The Book

Labix

January 15, 2024

Abstract

Contents

Part I

The Foundations of Mathematics

Chapter 1

Set Theory

1.1 The first 3 Axioms

1.1.1 Introduction

Sets are widely used in a variety of topics in Mathematics. It shows up naturally whenever we are trying to ”group up” a bunch of related stuff. Examples would be perhaps the set of all natural numbers, the set of all functions with a fixed point 1, or even the set of all matrices such that its determinant, if defined, is 1.

However, the concept of sets was only introduced in the 1900’s. This was because long ago, there was simply no need for vigorous mathematics. There was no need to ”prove” a theorem vigorously with logic and axioms until Gödel came along and seek to unify separate branches of mathematics. Nowadays, set theory has become a universal language that every Mathematician knows by heart.

We start with the concept of belonging. There is no good definition of a set except by allowing elements to belong to a set. This is how we shall characterize sets, as well as identify sets.

Definition 1.1.1.1: The Concept of Belonging

If x belongs to a set A we write $x \in A$.

So we have the concept of belonging. However, we don’t even know if sets even exists in our world! This is how obnoxious mathematicians are. If not given a proof, they would not rest on the notion that a set could even exists. Unfortunately, there is no real tool for us to even prove that a set, any set exists at all.

Axiomatic set theory sort of solves this problem by stating some universally agreed assumptions. These are statements that cannot be proved and is a compromise for all mathematicians that they will work on theorems and definitions under these assumptions.

Currently, the only way we can construct sets are by simply listing elements that belong to that set. For example, if I say that $1, 2, 3$ belongs to the set S and nothing more, then we can write S as

$$S = \{1, 2, 3\}$$

We use the open and close brackets at the begining and the end to indicate the start and the end of the contents of the set.

1.1.2 Axiom of Extensionality

Whenever presented an axiom, the reader should be prompt automatically to think about the reason for introducing this axiom. Could it be formulated in another way? Could it not be an axiom?

Axiom 1.1.2.1: Axiom of Extensionality

Two sets A, B are equal if and only if they have the same elements. We write $A = B$ in this case.

Later as we shall see, sets can be constructed using predicates or statements. The axiom of extensionality allows to different predicates to result in the same set, provided that they have the same elements. Moreover, the number of copies of the same element does not matter, as long as there is at least one. Therefore, we can shorten long sets as follows:

$$\{1, 2, 3, 1, 2, 3\} = \{1, 2, 3\}$$

Finally, order in set also does not matter because once again, the requirement that two sets are equal are simply that they share the same elements. Therefore $\{\text{Tom}, \text{Mary}\}$ and $\{\text{Mary}, \text{Tom}\}$ are also equal.

We then give the definition of subsets.

Definition 1.1.2.2: Subsets

Let A, B be sets. A is a subset of B means that every elements in A is contained in B. We denote it by

$$A \subseteq B$$

if A is possibly equal to B and

$$A \subset B$$

if it is a proper subset, meaning A cannot be B.

We then give a crucial theorem that we will use throughout the entirety of the book.

Theorem 1.1.2.3

Let A, B be sets. Then

$$A = B \iff A \subseteq B \text{ and } B \subseteq A$$

Proof. Suppose that $A = B$. Then $x \in A \implies x \in B$ and $x \in B \implies x \in A$. Suppose that $A \subset B$ and $B \subset A$. Then $x \in A \implies x \in B$ and $x \in B \implies x \in A$. □

This theorem is particularly useful in proving sets are equal. Often we will use this theorem in its reverse direction. It is a characterization of equal sets. Whenever we are given to prove two sets are equal, the reader should immediately be able to refer to this theorem.

Readers who are already equipped with further concepts such as relations will realize that $\subseteq$ is a relation on sets. In fact, it is reflexive and transitive, as seen in the following theorem.

Theorem 1.1.2.4

Let, A, B, C be sets.

- $A \subseteq A$
- $A \subseteq B$ and $B \subseteq C \implies A \subseteq C$

Proof.

- $x \in A \implies x \in A$
- $x \in A \implies x \in B \implies x \in C$

□

1.1.3 Axiom of Regularity

The axiom of regularity, similar to the axiom schema of specification in the next chapter is meant to prevent paradoxes rather than construct new sets.

Axiom 1.1.3.1: Axiom of Regularity

Every non-empty set x contains a member y such that x and y are disjoint sets.

With this axiom, expressions such as $S \in S$ does not make sense anymore. Which is good, because it prevents self referencing as a potential paradox.

1.1.4 Axiom Schema of Specification

Undoubtedly one of the weirdest axioms to get a hold off is the axiom of specification. Essentially, the axiom means that we cannot create sets that are too big. Readers are free to look up more related information on this axiom, especially with regards to Russell's Paradox.

Axiom 1.1.4.1: Axiom Schema of Specification

To every set A and every condition set S(x) there corresponds a set B whose elements are exactly those elements of x of A for which S(x) holds. We write $B = \{x \in A | S(x)\}$

This axiom basically prevents us from creating arbitrarily large sets from thin air. Try and compare the statements $\{x \in A|S(x)\}$ and $\{x|S(x)\}$. For the first one we are essentially pulling things out from a set A, so inherently the new set is "smaller" than A, while the latter we are basically pulling things out of thin air. Here we assumed that there is some universal set that contains every single thing in the universe. And the latter statement is basically a condition towards this universal set. This universal set creates a lot of problems for us. By introducing the Axiom Schema of Specification, it allows us to route our way around Russell's Paradox:

If we define

$$R = \{x | x \notin x\}$$

which means R is the set of all elements that does not contain itself, then we reach a contradiction. Notably,

$$R \in R \iff R \notin R$$

We prove this as usual, assuming one side then reaching the other.

Assume that $R \in R$. Then since R contains itself, by definition of R, R should only contains elements that does not contain itself thus $R \notin R$. Now assume that $R \notin R$, then by definition of R, $R \in R$.

The reason that the Axiom Schema of Specification avoids this problem is that the predicate (condition) that we specify must be applied on a set (that is not the universal set). In other words, the underlying set for the predicate cannot be too vague since the new set created from the axiom of specification is necessarily a subset of the underlying set.

With this axiom in place, we can finally investigate our first concrete set!

Definition 1.1.4.2: Empty Set

Let A be a set. The set which contains no elements is the empty set, denoted by

$$\emptyset = \{x \in A : x \neq x\}$$

The axiom of extensionality guarantees that this empty set is unique. Although we do not have an all-encompassing universal set (the larger set possible), we still have the smallest set possible, characterized by the following theorem.

Theorem 1.1.4.3

$\emptyset \subseteq A$ for every set A.

Proof. Every element in $\emptyset$ is also in A. □

Its quite hard to wrap your head around this proof. Using the definition of a subset, we want to show that $x \in \emptyset$ implies $x \in A$. Do you think this is true for all $x \in \emptyset$? Well since there are no elements in the emptyset, this would be true! Can you see why?

1.2 Unions and Intersections

1.2.1 Axiom of Unions

Axiom 1.2.1.1: Axiom of Unions

For every collection of sets A there exists a set that contains all the elements that belong to at least one set of the given collection. In other words, there exists a set B such that

$$B = \{x \in a | a \in A\}$$

One must wonder: why go the long away round and say that a unions are elements of a set living in a set of sets? Essentially this allows us to proceed with the Axiom of Pairing. The axiom of pairing allows new sets to be created from old sets. And in fact, the Axiom of Unions is simply recovering the elements hidden from the sets given by the axiom of pairing.

Definition 1.2.1.2: Union

Let A, B be sets. Define the union of A and B to be

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

At this point the reader should not be confused between operations and statements. I assume the readers are crystal clear about it but I would elaborate on it as a reminder. Unions and the upcoming intersections are operations on sets that produce new sets. They are by no means able to be evaluated to be true or false. This property is preserved exclusively for statements and statements alone. And statements are given by symbols such as $=$ and $\subseteq$. If you think about it, one should be able to judge whether $A \subseteq B$ is true or false, depending on the contents of A and B while it does not make sense to judge the validity of $A \cup B$.

Proposition 1.2.1.3

Let A, B, C be sets.

- Identity: $A \cup \emptyset = A$
- Commutativity: $A \cup B = B \cup A$
- Associativity: $(A \cup B) \cup C = A \cup (B \cup C)$
- Idempotent: $A \cup A = A$
- $A \subset B \iff A \cup B = B$

Proof. We prove it in order.

- $A \cup \emptyset = \{x : x \in A \text{ or } x \in \emptyset\} = \{x : x \in A\} = A$
- $A \cup B = \{x : x \in A \text{ or } x \in B\} = \{x : x \in B \text{ or } x \in A\} = B \cup A$
- Proved similarly by expanding the definition of union and using the fact that the logic operator ”or” is commutative.
- $A \cup A = \{x : x \in A \text{ or } x \in A\} = \{x : x \in A\} = A$
- Suppose that $A \subset B$. $x \in A \cup B \implies x \in A$ or $x \in B \implies x \in B$ or $x \in B \implies x \in B$. Thus $A \cup B \subset B$. $x \in B \implies x \in B$ or $x \in B \implies x \in A$ or $x \in B \implies x \in A \cup B$. Thus $B \subset A \cup B$. Now suppose that $A \cup B = B$. $x \in A \implies x \in B$ thus $A \subset B$.

□

There is a bunch of fancy names in front of the properties of the operations. They are simply jargons for the properties of any operations in general. Do not be frightened by it since I will mostly like not recall properties from their fancy names unless I want to shortent the length of a proof.

We are shown an operation which is some what equivalent to the ”or” operation in logic theory. We shall present a similar notion for ”and” as well.

Definition 1.2.1.4: Intersection

Let A, B be sets. Define the intersection of A and B to be

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

The intersection is a another way to produce new sets from old, except that smaller sets are created, instead of larger sets. However this does not mean that the number of sets it procures is limited.

Proposition 1.2.1.5

Let A, B, C be sets.

- Identity: $A \cap \emptyset = \emptyset$
- Commutativity: $A \cap B = B \cap A$
- Associativity: $(A \cap B) \cap C = A \cap (B \cap C)$
- Idempotent: $A \cap A = A$
- $A \subseteq B \iff A \cap B = A$

Proof. The first four are proved in similarity to the counterpart with unions. We prove the last item. Suppose that $A \subset B$. $x \in A \cap B \implies x \in A$ and $x \in B$. Thus $x \in A$ and $A \cap B \subset A$. $x \in A \implies x \in A$ and $x \in A \implies x \in A$ and $x \in B$. Thus $x \in A \cap B$ and $A \subset A \cap B$. Now suppose that $A \cap B = A$. $x \in A \implies x \in A \cap B \implies x \in B$. Thus $A \subset B$. □

The following definition is a condition that can be satisfied by two sets.

Definition 1.2.1.6: Disjoint

Let A, B be sets. A and B are disjoint if and only if $A \cap B = \emptyset$.

Sometimes it is more convenient to simply say two sets are disjoint rather than stating that their intersection is empty.

Finally we have the distributive law that links the two operators between sets.

Theorem 1.2.1.7: Distributive Law

Let A, B, C be sets.

- $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
- $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

Proof. The two are proved by supposing x in the left and the right and proving that the left and the right are subsets of each other by logic. □

1.2.2 Complements

Another important concept in set theory is complements of sets. However one shall see that there are exactly two notions of complements. One that involves complements from the universe and another complement from a relative set.

Definition 1.2.2.1: Relative Complements

Let A, B be sets. Define

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

to be the relative complement of B in A.

This complement depends on what the larger set is, therefore suggesting the name relative.

Definition 1.2.2.2: Absolute Complements

Let E be the set that contains all elements under study. Define the absolute complement of A in E to be

$$A^C = E \setminus A$$

From Russell's Paradox we already know that sets too large and vague is unacceptable, so we usually define what the universal set is. For example, if the universal set is natural numbers, them the absolute coplements of the even numbers is the odd numbers including zero (We have not formally talked about number systems, which we will in formal set theory courses). If the universal set is the real numbers then clearly the absolute complement is not only the odd numbers.

Notice that the concepts of relative complement and absolute complement coincide when $B \subseteq A$ in the definition of relative complements. The reason for two different notations is because relative complements does not necessarily require that $B \subseteq A$. It simply rules out elements of B the coincide with A, while in the case of absolute complements, the universal set is always a superset of A.

If we assume that there is a universal set that is a superset of both A and B, then we have the relation

$$B \cap A^C = B \setminus A$$

This can be proven once again using first order logic:

Proposition 1.2.2.3

Let A, B be sets. Then

$$A \setminus B = A \cap B^C$$

Proof. Let $x \in A \setminus B$. Then $x \in A$ and $x \notin B$. But $x \notin B$ implies $x \in B^C$. Thus $x \in A$ and $x \in B^C$ implies $x \in A \cap B^C$. This proves that $A \setminus B \subseteq A \cap B^C$.

The entire argument is reversible where implications are double sided. Thus also we have $A \cap B^C \subseteq A \setminus B$ and thus $A \setminus B = A \cap B^C$. □

Proposition 1.2.2.4

Let A, B be sets and subsets of E. The following four with respect to the absolute complement.

- $(A^C)^C = A$
- $\emptyset^C = E$ and $E^C = \emptyset$
- $A \cap A^C = \emptyset$ and $A \cup A^C = E$
- $A \subset B \iff B^C \subset A^C$

Proof. The four are proved by expanding on the definition of complement and proved by logic. □

Similar to the distributive law, we can associate the complement operator with the previous two operators, namely union and intersection.

Theorem 1.2.2.5: De Morgans Laws

Let A, B be sets,

- $(A \cup B)^C = A^C \cap B^C$
- $(A \cap B)^C = A^C \cup B^C$

Proof. Both are proved by expansion of the set language into logic. □

The following are a list of operations that are true with respect to the relative complements.

Proposition 1.2.2.6

Let A, B, C be sets.

- $A \subset B$ if and only if $A \setminus B = \emptyset$
- $A \setminus (A \setminus B) = A \cap B$
- $A \cap (B \setminus C) = (A \cap B) \setminus (A \cap C)$
- $A \cap B \subset (A \cap C) \cup (B \cap C^C)$
- $(A \cup C) \cap (B \cup C^C) \subset A \cup B$

Proof. Exercise. □

These obscure expressions are often explained clearly by drawing venn diagrams, which we will not discuss here because I do not know how to draw.

Finally, the rarely used symmetric difference will be defined here although there is no real usage for now.

Definition 1.2.2.7: Symmetric Difference

Let A, B be sets. Define the symmetric difference of A and B to be

$$A + B = (A \setminus B) \cup (B \setminus A)$$

1.2.3 Axiom of Powers

Finally, the axiom of power is simple: to assert the existence of power sets of a set.

Axiom 1.2.3.1: Axiom of Power Set

For each set there exists a collection of sets that contains among its elements all the subsets of the given set. Define that collection to be $\mathcal{P}(A)$, where A is any set.

The following is a proposition regarding integrating the operation of taking power sets and unions and intersections. It serves as a good exercise.

Proposition 1.2.3.2

Let E be a collection of sets. Then the following are true.

- $\bigcap_{X\in E}\mathcal{P}(X)=\mathcal{P}(\bigcap_{X\in E}X)$
- $\bigcup_{X\in E}\mathcal{P}(X)\subseteq\mathcal{P}(\bigcup_{X\in E}X)$

Proof. Let $A\in\bigcap_{X\in E}\mathcal{P}(X)$. Then $A\subseteq X$ for all $X\in E$. Thus $A\subseteq\bigcap_{X\in E}X$ and $A\in\mathcal{P}(\bigcap_{X\in E}X)$. For the reverse inclusion, the above implications are all double sided thus we are done.

Let $A\in\bigcup_{X\in E}\mathcal{P}(X)$. Then $A\subseteq\mathcal{P}(X)$ for some $X\in E$. Thus $A\subseteq\bigcup_{X\in E}X$ and $A\in\bigcup_{X\in E}\mathcal{P}(X)$ and we are done. Note that the reverse inclusion does not hold since we only have $A\subseteq\mathcal{P}(X)$ for some $X\in E$ as compared to the intersection of power sets. □

1.3 Functions and Relations

1.3.1 Cartesian Products

Ordered pairs are used to define cartesian products, which in turn gives a formal definition to functions.

Definition 1.3.1.1: Ordered Pairs

Define the ordered pair of a and b to be

$$(a,b)=\{\{a\},\{a,b\}\}$$

Note the cleverness here, the ordered pair is defined as a set of two elements, one of which is a singleton which defines which element in the ordere pair comes first. This allows order to be defined in a set which are a collection of unordered elements. Can you think of a way to extend this notion into triple, quadruples and even n-tuples?

Let us see an immediate consequence of this definition which proves that this notion indeed works out nicely.

Proposition 1.3.1.2

Let (a,b) and (c,d) be ordered pairs. Then $(a,b)=(c,d)$ if and only if $a=c$ and $b=d$.

Proof. Suppose that $(a,b)=(c,d)$. Then we have by definition,

$$\{\{a\},\{a,b\}\}=\{\{c\},\{c,d\}\}$$

There are two cases: $a=b$ and $a\neq b$. Suppose that $a=b$, then $\{\{a\}\}=\{\{c\},\{c,d\}\}$. This forces $\{a\}=\{c\}=\{c,d\}$ and $a=c=d$. Suppose that $a\neq b$. Then $\{a\}=\{c\}$ and $\{a,b\}=\{c,d\}$. Thus $a=c$ and $b=d$. b cannot be c here since $a=b=c$ is a contradiction.

Now suppose that $a=c$ and $b=d$, then $\{a\}=\{c\}$ and $\{a,b\}=\{c,d\}$ thus $\{\{a\},\{a,b\}\}=\{\{c\},\{c,d\}\}$. □

We now define the cartesian product of two sets, whose elements are ordered pairs.

Definition 1.3.1.3: Cartesian Product

Let A,B be sets. Define the Cartesian Product of A and B to be

$$A\times B=\{(a,b)|a\in A,b\in B\}$$

Clearly by definition a cartesian product gives rise to a set of ordered pairs. But the converse is also true, where a set of ordered pairs can give rise to a cartesian product.

Proposition 1.3.1.4

Suppose that R is a set of ordered pairs. Then there exists A,B such that $R\subseteq A\times B$.

Proof. Simply define $A=\{a|(a,b)\in R\}$ and $B=\{b|(a,b)\in R\}$. These sets in fact have names as we will see in the next definition. □

Readers should make sure that the predicate $(a,b)\in R$ is a valid construct in the language of set theory, just as a sanity check.

As promised, we give names to the sets defined in the proof above.

Definition 1.3.1.5: The First and Second Projection

Let R be a set of ordered pairs. Define the first and second projections of R to be

$$A = \{a|(a,b) \in R\}$$

and

$$B = \{b|(a,b) \in R\}$$

Unfortunately these definition will be forgotten and rarely be used again.

We end the section with properties of the cartesian product when used in conjunction with other set operators.

Proposition 1.3.1.6

Let A, B, X, Y be sets.

- $(A \cup B) \times X = (A \times X) \cup (B \times X)$
- $(A \cap B) \times (X \cap Y) = (A \times X) \cap (B \times Y)$
- $(A \setminus B) \times X = (A \times X) \setminus (B \times X)$

Proof. Again just a practise of expansion into logic. □

1.3.2 Relations

Relations appear naturally in the course of studying mathematics. Some simple relations that are used commonly include $=$, $\leq$ and $\geq$ and $<, >$. In fact, $\subset$ and $\subseteq$ are also relations.

More generally, relations are used to associate two elements of a set. This could be done because that they share similar properties or used for comparison.

Definition 1.3.2.1: Relation

Define a relation R to be a set of ordered pairs. If $(a,b) \in R$, we write aRb instead.

The definition of relations with ordered pairs seems unnatural especially when considering the usual relations like $\leq$ and $\geq$. In practise we will rarely think about relations set theoreticly. This is only meant for formalization into the language of set theory, which in turn is the basis of modern mathematics.

Definition 1.3.2.2: Domain and Range

The first projection of a relation R is called $\text{dom}(R)$ and the second projection is called $\text{ran}(R)$. In other words,

$$\text{dom}(R) = \{a|(a,b) \in R\} \text{ and } \text{ran}(R) = \{b|(a,b) \in R\}$$

Bear with the notation for now. This is a natural extension of names from definitions of functions since functions are defined with relations.

We now classify relations based on some properties it exhibits. Note that this is not all the properties of relations but instead only the first few important ones.

Definition 1.3.2.3: Classification of Relations

Let R be a relation. We say that

- R is reflexive if $(x,x) \in R$ for all $x \in R$
- R is symmetric if $(x,y) \in R \implies (y,x) \in R$ for all $x,y \in R$
- R is transitive if $(x,y) \in R$ and $(y,z) \in R$ implies $(x,z) \in R$ for all $x,y,z \in R$
- R is an equivalence relation if it is reflexive, symmetric and transitive

Readers can check that while $=$ is an equivalence relation, $<, >$ only satisfies transitivity while $\leq, \geq$ satisfies reflexivity in addition to transitivity.

Finally, we have inverse relations which proves itself useful when dealing with inverse functions.

Definition 1.3.2.4: Inverse Relations

Let R be a relation from A to B. Define the inverse relation to be

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

Notice that here, every relation is guaranteed to have an inverse. But for function, additional requirements have to be satisfied in order for functions to be reversed.

The remainder of the section is mainly for exhibiting the relation between partitions and equivalence relations. This is often useful in the sense that whenever an equivalence relation appears, a partition will be possible from the set.

Firstly we define partitions.

Definition 1.3.2.5: Partition

Let X be a set. A partition of X is a disjoint collection E of subsets of X such that

- $A, B \in E$ and $A \neq B$ implies $A \cap B = \emptyset$
- $\bigcup_{A \in E} A = X$

Although we have yet to define the order of a set (number of elements in a set), it is important to note that partitions does not mean that the number of elements in each partition is the same.

