \in \mathbb{R}$, then*

1. *(Linearity)* $\nabla \cdot (C_1\vec{F} + C_2\vec{G}) = C_1\nabla \cdot \vec{F} + C_2\nabla \cdot \vec{F}$
2. *(Product rule)* $\nabla \cdot (f\vec{F}) = \nabla f \cdot \vec{F} + f\nabla \cdot \vec{F}$
3. *(Laplacian)* $\Delta f = \nabla \cdot \nabla f = \partial^2_{xx} f + \partial^2_{yy} f + \partial^2_{zz} f$

**Proposition .1.16** (Properties of Divergence)**.** *If $\vec{F}: D \subset \mathbb{R}^3 \to \mathbb{R}^3$ and $\vec{G}: D \subset \mathbb{R}^3 \to \mathbb{R}^3$ vector fields and $f: D \subset \mathbb{R}^3 \to \mathbb{R}$ is a scalar field, that are all defined and differentiable in $D$. Let $C_1, C_2 \in \mathbb{R}$, then*

1. *(Linearity)* $\nabla \times (C_1\vec{F} + C_2\vec{G}) = C_1\nabla \times \vec{F} + C_2\nabla \times \vec{F}$
2. *(Product rule)* $\nabla \times (f\vec{F}) = \nabla f \times \vec{F} + f(\nabla \times \vec{F})$
3. *(Conservative Property)* $\nabla \times (\nabla f) = 0$
4. $\nabla \times (\nabla \times \vec{F}) = \nabla(\nabla \cdot \vec{F}) - (\nabla \cdot \nabla)\vec{F}$ *(provided $\vec{F}$ has continuous second partial derivatives, and where $(\nabla \cdot \nabla)\vec{F} = (\Delta P, \Delta Q, \Delta R)$)*

**Definition .1.47.** *Let $E \subseteq \mathbb{R}^n$*

1. *We say that $E$ is* **path connected** *if for any two points $A$ and $B$ in $E$ if there exists a continuous function $f: [0, 1] \to E$ such that $f(0) = A$ and $f(1) = B$.*
2. *We say $E$ is* **simply connected** *if $E$ is path connected and any simple closed curve $C$ that completely lies in $E$ can be continuously deformed into a single point without leaving $E$.*

**Theorem .1.17.** *Let $\vec{F}$ be a class $C^1$ in an open region $E \subseteq \mathbb{R}^3$. If $\vec{F}$ is conservative in $E$, then $\nabla \times \vec{F} = \vec{0}$ at every point of $E$. Moreover, if $\nabla \times \vec{F} = \vec{0}$ at every point of $E$ and $E$ is simply connected, then $\vec{F}$ is conservative in $E$.*

## Line Integrals

**Definition .1.48.** *Let $C$ be a bounded continuous parametric curve in $\mathbb{R}^n$. Recall that $C$ is a* **smooth curve** *if it has a parameterization $\vec{r} : I \subset \mathbb{R} \to \mathbb{R}^n$ such that $\frac{d\vec{r}}{dt}$ is continuous and nonzero in $I$. We say $C$ is a* **smooth arc** *if it is a smooth curve with finite parameter interval $I = [a,b]$.*

**Definition .1.49.** *Given a smooth curve $C$ with parameterization $\vec{r} : [a,b] \to \mathbb{R}^n$ we have*

$$l_C = \int_C ds = \int_a^b \left|\frac{d\vec{r}}{dt}\right| dt \tag{.1.81}$$

*In general, for a scalar field $f : \mathbb{R}^n \to \mathbb{R}$, we define the* **line integral along** *$C$ to be*

$$\int_C f(\vec{x}) ds = \int_a^b f(\vec{r}(t)) \left|\frac{d\vec{r}}{dt}\right| dt \tag{.1.82}$$

*This definition is parameterization independent.*

**Definition .1.50.** *If $\vec{F}$ is a continuous vector field and $C$ is an oriented smooth curve, then the* **line integral of the tangential component of $\vec{F}$ along** *$C$ is*

$$\int_C \vec{F} \cdot d\vec{r} = \int_C \vec{F} \cdot \hat{T} ds \tag{.1.83}$$

**Definition .1.51.** *If $C$ is a closed curve we also call this line integral the* **circulation** *of $\vec{F}$ around $C$, and we denote it by*

$$\oint_C \vec{F} \cdot d\vec{r} \tag{.1.84}$$

**Remark .1.21.** A line integral over a piecewise smooth path is the sum of the line integrals over the individual smooth arcs