The following notion will only be used in the main theorem of the section for convenience. Technically it is also used in number theory and group theory but they will not appear here.

Definition 1.3.2.6: Equivalence Class

Let R be an equivalence relation on a set X. Denote $[x] = \{y \in X|(x,y) \in R\}$ the equivalence class of $x \in X$ and $X/R = \{[x]|x \in X\}$ the set of all equivalence classes.

Now comes the main theorem, it is split into two parts for readability. The proof is not particularly long but may take some time to digest and understand.

Theorem 1.3.2.7

An equivalence relation R on a set X induces a partition on X.

Proof. For every $x \in X$, $x \in x/R$ thus $\bigcup_{x \in X} x/R = X$. Now suppose that $z \in x/R \cap y/R$, then $(x,z) \in R$ and $(y,z) \in R$. By the symmetric property $(z,y) \in R$. By transitivity $(x,y) \in R$.

Thus $x/R = y/R$. This proves that X/R is a partition. □

Theorem 1.3.2.8

A partition on X induces an equivalence relation R on X.

Proof. Suppose that X is partitioned. Define a relation R to be $(x, y) \in R$ if and only if x, y are in the same partition.

We now prove R is an equivalence relation. We have that $(x, x) \in R$ for every x since they necessarily appear in the same set and none others (by defition of partition) thus the reflexive property holds. If $(x, y) \in R$ then x, y are in the same partition thus naturally $(y, x) \in R$ holds thus the symmetric property holds. Finally if $(x, y) \in R$ and $(y, z) \in R$ then x, y, z are all in the same parition thus $(x, z) \in R$ holds which proves transitivity. We can now conclude that R is an equivalence relation and we are done. □

1.3.3 Functions

Functions play an integral role in all of mathematics. Therefore it is important to be able to express this concept in terms of set theoretic language.

Definition 1.3.3.1: Functions

Let X, Y be sets. A function from X to Y is a relation $f \subseteq X \times Y$ such that

- $\text{dom}(f) = X$
- $\exists y \in Y$ such that $(x, y) \in f$ for all $x \in X$ (existence of an output)
- $(x, y) \in f$ and $(x, z) \in f$ implies $y = z$ (uniqueness of an output)

In this case, we say that X is the domain of f and Y is the codomain of f. We often write f as $f : X \to Y$ to indicate the domain and codomain of f.

As one can see, functions are defined based on relations which is why the study of relations is also important. The additional rules the a relation has to satisfy in order to be a function is simply that all of X must be associated to exactly one thing in y, not zero, not two, meaning everything in the domain must be linked to something in Y.

An immediate consequence is that since relations are subsets of the cartesian product, all functions from X to Y must be encapsulated by the cartesian product in some way.

Proposition 1.3.3.2

The set of all functions from X to Y is a subset of $Y^X = \mathbb{P}(X \times Y)$.

Proof. Note that every function is a relation thus this induces a subset relation betweeen functions on a set and relations on a set. Then since we have shown that $R \subseteq X \times Y$, any function must be an element of $\mathbb{P}(X \times Y)$ and thus the set of all functions is a subset of $\mathbb{P}(X \times Y)$. □

Below we give three crucial functions which arises in most of the areas of mathematics.

Definition 1.3.3.3: Inclusion Map

Let $X \subset Y$ and $f : X \to Y$ where f is defined as $f(x) = x$. f is called the inclusion map of X into Y.

Notice the strict inclusion on $X \subset Y$. The inclusion map is meant to encorporate and identify X inside of Y. This is most often used when we want to extend the domain of a function. When the strict inclusion is relaxed and we have $X = Y$, it is called the identity map.

Definition 1.3.3.4: Identity Map

The inclusion map from X to X is called the identity map on X.

The name identity map will become clear once we reach inverse functions.

Definition 1.3.3.5: Restriction Map

Let $f : Y \to Z$ and $X \subset Y$. The restriction map of f is the function $g : X \to Z$ such that $g(x) = f(x)$ for all $x \in X$. Conversely, the extension map of $g \to Y$ is f. We write $g = f|X$.

Restriction maps are simply copies of the original map that is restricted to a certain subset of the original domain.

Finally, there are three important properties that a function can take.

Definition 1.3.3.6: Bijective Functions

Let $f : X \to Y$ be a function.

- f is injective if $f(x_1) = f(x_2) \implies x_1 = x_2$
- f is surjective if for all $y \in Y$ there exists $x \in X$ such that $f(x) = y$
- f is bijective if it is both injective and surjective

This properties of a function does not only depend on how the function/relation is defined, but the domain and codomain also plays a part.

1.3.4 Compositions

Functions are allowed to be composed. By doing this we are essentially creating a new function.

Definition 1.3.4.1: Composition of Functions

Let $f : A \to B$ and $g : B \to C$ be two functions. Define the composition of f and g to be

$$g \circ f : A \to C$$

If $a \in A$ then

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

Readers should verify that the new object $g \circ f$ is a function.

Lemma 1.3.4.2

Composition of functions result in a new function.

Proof. Easy exercise. □

While composition of functions is in general not commutative, it is fortunately associative.

Proposition 1.3.4.3: Associativity of Functions

Let $f : A \to B$, $g : B \to C$ and $h : C \to D$ be functions. Then the following is true.

$$(h \circ g) \circ f = h \circ (g \circ f)$$

Proof. Simple manipulation of definition of composition. □

In general, composition preserves injectivity and surjectivity, as seen by the folllowing proposition.

Proposition 1.3.4.4

Let $f : A \to B$ and $g : B \to C$ be functions.

- If f and g are injective then $g \circ f$ is injective
- If f and g are surjective then $g \circ f$ is surjective
- If f and g are bijective then $g \circ f$ is bijective

Proof. Easy exercise involving the use of definition of injectivity and surjectivity. □

1.3.5 Inverses

Inverse functions serve the important role of inverting functions so that we know what element in the domain is mapped to a fixed element in the codomain. However the condition that the inverse exists must be studied.

Definition 1.3.5.1: Inverse Functions

Let $f : X \to Y$ be a function. If the inverse relation $f^{-1} : Y \to X$ is also a function, then we say that f^{-1} is the inverse function of f.

The main criterion for inversability of a function is given by the below characterization.

Theorem 1.3.5.2

Let $f : X \to Y$ be a function. The inverse relation f^{-1} is a function from Y to X if and only if f is bijective.

Proof. We first suppose that the inverse of f is a function. We aim to prove injectivity and surjectivity.
Injectivity: Suppose that $f(x_1) = f(x_2) = y$. In terms of relations, this means that $(x_1, y) \in f$ and $(x_2, y) \in f$. By definition of inverse relations, $(y, x_1) \in f^{-1}$ and $(y, x_2) \in f^{-1}$. But since f^{-1} is a function, $x_1 = x_2$ and we are done.

Surjectivity: We want for every $y \in Y$ there exists $x \in X$ such that $f(x) = y$. So choose an arbitrary $y \in Y$. Since f^{-1} is a function from Y to X, $f^{-1}(y) = x$ lies in X. Then $(y, x) \in f^{-1}$ implies that $(x, y) \in f$, which we are done.

Finally, suppose now that f is bijective. We aim to show that f^{-1} is a function. The fact that $\text{dom}(f^{-1}) = Y$ is trivially satisfied. There are two items to show: that for every $y \in Y$, there exists $x \in X$ such that $f^{-1}(y) = x$ and that $(y, x) \in f^{-1}$ and $(y, z) \in f^{-1}$ implies $x = z$. For the first item, we use the fact that f is surjective. Let $y \in Y$. By surjectivity, there exists $x \in X$ such that $(x, y) \in f$. Then by definition of inverse relation $(y, x) \in f^{-1}$ and we are done.

Now for the second item, suppose that $(y, x) \in f^{-1}$ and $(y, z) \in f^{-1}$. Then by definition of inverse relation $(x, y) \in f$ and $(z, y) \in f$. Then using injectivity, we see that this should imply $x = z$ and so we are done. □

As seen in the proof, for f^{-1} to be a function from Y to X, injectivity and surjectivity plays a crucial role.

From the theorem, inverses of a function exists if and only if f is bijective. We often call bijectivity a necessary and sufficient condition for inverses. Necessary here means that without this property, inverses would not exist. Sufficiency here means that with this property, inverses can exists. Both of which when used together, exactly means ”if and only if”.

Proposition 1.3.5.3

If $f : X \to Y$ is a bijective function then f^{-1} is bijective. Moreover, $f^{-1} \circ f$ is the identity function on X and $f \circ f^{-1}$ is the identity function on Y.

Proof. We know that since $f : X \to Y$ is bijective, $f^{-1} : Y \to X$ is a proper function. We first prove injectivity. Suppose that $(y, x) \in f^{-1}$ and $(z, x) \in f^{-1}$. By definition of inverse relation we have that $(x, y) \in f$ and $(x, z) \in f$. This means that $y = z$ by definition of a function. For surjectivity, suppose that $x \in X$. Since f is a function from X, there exists y such that $f(x) = y$. But this means that $(x, y) \in f$ and by definition of inverse relations, $(y, x) \in f^{-1}$ and so we are done.

Making sure that applying the inverse and the function itself is just the identity function is an easy exercise. □

Definition 1.3.5.4: Images and Preimages

Let $f : X \to Y$ be a function. Let $A \subseteq X$ and $B \subseteq Y$. Denote the image of A under f to be

$$f(A) = \{y \in Y | f(x) = y, \forall x \in A\} \subseteq Y$$

Define the preimage of B under f to be the set

$$f^{-1}(B) = \{x \in X | f(x) \in B\} \subseteq X$$

This is such an abuse of notation that I cannot stress enough. Try not to be confused with the usual inverse of f, which does not always exists as a function, while the preimage, always exists regardless of whether f has an inverse function.

Now that images and preimages have come into play, can you express surjectivity in terms of images and/or preimages? Can you also express the condition for bijecitivity and inverses in terms of images and/or preimages?

Next proposition is a crucial one because it tells us what happens by applying f and its inverse relation when inverses of the function may not exist.

Proposition 1.3.5.5

Let A, B, X, Y be sets. Let $f : X \to Y$.

- If $B \subseteq Y$ then $f(f^{-1}(B)) \subseteq B$
- If $B \subseteq Y$ and f is surjective then $f(f^{-1}(B)) = B$
- If $A \subseteq X$ then $A \subseteq f^{-1}(f(A))$

- If $A \subseteq X$ and f is injective then $f^{-1}(f(A)) = A$

Proof. Let $f : X \to Y$ be a function.

- Let $y \in f(f^{-1}(B))$. Then there exists $x \in f^{-1}(B)$ such that $y = f(x)$. But $x \in f^{-1}(B)$ implies $f(x) \in B$ by definition. Thus $y \in B$ and we are done.
- We just have to show that $B \subseteq f(f^{-1}(B))$. Let f be surjective and $y \in B$. Then by surjectivity there exists $x \in f^{-1}(B)$ such that $y = f(x)$. But $x \in f^{-1}(B)$ means that $f(x) \in f(f^{-1}(B))$. Thus $y \in f(f^{-1}(B))$ and we are done.
- Let $x \in A$. Then $f(x) \in f(A)$. But by definition of $f^{-1}(f(A))$, $x \in f^{-1}(f(A))$ thus we are done.
- We just have to show that $f^{-1}(f(A)) \subseteq A$. Let f be injective and $x \in f^{-1}(f(A))$. Then by definition, $f(x) \in f(A)$. By injectivity, there exists only one element in A that maps to $f(x)$, and that is precisely x. Thus $x \in A$ and we are done.

□

Interested readers may look up left and right inverses of a function. They are characterized by the fourth and second item respecrtively. By combining these two properties, left and right inverses combine to become an inverse of a function in the sense that it maps Y to X.

1.4 Number Systems

1.4.1 Axiom of Infinity

Before we formulate the natural numbers via set theoretic language, we need one more axiom to guarantee the existence of the set of natural numbers.

Definition 1.4.1.1: Successor

Define the successor of a set x to be

$$x^+ = x \cup \{x\}$$

Notice that the successor of a set is also a set. With this notion we can define numbers as sets. For example, 0 is simply the empty set $\emptyset$, 1 would be $\{\emptyset\}$, 2 would be $\{\emptyset, \{\emptyset\}\}$ and so on. This gives a unique code for every natural number, as well as giving order to the set of natural numbers, as we will see soon.

Finally, we need to guarantee that such a set exists.

Axiom 1.4.1.2: Axiom of Infinity

There exists a set containing the empty set $\emptyset$ and containing the successor of each of its elements.

As with all the other axioms, to argue that an axiom is necessary and unprovable by other axioms is a very advanced topic. For now we will only state that it is needed.

1.4.2 The Peano Axioms

Definition 1.4.2.1: The Peano Axioms

We say that ω is a successor set if

1. $\emptyset \in \omega$
2. $n \in \omega \implies n^+ \in \omega$
3. $n^+ \neq \emptyset$ for all $n \in \omega$
4. $n, m \in \omega$ and $n^+ = m^+ \implies n = m$
5. $S \subset \omega$ and $\emptyset \in S$ and $n^+ \in S$ for all $n \in S$ implies $S = \omega$

Each of these axioms plays an important role in the formulation, without any one of them the natural numbers would not come into play.

The first item guarantees that the natural numbers is a nonempty set. The second item guarantees that all successors, as required to define natural numbers, lie in the successor set. The third item guarantees that the natural numbers is not a huge loop of numbers that circulates back to 0. The fourth item guarantees that no two numbers has a common successor. The final axiom prevents extra loops to appear other than the natural numbers itself. For example, there cannot be loops such as $y = x^+$, $z = y^+$ and $x = z^+$. Notice that the third item does not guarantees this since it onlys prevents the natural numbers to be one big loop. The fifth item in turn guarantees this since the natural numbers has to satisfy that when elements near the number is a natural number, it will also be a natural number. Since x, y, z is a closed loop, they will never be natural numbers since the fifth item states the induction basis that starts from 0.

From now on, in a successor set, we now say 0 in place of $\emptyset$. If the readers wish to, we can now say that the successor of 0 is 1, the successor of 1 is 2 and vice versa.

The following theorem allows functions that work recursively to be defined.

Theorem 1.4.2.2: Recursion Theorem

Let X be a set and $a \in X$. Let $f : X \to X$. There exists a function $u : \omega \to X$ such that

- $u(0) = a$
- $u(n^+) = f(u(n))$ for all $n \in \omega$

Proof. Let
$$E = \{A \subset \omega \times X | (0, a) \in A \text{ and } (n, x) \in A \implies (n^+, f(x)) \in A\}$$
Since $\omega \times X \in E$, $E \neq \emptyset$. Define $u = \bigcap_{B \in E} B$. Then $u \in E$. We want to show that u is a function. We prove that the set of all n such that $(n, x) \in u$ and $(n, y) \in u \implies x = y$ is ω Let

$$S = \{n \in \omega | (n, x) \in u, (n, y) \in u \implies x = y\}$$

We prove it by the fifth item of Peano Axiom.

We first show that $0 \in S$. Suppose for a contradiction that $0 \notin S$ and $(0, b) \in u$ and $a \neq b$. Consider $v = u - \{(0, b)\}$. $(0, a) \in v$ and $(n, x) \in A \implies (n^+, f(x)) \in A$ since $n^+ \neq 0$ for all $n \in \omega$. Thus $v \in E$ contradicts the fact that $u = \bigcap_{B \in E} B$.

We now show that $n^+ \in S$ if $n \in S$. Let $n \in S$. Then there exists a unique $x \in X$ such that $(n, x) \in u$. Suppose that $n^+ \notin S$. Then there exists $y \neq f(x)$ such that $(n, y) \in u$. Consider $v = u - \{(n, y)\}$. $(0, a) \in v$ and $(n, x) \in v \implies (n^+, f(x)) \in v$. Thus $v \in E$ contradicts the fact that $u = \bigcap_{B \in E} B$. We thus have $S = \omega$, finishing the proof. □

Expanding things out, we see that $u(0) = a$, $u(1) = f(a)$, $u(2) = f(f(a))$ and so on. We have constructed a function u such that input the number in u means that you are compositing that number of f to the initial element. This is why through this theorem, we allowed the existence of recursive functions. The application of this immediate as we will use it to define addition and multplication through this theorem.

1.4.3 Arithmetic

We begin by defining the notion of addition in natural numbers.

Proposition 1.4.3.1

For every natural number m there exists a function $s_m : \mathbb{N} \to \mathbb{N}$ such that

- $s_m(0) = m$
- $s_m(n^+) = (s_m(n))^+$

$s_m(n)$ is by definition, the sum $m + n$.

Proof. Set X as $\mathbb{N}$, f as the successor function in recursion theorem and we are done. □

The recursion here used is the repeated use of successors, meaning we are applying +1 a certain amount of times.

We can now prove properties of addition in set theoretic language.

Proposition 1.4.3.2: Properties of Addition

Let $x, y, z \in \mathbb{N}$.

- (A1) $x + y \in \mathbb{N}$

- (A2) $(x+y)+z = x+(y+z)$
- (A3) $0+x = x = x+0$
- (A4) $x+y = y+x$

Proof. We prove associativity, identity and commutativity in order.

The closure under addition is direct from the definition of addition.

- For associativity, we induct on z. When $z=0$, we have

  $$\begin{aligned}(x+y)+0 &= x+y \\ &= x+(y+0)\end{aligned}$$

  Suppose that $(x+y)+n = x+(y+n)$, we have

  $$\begin{aligned}(x+y)+n^+ &= ((x+y)+n)^+ \\ &= (x+(y+n))^+ && \text{(Induction Hypothesis)} \\ &= x+(y+n)^+ \\ &= x+(y+n^+)\end{aligned}$$

  Thus by the principle of induction, we have associativity.

- For identity, we induct on x. When $x=0$, we have that

  $$0+0 = 0 = 0+0$$

  Suppose that $0+n = n = n+0$, we have

  $$\begin{aligned}0+n^+ &= (0+n)^+ \\ &= n^+ && \text{(Induction Hypothesis)} \\ &= n^+ + 0\end{aligned}$$

  Thus by the principle of induction, we have identity.

- For commutativity, we induct on y first to show that $x^+ + y = (x+y)^+$. When $y=0$, we have $x^+ + 0 = x^+ = (x+0)^+$. Now suppose that $x^+ + n = (x+n)^+$

  $$\begin{aligned}x^+ + n^+ &= (x^+ + n)^+ \\ &= ((x+n)^+)^+ && \text{(Induction Hypothesis)} \\ &= (x+n^+)^+\end{aligned}$$

  Thus our first induction is complete. Now we prove commutativity by induction on x. When $x=0$, we have $0+y = y+0$ from identity. Now suppose that $x+y = y+x$.

  $$\begin{aligned}x^+ + y &= (x+y)^+ && \text{(First induction)} \\ &= (y+x)^+ && \text{(Induction Hypothesis)} \\ &= y+x^+\end{aligned}$$

  Thus by the principle of induction, we have commutativity.

□

We then proceed to multiplication.

Proposition 1.4.3.3

For every natural number m there exists a function $p_m : \mathbb{N} \to \mathbb{N}$ such that

- $p_m(0) = 0$
- $p_m(n^+) = p_m(n) + m$

$p_m(n)$ is by definition, multiplication $m \times n$.

Proof. Take X as $\mathbb{N}$ and f as $f(x) = x + m$. Then using the recursion theorem we are done. □

We note here, that the successor function n^+ is in fact equivalent to $n + 1$ (using the definition of addition). This may be seen inherently when the successor notion is introduced, but it is only now that we can formulate it properly since addition is defined.

Similarly, the recursion here is that repeated addition of $+m$. And we can also prove remaining properties of multiplication in set theoretic language.

Proposition 1.4.3.4

Let $x, y, z \in \mathbb{N}$. Addition and multiplication follow the distributive law $x \cdot (y + z) = x \cdot y + x \cdot z$.

Proof. We prove the distributive law by induction on z. When $z = 0$, we have $x \cdot (y + 0) = x \cdot y$ and $x \cdot y + x \cdot 0 = x \cdot y$. Now suppose that $x \cdot (y + n) = x \cdot y + x \cdot n$.

$$\begin{aligned}
x(y + n^+) &= x(y + n)^+ && \text{(Definition of Addition)} \\
&= x(y + n) + x && \text{(Definition of Multiplication)} \\
&= (xy + xn) + x && \text{(Induction Hypothesis)} \\
&= xy + xn + x && \text{(Associativity of Addition)} \\
xy + xn^+ &= xy + (xn + x) && \text{(Definition of Multiplication)} \\
&= xy + xn + x && \text{(Associativity of Addition)}
\end{aligned}$$

Thus by the principle of induction, we have the distributive law. □

Proposition 1.4.3.5

Let $x, y, z \in \mathbb{N}$.

- (M1) $x \cdot y \in \mathbb{N}$
- (M2) $(x \cdot y) \cdot z = x \cdot (y \cdot z)$
- (M3) $1 \cdot x = x = x \cdot 1$
- (M4) $x \cdot y = y \cdot x$

Proof. We prove associativity, commutativity, identity in this order. Note we first prove commutativity then prove identity.

The closure under Multiplication is direct from the definition of multiplication.

- We prove associativity by induction on z. When $z = 0$, $(x \cdot y) \cdot 0 = 0$ and

$x \cdot (y \cdot 0) = x \cdot 0 = 0$. Now suppose that $(x \cdot y) \cdot n = x \cdot (y \cdot n)$. We have

$$\begin{aligned}
(x \cdot y) \cdot n^+ &= (x \cdot y) \cdot n + x \cdot y && \text{(Definition of Multiplication)} \\
&= x \cdot y + (x \cdot y) \cdot n && \text{(Commutativity of Addition)} \\
&= x \cdot y + x \cdot (y \cdot n) && \text{(Induction Hypothesis)} \\
&= x \cdot (y + y \cdot n) && \text{(Distributivity)} \\
&= x \cdot (y \cdot n + y) && \text{(Commutativity of Addition)} \\
&= x(y \cdot n^+) && \text{(Definition of Multiplication)}
\end{aligned}$$

Thus by the principle of induction, we have the associative law.

- We prove commutativity by induction on y. When $y = 0$, we have $x \cdot 0 = 0 = 0 \cdot x$. Now suppose that $x \cdot y = y \cdot x$, we have

$$\begin{aligned}
x \cdot y^+ &= (x \cdot y) + x && \text{(Definition of Multiplication)} \\
&= y \cdot x + x && \text{(Induction Hypothesis)} \\
&= x + \cdot (y + x) && \text{(Commutativity of Addition)} \\
&= y^+ \cdot x && \text{(Distributivity)}
\end{aligned}$$

Thus by the principle of induction, we have the commutative law.

- The identity is a special case of the commutative law, taking $y = 1$.

□

Now that we have the natural numbers, we can also extend the numbers into integers, rationals, real numbers and even complex numbers. Most of them without the use of set theoretic language. However, this is another topic unrelated to set theory mostly.

1.5 Comparing Sets

1.5.1 Order in the Natural Numbers

Another important aspect of the natural numbers is that they are comparable. We have the notion of size in them. The Peano axioms for the natural numbers also encapusulates this nicely.

Definition 1.5.1.1: Comparable

Two natural numbers m, n are comparable if either $m \in n$ or $m = n$ or $n \in m$.

This is called a trichotomy, which means that two natural numbers are comparable when exactly one out of the three conditions are fulfilled. Since we do not have the concept of size yet, we use belonging symbols because if one remembers, natural numbers are defined recursively.

The following theorem states that the trichotomy holds for any two natural numbers, which leads us to being able to compare them.

Theorem 1.5.1.2

Any two natural numbers are comparable.

Proof. □

Now that we have seen that any two natural numbers are comparable, we can essentially order them according to whether one is a subset of another, or in terms of magnitude, whether one is larger than the other or not.

Definition 1.5.1.3: Order in $\mathbb{N}$

Define the relation $<$ for two natural numbers m, n such that if $m \in n$, then $m < n$.

As we will see later, even comparisons have different types, namely total orders and partial orders. Total order means that everything in the set can be rearranged into one long queue according to the way we want to order it. Partial order means that not everything in the set is comparable against each other, leading to a tree like appearance when we try and sort the elements.

Proposition 1.5.1.4

The relation $<$ is transitive.

It is worth noting here that $<$ is neither reflexive nor transitive which the readers can check. Can you think of the inverse relation for $<$?

1.5.2 Cardinality

We begin not from the definition of cardinality, but rather by introducing comparison between the cardinality of two sets. Inherently, cardinality simply means the number of elements in a set (for the finite case).

Definition 1.5.2.1: Cardinality of a Set

Two sets A and B have the same cardinality if there exists a bijection between A and B. In this case, we write

$$|A| = |B|$$

where $|A|$ stands for the cardinality of A.

We say that A has cardinality strictly less than B if there exists an injective function from A to B, but no bijective function exists. In this case, we write $|A| < |B|$.

Basically we compare the size of two sets simply by attempting to define a bijective function between of them. Naturally, we have the following proposition.

Proposition 1.5.2.2

Every natural number has cardinality strictly less than $|\mathbb{N}|$.

Definition 1.5.2.3: Finite Sets

A set A is called finite if it has the same cardinality as a natural number represented as a set. In that case, the cardinality of A is precisely the natural number.

For example, if $A = \{a, b, c, d\}$, then it has cardinality 4 and we write it as $|A| = 4$. From the finiteness of the sets, we can perform the usual addition and multiplication on cardinality, provided it makes sense.

Proposition 1.5.2.4

Let E, F be sets with finite cardinality.

- If E, F are disjoint, then $|E \cup F| = |E| + |F|$
- $|E \times F| = |E| \cdot |F|$

Definition 1.5.2.5: Countably Infinite Sets

We say that a set A is countably infinite if it has cardinality equal to $\mathbb{N}$. In other words, if

$$|A| = |\mathbb{N}|$$

In particular, mathematicians also say that a set is countable if it is either finite or countably infinite. This is due to the fact that there is also a notion of uncountability.

Definition 1.5.2.6: Uncountable Sets

We say that a set A is uncountable if it has cardinality strictly greater than $\mathbb{N}$. In other words, if

$$|A| > |\mathbb{N}|$$

1.5.3 Posets

Definition 1.5.3.1: Antisymmetric

A relation R in X is antisymmetric if for every $x, y \in X$, $(x, y) \in R$ and $(y, x) \in R$ implies $x = y$.

Definition 1.5.3.2: Partial Order

A partial order is a relation R in X such that R is reflexive, antisymmetric and transitive in X. We use the symbol $\leq$ for a partial order. We say that X is partially ordered.

Definition 1.5.3.3: Total Order

$\leq$ is a total order in X if for every $x, y \in X$ either $x \leq y$ or $y \leq x$. In this case we say that X forms a chain. We also say that X is totally ordered.

Definition 1.5.3.4: Initial Segments

Let X be a partially ordered set and $a \in X$, the set

$$s(a) = \{x \in X : x < a\}$$

is the initial segment determined by a. The weak initial segment is denoted

$$\bar{s}(a) = \{x \in X : x \leq a\}$$

Definition 1.5.3.5: Least and Greatest Element

Let X be a partially ordered set and $a \in X$ such that $a \leq x$ for all $x \in X$. Then a is the least element of X. If $b \in X$ such that $x \leq b$ for all $x \in X$ then b is the greatest element of X.

Definition 1.5.3.6: Minimal and Maximal Elements

Let X be a partially ordered set and $a \in X$ such that $x \leq a$ implies $x = a$, then a is called the minimal element of X. If $b \in X$ such that $b \leq x$ implies $x = b$, then b is called the maximal element of X.

1.5.4 Axiom of Choice and Zorn's Lemma

Definition 1.5.4.1: Families

A family is a function $f : I \to X$ where I is an index set, usually ω or a natural number, such that every element of the range of f is of the form x_i, $i \in \omega$.

Axiom 1.5.4.2: Axiom of Choice

The Cartesian Product of a non-empty family of non-empty sets is non-empty.

Theorem 1.5.4.3

If a set is infinite, then it has a subset equivalent to ω.

Theorem 1.5.4.4: Zorn's Lemma

If X is a partially ordered set such that every chain in X has an upper bound, then X contains a maximal element.

1.5.5 Well Ordering

Definition 1.5.5.1: Well Ordered Set

A partially ordered set is well ordered if every non-empty subset of it has a smallest element.

Theorem 1.5.5.2

Every well ordered set is totally ordered.

Theorem 1.5.5.3: The Principle of Transfinite Induction

Suppose that S is a subset of a well ordered set X, and suppose that $x \in X$ such that $s(x) \subset S$, then $x \in S$.

Definition 1.5.5.4: Continuation

A well ordered set A is a continuation of a well ordered set B if

- $B \subset A$
- B is an initial segment of A
- B and A have the same ordering

Theorem 1.5.5.5

If E is an arbitrary collection of initial segment of a well ordered set, E is a chain with respect to continuation.

Theorem 1.5.5.6

If a collection E of well ordered sets is a chain with respect to continuation and if $U = \bigcup_{X \in E} X$, then there is a unique well ordering of U such that U is a continuation of each set.

Theorem 1.5.5.7: Well Ordering Theorem

Every set can be well ordered.

Proposition 1.5.5.8

The well ordering theorem implies the axiom of choice.

1.5.6 Transfinite Recursion

Definition 1.5.6.1: A sequence of type a in X

Let a be an element in a well ordered set W. Let X be an arbitrary set. The sequence of type a in X means a function from $s(a) \subset W$ into X.

Definition 1.5.6.2: A sequence of type W in X

A sequence of type W in X is a function f whose domain consists of all sequences of type a in X, for all elements a in W and range is included in X.

Theorem 1.5.6.3: Transfinite Recursion Theorem

If W is a well ordered set and if f is a sequence function of type W in a set X, then there exists a unique function U from W into X such that $U(a) = f(U^a)$ for each a in W, where U^a is the restriction of $U : W \to X$ to the initial segment $s(a)$.

Definition 1.5.6.4: Similarity

Two partially ordered sets are similar if there is a one to one correspondence that preserves order, or $f(a) \leq f(b) \implies a \leq b$.

Proposition 1.5.6.5

Let f be a similarity from X to Y.

- f^{-1} is a similarity from Y to X
- gf is a similarity from X to Z if g is a similarity.

Theorem 1.5.6.6

If f is a similarity of a well ordered set X to itself, then $a \leq f(a)$ for all $a \in X$.

Theorem 1.5.6.7

Let X, Y be well ordered sets. If X and Y are similar, then the correspondence function is unique.

Theorem 1.5.6.8

A well ordered set is never similar to one of its initial segments.

Theorem 1.5.6.9

[Compatibility Theorem] Let X and Y be well ordered sets. Either X and Y are similar or one of them is similar to an initial segment of the other.

1.6 Beyond Infinity

1.6.1 Ordinal Numbers

Definition 1.6.1.1: ω Successor Function

Let f be a function from strict predecessors of some natural number n, and $f(0) = \omega$ and $f(m^+) = f(m)^+$ whenever $m^+ < n$. Then this function is an ω Successor Function.

Axiom 1.6.1.2: Axiom of Substitution

If $S(a, b)$ is a sentence such that for each a in a set A the set $\{b : S(a, b)\}$ can be formed, then there exists a function F with domain A such that $F(a) = \{b : S(a, b)\}$ for each a in A.

Definition 1.6.1.3

An ordinal number is defined as the well ordered set α such that $s(\xi) = \xi$ for all $\xi \in \alpha$, where $s(\xi) = \{\eta \in \alpha : \eta < \xi\}$

Theorem 1.6.1.4

If α is an ordinal number, α^+ is also an ordinal number.

Definition 1.6.1.5: Ordinal Numbers after ω

Define F from the axiom of substitution such that $F(0) = \omega$ and $F(n^+) = F(n)^+$ for each natural number n. We write $F(n) = \omega + n$ and ω^2 as the set consisting of all n and all $\omega + n$ with $n \in \omega$.

Theorem 1.6.1.6

If ξ is an element of an ordinal number α, ξ is also an ordinal number.

Theorem 1.6.1.7

If two ordinal numbers are similar, than they are equal.

Theorem 1.6.1.8

Every set of ordinal numbers is totally ordered. Every set of ordinal numbers is well ordered.

Theorem 1.6.1.9: Transfinite Ordinal Numbers

The natural numbers are finite ordinal numbers, the others are called transfinite. Each finite ordinal numbers other than 0 has an immediate predecessor. Transfinite ordinal numbers that do not have a predecessor is called limit numbers.

Theorem 1.6.1.10: Supremum of Ordinal Numbers

Every set of ordinal numbers have a supremum.

Theorem 1.6.1.11: Burali-Forti Paradox

No set contains all ordinal numbers.

Theorem 1.6.1.12: Counting Theorem

Each well ordered set is similar to a unique ordinal number.

Chapter 2

Number Systems

2.1 Natural Numbers

2.1.1 Order of Natural Numbers

Definition 2.1.1.1: Natural Numebers

The set of natural numbers $\mathbb{N}$, formulated in ZFC set theory via peano axioms, is the set

$$\mathbb{N} = \{0, 1, 2, \dots\}$$

of natural numbers. Addition and multiplication is also defined.

Definition 2.1.1.2: Order

Let $a, b \in \mathbb{N}$. We say $a < b$ if there exists some $c \in \mathbb{N}$ and $c \neq 0$ so that $a + c = b$. $<$ is a relation in the set $\mathbb{N}$.

Proposition 2.1.1.3: Trichotomy

Let $a, b \in \mathbb{N}$. Then either $a = b$ or $a < b$ or $b < a$.

Proof. From set theory, we have that $a \in b$ or $a = b$ or $b \in a$, which corresponds to $a < b$, $a = b$, $b < a$ respectively. □

Proposition 2.1.1.4

Suppose that $a, b, c \in \mathbb{N}$. The relation $<$ has the below properties.

- $a < b$ and $b < c \implies a < c$
- $a < b \implies a + c < b + c$
- $a < b \implies ac < bc$

Proof. We prove the three using the definition of $<$.

- Suppose $a < b$ and $b < c$. There exists some $x, y \in \mathbb{N}$ such that $a + x = b$ and $b + y = c$. Then $a + x + y = c$ thus $a < c$
- Suppose $a < b$. Then $a + c = b$ for some $c \in \mathbb{N}$. Then $b + d = (a + c) + d = (a + d) + c$ thus $a + d < b + d$.
- Suppose $a < b$. Then $a + d = b$ for some $d \in \mathbb{N}$. Then $ac + dc = bc$ thus $ac < bc$

□

Definition 2.1.1.5: Partial Order

Let $a, b \in \mathbb{N}$. We say $a \leq b$ if either $a < b$ or $a = b$.

Theorem 2.1.1.6

The relation $\leq$ in the natural numbers are partial order.

Proof. Recall from set theory that a partial order is a relation such that it is reflexive, antisymmetric and transitive.

- Since we have $a \leq a$, $\leq$ is reflexive.
- $a \leq b$ and $b \leq a \implies a = b$ by the trichotomy of natural numbers.
- $a \leq b$ and $b \leq c \implies a \leq c$ from the properties of the relation $<$.

□

Theorem 2.1.1.7

The set of natural numbers is totally ordered.

Proof. For any two numbers in $\mathbb{N}$, we have the trichotomy, thus we have either $a \leq b$ or $b \leq a$ for all $a, b \in \mathbb{N}$. □

2.2 Integers

2.2.1 Introduction to Integers

Definition 2.2.1.1: Integers

For every natural number n except 0, we introduce a number $-n$ such that $n + (-n) = 0$. It is called the additive inverse of n. The set of all natural numbers and additive inverses is the set

$$\mathbb{Z} = \{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\}$$

Definition 2.2.1.2: Subtraction

We define subtraction of $a, b \in \mathbb{Z}$ to be $a - b$, which is $a + (-b)$.

Proposition 2.2.1.3

The set of integers is totally ordered.

2.2.2 Divisibility

We begin our study of number theory with divisibility.

Definition 2.2.2.1

[Divisibility] Let $a, b \in \mathbb{Z}$. We define the relation

$$a|b$$

if and only if there exists some $k \in \mathbb{Z}$ such that $b = ak$. We say that a divides b in this case.

The definition is vey simple. The intuition is straight forward as well. Savour this moment as the subject increases its difficulty exponentially.

Proposition 2.2.2.2

Let $d, m, n \in \mathbb{Z}$. The relation $|$ has the following properties and thus is a partial order in $\mathbb{N}$.

- (Reflexivity) $n|n$
- (Antisymmetry) $m|n$ and $n|m \implies m = n$
- (Transitivity) $d|n$ and $n|m \implies d|m$
- (Linearity) $d|n$ and $d|m \implies d|(an + bm)$ for any $a, b \in \mathbb{Z}$
- $1|n$
- $n|0$

Proof. We prove antisymmetry and transitivity and leave the others for the reader. Let $m, n, d \in \mathbb{Z}$.

- (Antisymmetry) If $m|n$ and $n|m$ then there exists some $k_1, k_2 \in \mathbb{N}$ such that $n = k_1 m$ and $m = k_2 n$ thus $n = k_1 k_2 n$. Then $k_1 k_2 = 1 \implies k_1 = k_2 = 1$ and $m = n$
- (Linearity) If $d|n$ and $n|m$ then there exists $k_1 k_2 \in \mathbb{N}$ such that $n = k_1 d$ and $m = k_2 n$. Then $m = k_2 k_1 d$ thus $d|m$

□

These properties will come up again and again and will be the foundation of number theory. It is safe to say that number theory is built upon the notion of divisibility.

2.2.3 The Division Algorithm

This section is dedicated to develop the Euclidean algorithm, a means to find the greatest common divisor. The gcd is a central notion in number theory as well.

Definition 2.2.3.1

[Greatest Common Divisor] Suppose that $m, n \in \mathbb{Z}$. A number $d \in \mathbb{N}$ such that

- $d \geq 0$
- $d|m$ and $d|n$
- $e|a$ and $e|b \implies e|d$

is called the greatest common divisor of m and n, denoted $\gcd(m, n)$.

In contrast to the greatest common divisor, we also have the lowest common multiple. Although they work as a pair, we often see the notion of gcd come up more than lcm.

Definition 2.2.3.2

[Lowest Common Multiple] Suppose that $m, n \in \mathbb{Z}$. A number $l \in \mathbb{N}$ such that

- $l \geq 0$
- $m|l$ and $n|l$
- $m|e$ and $n|e \implies l|e$

is called the lowest common multiple of m and n, denoted $\text{lcm}(m, n)$.

Beware that both of these definitions does not imply the uniqueness of such a number. However, with a little work, we will see that both of them are indeed unique. Readers should think about whether the existence of these numbers is guaranteed as well.

Proposition 2.2.3.3

Let $m, n \in \mathbb{Z}$. $\gcd(m, n)$ and $\text{lcm}(m, n)$ are unique.

Proof. By the third property of both numbers, we must have if c, d are $\gcd(m, n)/\text{lcm}(m, n)$, then $c|d$ and $d|c$ thus $c = d$ and $\gcd(m, n)/\text{lcm}(m, n)$ is unique. □

We will see more on gcd and lcm when we deal with factorization. For now, we turn our heads to the division algorithm. This algorithm proves to us that upon dividing two integers, as long as they are not divisible by one or the other, you can always guarantee a remainder smaller than the divident.

Theorem 2.2.3.4

[The Division Algorithm] Let $a \in \mathbb{N}$ and $b \in \mathbb{Z}$ with $b \neq 0$. Then there exists unique $q, r \in \mathbb{Z}$ such that

$$b = aq + r$$

with $0 \leq r < a$.

Proof. We prove existence first by considering three cases.
Cases 1: b is divisible by a. If b is divisible by a then there exists $k \in \mathbb{Z}$ such that $b = ka$ thus $k = q$ and $r = 0$.

Case 2: b is positive and a does not divide b. Let

$$S = \{b - ka \in \mathbb{N} | k \in \mathbb{N}\}$$

Then $S \subseteq \mathbb{N}$ thus we can apply the well-ordering principle to S. Let r be the least natural number in S. Then $r \in S$ implies $r = b - ka$ for some $k \in \mathbb{N}$. Thus $b = ka + r$ for some k and r. We show that $r < a$. Suppose for a contradiction that $r \geq a$. Then $u = r - a \in \mathbb{N}$ and

$$b = ka + r \implies b = ka + (u - a) \implies b = (k-1)a + u$$

thus $u \in S$ and $u < r$, contradicting the fact that r is the least element in S. Thus $r \leq a$. If $r = a$, then

$$b = ka + a \implies b = (k+1)a$$

which means that $a|b$ which is false in our case. Thus we must have $r < a$.

Case 3: b is negative and a does not divide b. Then apply the exact same argument to the number $-b$ to get $(-b) = ka + r$ and $b = -ka - r$. Let $k' = -k - 1$ and $r' = -r + a$. Then

$$b = -ka - r = k'a + a + r' - a = k'a + r'$$

Since we have $0 \leq r < a$, we have $-a < -r \leq 0$ and $0 < r' \leq a$. Again $r' \neq a$ or else $a|b$ which contradicts our assumption.

We now prove uniqueness. Suppose that $b = aq_1 + r_1$ and $b = aq_2 + r_2$. Then $r_1 - r_2 = a(q_2 - q_1)$. We know that $-a < r_1 - r_2 < a$ thus $-a < a(q_2 - q_1) < a$ and $-1 < q_2 - q_1 < 1$ which is impossible for integers q_1, q_2 unless $q_1 = q_2$. If $q_1 = q_2$ then $r_1 = r_2$ and we are done. □

The division algorithm does not require b to be larger than a. In fact, if a is larger than b, then the division algorithm simply gives a itself as the remainder. Before we reach our conclusion, we need one more proposition.

Proposition 2.2.3.5

Suppose that $m \geq n > 0$ are natural numbers with $m = qn + r$ for some $q, r \in \mathbb{N}$. Then

$$\gcd(m, n) = \gcd(n, r)$$

Proof. Suppose that $d = \gcd(m, n)$. Then we know that $d < n$ from definition. We want to show that d satisfies the three results of a gcd but in terms of n and r. Since $d|n$ and $d|m$, by linearity we must have $d|r$.

Now suppose for a contradiction that there exists e such that e is a common divisor of n and r and $e > d$. Then $e|n$ and $e|r$ by definition thus $e|m$ by linearity. $e|m$ and $e|n$ implies that e is a larger common divisor of m and n than d. However this is not possible since d is assumed to be the largest among the common divisors. This is a contradiction thus $d = \gcd(n, r)$ and we are done. □

Theorem 2.2.3.6

[Euclid's Algorithm] Suppose that $m \geq n > 0$ are natural numbers. We have the following inequalities.

$$m = nq_1 + r_1 \text{ with } 0 < r_1 < n$$
$$n = r_1q_2 + r_2 \text{ with } 0 < r_2 < n$$
$$r_1 = r_2q_3 + r_3 \text{ with } 0 < r_3 < n$$
$$\dots\dots\dots\dots\dots\dots\dots\dots\dots$$
$$r_{k-2} = r_{k-1}q_k + r_k \text{ with } 0 < r_k < r_{k-1}$$
$$r_{k-1} = r_kq_{k-1}$$

From this, we have $r_k|r_{k-1}$, $r_k|r_{k-2} \dots r_k|n$ and $r_k|m$.

Proof. The first part of the results is due to the repeated use of the division algorithm. For the second part, we have

$$\gcd(m, n) = \gcd(n, r_1) = \gcd(r_1, r_2) = \cdots = \gcd(r_{k-1}, r_k) = r_k$$

and we are done. □

Lemma 2.2.3.7

[Bezout's Lemma] Let $a, b \in \mathbb{Z}$ such that they are not both 0. Then there exists $x, y \in \mathbb{Z}$ such that

$$ax + by = \gcd(a, b)$$

Proof. Reconstruct x and y using the Euclidean Algorithm. This is possible since $\gcd(m, n) = r_k$ and every $r_1, \dots, r_{k-1}$ has a factor of r_k in it. □

Lemma 2.2.3.8

Let $a, b \in \mathbb{Z}$ such that they are not both 0. Then the equation

$$ax + by = \gcd(a, b)$$

has an infinite number of integer solutions.