$$\int_{\bigcup_{i=1}^n C_i} ds = \sum_{i=1}^n \int_{C_i} ds \tag{.1.85}$$

## Line Integral Theorems

# Theorem 38 (Fundamental Theorem of Line Integrals).

*Let $C$ be a piecewise smooth parametric curve with initial point $A$ and terminal point $B$. If $f : D \subseteq \mathbb{R}^n \to \mathbb{R}$ is a scalar function with continuous first partial derivatives in an open region containing $C$, then*

$$\int_C \nabla f \cdot d\vec{r} = f(B) - f(A) \tag{.1.86}$$

**Corollary .1.18.** *If $C$ is a piecewise smooth closed curve contained in a region $D$ where the vector field $\vec{F} : D \subseteq \mathbb{R}^n \to \mathbb{R}^n$ is conservative, then*

$$\oint_C \vec{F} \cdot d\vec{r} \tag{.1.87}$$

**Definition .1.52.** *A vector field $\vec{F}$ is said to be* **path-independent** *in a region $\Omega$ if for every pair of points $A$ and $B$ in $\Omega$ and every pair of piecewise smooth curves $C_1$ and $C_2$ with initial point $A$ and terminal point $B$, we have*

$$\int_{C_1} \vec{F} \cdot d\vec{r} = \int_{C_2} \vec{F} \cdot d\vec{r} \tag{.1.88}$$

**Theorem .1.19.** *Let $D$ be an open connected domain in $\mathbb{R}^n$ and let $\vec{F}$ be a smooth vector field defined on $D$. Then the following properties are equivalent:*

1. *$\vec{F}$ is conservative in $D$*
2. *$\oint_C \vec{F} \cdot d\vec{r} = 0$ for every piecewise smooth closed curve $C \subset D$*
3. *$\vec{F}$ is path independent in $D$.*

## Theorem 39 (Green's Theorem).

*Let $R$ be a closed region in the $xy$-plane whose boundary $\partial R$ consists of a finite number of piecewise smooth simple closed curves that are positively oriented with respect to $R$. If $\vec{F}$ is a smooth vector field on $R$, then*

$$\oint_{\partial R} \vec{F} \cdot d\vec{r} = \int\int_R (\nabla \times \vec{F}) \cdot \hat{k} dA \tag{.1.89}$$

*In particular $\hat{k}$ is the unit normal field specifying the orientation of $R$, and $\partial R$ is oriented such that its principal normal field $\vec{N}$ points away from the region and*

$$\hat{N} = \hat{T} \times \hat{k} \tag{.1.90}$$

**Remark .1.22.** You don't need anything past this point yet.

## Theorem 40 (Plane Divergence Theorem).

*Let $R$ be a closed region in the $xy$-plane whose boundary $\partial R$ consists of a finite number of piecewise smooth simple closed curves. Let $\vec{N}$ denote the unit outward (from $R$) normal field on $C$. If $\vec{F}$ is a smooth vector field on $R$, then*

$$\oint_{\partial R} \vec{F} \cdot \hat{N} ds = \int\int_R \nabla \cdot \vec{F} dA \tag{.1.91}$$

## Theorem 41 (Stoke's Theorem).

*Let $S$ be a piecewise smooth oriented surface in 3-space having a unit normal field $\hat{N}$ and* *boundary C consisting of a finite number of piecewise smooth closed curves with orientation inherited from $\mathcal{S}$. If $\vec{F}$ is a smooth vector field defined on an open set containing $\mathcal{S}$, then*

$$\oint_C \vec{F} \cdot d\vec{r} = \int\int_S (\nabla \times \vec{F}) \cdot \hat{N} dS \tag{.1.92}$$

## Surface Integrals

**Definition .1.53.** *A* **parametric surface** *in 3-space is a continuous function $\vec{r} : R \subseteq \mathbb{R}^2 \to \mathbb{R}^3$ for some rectangle R*

$$R = \{(u, v) \in \mathbb{R}^2 : a \leqslant u \leqslant b, c \leqslant v \leqslant d\} \tag{.1.93}$$

*in the uv-plane having values in 3-space:*

$$\vec{r}(u, v) = \langle x(u, v), y(u, v), z(u, v)\rangle, (u, v) \in R \tag{.1.94}$$

**Remark .1.23.** If $\vec{r}$ is one-to-one the surface does not intersect itself. In this case $\vec{r}$ maps the boundary of $R$ onto a curve in 3-space called the ***boundary of the parametric surface***. A surface with no boundary is called a ***closed surface***.