Proof. Using Bezout's Lemma, we conclude that (x_0, y_0) is a solution to the equation. But then

$$(x_0 - bt, y + at)$$

are also solutions for $t \in \mathbb{Z}$ since

$$a(x_0 - bt) + b(y + at) = ax + by = \gcd(a, b)$$

□

Corollary 2.2.3.9: L

t $a, b \in \mathbb{Z}$ such that they are not both 0 and $d \in \mathbb{Z}$. Then d divides a and b if and only if $d | \gcd(a, b)$.

2.2.4 Unique Factorization

Definition 2.2.4.1

[Prime Numbers] We say that $n \in \mathbb{N}$ is a prime number if and only if it has exactly two factors, which is 1 and n. Else n is composite.

Lemma 2.2.4.2

Every integer is divisible by a prime.

Lemma 2.2.4.3

Every integer $n > 1$ can be written as a product of primes.

Theorem 2.2.4.4

There is an infinite number of primes.

Proposition 2.2.4.5

[Euclid's Lemma] Suppose that $p, m, n \in \mathbb{N}$, with p prime and $m, n > 1$. Suppose that $p | mn$. Then p divides at least one of m or n.

Proposition 2.2.4.6

Suppose that p is a prime such that $p|a_1a_2\cdots a_k$. Then $p|a_i$ for some $i \in \{1,2,\dots,k\}$

Theorem 2.2.4.7

[Fundamental Theorem of Arithmetic] Suppose that $n \neq 0$ is a natural number. Then there exists exactly one prime factorization for every n, meaning that the decomposition

$$n = \prod_{k=1}^{n} p_k^{s_k}$$

where p_k is prime exists and is unique.

Theorem 2.2.4.8

Suppose that $m, n \in \mathbb{N}$. Suppose that

$$m = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_r^{\alpha_r}$$

$$n = p_1^{\beta_1} p_2^{\beta_2} \cdots p_q^{\beta_q}$$

with $p_1 = 2$, $p_2 = 3$, $p_3 = 5 \dots$. Without loss of generality $r \leq q$. Then

$$\gcd(m,n) = p_1^{\min(\alpha_1,\beta_1)} p_2^{\min(\alpha_2,\beta_2)} \cdots p_q^{\min(\alpha_q,\beta_q)}$$

$$\operatorname{lcm}(m,n) = p_1^{\max(\alpha_1,\beta_1)} p_2^{\max(\alpha_2,\beta_2)} \cdots p_q^{\max(\alpha_q,\beta_q)}$$

Theorem 2.2.4.9

Suppose that m and n are natural numbers. Then

$$\gcd(m,n) \times \operatorname{lcm}(m,n) = m \times n$$

Proof. Since $\min\{a,b\} \cdot \max\{a,b\} = ab$, from the above theorem, we have that $\gcd(m,n) \times \operatorname{lcm}(m,n) = m \times n$ and we are done. □

2.3 Rational Numbers

We now produce multiplicative inverses to form the the set of rationals.

2.3.1 Introduction to Rationals

Definition 2.3.1.1

For every integer $n \in \mathbb{Z}$ except 0 and 1, we introduce a number n^{-1} such that $n \cdot n^{-1} = 1$. It is called the multiplicative inverse of n. The set of all combination of integers and multiplicative inverses is the set

$$\mathbb{Q} = \{\frac{a}{b} | a, b \in \mathbb{Z}, b \neq 0\}$$

Numbers in $\mathbb{Q}$ are called rational numbers.

Definition 2.3.1.2: Divsion

We define division of a and b to be $\frac{a}{b}$, which is $a \cdot b^{-1}$.

Definition 2.3.1.3: Reduced Form

Suppose that $a \in \mathbb{Q}$. $a = \frac{r}{s}$ is a reduced form if

- $s > 0$
- $\gcd(r, s) = 1$

Theorem 2.3.1.4

For every $x \in \mathbb{Q}$, x has exactly one reduced form.

2.3.2 Arithmetic and Order of Rationals

Definition 2.3.2.1

We define equality $\frac{a}{b} = \frac{c}{d}$ if and only if $ad = bc$.

Definition 2.3.2.2

[Arithmetic of Rationals] We define the four basic operators on rationals as follows.

- $\frac{a}{b} + \frac{c}{d} = \frac{ad+bc}{bd}$
- $\frac{a}{b} - \frac{c}{d} = \frac{ad-bc}{bd}$
- $\frac{a}{b} \cdot \frac{c}{d} = \frac{ac}{bd}$
- $\frac{a}{b} / \frac{c}{d} = \frac{ad}{bc}$

Proposition 2.3.2.3

The additive inverse of $\frac{a}{b}$ is $-\frac{a}{b}$. Its multiplicative inverse is $\frac{b}{a}$

2.4 Real Numbers

2.4.1 Dedekind Cuts

Definition 2.4.1.1: Dedekind Cuts

A dedekind cut is a partition of $\mathbb{Q}$ into two subsets A, B such that

- A is non-empty
- $A \neq \mathbb{Q}$
- $x, y \in \mathbb{Q}$ and $x < y$ and $y \in A \implies x \in A$
- $x \in A \implies$ there exists a $y \in A$ such that $y > x$

We use A to denote this cut since B is determined by A

Definition 2.4.1.2: Order

If A, B are dedekind cuts then we say that $A < B$ if and only if $A \subset B$.

Definition 2.4.1.3: Real Numbers

We define the set of real numbers $\mathbb{R}$ as the set of all dedekind cuts of $\mathbb{Q}$.

Proposition 2.4.1.4

Define addition, subtraction, multiplication, division as follows. Then the resulting set is also a dedekind cut.

- $A + B = \{a + b : a \in A \text{ and } b \in B\}$
- $A - B = \{a - b : a \in A \text{ and } b \in \mathbb{Q} \setminus B\}$
- $A \times B = \{a \times b : a \in A \text{ and } b \in B\}$ if $A, B \geq 0$ or $A, B \leq 0$. If at one of $A, B < 0$ then use the identity $-(-A \times B)$ or $-(A \times -B)$ depending on whether $A < 0$ or $B < 0$ respectively.
- $A/B = \{a/b : a \in A \text{ and } b \in \mathbb{Q} \setminus B\}$ if $A, B \geq 0$ or $A, B \leq 0$. Use the similar approach as multiplication when one of $A, B < 0$.

Proposition 2.4.1.5

The set of all rational numbers $\mathbb{Q}$ is a subset of the real numbers $\mathbb{R}$.

Proof. We define $A = \{x \in \mathbb{Q} : x < q\}$ for every $q \in \mathbb{Q}$. Thus the set A satisfies a dedekind cut. □

Definition 2.4.1.6: Irrational Numbers

Any dedekind cut which is not a rational number is called an irrational number.

Theorem 2.4.1.7

There exists an irrational number.

Proof. We want to show that there is an irrational number represented by a dedekind cut such that its square is 2. Consider the set $A = \{x \in \mathbb{Q} : x < 0 \text{ or } x^2 < 2\}$. A is non-empty since $0 \in A$. $A \neq \mathbb{Q}$ since $3 \notin A$. Suppose $p \in A$. We then need to show that $q \in A$ whenever $q < p$. When $0 \leq q < p$, we have $0 \leq q^2 < p^2$ from ordering of rationals. When $q < 0$, then $q \in A$ by definition of A. Thus this is true. Now we need to show that there is always a rational q larger than p which is in A. Choose $q = \frac{2p+2}{p+2}$, then $p < q$ and $q^2 < 2$. Thus A is a dedekind cut.

Now consider $A \times A$. We have $A \times A \leq 2$ since for all $x, y \in A$, we have $x^2 < 2$ and $y^2 < 2$ and thus $xy < 2$ whenever $xy \geq 0$. Thus the set
$A \times A = \{r \in \mathbb{Q} : r < 0 \text{ or } r = xy \text{ for some } x, y \in A \text{ and } x, y > 0\}$ is less than or equal to 2. We know that $A \times A$ is a dedekind cut. But we want to know if $A \times A$ represents the number 2. Suppose that $u \in A \times A$. Then we know that from A, there exists a number $v \in A$ such that $u < v^2 < 2$. And this applies for every u. Then we know that $A \times A = 2$ since
$A \times A = \{x \in \mathbb{Q} : x < 2\}$, which is our definition of rational numbers with dedekind cut.

We have proved that there exists a dedekind cut such that its square is 2. But is that dedekind cut irrational? We now represent A with $\sqrt{2}$. Suppose that $\sqrt{2}$ is rational. Then we can write it is as $\frac{m}{n}$ in reduced form. Then we have $2n^2 = m^2$. Then $2|m^2$ thus $2|m$. Let $m = 2k$ for some $k \in \mathbb{N}$. Then $2k^2 = n^2$ which similarly implies that $2|n$. This contradicts the fact that $\frac{m}{n}$ is in reduced form, thus $\sqrt{2}$ is in fact not rational, and is an irrational number. □

Proposition 2.4.1.8

Suppose that A, B, C are dedekind cuts.

- (O1) $A < B$ or $A = B$ or $B < A$
- (O2) $A < B$ and $A < C \implies A < C$
- (O3) $A < B \implies A + C < B + C$
- (O4) $A < B$ and $z > 0 \implies AC < BC$

Proposition 2.4.1.9

Let x, y, $z \in \mathbb{R}$.

- (A1) $x + y \in \mathbb{R}$
- (A2) $(x + y) + z = x + (y + z)$
- (A3) $\exists 0 \in \mathbb{R}$ such that $x + 0 = 0 = 0 + x$
- (A4) $x + y = y + x$
- (A5) $\exists(-x) \in \mathbb{R}$ such that $x + (-x) = 0 = (-x) + x$
- (M1) $xy \in \mathbb{R}$
- (M2) $(xy)z = x(yz)$
- (M3) $\exists 1 \in \mathbb{R}$ such that $x \cdot 1 = x = 1 \cdot x$
- (M4) $xy = yx$
- (M5) $\exists(x^{-1}) \in \mathbb{R}$ such that $x(x^{-1}) = 1 = (x^{-1})x$
- (D1) $x(y + z) = xy + xz$
- (O1) $x < y$ or $x = y$ or $y < x$
- (O2) $x < y$ and $y < z \implies x < z$
- (O3) $x < y \implies x + z < y + z$
- (O4) $x < y$ and $z > 0 \implies xz < yz$

The absolute value is an important function when it comes to defining useful concepts such as distances in the field of real.

Definition 2.4.1.10

[The Absolute Value] The absolute value of a real number x is defined by

$$|x| = \begin{cases} x & \text{if } x \geq 0 \\ -x & \text{if } x \leq 0 \end{cases}$$

The absolute value has some properties that are extremely useful in certain circumstances, notably number 4 and 5.

Proposition 2.4.1.11

The absolute Value has the folowing properties

1. $|x| \geq 0$
2. $|xy| = |x||y|$
3. $|\frac{x}{y}| = \frac{|x|}{|y|}$
4. $|x + y| \leq |x| + |y|$
5. $||x| - |y|| \leq |x - y|$

Proof. I left out the proofs of (2) and (3) since they are simplay obtained via case by case analysis.

1. When $x \geq 0$ we have $|x| = x \geq 0$. When $x < 0$ we have $|x| = -x > 0$
2. We start by squaring the left hand side of the inequality.
$$\begin{aligned} |x + y|^2 &= (x + y)^2 \\ &= x^2 + 2xy + y^2 \\ &\leq |x|^2 + 2|x||y| + |y|^2 \\ &= (|x| + |y|)^2 \end{aligned}$$
Since the both sides of the inequality is non-negative, we can take the square root on both sides, thus obtaining $|x + y| \leq |x| + |y|$.
3. Choose x to be $x - y$ in (4) and we obtain $|x| - |y| \leq |x - y|$. Similarly, choosing y to be $y - x$ in (4), we find that $|y| - |x| \leq |y - x| = |x - y|$. Thus we have $||x| - |y|| \leq |x - y|$.

□

2.4.2 The Binomial Theorem

Definition 2.4.2.1: The Binomial Coefficient

Let $n, r \in \mathbb{N}$ with $n > 0$. We define the binomial coefficient $\binom{n}{r}$ to mean the number $\frac{n!}{r!(n-r)!}$ when $r \leq n$. When $r > m$ then $\binom{n}{r} = 0$.

Proposition 2.4.2.2

Let $n, r \in \mathbb{N}$ with $0 < r < n$, we have $\binom{n}{r} = \binom{n}{n-r}$.

Proposition 2.4.2.3

Let $n, r \in \mathbb{N}$ with $0 < r < n$, we have $\binom{n}{r-1} + \binom{n}{r} = \binom{n+1}{r}$.

Theorem 2.4.2.4: The Binomial Theorem

Suppose $a, b \in \mathbb{R}$. Then

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k$$

Theorem 2.4.2.5

[Vandermonde's Theorem] Suppose that $a, b, n \in \mathbb{N}$. Then

$$\binom{a+b}{n} = \sum_{k=0}^{n} \binom{a}{k} \binom{b}{n-k}$$

2.5 Complex Numbers

2.5.1 Introduction to Complex Numbers

Definition 2.5.1.1: Complex Numbers

Define the number $i = \sqrt{-1}$. Define $z = a + bi$ to be a complex number. Every complex number is uniquely determined by an ordered pair $(a, b) \in \mathbb{R}^2$. $a = Re(z)$ is called the real part of z. $b = Im(z)$ is called the imaginary part of z. The set of all complex numbers is denoted $\mathbb{C}$.

Definition 2.5.1.2: Equality

We define the relation equality in $\mathbb{C}$ as $z_1 = z_2$ with $z_1 = a + bi$ and $z_2 = c + di$ if and only if $a = c$ and $b = d$.

Definition 2.5.1.3

Let $z = a + bi$, $w = c = di$. We define the four arithmetic operations in $\mathbb{C}$ as follows.

- $z + w = (a + c) + (b + d)i$
- $z - w = (a - c) + (b - d)i$
- $zw = (ac - bd) + (ad + bc)i$
- $\frac{z}{w} = \frac{(ac+bd)+(bc-ad)i}{a^2+b^2}$

The four operations gives another number in $\mathbb{C}$.

Proposition 2.5.1.4

Let x, y, $z \in \mathbb{C}$.

- (A1) $x + y \in \mathbb{C}$
- (A2) $(x + y) + z = x + (y + z)$
- (A3) $\exists 0 \in \mathbb{C}$ such that $x + 0 = 0 = 0 + x$
- (A4) $x + y = y + x$
- (A5) $\exists(-x) \in \mathbb{C}$ such that $x + (-x) = 0 = (-x) + x$
- (M1) $xy \in \mathbb{C}$
- (M2) $(xy)z = x(yz)$
- (M3) $\exists 1 \in \mathbb{C}$ such that $x \cdot 1 = x = 1 \cdot x$
- (M4) $xy = yx$
- (M5) $\exists(x^{-1}) \in \mathbb{C}$ such that $x(x^{-1}) = 1 = (x^{-1})x$
- (D1) $x(y + z) = xy + xz$

Lemma 2.5.1.5

$\mathbb{R}$ is a subset of $\mathbb{C}$.

Proof. For every real number $a \in \mathbb{R}$ you can associate it with the complex number $a + 0i$. □

Definition 2.5.1.6: Conjugates

For every complex number $z = a + bi$ there exists a conjugate $\bar{z} = a - bi$

Proposition 2.5.1.7

Suppose that $z, w \in \mathbb{C}$.

- $\bar{\bar{z}} = z$
- $z \bar{+} w = \bar{z} + \bar{w}$
- $z\bar{w} = \bar{z}\bar{w}$

2.5.2 Modulus and Argument

Definition 2.5.2.1: Modulus

Define the modulus of $z = a + bi$ to be $|z| = \sqrt{a^2 + b^2}$.

Proposition 2.5.2.2

Suppose that $z, w \in \mathbb{C}$.

- $|z|^2 = |z||\bar{z}|$
- $|\bar{z}| = |z|$
- $|zw| = |z||w|$
- $z\bar{z} = |z|^2$
- $|z + w| \leq |z| + |w|$
- $|z - w| = ||z| - |w||$

Definition 2.5.2.3: Argument

Let $z = a + bi$. Define the argument of z to be the θ such that

$$\cos{(\theta)} = \frac{a}{a^2 + b^2} \text{ and } \sin{(\theta)} = \frac{b}{a^2 + b^2}$$

If $-\pi < argz \leq \pi$, we call it the principal argument of z.

Proposition 2.5.2.4

Suppose that $z, w \in \mathbb{C}$.

- $\arg{(zw)} = \arg{(z)} + \arg{(w)}$
- $\arg{(\frac{z}{w})} = \arg{(z)} - \arg{(w)}$
- $\arg{(\bar{z})} = -\arg{(z)}$

Proposition 2.5.2.5: Polar Form

Using the modulus and the argument, a complex number can be uniquely determined by $|z|$ and $\arg z$. It can be written as $z = r(\cos{(theta)} + i\sin{(\theta)})$ where $r = |z|$ and $\cos{(\theta)} = \frac{a}{a^2+b^2}$ and $\sin{(\theta)} = \frac{b}{a^2+b^2}$ with $-\pi < \theta \leq \pi$ This is the polar form of a complex number.

Proposition 2.5.2.6

Suppose that $z = r(\cos(\theta) + i\sin(\theta))$ and $w = s(\cos(\phi) + i\sin(\phi))$.

- $zw = rs(\cos(\theta + \phi) + i\sin(\theta + \phi))$
- $\frac{1}{z} = \frac{1}{r}(\cos(-\theta) + i\sin(-\theta))$
- $\frac{z}{w} = \frac{r}{s}(\cos(\theta - \phi) + i\sin(\theta - \phi))$
- $\bar{z} = r(\cos(-\theta) + i\sin(-\theta))$

Theorem 2.5.2.7

[De Moivre's Theorem] Suppose that $r \in \mathbb{R}$. Then

$$(\cos\theta + i\sin\theta)^r = \cos(r\theta) + i\sin(r\theta)$$

for any $\theta \in \mathbb{R}$

2.5.3 Exponential Form of Complex Numbers

This part of the complex numbers requires development from the real analysis. In particular, we use the fact that the taylor series for e^x, $cos(x)$, $\sin(x)$ is convergent for all real values of x. It also requires complex analysis to prove that substituting x with a complex number is convergent. Hence we will assume that it is already true for developing the complex numbers.

Definition 2.5.3.1: Exponentiation

$e^{i\theta} = cos\theta + i\sin\theta$

Proposition 2.5.3.2

For every $z \in \mathbb{C}$, $z = re^{i\theta}$.

Proof. The proof is direct from the definition of $e^{i\theta}$. □

Proposition 2.5.3.3

Suppose that $z = re^{i\theta}$ and $w = se^{i\phi}$.

- $zw = rse^{i(\theta+\phi)}$
- $\frac{r}{s} = \frac{r}{s}e^{i(\theta-\phi)}$
- $\bar{z} = re^{-i\theta}$
- $z^n = r^n e^{in\theta}$ where $n \in \mathbb{N}$

Proposition 2.5.3.4

Suppose that $\theta, \phi \in \mathbb{R}$. Then

$$e^{i\phi} = e^{i\theta} \text{ if and only if there exists some } k \in \mathbb{Z} \text{ such that } \phi = \theta + 2\pi k$$

Proposition 2.5.3.5

Suppose that $z = re^{i\theta}$ and $w = se^{i\phi}$. Suppose that $n \in \mathbb{N}$. Then $z^n = w$ if and only if $r = s^{\frac{1}{n}}$ and $\theta = \frac{\phi+2\pi k}{n}$ for some $k \in \mathbb{Z}$.

Theorem 2.5.3.6

[Roots of Unity] Suppose that $z = re^{i\theta}$. Then the nth roots of z are $r^{\frac{1}{n}}e^{\frac{(\theta+2\pi k)i}{n}}$ where $k = 0, 1, \dots, n-1$.

2.6 Algebraic Inequalities

Theorem 2.6.0.1: Bernoulli's Inequality

For all $x \geq -1$ and $n \in \mathbb{N}$,

$$(1+x)^n \geq 1+nx$$

Proof. We prove the inequality by induction on n. In the case of $n=1$, we have $1+x \geq 1+x$, which is true for all x. Now suppose that the inequality works for some $n \in \mathbb{N}$. We have

$$\begin{aligned}(1+x)^{n+1} &\geq (1+x)(1+nx) && \text{(Induction Hypothesis and } x \geq -1\text{)}\\ &= 1+(n+1)x+nx^2 \\ &\geq 1+(n+1)x && \text{(since } x^2 \geq 0\text{)}\end{aligned}$$

Thus we have the Bernoulli's Inequality by the principle of mathematical induction. □

Theorem 2.6.0.2: Weierstrass' Inequality

Let $a_1, \dots, a_n$ be positive numbers. Then when $n \geq 2$,

$$(1+a_1)\dots(1+a_n) > 1+a_1+\dots+a_n$$

Theorem 2.6.0.3: AMGM

Let $a_1, \dots, a_n$ be positive numbers. Then

$$\frac{a_1+\dots+a_n}{n} \geq (a_1a_2\dots a_n)^{\frac{1}{n}}$$

Theorem 2.6.0.4: Cauchy-Schwarz Inequality

Let $x_1, \dots, x_n, y_1 \dots, y_n \in \mathbb{R}$. Then

$$\left(\sum_{k=1}^n x_ky_k\right) \leq \left(\sum_{k=1}^n x_k^2\right)\left(\sum_{k=1}^n y_k^2\right)$$

Theorem 2.6.0.5: Tchbychef's Inequality

Let $x_1, \dots, x_n, y_1 \dots, y_n \in \mathbb{R}$ such that $x_1 \leq x_2 \leq \dots \leq x_n$ and $y_1 \leq y_2 \leq \dots \leq y_n$. Then

$$n\left(\sum_{k=1}^n x_ky_k\right) \geq \left(\sum_{k=1}^n x_k\right)\left(\sum_{k=1}^n y_k\right)$$

2.7 Basics of Matrices

2.7.1 Matrices and its Operations

Definition 2.7.1.1: Matrix

A rectangular array of $m \times n$ real numbers, called the elements, or entries,

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

is called an $m \times n$ matrix over $\mathbb{R}$. For $i = 1, \dots, m$, let

$$r_i = \begin{pmatrix} a_{i1} & a_{i2} & \cdot & a_{in} \end{pmatrix}$$

and for $j = 1, \dots, n$, let

$$c_j = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix}$$

then r_i is called the ith row of A and c_j is called the jth row of A. The element of A at the intersection of the ith row and jth column is called the (i, j)th entry of A. The set of all $m \times n$ matrices over $\mathbb{R}$ is denoted by $M_{m \times n}(\mathbb{R})$. We sometimes denote A as $(a_{i,j})_{m \times n}$

Definition 2.7.1.2: Matrix Addition

Let A, B be $m \times n$ matrices. We define the binary operation $+ : M_{m \times n}(\mathbb{R}) \times M_{m \times n}(\mathbb{R}) \to M_{m \times n}(\mathbb{R})$ to be

$$A + B = \begin{pmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \cdots & a_{1n} + b_{1n} \\ a_{21} + b_{21} & a_{22} + b_{22} & \cdots & a_{2n} + b_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} + b_{m1} & a_{m2} + b_{m2} & \cdots & a_{mn} + b_{mn} \end{pmatrix}$$

Proposition 2.7.1.3

Let $A, B, C \in M_{m \times n}(\mathbb{R})$. Then

- $(A + B) + C = A + (B + C)$
- $(A + B) = (B + A)$
- $A + 0 = 0 + A = A$
- There exists a unique $M \in M_{m \times n}(\mathbb{R})$ such that $A + M = M + A = 0$

Proof. Addition is associative, commutative and has an identity in $\mathbb{R}$. □

Definition 2.7.1.4: Scalar Multiplication

Let $A = (a_{i,j})_{m \times n}$ and $\lambda \in \mathbb{R}$. We define the scalar multiple $\cdot : \mathbb{R} \times M_{m \times n}(\mathbb{R}) \to M_{m \times n}(\mathbb{R})$ as $\lambda A = (\lambda a_{i,j})_{m \times n}$.

Proposition 2.7.1.5

Let $A, B \in M_{m\times n}(\mathbb{R})$ and $\lambda, \mu \in \mathbb{R}$. Then

- $(\lambda\mu)A = \lambda(\mu A)$
- $\lambda(A + B) = \lambda A + \lambda B$
- $(\lambda + \mu)A = \lambda A + \mu A$

Proof. Simple proof by using the definition of scalar multiplication directly. □

Definition 2.7.1.6: Matrix Multiplication

Let $A \in M_{m\times p}(\mathbb{R})$, $B \in M_{p\times n}(\mathbb{R})$. We define matrix multiplication as $\cdot : M_{m\times p}(\mathbb{R}) \times M_{p\times n}(\mathbb{R}) \to M_{m\times n}(\mathbb{R})$ where

$$A \cdot B = (c_{i,j})_{m\times n}$$

with

$$c_{i,j} = \sum_{k=1}^{p} a_{ik}b_{kj}$$

Proposition 2.7.1.7

Let A, B, C be matrices over R, with matrix multiplication assumed possible below.

- $(AB)C = A(BC)$
- $(A + B)C = AB + AC$
- $A(B + C) = AB + AC$

Proof. Once again an easy proof exploiting the definition of matrix multiplication □

Definition 2.7.1.8: Invertible Matrices

A square matrix A is said to be invertible or non-singular if there is a square matrix B such that $AB = BA = I$. In this case B is the inverse of A. A matrix that is not-invertible is a singular matrix.

Theorem 2.7.1.9

If A is invertible, then it has a unique inverse.

Proof. Suppose that A is invertible. Then there exists B such that $AB = I$. Thus the inverse is exactly B. Suppose that C is also an inverse of A. Then $AB = I = AC$. Thus $BAB = BAC$ implies $B = C$. □

Definition 2.7.1.10: Upper Triangular Matrices

A matrix is called upper triangular if all of its entries below the main diagonal are zero.

Definition 2.7.1.11: Diagonal Matrices

A square matrix is said to be a diagonal matrix if $d_{ij} = 0$ whenever $i \neq j$

Definition 2.7.1.12: Transpose

Let $A = (a_{ij})_{m\times n}$. The transpose of A is the $n \times m$ matrix denoted by A^T obtained by interchanging the row and columns of A, that is, $A^T = (a_{ji})_{n\times m}$

2.7.2 Elementary Matrices

Definition 2.7.2.1

[Recombine Matrix] The $n \times n$ recombine matrix is given by

$$R_{i,j,a} = \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & a & & \\ & & & \ddots & & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix}$$

where the diagonals are all 1 and other elements that are not shown is 0.

Definition 2.7.2.2

[Scale Matrix] The $n \times n$ scale matrix is given by

$$R_i(a) = \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & a & & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix}$$

where the diagonals are all 1 except for the i, ith element and other elements that are not shown is 0.

Definition 2.7.2.3

[Transposition Matrix] The $n \times n$ transposition matrix is given by

$$R_{i,j} = \begin{pmatrix} 1 & & & & & & & & \\ & \ddots & & & & & & & \\ & & 0 & & & & 1 & & \\ & & & 1 & & & & & \\ & & & & \ddots & & & & \\ & & & & & 1 & & & \\ & & 1 & & & & 0 & & \\ & & & & & & & \ddots & \\ & & & & & & & & 1 \end{pmatrix}$$

where the diagonals are all 1 except for the i, ith and j, jth element and other elements that are not shown is 0.

Theorem 2.7.2.4

The inverse of the three elementary matrices exists and are also their respective elementary matrices.

Proof. Note that
$$R_{i,j,a}R_{i,j,-a} = I$$
and
$$R_i(a)R_i(a^{-1}) = I$$
and
$$R_{i,j}R_{i,j} = I$$
□

2.7.3 Row Operations

Definition 2.7.3.1

[Row Operations] Let $A_{m\times n}$ with rows $r_1, \dots, r_m$. There are three types of row operations available on A.

- For some $i \neq j$, add a multiple of r_j to r_i
- Interchange r_i and r_j
- Multiply a row by a non-zero scalar

Theorem 2.7.3.2

The three row operations are in fact matrix left multiplications of the elementary matrices. Namely, the recombine matrix, the scale matrix and the transposition matrix corresponds to the above row operations respectively.

Proof. It suffices to check for yourselves that each row operation corresponds to an elementary matrix by simple matrix multiplication. □

Theorem 2.7.3.3

Let $A_{m\times n}$ be a matrix and $P_{m\times m}$ a product of $m \times m$ elementary matrices. Then the equations $Ax = 0_m$ and $(PA)x = 0_m$ has the same solution set.

Proof. Since P is invertible, we have that $Ax = 0_m$ if and only if $(PA)x = 0_m$. Thus they have the same solutions. □

Definition 2.7.3.4

[Upper Echelon Form] A matrix satisfying the below properties is said to be in upper echelon form.

- All zero rows are below all non-zero rows.
- The first non-zero entry of a row is to the right of the first non-zero entry of the row above.
- The first non-zero entry of every row is 1

Definition 2.7.3.5

[Row-reduced Form] We say that a matrix in upper echelon form is in row-reduced form if above and below every first non-zero entry of a row, all entries are 0.

Definition 2.7.3.6

[Reduction Procedure] Let $A = (a_{ij})_{m\times n}$. Start with a_{11}.

1. If a_{ij} and all entries below are 0, move on pivot to the right to $a_{i,j+1}$ and repeat step 1, or terminate if $j = n$
2. If $a_{ij} = 0$ but $(a_k j) \neq 0$ for some $k > i$, apply $R_{i,j}$
3. If $a_{ij} \neq 1$, apply $R_{a_{i,j}^{-1}}$
4. If for any $k \neq i$, $a_{kj} \neq 0$, apply $R_{k,i,-a_{kj}}$
5. If $i = m$ or $j = n$ then terminate, else move pivot to $i + 1, j + 1$ and go back to step 1.

Theorem 2.7.3.7

Every matrix is row equivalent to one and only one matrix in row reduced form.

Proof. The above reduction procedure provides the existence of a row reduced form. We prove by induction on the number of columns of A the uniqueness of the matrix. Let A be an $m - \times n$ matrix. When $n = 1$, there are only two possible row reduced forms. $a_{i1} = 0$ for all $i > 1$, and a_{11} i either 0 or 1. We have $a_{11} = 0$ if and only if the original matrix is the 0 matrix. So any non-zero matrix $m \times 1$ has only one possible row reduced form.

Now suppose that $n > 1$, and the theorem is true for smaller n. Let A' be the $m \times (n - 1)$ matrix obtained by deleting the last column from A. This means that $A = (A'|\mathbf{k})$. By induction the row reduced form of A' is unique. Now if any sequence of row operations that places A into row reduced form, it also places A' into row reduced form, so if B and C are two row reduced form of A, they differ only by the last column. Now since row operations conserve the set of solutions to $A\mathbf{x} = 0$, we have that if $\mathbf{c}$ is the solution, then $B\mathbf{c} = 0$ and $C\mathbf{c} = 0$ and $(B - C)vbc = 0$. Since the first $n - 1$ columns of B and C are the same, if $B \neq C$, we have $B - C$ is of the form $(\mathbf{0}_{m,n-1}|\mathbf{u})$ and $\mathbf{u} \neq 0$. Since the last column is nonzero, there must be at least one element in $\mathbf{u}$ nonzero, meaning there is at least one row in the form $(\mathbf{0}|p)$ for some $p \neq 0$.

Now $(B - C)\mathbf{c} = 0 \implies (\mathbf{0}|\mathbf{u})\mathbf{c} = 0$ and thus $(\mathbf{0}|p) \cdot \mathbf{c} = 0$ and $pc_n = 0$. If $p \neq 0$ then naturally $c_n = 0$ by cancellation law. This implies that there is a leading one in the nth column of B. Because if this is not true, any choice of $\alpha \in \mathbb{R}$ and setting $c_n = \alpha$ could lead to a solution to $B\mathbf{c} = 0$, contradicting the fact that $c_n = 0$.

Now the leading one in the nth column must occur in the first zero row of A' in both B and C. And since other every other entry in the column of a leading one is zero, we finally have that the nth column of B and C is equal. Thus $B = C$. □

2.7.4 Determinants

We borrow notations from group theory.

Definition 2.7.4.1

[Odd Even Permutations] A permutation is said to be even, and to have sign +1 if ϕ is a composition of an even number of transpositions, and ϕ is said to be odd, and to have sign −1 if ϕ is a composition of an odd number of transpositions.

Definition 2.7.4.2

[Determinants] The determinant of a $n \times n$ matrix $A = (a_{ij})$ is the scalar quantity

$$\det(A) = \sum_{\phi \in S_n} \operatorname{sign}(\phi) a_{1\phi(1)} a_{2\phi(2)} \dots a_{n\phi(n)}$$

Lemma 2.7.4.3

$\det(I_n) = 1$.

Proof.

$$\begin{aligned} \det(I_n) &= \sum_{\phi \in S_n} \operatorname{sign}(\phi) a_{1\phi(1)} a_{2\phi(2)} \dots a_{n\phi(n)} \\ &= a_{11} a_{22} \dots a_{nn} \\ &= 1 \end{aligned}$$

□

Proposition 2.7.4.4

If A has two equal rows then $\det(A) = 0$

Proposition 2.7.4.5

Applying elementary row operations does the following to the determinant of a matrix.

- $\det(R_{i,j,a}A) = \det(A)$
- $\det(R_i(a)A) = a\det(A)$
- $\det(R_{i,j}) = -\det(A)$

Proposition 2.7.4.6

If $A = (a_{ij})_{n \times n}$ is upper triangular, then

$$\det(A) = a_{11} a_{22} \dots a_{nn}$$

Proposition 2.7.4.7

Let $A = (a_{ij})_{n \times n}$, $B = (b_{ij})_{n \times n}$. Then

- $\det\left(A^T\right) = \det(A)$
- $\det(AB) = \det(A)\det(B)$

2.7.5 Inverses of Matrices

Definition 2.7.5.1

[Minor] Let $A \in M_{n\times n}(\mathbb{R})$. The minor M_{ij} of the element a_{ij} of A is the determinant of the submatrix obtained by deleting the ith row and the jth column of A.

$$M_{ij} = \begin{vmatrix} a_{1,1} & \cdots & a_{1,j-1} & a_{1,j+1} & \cdots & a_{1,n} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ a_{i-1,1} & \cdots & a_{i-1,j-1} & a_{i-1,j+1} & \cdots & a_{i-1,n} \\ a_{i+1,1} & \cdots & a_{i+1,j-1} & a_{i+1,j+1} & \cdots & a_{i+1,n} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & \cdots & a_{m,j-1} & a_{m,j+1} & \cdots & a_{m,n} \end{vmatrix}$$

Definition 2.7.5.2

[Cofactor] The cofactor of a_{ij} in $A \in M_{n\times n}(\mathbb{R})$ is defined as

$$A_{ij} = (-1)^{i+j} M_{ij}$$

Definition 2.7.5.3

[Adjoint] Let $A \in M_{n\times n}(\mathbb{R})$. The adjoint of A is defined as

$$\text{adj}(A) = (A_{ij})^T_{m\times n}$$

Theorem 2.7.5.4

Let $A \in M_{n\times n}(\mathbb{R})$. Then

$$A(\text{adj}(A)) = (\text{adj}(A))A = \det(A)I$$

Proposition 2.7.5.5

Let $A \in M_{n\times n}(\mathbb{R})$. Then $\det(A) = \sum_{k=1}^{n} a_{ik}A_{ik}$ for $i = 1, 2, 3$ and $\det(A) = \sum_{k=1}^{n} a_{kj}A_{kj}$ for $j = 1, 2, 3$

Theorem 2.7.5.6

[Inverse of a Matrix] Let $A \in M_{n\times n}(\mathbb{R})$. Then A is invertible if and only if $\det(A) \neq 0$. In this case,

$$A^{-1} = \frac{1}{det(A)} \text{adj}(A)$$

Proposition 2.7.5.7

Let $A \in M_{n\times n}(\mathbb{R})$. If A is invertible then $\det\left(A^{-1}\right) = \frac{1}{det(A)}$

Theorem 2.7.5.8

Suppose that A, B are invertible.

- AB is also invertible and $(AB)^{-1} = B^{-1}A^{-1}$
- A^n is also invertible where $n \in \mathbb{N}$ and $(A^n)^{-1} = (A^{-1})^n$
- A^{-1} is also invertible and $(A^{-1})^{-1} = A$
- A^T is also invertible and $(A^T)^{-1} = (A^{-1})^T$

2.7.6 System of Linear Equations

Definition 2.7.6.1

[System of Linear Equations] We say that

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases}$$

is a system of linear equations in n unknowns $x_1, \ldots, x_n$.

2.7.7 System of Linear Equations

Definition 2.7.7.1

olution Sets[] A solution of a system of equations is a list of numbers $x_1, \ldots, x_n$ that makes all of the equations true simultaneously. The solution set of a system of equations is the collection of all solutions.

Definition 2.7.7.2: Consistent Systems

[] We say that the system of linear equations is consistent if it has at least one solution. We say that the system of linear equations is inconsistent if it has no solutions.

Definition 2.7.7.3: Homogenous Systems

[] We say that the system of linear equations is a homogenous system if all the constant coefficients are 0.

Definition 2.7.7.4: Representation of System of Linear Equations

The matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

and

$$B = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$$

can represent a system of linear equations by $Ax = B$ with $x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$. A is said to be the coefficient matrix. X is said to be a solution if it satisfies the above m equations simultaneously.

Part II

The Fundamentals of Analysis

Chapter 3

Real Analysis

3.1 Developing the Real Numbers

In this section, we will give the very first properties of the real numbers for readers to familiarize themselves with so that when we go on to the main discussion of sequences, we are well equipped with theorems at our disposal.

3.1.1 Properties of the Real Numbers

The study of calculus begins with the study of real numbers. In this chapter we will discuss the properties of the real numbers, including its density, inequalities and supermums and infinums.

Let us have a look at binary operations on the real numbers.

Definition 3.1.1.1: Axioms of Reals

$\mathbb{R}$ is an ordered field. Let x, y, $z \in \mathbb{R}$.
($\mathbb{R}$,+) is an abelian group.

- $x + y \in \mathbb{R}$
- $(x + y) + z = x + (y + z)$
- There exists $0 \in \mathbb{R}$ such that $x + 0 = 0 + x = x$
- There exists $-x \in \mathbb{R}$ such that $x + (-x) = (-x) + x = 0$
- $x + y = y + x$

($\mathbb{R}/\{0\}$,$\times$) is an abelian group.

- $xy \in \mathbb{R}$
- $(xy)z = x(yz)$
- There exists $1 \in \mathbb{R}$ such that $x \cdot 1 = x = 1 \cdot x$
- There exists $(x^{-1}) \in \mathbb{R}/\{0\}$ such that $x(x^{-1}) = 1 = (x^{-1})x$
- $xy = yx$

The distributive law holds.

- $x(y + z) = xy + xz$
- $(x + y)z = xz + yz$

The above axioms allow $\mathbb{R}$ to be a field. $\mathbb{R}$ is also an ordered field.

- $x < y$ or $x = y$ or $y < x$

- $x < y$ and $y < z \implies x < z$
- $x < y \implies x + z < y + z$
- $x < y$ and $z > 0 \implies xz < yz$

Although technically these are not axioms and these properties can be deduced from the dedekind definition of real numbers, we will take these for granted in order to facilitate further theorems and definitions.

The absolute value is an important function when it comes to defining useful concepts such as distances in the field of real. It simply measures the shortest distance between two points.

Definition 3.1.1.2: The Absolute Value

The absolute value of a real number $x \in \mathbb{R}$ is defined by

$$|x| = \begin{cases} x & \text{if } x \geq 0 \\ -x & \text{if } x \leq 0 \end{cases}$$

Readers should already be familiar with the absolute value in high school. In particular, one should be able to draw graphs related to the absolute function. However, there are properties of the absolute value that are yet to be seen or proved. The following properties will be a list of its useful properties.

Proposition 3.1.1.3

Let $x, y \in \mathbb{R}$. Then the following are true of the absolute value.

- $|x| \geq 0$ with equality if and only if $x = 0$
- $|xy| = |x||y|$
- $|x + y| \leq |x| + |y|$
- $||x| - |y|| \leq |x - y|$

Proof. I left out the proofs of the first and second property since they are simplay obtained via case by case analysis.

- We start by squaring the left hand side of the inequality.

$$\begin{aligned} |x+y|^2 &= (x+y)^2 \\ &= x^2 + 2xy + y^2 \\ &\leq |x|^2 + 2|x||y| + |y|^2 \\ &= (|x| + |y|)^2 \end{aligned}$$

  Since the both sides of the inequality is non-negative, we can take the square root on both sides, thus obtaining $|x + y| \leq |x| + |y|$.

- Choose x to be $x - y$ in (4) and we obtain $|x| - |y| \leq |x - y|$. Similarly, choosing y to be $y - x$ in (4), we find that $|y| - |x| \leq |y - x| = |x - y|$. Thus we have $||x| - |y|| \leq |x - y|$.

□

These properties will also be used extensively throguhout the entire notes. The reader should be absolutely familiar with its properties so that they can understand the definitions and the proofs more smoothly. The geometric interpretation of the absolute value will serve as an extremely important visualization on limits and convergences.

Often we use intervals for our domain of real valued functions. We will see a definition of it below.

Definition 3.1.1.4: Intervals

An interval in $\mathbb{R}$ is a non-empty set I with the property that $a, b \in I$ and $a < x < b \implies x \in I$. An interval containing more than one point is called non-degenerate. We use the following notations to describe intervals.

- $(a, b) = \{x \in \mathbb{R} | a < x < b\}$
- $(a, b] = \{x \in \mathbb{R} | a < x \leq b\}$
- $[a, b) = \{x \in \mathbb{R} | a \leq x < b\}$
- $[a, b] = \{x \in \mathbb{R} | a \leq x \leq b\}$

The important to take away from this definition is that it is sort of continuous. Although we have not fully developed the notion of continuity yet, you could see that they are connected in the sense that a slight perturbation of x in an interval I, it will stay in the interval.

3.1.2 The Completeness Axiom

What is special about the real numbers from most of the other number systems is the existence of the completeness axiom that says a bounded subset must have a least upper bound. But before we reach our main result, we develop the notion of supremum and infinum which is especially important when considering convergences and continuity in future chapters.

Let us give a proper definition for bounds.

Definition 3.1.2.1: Upper and Lower Bounds

Let S be a non empty subset of $\mathbb{R}$

- S is said to be bounded above if there exists $u \in \mathbb{R}$, called an upper bound of S such that $x \leq u$ for all $x \in S$
- S is said to be bounded below if there exists $l \in \mathbb{R}$, called a lower bound of S such that $l \leq x$ for all $x \in S$
- S is said to be bounded if there exists a $M \in \mathbb{R}$ such that $|x| \leq M$ for all $x \in S$

Here is a trivial proposition involving bounds to exercise your minds.

Proposition 3.1.2.2

A non-empty subset S of $\mathbb{R}$ is bounded if and only if it is bounded above and bounded below.