**Definition .1.54.** *If a finite number of parametric surfaces are joined pairwise along their boundaries one obtains a* **composite surface***, or just a surface thinking geometrically.*

**Definition .1.55.** *A set $S \subseteq \mathbb{R}^3$ is a* **smooth surface** *if any point $P \in S$ has a neighborhood N that is the domain of a smooth function $g : N \to \mathbb{R}$ satisfying*

1. *$N \cap S = \{Q \in N : g(Q) = 0\}$*
2. *$\nabla g(Q) \neq \vec{0}$ if $Q \in N \cap S$*

**Remark .1.24.** This means the surface has a unique tangent plane at any non-boundary point $P$.

**Definition .1.56.** *If $\vec{r} : D \subseteq \mathbb{R}^2 \to \mathbb{R}^3$ is a parameterization of a smooth surface S, the* **normal vector** *to S at $\vec{r}(u, v)$ is*

$$\vec{n} = \frac{\partial \vec{r}}{\partial u} \times \frac{\partial \vec{r}}{\partial v} \tag{.1.95}$$

**Definition .1.57.** *The* **area element** *at $\vec{r}(u, v)$ on S is given by*

$$dS = \left|\frac{\partial \vec{r}}{\partial u} \times \frac{\partial \vec{r}}{\partial v}\right| dudv \tag{.1.96}$$

*Then if $f(\vec{r})$ is continuous on $\mathcal{S}$ and the domain of $\vec{r}$ is D in the uv-plane*

$$\int\int_S f dS = \int\int_D f(\vec{r}(u, v)) \left|\frac{\partial \vec{r}}{\partial u} \times \frac{\partial \vec{r}}{\partial v}\right| dudv \tag{.1.97}$$

**Definition .1.58.** *A smooth surface $\mathcal{S}$ in 3-space is said to be* **orientable** *if there exists a unit vector field $\hat{N}$ defined on $\mathcal{S}$ that varies* **continuously** *over $\mathcal{S}$, and is everywhere normal to $\mathcal{S}$. Any such vector field $\hat{N}$ induces an orientation on $\mathcal{S}$. The side of $\mathcal{S}$ out of which $\hat{N}$ points is the* **positive side***, and the other side is the* **negative side***. An* **oriented surface** *is a smooth surface with a particular choice of orienting unit normal vector field $\hat{N}$.*

**Remark .1.25.** An oriented surface $\mathcal{S}$ ***induces an orientation*** on any of its boundary curves $C$; if we stand on the positive side of the surface $\mathcal{S}$ and walk around $C$ in the direction of its orientation, then $\mathcal{S}$ will be on our left side.

**Definition .1.59.** *Given any continuous vector field $\vec{F}$, the* **flux** *of $\vec{F}$ across the orientable surface $\mathcal{S}$ is the surface integral of the normal component of $\vec{F}$ over $\mathcal{S}$*

$$\int\int_S \vec{F}\cdot\hat{N}dS = \int\int_S \vec{F}\cdot d\vec{S} \tag{.1.98}$$

*and when the surface is closed we write*

$$\oiint_S \vec{F}\cdot\hat{N}dS = \oiint_S \vec{F}\cdot d\vec{S} \tag{.1.99}$$

**Remark .1.26.** If $\vec{r}(u,v)$ parametrizes $\mathcal{S}$ with domain $D$, we have normal

$$\vec{n} = \frac{\partial\vec{r}}{\partial u}\times\frac{\partial\vec{r}}{\partial v} \tag{.1.100}$$

and $dS = |\vec{n}|dudv$. Hence

$$d\vec{S} = \hat{N}dS = \pm\frac{\vec{n}}{|\vec{n}|}|\vec{n}|dudv = \pm\vec{n}dudv \tag{.1.101}$$

where the sign reflects the orientation of the surface and parameterization.

## Theorem 42 (Divergence Theorem).

*Let $D$ be a three dimensional domain bounded by piecewise smooth closed surfaces. Suppose its boundary $\mathcal{S}$ is an oriented closed surface with unit normal field $\hat{N}$ pointing out of $D$. If $\vec{F}$ is a smooth vector field defined on $D$, then*

$$\oiint_S \vec{F}\cdot\hat{N}dS = \int\int\int_D \nabla\cdot\vec{F}dV \tag{.1.102}$$