Proof. If S is bounded, then $-M \leq x \leq M$ thus it is bounded above and below. If S is bounded above by M and bounded below by N, then $|x| \leq \max(|M|, |N|)$. □

The notion of supremum and infinum is closely tied to that of boundedness. In particular, there is only one number in $\mathbb{R}$ that can be called the supremum per set, similarly for the infinum. But for boundedness, you can take any number in $\mathbb{R}$. As long as its absolute value is big enough to cover up for all elements in S, it could be a bound. So if $M > 0$ is a bound for a set S then naturally $M + 1$, $1.1M$ and more would also be bounds for S.

Definition 3.1.2.3: Supremum

Let S be a subset of $\mathbb{R}$. We say that $U \in \mathbb{R}$ is the supremum of S when

- $x \leq U$ for all $x \in S$
- If u is an upper bound of S then $U \leq u$

We denote the supremum of S as $\sup(S) = U$

Definition 3.1.2.4: Infinum

Let S be a non-empty subset of $\mathbb{R}$. We say that $L \in \mathbb{R}$ is the infinum of S when

- $L \leq x$ for all $x \in S$
- If l is an upper bound of S then $l \leq L$

We denote the infinum of S as $\inf(S) = L$

Readers can easily check that the supremum and infinum, should it exists, is unique. One can also think about the criteria for the supremum and infinum to exists. Naturally, since the notion of supremum and infinum involves comparison, there will be no such thing in complex numbers.

Theorem 3.1.2.5

Let S be a non-empty subset of $\mathbb{R}$. Then

$$-\sup(S) = \inf(-S)$$

where $-S = \{-x | x \in S\}$.

Proof. Since for all $x \in S$ $x \leq \sup(S)$. But this implies that $-\sup(S) \leq -x$ for all $x \in S$ thus for all $-x \in -S$. Then by definition $-\sup(S) = \inf(-S)$. □

The above theorem allows us to translate properties of the supremum to properties of the infinum by simply "inverting" the set. That way to prove things about the infinum we could invert the set, use the proof by the supremum and invert it back to prove it for the infinum. In practise however, we barely use this theorem to prove anything.

The following property is an important characterization of supremums and infinums that will be used later when we need it. It is called the approximation property because it allows some sort of wiggle room between the supremum and our approximation in the sense that we can always make a better approximation for the supremum.

Theorem 3.1.2.6: Approximation Property

Let S be a non-empty subset of $\mathbb{R}$.

- If $\sup(S)$ exists then for all $\epsilon > 0$ there exists $x \in S$ such that $\sup(S) - \epsilon < x \leq sup(S)$.
- If $\inf(S)$ exists then for all $\epsilon > 0$ there exists $x \in S$ such that $\inf(S) \leq x < inf(S) + \epsilon$.

Proof. Notice that since $\sup(S)$ is the supremum of S, we have that $\sup(S) - \epsilon$ is not the supremum of S Thus there exists some element $x \in S$ such that $\sup(S) - \epsilon < x$. Mirror the proof for the infinum version. □

We finally reach the goal of this chapter, which is the completeness axiom. This is called an axiom because it is inherent in the construction of the real numbers. Obviously one can prove it from the very definition of the real numbers, which is just dedekind cuts. However since we are less interested in the formation of the real numbers, we take this completeness as granted and treat it as axiom rather than the theroem.

Axiom 3.1.2.7: Completeness Axiom

Every non-empty subset of $\mathbb{R}$ that is bounded above has a least upper bound.

It is only the real number field that consists of the completeness axiom. This means that there are no gaps in the number line when the real numbers are introduced. Once again we set foot to a seemingly intuitive and straight forward theorem that in fact has a lot of use.

3.1.3 Density of the Real Numbers

In this chapter we will investigate another important property of the real numbers, albeit shared by the complex numbers as well. This is the density of the real numbers. We try to mathematically prove and formulate the fact that the real numbers are very rich. It is dense in the sense that for every two real numbers, you can always find another real number in between. Obviously if you simply consider the rationals or the irrationals, this is also true. But the main thing to take away here is that there would be an uncountably amount of rationals and irrationals each, between any two real numbers. This distribution can not be measured as well. The rationals and the irrationals are so interspersed and intertwined that there is no notion of "next" number in the reals. And you would not even know whether it is rational or irrational.

To start off this chapter, we need an important where it use throughout these notes will only be of this section.

Theorem 3.1.3.1: Archimedean Property

For any real numbers a and b with $a > 0$ there exists $n \in \mathbb{N}$ such that $na > b$.

Proof. Suppose that $na \leq b$ for all n. Then the set of all na, denoted by S has an uppper bound b, thus also has a least upeer bound, say u. We have that there exists some $na \in S$ such that $u - a < na \leq u$. But this implies that $u < (n+1)a \in S$, a contradiction. □

The Archimedean property, as stated by the name, is given by Archimedes. While the theorem does look somewhat trivial, often it is the trivial theorems that are the hardest to prove, especially by new students. This is also quite a standard example for proof of contradiction.

Before we move on, we also need the floor function in our proof of density, which is formalized below.

Theorem 3.1.3.2: Floor Function

For each $x \in \mathbb{R}$ there exists a unique integer $\lfloor x \rfloor$ such that

$$\lfloor x \rfloor \leq x < \lfloor x \rfloor + 1$$

Proof. Let $A = \{n \in \mathbb{Z} : n \leq x\}$. By the archimedean property there exists $k > -x$ and thus $-k < x$ with $-k \in A$ hence A is non-empty. Let $a = sup(A)$. We have some $n \in A$ such that $a - 1 < n \leq a \leq x$ and thus $a + 1 < n + 1 \leq a + 1 \leq x + 1$. From this we have $n \leq x$ as well as $n + 1$ not in A thus $x < n + 1$.

Now suppose that $m \leq x < m + 1$ and $n \leq x < n + 1$. From these we derive that $m \leq x < n + 1$ and $n \leq x < m + 1$. Thus $m - n < 1$ and $m - n > -1$ and $m - n \in \mathbb{Z}$. and we have $m - n = 0$ □

Now who said that the irrational numbers are not just some fantasy of a mathematician. Let us give birth to the first ever irrational mankind every thought of! Pythagoras and his disciples were quite confused by this phenomenon that they deemed their mathematics must be wrong somewhere.

Lemma 3.1.3.3

The positive solution to the equation $x^2 - 2 = 0$, denoted $\sqrt{2}$ is irrational.

Proof. Suppose that $\sqrt{2}$ is rational, then it can be represented as $\frac{m}{n}$ with $hcf(m, n) = 1$. Then we have $m^2 = 2n^2$, implying that $2|m^2$ and $2|m$. Thus we can say that $m = 2k$ for some $k \in \mathbb{Z}$. Substituting $m = 2k$, we have $2k^2 = n^2$ which similarly implies that $2|n$ which is a contradiction since they have a common factor. □

Congratulations we showed that irrational numbers do exist! Finally we can move on to the core of this section, the density theorems.

Theorem 3.1.3.4: Density of the Rationals

Between any pair of distinct real numbers there is a rational number.

Proof. Suppose $a, b \in \mathbb{R}$ and $a < b$. By the archimedean property we have that $n(b - a) > 1$ for some $n \in \mathbb{N}$. Let $m = \lfloor na \rfloor + 1$. We have $na < m \leq na + 1 < nb$. Thus $a < \frac{m}{n} < b$. □

Theorem 3.1.3.5: Density of the Irrationals

Between any pair of distinct real numbers there is an irrational number.

Proof. Suppose $a, b \in \mathbb{R}$ and $a < b$. By the density of the rationals we have that $\frac{a}{\sqrt{2}} < r < \frac{b}{\sqrt{2}}$ for some $r \in \mathbb{Q}$ thus $a < r\sqrt{2} < b$. □

Once we have the density in the rationals and irrationals, we now have an infinite number of rationals and irrationals betweena any two real numbers.

Theorem 3.1.3.6

Between any pair of distinct real numbers there is an infinite number of rationals and irrationals.

Proof. Recursively apply density of rationals and density of irrationals. □

Beware that there is no pattern as to how the rationals and irrationals line up in the real number line. It is unknown whether a rational or an irrational ”follows” a real number. In fact, it may not be possible to define what ”the next number” of a real number is.

3.2 Sequences

Sequences have not only proven themselves to be useful in real analysis, but also in several other areas of mathematics as well including complex analysis and topology and metric spaces (they are all just analysis) and differential equations and more. They are the intermediate step for the notion of continuity and differentiability since those definitions can be reduced by an equivalence characterization via sequences. Therefore it is important to investigate properties of the sequences to better understand and prove theorems of continuity and limits.

3.2.1 Sequences and Convergences

We begin our discussion we the definition and the first properties of sequences, as well as convergence.

Definition 3.2.1.1: Sequences

A sequence in $\mathbb{R}$ is a function $f : \mathbb{N} \to \mathbb{R}$ that assigns to each natural number a real number. We write often write $f(n) = a_n$ and say that it is the nth term of the sequence. We use $(a_n)_{n\in\mathbb{N}}$ to represent a sequence indexed by n.

Real analysis is all about closeness of things. The definition of convergence below will be one such. It is important here to not only understand the geometric meaning of convergence, which is closeness towards the limit, but to also be able to apply the definition to prove convergence. While we will see later that there are more ways to prove convergences, when all else fails, we must return to this definition and therefore we should be well trained with its definition.

Definition 3.2.1.2: Covergence of Sequences

A sequence $(a_n)_{n\in\mathbb{N}}$ tends to a if and only if for every $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that

$$n > N \implies |a_n - a| < \epsilon$$

We write $a_n \to a$ in this case.

While generally not true in other spaces, one neat property of the real numbers is that we know that if it converges, it will only converge to exactly one number. While we will barely use this theorem anywhere, it is important to know that this is true and the reasoning behind it. This will be very first theorem involving the notion of convergence and therefore it is helpful for us to truely understand what convergence means.

Theorem 3.2.1.3: Uniqueness of Limits

A sequence in $\mathbb{R}$ cannot converge to more than one limit.

Proof. Suppose that $a_n \to a$ and $a_n \to b$. Choose ϵ to be $\frac{|a-b|}{10}$. Then there exists N_1 and N_2 such that $|a_n - a|\epsilon$ for all $n > N_1$ and $|a_n - b| < \epsilon$ for all $n > N_2$ respectively. When $n > \max(N_1, N_2)$ we have both inequalies hold together. Then

$$\begin{aligned} |a-b| &\leq |a_n - a| + |a_n - b| \\ &\leq \frac{|a-b|}{10} + \frac{|a-b|}{10} \\ &= \frac{|a-b|}{5} \end{aligned}$$

Which is a contradiction. □

The main idea here is that since convergence means closeness, if we have two numbers to converge to then there will be two closeness. But then numbers between these two closeness will not be close to each other anymore!

Another important property of convergence sequecnes is that it is bounded.

Proposition 3.2.1.4

A convergent sequence in $\mathbb{R}$ is bounded.

Proof. Suppose that $a_n \to a$. Fixing ϵ to be any number larger than 0, we have that $|a_n - a| < \epsilon \implies a - \epsilon < a_n < a + \epsilon$ for all n larger than some number N. At the same time, all the terms less than N are bounded by $M = \max(|a_1|, |a_2|, \dots, |a_N|)$. Thus if we take

$$C = \max(M, |a + \epsilon|, |a - \epsilon|)$$

C would bound all the terms of a_n. □

The next two theorems demonstrate inequalities between convergent sequences.

Theorem 3.2.1.5

Let $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ be sequences with $a_n \to a$ and $b_n \to b$. If $\exists N \in \mathbb{N}$ such that $a_n \leq b_n$ for all $n > N$, then $a \leq b$.

Proof. Suppose that $b < a$. Then we have $b < \frac{a+b}{2} < a$. Choosing $\epsilon = \frac{a-b}{2}$, then whenever $n > N_1$, we have

$$b_n - b < \frac{a-b}{2}$$

Similarly, choose $\epsilon = \frac{b-a}{2}$, then whenever $n > N_2$, we have

$$a_n - a < \frac{b-a}{2}$$

Then we have

$$b_n < \frac{a+b}{2} < a_n$$

which is a contradiction. □

Theorem 3.2.1.6: Sandwich Theorem

Let $(a_n)_{n\in\mathbb{N}}$, $(b_n)_{n\in\mathbb{N}}$, $(c_n)_{n\in\mathbb{N}}$ be sequences in $\mathbb{R}$ such that

$$a_n \leq b_n \leq c_n$$

for all n larger than some $N \in \mathbb{N}$. If $a_n \to L$ and $c_n \to L$ then $b_n \to L$.

Proof. Suppose $a_n \to L$ and $c_n \to L$. Then choose an ϵ that is greater than 0. Then there exists N_1 and N_2 such that $|a_n - L| < \epsilon$ whenever $n > N_1$, and $|c_n - L| < \epsilon$ whenever $n > N_2$. Then whenever $N = \max(N_1, N - 2)$, we have $-\epsilon < a_n - L < b_n - L < c_n - L < \epsilon$, and thus we have $|b_n - L| < \epsilon$ whenever $n > N$. □

The sandwich theorem is useful in detecting the convergence to 0 of a positive sequence. Whenever one is given a very complicated positive sequence that does not have an easily determinable limit, one can use a sequence with larger values that tends to 0 show that that sequence converges to 0. This however requires more care and exercise to get familiarize.

As one accummulate more examples, it can be made clear that even divergence has different kinds. This is made explicitly true when considering the sequence $a_n = \sin(n)$ and $b_n = n$. One oscillates while the other blows up. We therefore give a definition of blowing up in technicallity.

Definition 3.2.1.7: Divergence to Infinity

A sequence $(a_n)_{n\in\mathbb{N}}$ in $\mathbb{R}$ is said to

- diverge to ∞ if for every real number $C > 0$ there exists a number $N \in \mathbb{N}$ such that
$$n > N \implies a_n > C$$
- diverge to $-\infty$ if for every real number $C < 0$ there exists a number $N \in \mathbb{N}$ such that
$$n > N \implies a_n < C$$

This definition should be much easier to swallow than that of convergence sequences. In practise it is also not hard to prove sequences that tends to infinity. Usually showing that a limit convergences requires you to find out the limit explicitly (except that if you are sandwiching) and that is often harder than simply showing that a sequence blows up.

But in case it is not obvious to know that it diverges to infinity, we can apply a comparison test as stated below.

Proposition 3.2.1.8

Let $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ be sequences in $\mathbb{R}$. If $b_n \geq a_n$ for all n larger than some $N \in \mathbb{N}$ and $a_n \to \infty$ then $b_n \to \infty$.

Proof. Suppose $a_n \to \infty$. Then we have for every $C > 0$ there exists some $N \in \mathbb{N}$ such that $a_n > C$ whenever $n > N$. Then we have $b_n > a_n > C$ for all $n > N$ thus $b_n \to \infty$. □

This proposition bears similarity to that of the convergence version. Often the hardest part of using these two comparisons is to find an appropriate sequence to demonstrate the limit of the main sequence.

The collection of all convergent sequences also works similarly to $\mathbb{R}$ in the sense that aside from comparison, there is also a notion of addition, scalar multplication and sequence multiplication. However the main usage of this proposition is not to construct new sequences from old, but to evaluate limits from know sequences.

Proposition 3.2.1.9: Algebra of Sequences

Let $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ be sequences in $\mathbb{R}$. Let $a_n \to a$ and $b_n \to b$. Then the following are true.

- Sum Rule: $sa_n + tb_n \to sa + tb$ for any $s, t \in \mathbb{R}$
- Product Rule: $a_n b_n \to ab$
- Quotient Rule: $\frac{a_n}{b_n} \to \frac{a}{b}$ if $b \neq 0$
- $|a_n| \to |a|$

Proof. Suppose that $(a_n) \to a$ and $(b_n) \to b$

- Choosing any $\frac{\epsilon}{|s|} > 0$, we have that whenever $n > N_1$ for some N_1, $|a_n - a| < \frac{\epsilon}{|s|}$, thus $|sa_n - sa| < \epsilon$, thus $(sa_n) \to sa$. Similarly, $(tb_n) \to tb$. Now for every $\frac{\epsilon}{2} > 0$, we have that whenever $n > \max(N_1, N_2)$, $|sa_n - sa|$ and $|tb_n - tb|$ whenever $n > \max(N_1, N_2)$. Then
$$\begin{aligned}|(sa_n + tb_n) - (sa + tb)| &\leq |sa_n - sa| + |tb_n - tb|\\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2}\\ &= \epsilon\end{aligned}$$

Thus we have the desired result.

- Choose $M \geq |a|$ such that $|b_n| \leq M$ for all n. This is possible since both sequecnes are bounded. Then there exists $N \in \mathbb{N}$ we have $|a_n - a| < \frac{\epsilon}{2M}$ and $|b_n - b| < \frac{\epsilon}{2M}$ for all $n > N$. We then have

$$\begin{aligned} |a_n b_n - ab| &= |(a_n - a)b_n + a(b_n - b)| \\ &\leq |a_n - a||b_n| + |a||b_n - b| \\ &\leq M|a_n - a| + M|b_n - b| \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2} \\ &= \epsilon \end{aligned}$$

Thus we have the desired result.

- We want to show that $\frac{1}{b_n} \to \frac{1}{b}$ since we can apply the product rule to produce the quotient rule. From the product rule we have that $bb_n \to b^2$. Choosing $\epsilon = \frac{b^2}{2} > 0$, we have a fixed N such that $\left|bb_n - b^2\right| < \frac{b^2}{2}$ whenever $n > N$. Thus we have $\frac{b^2}{2} < bb_n$ whenever $n > N$. Then we have

$$\begin{aligned} \frac{b^2}{2} < |bb_n| &\implies \frac{1}{|bb_n|} < \frac{2}{b^2} \\ &\implies \left|\frac{b_n - b}{bb_n}\right| < \frac{2|b_n - b|}{b^2} \\ &\implies \left|\frac{1}{b_n} - \frac{1}{b}\right| < \frac{2|b_n - b|}{b^2} \end{aligned}$$

Since we have $0 < \left|\frac{1}{b_n} - \frac{1}{b}\right| < \frac{2|b_n - b|}{b^2}$ then by sandwich theorem, we have that $\frac{1}{b_n} \to \frac{1}{b}$. Then by product rule, we have the desired result.

- For every $\epsilon > 0$, there exists some N such that $||a_n| - |a|| \leq |a_n - a| < \epsilon$ whenever $n > N$. Thus trivially $|a_n| \to a$.

□

We will return to this property more often than you think as we will often go back to sequences when further defining new notions such as continuity and differentiability. This is why developing the notion of sequences would be useful.

Next up we also have the ability to invert sequences that tends to 0 and infinity, given sufficient conditions.

Proposition 3.2.1.10

The following two statements are true about a sequence $(a_n)_{n\in\mathbb{N}}$.

- If $a_n \to \infty$ then $\frac{1}{a_n} \to 0$
- If $a_n \to 0$ and there exists $N \in \mathbb{N}$ such that $a_n > 0$ for all $n > N$, then $\frac{1}{a_n} \to \infty$

Proof. We prove the following with care.

- Since $a_n \to \infty$, we have for every $C > 0$ there exists some N such that $a_n > C$ whenever $n > N$. Then we have for every $\epsilon = \frac{1}{C} > 0$ there exists N such that $\left|\frac{1}{a_n}\right| < \frac{1}{C} = \epsilon$. Thus we have the desired result.

- We have that for every $\epsilon > 0$, there exists some N_1 such that $|a_n| < \epsilon$ whenever $n > N_1$, also we have $a_n > 0$ when $n > N_2$. Choosing $C = \frac{1}{\epsilon}$, then for all $n > \max(N_1, N_2)$, we have $|a_n| < \epsilon \implies \left|\frac{1}{\epsilon}\right| > \frac{1}{\epsilon} = C$ Thus we have the desired result.

□

Finally in this exceptionally long section, we will give special names for sequences that looks more timid. Sequences that only goes up or down will have special names, and then we show an application of the completeness axiom.

Definition 3.2.1.11: Categorization of Sequences

A sequence (a_n) in $\mathbb{R}$ may have different properties as below.

- $(a_n)_{n\in\mathbb{N}}$ is strictly increasing if $a_n < a_{n+1}$ for all $n \in \mathbb{N}$
- $(a_n)_{n\in\mathbb{N}}$ is increasing if $a_n \leq a_{n+1}$ for all $n \in \mathbb{N}$
- $(a_n)_{n\in\mathbb{N}}$ is strictly decreasing if $a_n > a_{n+1}$ for all $n \in \mathbb{N}$
- $(a_n)_{n\in\mathbb{N}}$ is decreasing if $a_n \geq a_{n+1}$ for all $n \in \mathbb{N}$
- $(a_n)_{n\in\mathbb{N}}$ is monotone if it is either increasing or decreasing

Theorem 3.2.1.12: Monotone Sequence Theorem

Every bounded monotone sequence converges in $\mathbb{R}$.

Proof. Suppose $(a_n)_{n\in\mathbb{N}}$ is increasing and bounded. Then $S = \{a_n | n \in \mathbb{N}\}$ is bounded thus has a supremum. We have that

$$\sup(S) - \epsilon < a_N \leq a_{N+1} \leq \cdots \leq \sup(S) < \sup(S) + \epsilon$$

for all n larger than some $N \in \mathbb{N}$. Thus we have the for all $n > N$,

$$\sup(S) - \epsilon < a_n < \sup(S) + \epsilon \implies |a_n - \sup(S)| < \epsilon$$

The proof is similar for the decreasing version. □

The monotone sequence theorem is in fact equivalent to the completeness axiom. However we will not pursue this notion further as it does not deem particularly useful.

3.2.2 Subsequences

Subsequences, as the name suggests, are new sequences extracted from old sequences. Since they came from another sequences, some properties of the old sequence should retain, as you will see. Of all the subsequences, the most important would be the limit inferior and limit superior.

Definition 3.2.2.1: Subsequences

A subsequence of a sequence $(a_n)_{n\in\mathbb{N}}$ in $\mathbb{R}$ is a sequence $(a_{n_k})_{k\in\mathbb{N}}$, where

$$0 \leq n_1 < n_2 < \dots$$

An immediate property of subsequences is the following.

Proposition 3.2.2.2

Let $(a_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}$ and let $a_n \to a$. Then any subsequence of $(a_{n_k})_{k\in\mathbb{N}}$ converges to a.

Proof. Given $\epsilon > 0$ there exists some N such that $|a_n - a| < \epsilon$ whenever $n > N$. Since $n_k > k$, we have that $|a_{n_k} - a| < \epsilon$ for all $k > N$, thus we have $a_{n_k} \to a$ □

There is in fact an inverse saying that if every subsequence convergesn to the same number, then the sequence also converges to that number. However it is impractical since I doubt there are any methods that can find out what all the subsequences converge to, all without knowing the limit of the original sequence.

Next we have a very important theorem that constructs very useful subsequences.

Theorem 3.2.2.3: Bolzano-Weierstrass Theorem

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

Proof. Let $(a_n)_{n\in\mathbb{N}}$ be a bounded sequence, say $c_0 \leq a_n \leq d_0$ for all n. Bisect the interval $I_0 = [c_0, d_0]$. Observe that if the union of two sets contains infinitely many terms, than at least one of the sets must contain infinitely many terms of the sequence.

Suppose I_1 is the said set. From I_1, choose one such term, say a_{n_1}. Bisect I_1 again and call I_2 the interval with infinitely many terms of the sequence. Choose one such term say a_{n_2}, with $n_2 > n_1$, by repeating this procedure, we obtain a subsequence (a_{n_k}) and a sequence of intervals $I_k = [c_k, d_k]$ such that

$$c_0 \leq c_{k-1} \leq c_k \leq a_{n_k} \leq d_k \leq d_{k-1} \leq d_0$$

and $d_{k+1} - c_{k+1} = \frac{1}{2}(d_k - c_k)$.

Observe that (c_k) and (d_k) are monotonic and bounded, thus converges to c and d respectively. Since

$$d_k - c_k = 2^{-k}(d_0 - c_0) \to 0$$

we have that $c = d$. By the sandwich theorem, we have $a_{n_k} \to c$ □

As you see, the monotone convergence theorem is applied thus the completeness axiom actually implies the Bolzano-Weierstrass theorem. Once again, the converse is also true but is too much of a nuisance to prove, given that there is virtually no good application. However, the Bolzano-Weierstrass theorem will prove itself extremely useful in future proofs. It is one of the rare theorems, that returns you a sequence will useful properties.

Finally, we have the two important subsequences.

Definition 3.2.2.4: Limit Superior and Limit Inferior

Let $(a_n)_{n\in\mathbb{N}}$ be a sequence. Define the limit superior of the sequence to be the the limit of

$$s_n = \sup_{m\geq n} a_m$$

if it exists.

Define the limit inferior of the sequence to be the limit of

$$t_n = \inf_{m\geq n} a_m$$

if it exists.

They will not serve much of a purpose except to prove a test that involves series later. However, I will collect a few of its properties here. There are much nicer properties involving inequalities but is rarely useful unless you are taking an exam.

Proposition 3.2.2.5

Let $(a_n)_{n\in\mathbb{N}}$ be a bounded sequence. Then s_n and t_n converges.

Proof. Trivially we must have s_n a decreasing sequence and t_n an increasing sequence since $\{a_m|m \geq n_2\} \subset \{a_m|m \geq n_1\}$ as long as $n_1 < n_2$. Since $(a_n)_{n\in\mathbb{N}}$ is bounded, so is the subsequences s_n and t_n. Thus s_n and t_n is convergent by the monotone convergent theorem. □

Finally the below proposition is a characterization of limit superior and limit inferior. This would be quite useful if combined with the monotone convergence theorem.

Proposition 3.2.2.6

Let $(a_n)_{n\in\mathbb{N}}$ be a sequence. Then $a_n \to a$ if and only if

$$\lim_{n\to\infty} \sup_{m\geq n} a_m = \lim_{n\to\infty} \inf_{m\geq n} a_m = a$$

Proof. Firstly suppose that $a_n \to a$. Then since $(a_n)_{n\in\mathbb{N}}$ is convergent it is bounded. Then by the above theorem the subsequences converges. By theorem 2.2.2 we must have the limit suprerior and limit inferior converge to a.

Now suppose that the limit superior and limit inferior both converges to $a \in \mathbb{R}$. Then this means that fixing $\epsilon > 0$, there exists $N_1 \in \mathbb{N}$ such that $a - \epsilon < \sup_{m\geq n} a_m < a + \epsilon$ for all $n > N_1$. Similarly there exists N_2 such that $a - \epsilon < \inf_{m\geq n} a_m < a + \epsilon$ for all $n > N_2$. This means that

$$a - \epsilon < \inf_{m\geq n} a_m < a_m < \sup_{m\geq n} a_m < a + \epsilon$$

for all $m \geq n > \max N_1, N_2$. Thus $a_n \to a$ and we are done. □

3.2.3 Cauchy Sequences

Definition 3.2.3.1: Cauchy Sequence

A sequence $(a_n)_{n\in\mathbb{N}}$ is said to be Cauchy if for every $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that

$$n, m > N \implies |a_n - a_m| < \epsilon$$

Proposition 3.2.3.2

Every convergent sequence is Cauchy.

Proof. Suppose that $(a_n)_{n\in\mathbb{N}}$ is convergent. Choose $\frac{\epsilon}{2} > 0$, then there exists N such that $|a_n - a| < \frac{\epsilon}{2}$ and $|a_m - a| < \frac{\epsilon}{2}$ whenever $n, m > N$. Then

$$\begin{aligned} |a_n - a_m| &\leq |a_n - a| + |a_m - a| \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2} \\ &= \epsilon \end{aligned}$$

Thus we have the desired result. □

Proposition 3.2.3.3

Every Cauchy Sequence is bounded.

Proof. Suppose that $(a_n)_{n\in\mathbb{N}}$ is cauchy. Then fixing $\epsilon > 0$ there exists some N such that $|a_n - a_m| < \epsilon$ whenever $n, m > N$. Then we have

$$\begin{aligned} |a_n| &\leq |a_n - a_N| + |a_N| \\ &< |a_N| + \epsilon \end{aligned}$$

whenever $n > N$. Thus we have that a_n is bounded whenever $n > N$. For $n \leq N$, it is bounded by $M = \max(a_1, a_2, \dots, a_N)$. So by taking the max of M and $a_N + \epsilon$, we have that all terms in the sequence are bounded. □

Proposition 3.2.3.4

Every Cauchy sequence is convergent.

Proof. Since a cauchy sequence is bounded, we have from the Bolzano Weierstrass theorem that there is subsequence, say $(a_{n_k})_{k\in\mathbb{N}}$ that converges to say, a. Then fixing $\frac{\epsilon}{2} > 0$, there exists some $N_1 \in \mathbb{N}$ such that $|a_{n_k} - a| < \frac{\epsilon}{2}$ whenever $k > N_1$. Also since it is cauchy, we have that when we have $\frac{\epsilon}{2} > 0$ there exists some $N_2 \in \mathbb{N}$ such that $|a_n - a_{n_k}| < \frac{\epsilon}{2}$ whenever $n, k > N_2$. Choosing $N = \max(N_1, N_2)$, then we have

$$\begin{aligned} |a_n - a| &\leq |a_n - a_{n_k}| + |a_{n_k} - a| \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2} \\ &= \epsilon \end{aligned}$$

Thus we have the desired result. □

Theorem 3.2.3.5: Geometric Sequences

$$x^n \to \begin{cases} \infty & \text{if } x > 1 \\ 1 & \text{if } x = 1 \\ 0 & \text{if } -1 < x < 1 \\ \text{diverges} & \text{if } x \leq -1 \end{cases}$$

Proof. We treat different cases separately.

- We have $x^n \geq (1 + n(x-1))$ by the Bernoulli's Inequality. Since $(1 + n(x-1))$ diverges to $+\infty$, we also have that (x^n) diverges to $+\infty$.
- Now when $x = 1$ we have the sequence $1, 1, 1, \dots$ thus it converges to 1.
- If $|x| < 1$, let $x = \frac{1}{1+t}$ with $t > 0$. Then by Bernoulli's Inequality, we have $(1+t)^n \geq 1 + nt > nt$, thus $\frac{1}{(1+t)^n} \leq \frac{1}{1+nt} < \frac{1}{nt}$. But since $nt \to \infty$, we have that $\frac{1}{nt} \to 0$. Then for every $\epsilon > 0$ there exists N such that $\left|\frac{1}{nt}\right| < \epsilon$ whenever $n > N$. Using this, we have $|x^n| = \left|\frac{1}{(1+t)^n}\right| < \epsilon$. Thus we have the desired result.
- It is easy to show that it neither goes to $\pm\infty$ nor converges.

□

Definition 3.2.3.6: Strictly Contracting

A sequence $(a_n)_{n\in\mathbb{N}}$ is strictly contracting if $|a_{n+2} - a_{n+1}| \leq l|a_{n+1} - a_n|$ for all $n \in \mathbb{N}$ where $l \in (0, 1)$

Theorem 3.2.3.7

Every strictly contracting sequence is Cauchy.

Proof. Suppose that $(a_n)_{n\in\mathbb{N}}$ is a strictly contracting sequence. Without loss of generality suppose $n > m$.

$$\begin{aligned}|a_n - a_m| &\leq |a_n - a_{n-1}| + |a_{n_1} - a_{n-2}| + \cdots + |a_{m+2} - a_{m+1}| + |a_{m+1} - a_m| \\ &\leq l^{n-m}|a_{m+1} - am| + l^{n-m-1}|a_{m+1} - am| + \cdots + l|a_{m+1} - a_m| + |a_{m+1} - a_m| \\ &= (l^{n-m} + l^{n-m-1} + \cdots + l + 1)|a_{m+1} - a_m| \\ &= \frac{1 - l^{n-m+1}}{1 - l}|a_{m+1} - a_m| \\ &= \frac{1 - l^{n-m+1}}{1 - l}l^m|a_2 - a_1| \\ &\leq \frac{l^m}{1 - l}|a_2 - a_1|\end{aligned}$$

As $m \to \infty$ we have $\left(\frac{l^m}{1-l}\right) \to 0$ by the geometric sequence. Then for every $\frac{\epsilon}{|a_2-a_1|}$, there exists N such that $\left|\frac{l^m}{1-l}\right| < \frac{\epsilon}{|a_2-a_1|}$ whenever $n > N$. Using this, we have that $|a_n - a_m| \leq \frac{\epsilon}{|a_2-a_1|} < \epsilon$. Thus we have the desired result. □

3.2.4 Standard Limits

Theorem 3.2.4.1: nth root

For every $x \in \mathbb{R}$ and $n \in \mathbb{N}$ there exists a unique nth root denoted by $x^{\frac{1}{n}}$.

Proof. Let $A = \{x > 0 : x^n > a\}$ We want to show that the infinum of this set is exactly the nth root of a. Note that $(1 + a)^n \geq 1 + na > a$ thus A is non-empty. Also since $x > 0$ we have that 0 is a lower bound and thus its infinum exists by the completeness axiom. We let $b = \inf(A)$. Since b is the infinum, we have some $a_k \in A$ such that $b \leq a_k < b + \frac{1}{k}$. By the sandwich theorem we have $a_k \to b$. By the product rule of series, we have $a_k^n \to b^n$. Recall that $a_k \in A$, then we have $a_k^n > a$, and $\lim_{k\to\infty} a_k^n \geq a$, thus $b^n \geq a$.

Assume now that $b^n > a$, we have $0 < \frac{a}{b^n} < 1$ so we may choose $\delta > 0$ so that $\delta < \frac{b}{n}(1 - \frac{a}{b^n})$. We now show that $b - \delta \in A$. Not that from our definition of δ, we have $\delta < \frac{b}{n}(1 - \frac{a}{b^n})$ and since $n > 1$ and $0 < 1 - \frac{a}{b^n} < 1$, we have $\delta < \frac{b}{n}\left(1 - \frac{a}{b^n}\right) < b$. Thus we have $\frac{\delta}{b} < 1$, and $-\frac{\delta}{b} > -1$. Thus we can apply Bernoulli's Inequality.

$$\begin{aligned}(b - \delta)^n &= b^n\left(1 - \frac{\delta}{b}\right)^n \\ &\geq b^n\left(1 - n\frac{\delta}{b}\right) && \text{(By Bernoulli's Inequality)} \\ &> a\end{aligned}$$

since $\delta < \frac{b}{n}\left(1 - \frac{a}{b^n}\right) \implies b^n\left(1 - n\frac{\delta}{b}\right) > a$. This proves that $b - \delta \in A$, contradicting the fact that b is the infinum. Hence we can only have $b^n = a$ This completes the proof of existence of the nth root.

Now we prove the uniqueness of the nth root. Suppose that $b^n = c^n = a$. Without loss of generality assume $b \leq c$We have

$$\begin{aligned} 0 &= c^n - b^n \\ 0 &= (c-b)(c^{n-1} + bc^{n-2} + \cdots + b^{n-2}c + b^{n-1}) \end{aligned}$$

But $c^{n-1} + bc^{n-2} + \cdots + b^{n-2}c + b^{n-1} > nb^{n-1}$. Thus we can only have $b = c$. □

Theorem 3.2.4.2

For all $x > 0$,

$$x^{\frac{1}{n}} \to 1$$

Proof. Consider the case with $x \geq 1$. We have $x^{\frac{1}{n}} \geq 1$. From the Bernoulli's Inequality, we have $\left(1 + x^{\frac{1}{n}} - 1\right) \geq 1 + n\left(x^{\frac{1}{n}} - 1\right) \implies x \geq 1 + n\left(x^{\frac{1}{n}} - 1\right)$. Thus we have $0 < x^{\frac{1}{n}} - 1 \leq \frac{x-1}{n}$. Since $(\frac{x-1}{n}) \to 0$, we have that $\left(x^{\frac{1}{n}} - 1\right) \to 0$ from the sandwich theorem and thus $\left(x^{\frac{1}{n}}\right) \to 1$. For the case $0 < x < 1$, note that $\left(\frac{1}{x^{\frac{1}{n}}}\right) \to 1$ since $\frac{1}{x} > 1$. Using the algebra for sequences, we have $\left(\frac{1}{\frac{1}{x}^{\frac{1}{n}}}\right) = \left(x^{\frac{1}{n}}\right) \to 1$. □

Theorem 3.2.4.3

The sequence $n^{\frac{1}{n}}$ converges to 1.

Proof. Note that $n \geq 1$ and $n^{\frac{1}{2n}} > 0$, we have

$$\begin{aligned} \sqrt{n} &= (1 + (n^{\frac{1}{2n}} - 1))^n \\ &\geq 1 + n(n^{\frac{1}{2n}} - 1) \\ &> n(n^{\frac{1}{2n}} - 1) \end{aligned}$$

from Bernoulli's Inquality. Thus we have that $1 \leq n^{\frac{1}{2n}} < \frac{1}{\sqrt{n}} + 1$. So by the sand wich theorem we have $(n^{\frac{1}{2n}}) \to 1$ and $(n^{\frac{1}{n}}) = (n^{\frac{1}{2n}})^2 \to 1$ by the arithmetic of sequences. □

Theorem 3.2.4.4: Ratio Lemma

Suppose $0 \leq l < 1$. Let $(a_n)_{n \in \mathbb{N}}$ be a sequence.

- If there exists some $N \in \mathbb{N}$ $\frac{a_{n+1}}{a_n} \leq l$ for all $n > N$, then $a_n \to 0$
- If $\frac{a_{n+1}}{a_n} \to l$ then $a_n \to 0$

Proof. Suppose that $0 \leq l < 1$

- Note that since $a_n \leq la_{n-1}$, we have $a_n \leq l^n a_N$. Since $0 < a_n \leq l^n a_N$, and $(l^n) \to 0$ by geometric sequences, we have $a_n \to 0$.
- Since $\frac{a_{n+1}}{a_n} \to l$, we have that for every $\epsilon > 0$, there exists N such that $\left|\frac{a_{n+1}}{a_n} - l\right| < \epsilon$ whenever $n > N$. Then choosing ϵ such that $l + \epsilon < 1$, we have that all the terms after a_N being less than $l + \epsilon < 1$, thus apply the ratio lemma to have $a_n \to 0$.

□

Theorem 3.2.4.5

Suppose $k \in \mathbb{N}$. Then

$$\frac{x^n}{n^k} \to \begin{cases} \infty & \text{if } x > 1 \\ 0 & \text{if } 0 < x \leq 1 \end{cases}$$

Proof. Consider $\frac{a_{n+1}}{a_n}$ with $a_n = \left(\frac{x^n}{n^k}\right)$, We have that

$$\begin{aligned} \frac{a_{n+1}}{a_n} &= \frac{x^{n+1}}{(n+1)^k}\frac{n^k}{x^n} \\ &= \left(1 - \frac{1}{n+1}\right)^k x \end{aligned}$$

We have that $\left(1 - \frac{1}{n+1}^k\right) \to 0$ so by the ratio lemma, it converges to 0 whenever $0 < x \leq 1$. Then suppose that $b_n = \frac{1}{a_n}$. Then $\frac{b_{n+1}}{b_n} = \frac{1}{\left(1-\frac{1}{n+1}\right)^k}\frac{1}{x}$ thus b_n tends to 0 whenever $x > 1$. Then by taking the reciprocal we have $a_n \to \infty$ when $x > 1$. □

3.2.5 Further Developing the Real Numbers

Definition 3.2.5.1: Decimal Representation

For every $x \in \mathbb{R}$,

$$x = \sum_{k=0}^{\infty} \frac{d_k}{10^k}$$

where $d_n \in 0, 1, \dots 9$ for all $n \in \mathbb{N}$ and d_0 an integer. We write $x = d_0.d_1d_2\dots$

Definition 3.2.5.2: Categorization of Decimals

An infinte decimal $\pm d_0.d_1d_2\dots$ is

- terminating if there exists a natural number N such that $d_n = 0$ for every $n > N$
- recurring if there exists natural numbers N and r such that $d_n = d_{n+r}$ for every $n > N$
- non-recurring if it is neither terminating nor recurring

Theorem 3.2.5.3

Every infinte decimal $\pm d_0.d_1d_2\dots$ represents a real number.

Proof. Suppose we have an infinite decimal $\pm d_0.d_1d_2\dots$. Define a sequence

$$s_n = \sum_{k=0}^{n} \frac{d_k}{10^k}$$

Then we have that (s_n) is increasing and is bounded by by the infinite decimal. Thus by the completeness axiom (d_n) converges and it converges to its supremum which is a real number. Thus the real number it represents is its supremum. □

Theorem 3.2.5.4

Every real number can be represented by an infinte decimal.

Proof. Define a sequence

$$s_n = \sum_{k=0}^{n} \frac{d_k}{10^k}$$

with d_k selected such that d_k is the largest in $0, \dots 9$ and s_n is less than $x \in \mathbb{R}$. Then we have that (s_n) is increasing and is bounded by x. Thus by the completeness axiom (s_n) converges and it converges to its supremum. Note that for however small ϵ there exists an $n \in \mathbb{N}$ such that $10^n \epsilon > 1$, and thus $\epsilon > \frac{1}{10^n}$. Then for every $\epsilon > 0$ there exists some N such that

$$\left| x - \sum_{k=0}^{n} \frac{d_k}{10^k} \right| < \frac{1}{10^n} < \epsilon$$

whenever $n > N$. Thus the condition for convergence is satisfied and $(s_n) \to x$. Thus the there exists an infinite decimal such that it converges to the selected real number. □

Lemma 3.2.5.5

For every real number x there is a sequence of rationals that converges to x.

Proof. By the construction of the last theorem the sequence used is a sequence of rationals that converges to x. □

Theorem 3.2.5.6

Every real number x is rational if and only if it can be written as a terminating or recurring decimal. This implies that x is irrational if and only if it is non-recurring.

Proof. We first prove the forward implication. Suppose that $x = \frac{p}{q}$ with $p, q \in \mathbb{N}$. To represent a negative real number simply append the minus operation in front of $\frac{p}{q}$. Now, remove the integer part of x so that the new number, y, is between 0 and 1. Then y is also a rational number since alegra under rationals are closed. Let $y = \frac{m}{n}$. Let $m = d_0 n + r_1$. Then let $10r_1 = d_1 n + r_2$. Recursively define $10r_{k-1} = d_{k-1} n + r_k$. Observe that $r_k \in \{o, \dots, n-1\}$. Thus the operation eventually repeats, and we obtain a recurring decimal. If the recurring block of the decimal is only 0, then it is in fact a terminating decimal, a subset of recurring decimals. Note that the proof here required knowledge on basic number theory.

We now prove the backward implication. Suppose that we have a recursive decimal x with $0 < x < 1$, a non-recurring block of length p and recurring block of length q. Then we have $(10^{p+q} - 1)x = n \in \mathbb{N}$. Then we have $x = \frac{n}{10^{p+q}-1}$. Append any integer to the front of the decimal then any recurring decimal can be represented by a rational number. □

Corollary 3.2.5.7

The series

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

converges.

Proof. Consider $\frac{a_{n+1}}{a_n}$. We have that $\frac{a_{n+1}}{a_n} = \left(1 + \frac{1}{n+1}\right)\left(1 - \frac{1}{(n+1)^2}\right)^n$. Note that since

$\frac{-1}{(n+1)^2} > -1$, we can apply the Bernoulli's Inequality can be applied. Then

$$\left(1+\frac{1}{n+1}\right)\left(1-\frac{1}{(n+1)^2}\right)^n \geq \left(1+\frac{1}{n+1}\left(1-\frac{n}{(n+1)^2}\right)\right) = 1+\frac{1}{(n+1)^3} \geq 1$$

Thus a_n is an increasing sequence.

Now note that $\left(1+\frac{1}{2n}\right)^n = \frac{1}{\left(1+\frac{1}{2n+1}\right)^n}$. Since $\frac{-1}{2n+1} > -1$. Thus we can apply Bernoulli's Inequality to have

$$\begin{aligned}\left(1-\frac{1}{2n+1}\right)^n &\geq 1-\frac{n}{2n+1}\\ \left(1-\frac{1}{2n+1}\right)^n &\geq \frac{n+1}{2n+1}\\ \frac{2n+1}{n+1} &\geq \frac{1}{\left(1-\frac{1}{2n+1}\right)^n}\\ 2-\frac{1}{n+1} &\geq \left(1+\frac{1}{2n}\right)^n\\ 2 &\geq \left(1+\frac{1}{2n}\right)^n\end{aligned}$$

Then we have $0 \leq a_{2n} \leq 4$, which is bounded. By the completeness axiom, this sequence converges. □

3.3 Series

3.3.1 Series and Convergences

Definition 3.3.1.1: Definition of Series

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

$$\sum_{k=1}^{\infty} a_n = a_1 + a_2 + \dots + a_k + \dots$$

as an infinite series. The nth partial sum of the series is defined by

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

Definition 3.3.1.2: Convergence of Series

The series is said to converge if the sequence of partial sums converge. The series diverges if the sequence of partial sums diverge.

Theorem 3.3.1.3: Algebra of Series

Let $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ be sequecnces in $\mathbb{R}$ and let $s, t \in \mathbb{R}$. If $\sum_{k=1}^{\infty} a_k$ and $\sum_{k=1}^{\infty} b_k$ converges, then

$$\sum_{k=1}^{\infty}(sa_k + tb_k)$$

converges.

Proof. Since series are also sequences, we can simply apply the arithmetic of sequences to obtain the results. □

Theorem 3.3.1.4: Geometric Series

The geometric series

$$\sum_{k=0}^{\infty} ar^k$$

where $a, r \in \mathbb{R}$ and $a \neq 0$, converges if and only if $|r| < 1$. In this case

$$\sum_{k=0}^{\infty} ar^k = \frac{a}{1-r}$$

Proof. Consider the partial sum $s_n = a + ar + \dots + ar^n$. We have $s_n = \frac{1-r^{n+1}}{1-r}$. If $|r| < 1$, then (r^{n+1}) converges to 0. Then for every $(1-r)\epsilon > 0$ there exists some N such that $n \geq N$ implies $\left|r^{n+1}\right| < (1-r)\epsilon$. Then

$$\begin{aligned} \left|\frac{1-r^{n+1}}{1-r} - \frac{1}{1-r}\right| &= \left|\frac{r^{n+1}}{1-r}\right| \\ &< \epsilon \end{aligned}$$

Thus the geometric series converges if and only if $|r| < 1$. □

Corollary 3.3.1.5

The infinite sum

$$\sum_{k=0}^{\infty} \frac{1}{k!}$$

converges.

Proof. For all $n > 0$, we have that $n! \geq 2^{n-1}$, we have that $\frac{1}{n!} \leq \frac{1}{2^{n-1}}$. Thus we have that

$$\sum_{k=0}^{n} \frac{1}{k!} \leq \sum_{k=0}^{n} \frac{1}{2^k}$$

$\sum_{k=0}^{n} \frac{1}{2^k}$ is a geometric series so it converges then it has an upper bound. By the completeness axiom, we have that the sum converges. □

Definition 3.3.1.6: Euler's Number

Define Euler's number to be

$$e = \sum_{k=0}^{\infty} \frac{1}{k!}$$

Proposition 3.3.1.7

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

Proof. We start by considering the nth term of both sequences. From the binomial theorem, we have that

$$\begin{aligned} \left(1 + \frac{1}{n}\right)^n &= \sum_{k=0}^{n} \frac{1}{k!} \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right) \\ &\leq \sum_{k=0}^{n} \frac{1}{k!} \end{aligned}$$

Thus letting $n \to \infty$, we have

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

However, if $n > m$,

$$\left(1 + \frac{1}{n}\right)^n = \sum_{k=0}^{m} \frac{1}{k!} \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n} \cdots \left(1 - \frac{k-1}{n}\right)\right)$$

Letting $n \to \infty$, we have $e \geq s_m$. Letting $m \to \infty$, we have $e \geq s$. Thus we have $e = \lim_{n\to\infty} \left(1 + \frac{1}{n}\right)^n$ □

Theorem 3.3.1.8: Null Sequence Test

If $(a_n)_{n\in\mathbb{N}}$ does not tend to 0 then $\sum_{k=0}^{\infty} a_k$ diverges.

Proof. Suppose that $\sum_{n=0}^{\infty} a_n$ converges. Then we have that for all $\epsilon > 0$, there exists some N

such that

$$\left|\sum_{k=0}^{m} a_k - \sum_{k=0}^{n} a_k\right| < \epsilon$$

whenever $n, m > N$. In particular, choosing $m = n+1$, we have

$$\begin{aligned}\left|\sum_{k=0}^{n+1} a_k - \sum_{k=0}^{n} a_k\right| &= |a_{n+1}| \\ &< \epsilon\end{aligned}$$

for all $n > N$. Thus we have that $(a_n) \to 0$. □

3.3.2 Series with Non-negative Terms

Theorem 3.3.2.1

If $a_n \geq 0$ for all $n > N$ with $N \in \mathbb{N}$ then the series $\sum_{k=1}^{\infty} a_k$ converges in $\mathbb{R}$ if and only if its partial sums are bounded.

Proof. We first prove the forward implication. Suppose that $\sum_{k=1}^{\infty} a_k$ converges, then by 2.1.4 it is bounded. Now suppose that $\sum_{k=1}^{\infty} a_k$ is bounded. Then since $a_n \geq 0$ then $\sum_{k=1}^{\infty} a_k$ is increasing. Then $\sum_{k=1}^{\infty} a_k$ is convergent by monotonic increasing theorem. □

Theorem 3.3.2.2: Direct Comparison Test

Suppose $0 \leq a_n \leq b_n$ for all $n > N$ with $N \in \mathbb{N}$.

- If $\sum_{k=1}^{\infty} b_k$ converges then $\sum_{k=1}^{\infty} a_k$ converges.
- If $\sum_{k=1}^{\infty} a_k$ diverges then $\sum_{k=1}^{\infty} b_k$ diverges.

Proof. Suppose $0 \leq a_n \leq b_n$ for all $n > N$.

- Suppose that $\sum_{k=1}^{\infty} b_k$ converges to b. Then since $a_n \leq b_n$ for all n then a_n is bounded above by b. Then by the monotonic increasing theorem a_n converges.
- Suppose that $\sum_{k=1}^{\infty} a_k$ diverges. Then for every $C > 0$ there exists N such that $\sum_{k=1}^{\infty} a_k > C$ for all $n > N$. Then

  $$C < \sum_{k=1}^{\infty} a_k \leq \sum_{k=1}^{\infty} b_k$$

  thus we also have $\sum_{k=1}^{\infty} b_k \geq C$. Thus $\sum_{k=1}^{\infty} b_k$ diverges.

□

Lemma 3.3.2.3

The harmonic series $\sum_{k=1}^{\infty} \frac{1}{k}$ diverges.

Proof. Let $s_n = \sum_{k=1}^{n} \frac{1}{k}$. Then

$$\begin{aligned} s_{2n} &= \frac{1}{n+1} + \frac{1}{n+2} + \cdots + \frac{1}{2n-1} + \frac{1}{2n} + s_n \\ &\geq \frac{1}{2n} + \frac{1}{2n} + \cdots + \frac{1}{2n} + \frac{1}{2n} + s_n \\ &= s_n + \frac{1}{2} \end{aligned}$$

Now, we have

$$\begin{aligned} s_{2^n} &\geq s_{2^{n-1}} + \frac{1}{2} \\ &\geq s_{2^{n-2}} + \frac{1}{2} + \frac{1}{2} \\ &\geq s_1 + \frac{1}{2} + \cdots + \frac{1}{2} && (n \text{ times}) \\ &= 1 + \frac{n}{2} \end{aligned}$$

By the direct comparison test, we now have s_n diverge since $1 + \frac{n}{2}$ diverges. □

Lemma 3.3.2.4

$\sum_{k=1}^{\infty} \frac{1}{k^2}$ converges.

Proof. We first prove that $\sum_{k=2}^{\infty} \frac{1}{k^2-k}$ converges. Consider $s_n = \sum_{k=2}^{n} \frac{1}{(k)(k-1)}$.

$$\begin{aligned} s_n &= \sum_{k=2}^{n} \left(\frac{1}{k-1} - \frac{1}{k} \right) \\ &= \left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n-1} \right) - \left(\frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n-1} + \frac{1}{n} \right) \\ &= 1 - \frac{1}{n} \end{aligned}$$

Thus we have that $s_n \to 1$. By the direct comparison test, we have that $\sum_{k=1}^{\infty} \frac{1}{k^2}$ converges. □

Lemma 3.3.2.5

The number e is irrational.

Proof. Suppose that e can be written in an irreducible fraction $\frac{p}{q}$.

$$\begin{aligned} e - \sum_{k=1}^{q+1} \frac{1}{(k-1)!} &= \frac{p}{q} - \left(\frac{1}{0!} + \frac{1}{1!} + \cdots + \frac{1}{q!} \right) \\ &= \frac{p(q-1)! - (q! + q! + \frac{q!}{2} + \cdots + 1)}{q!} \end{aligned}$$

Then $k = p(q-1)! - (q! + q! + \frac{q!}{2} + \dots + 1) \in \mathbb{N}$ and $e - \sum_{k=1}^{q+1} \frac{1}{(k-1)!} = \frac{k}{q!}$. Now consider

$$\begin{aligned}
e - \sum_{k=1}^{q+1} \frac{1}{(k-1)!} &= \sum_{k=0}^{\infty} \frac{1}{k!} - \sum_{k=0}^{q} \frac{1}{k!} \\
&= \frac{1}{(q+1)!} + \frac{1}{(q+2)!} + \dots \\
&= \frac{1}{q!} \left(\frac{1}{q+1} + \frac{1}{(q+1)(q+2)} + \dots \right) \\
&< \frac{1}{q!} \left(\frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} + \dots \right) && \text{(since } q > 1\text{)} \\
&= \frac{1}{q!} \left(\frac{1}{2} + \frac{1}{4} + \dots \right) \\
&= \frac{1}{q!} && \text{(by geometric series)}
\end{aligned}$$

Thus now we have that $\frac{k}{q!} < \frac{1}{q!}$ for some $k \in \mathbb{N}$ which is not possible for any choice of k. Thus we have arrived in a contradiction. □

In this section here we suppose that the techniques and origins of integration is already properly introduced, it would be weird to add the integral test to the section of integration since it is developed for testing series convergence.

Theorem 3.3.2.6: Integral Bounds

Let f be decreasing and non-negative and integrable on the interval $[1, \infty)$. Then

$$\int_{m+1}^{n+1} f(x)\,dx \leq \sum_{k=m+1}^{n+1} f(k) \leq \int_{m}^{n} f(x)\,dx$$

Proof. The inequality is obtained through drawing the graph of a decreasing function f and estimating the area of the curve with rectangles of width 1. □

Theorem 3.3.2.7: Integral Test

Let f be decreasing and non-negative and integrable on the interval $[1, \infty)$. Then $s_n = \sum_{k=1}^{n} f(k)$ converges if and only if $\int_1^\infty f(x)\,dx$ converges.

Proof. From the integral bounds, we have

$$\sum_{k=0}^{\infty} f(k) \leq \int_1^\infty f(x)\,dx$$

Thus when the integral is converges, the increasing sum is bounded. Thus by the monotone convergence theorem, the $\sum_{k=0}^{\infty}$ converges. When the integral diverges, then

$$\int_0^\infty f(x)\,dx \leq \sum_{k=0}^{\infty} f(k)$$

also diverges by the comparison test. □

Theorem 3.3.2.8: p series test

$\sum_{k=1}^{\infty} \frac{1}{k^p}$ converges when $p > 1$ and diverges when $0 < p \leq 1$.

Proof. When $0 < p \leq 1$, we have that $\sum_{k=1}^{\infty} \frac{1}{n^p}$ diverges by the direct comparison test with $\sum_{k=1} \frac{1}{n}$. When $p > 1$, it converges by the direct comparison test with $\sum_{k=1}^{\infty} \frac{1}{n^2}$. For $p \in (1, 2)$, we have that $\frac{1}{x^p}$ is decreasing. Consider

$$\begin{aligned}\int_1^{\infty} \frac{1}{x^p}\,dx &= \lim_{n\to\infty} \frac{1}{(1-p)n^{(p-1)}} - \frac{1}{1-p} \\ &= \frac{1}{p-1}\end{aligned}$$

Thus $\int_1^{\infty} \frac{1}{x^p}\,dx$ converges and by the integral test, $\sum_{k=1}^{\infty} \frac{1}{n^p}$ converges. □

Theorem 3.3.2.9: Limit Comparison Test

Let $a_n, b_n > 0$ for all $n > N$ with $N \in \mathbb{N}$. If $\frac{a_n}{b_n}$ converges to $x \in (0, \infty)$, then $\sum_{k=1}^{\infty} b_k$ converge if and only if $\sum_{k=1}^{\infty} a_k$ converges.

Proof. We know that for every $\epsilon > 0$ there exists N such that

$$\left|\frac{a_n}{b_n} - x\right| < \epsilon$$

whenever $n > N$. This means that

$$(x-\epsilon)b_n < a_n < (x+\epsilon)b_n$$

As $x > 0$, choose ϵ so that $x - \epsilon > 0$. Then $b_n < \frac{1}{x-\epsilon}a_n$ thus by the direct comparison test if $\sum_{k=1}^{\infty} a_k$ converges then $\sum_{k=1}^{\infty} b_k$ converges. Similarly, $\frac{1}{c+\epsilon} < b_n$ so if $\sum_{k=1}^{\infty} a_k$ diverges, then $\sum_{k=1}^{\infty} b_k$ diverges. □

Theorem 3.3.2.10: Root Test

Let $a_n \geq 0$ for all n and let $p = \lim_{n\to\infty} a_n^{1/n}$.

- If $p < 1$ then $\sum_{k=1}^{\infty} a_k$ converges
- If $p > 1$ then $\sum_{k=1}^{\infty} a_k$ diverges

Proof. We first proof the case with $p < 1$. Choose $\epsilon > 0$ such that $p + \epsilon < 1$. Then for all $n > N$, we have $a_n^{\frac{1}{n}} < (p+\epsilon)$ and $a_n \leq (p+\epsilon)^n$. Since $p + \epsilon < 1$, $\sum_{k=0}^{\infty}(p+\epsilon)^k$ converges and by comparison test, $\sum_{k=1}^{\infty} a_k$ converges.

Now suppose that $p > 1$ then choose $\epsilon > 0$ such that $p - \epsilon > 1$. Then for all $n > N$ we have $a_n^{\frac{1}{n}} > p + \epsilon$ thus $a_n > (p+\epsilon)^n$. Then $\sum_{n=0}^{\infty}(p+\epsilon)^n$ diverges by the geometric series and $\sum_{k=1}^{\infty} a_k$ diverges by comparison test. □

Theorem 3.3.2.11: Ratio Test

Let $a_n > 0$ for all $n > N$ with $N \in \mathbb{N}$. Let $r = \lim_{n\to\infty} \frac{a_{n+1}}{a_n}$

- If $r < 1$, then $\sum_{k=1}^{\infty} a_k$ converges.
- If $r > 1$, then $\sum_{k=1}^{\infty} a_k$ diverges.

Proof. Suppose that $r < 1$. Then choose

$$0 < \epsilon = \frac{1-r}{2} < 1$$

Then there exists some N such that for every $n > N$,

$$\left|\frac{a_{n+1}}{a_n} - r\right| < \frac{1-r}{2} \implies \frac{a_{n+1}}{a_n} < \frac{1+r}{2} < 1$$

This means that $a_{n+1} < (\frac{1+r}{2})a_n$ and $a_{n+1} < (\frac{1+r}{2})^n a_1$. Since $\frac{1+r}{2} < 1$, $\sum_{n=0}^{\infty}(\frac{1+r}{2})^n$ converges by geometric series and $\sum a_n$ converges by comparison test.

Similarly for the divergence case, choose

$$\epsilon = \frac{r-1}{2}$$

Then there exists some N such that for every $n > N$,

$$\left|\frac{a_{n+1}}{a_n} - r\right| < \frac{1-r}{2} \implies 1 < \frac{r+1}{2} < \frac{a_{n+1}}{a_n}$$

Then $a_n(\frac{r+1}{2}) < a_{n+1}$ and $a_1(\frac{r+1}{2})^n < a_{n+1}$. By geometric series $\sum_{n=0}^{\infty}(\frac{r+1}{2})^n$ diverges and by comparison test $\sum_{k=1}^{\infty} a_k$ diverges. □

3.3.3 Alternating Series

Definition 3.3.3.1: Alternating Series

An alternating series is one of the form

$$\sum_{k=1}^{\infty}(-1)^{k+1}a_k$$

where $a_n > 0$ for all $n \in \mathbb{N}$.

Theorem 3.3.3.2: Alternating Series Test

Suppose $a_n \geq a_{n+1}$ for all n and $a_n \to 0$. Then

$$\sum_{k=1}^{\infty}(-1)^k a_k$$

converges.

Proof. Firstly note that

$$s_{2n+2} = \sum_{k=1}^{2n}(-1)^k a_k - a_{2n+1} + a_{2n+2}$$

Since a_n is decreasing, $a_{2n+1} \geq a_{2n+1}$ thus $s_{2n+2} \geq s_{2n}$. Similarly, we have $s_{2n+1} \leq s_{2n-1}$.

Combining the two inequality, we have

$$s_2 \leq s_{2n} + a_{2n+1} = s_{2n+1} \leq s_1$$

Since s_{2n} is increasing and bounded, it converges. Similarly, s_{2n+1} is decreasing and bounded thus converges. Suppose that $(s_{2n}) \to s$. Then $s_{2n} + a_{2n+1} = s_{2n+1}$. Thus $(s_{2n+1}) \to s + 0 = s$. Suppose that for all ϵ, there exists N such that $|a_n| < \epsilon$ whenever $n > N$. Then, we have

$$\begin{aligned} s_{2n} &\leq s \\ s_{2n+1} - a_{2n+1} &\leq s \\ s_{2n+1} - s &\leq a_{2n+1} \\ |s_{2n+1} - s| &\leq a_{2n+1} \\ |s_{2n+1} - s| &< \epsilon \end{aligned}$$

Similarly,

$$\begin{aligned} s &\leq s_{2n-1} \\ s &\leq s_{2n} - a_{2n-1} \\ s - s_{2n} &\leq a_{2n-1} \\ |s_{2n} - s| &\leq a_{2n-1} \\ |s_{2n} - s| &< \epsilon \end{aligned}$$

Thus we have that $|s_n - s| < \epsilon$ whenever $n > N$ and (s_n) converges. □

Lemma 3.3.3.3

The alternating harmonic series

$$\sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k}$$

converges.

Proof. Simple application of the alternating series test. □

3.3.4 Absolute Convergence

Definition 3.3.4.1: Absolute Convergence

A series $\sum_{k=0}^{\infty} a_k$ in $\mathbb{R}$ is said to converge absolutely if

$$\sum_{k=0}^{\infty} |a_k|$$

converges.

Theorem 3.3.4.2

Every absolutely convergent series is convergent.

Proof. Consider $s_n = \sum_{k=1}^{n} a_k$ and $t_n = \sum_{k=1}^{n} |a_n|$. Without loss of generality suppose that

$n > m$

$$\begin{aligned} |s_n - s_m| &= |a_n + a_{n-1} + \cdots + a_{m+1}| \\ &\leq |a_n| + |a_{n-1}| + \cdots + |a_{m+1}| \\ &= t_n - t_m \\ &= |t_n - t_m| \end{aligned}$$

If (t_n) converges then for every $\epsilon > 0$ there exists N such that $|t_n - t_m| < \epsilon$ for all $n, m > N$. This implies that $|s_n - s_m| < \epsilon$ thus s_n is convergent. □

Theorem 3.3.4.3: Ratio Test for Series

Suppose $a_n \neq 0$ for all n and $\left|\frac{a_{n+1}}{a_n}\right| \to l$. Then $\sum_{k=1}^{\infty} a_k$ converges absolutely if $0 \leq l < 1$ and diverges if $l > 1$.

Proof. Since $\frac{a_{n+1}}{a_n} \to l < 1$, choose $\epsilon > 0$ such that $l + \epsilon < 1$. There exists an N such that $\left|\frac{a_{n+1}}{a_n} - l\right| < \epsilon$ whenever $n > N$.

$$\begin{aligned} \left|\frac{a_{n+1}}{a_n} - l\right| &< \epsilon \\ \left|\left|\frac{a_{n+1}}{a_n}\right| - l\right| &< \epsilon \\ \left|\frac{a_{n+1}}{a_n}\right| &< l + \epsilon \\ |a_{n+1}| &< |a_n|(l + \epsilon) \\ |a_{n+1}| &< |a_N|(l + \epsilon)^{n+1-N} \end{aligned}$$

Since

$$\sum_{k=1}^{\infty} |a_N|(l + \epsilon)^{k+1-N} = \frac{a_N}{1 - (l + \epsilon)}$$

We have that $\sum_{k=1}^{\infty} |a_k|$ converges by comparison test.

Now suppose that $l > 1$, there exists N $|a_{n+1}| > |a_n|$ for some $n > N$. This means that (a_n) does not converge to 0 and by the null sequence test $\sum_{k=1}^{\infty} |a_k|$ diverges. □

3.3.5 Rearrangement of Series

Definition 3.3.5.1: Rearrangement

$(b_n)_{n \in \mathbb{N}}$ is a rearrangement of $(a_n)_{n \in \mathbb{N}}$ if there is a bijection $\sigma : \mathbb{N} \to \mathbb{N}$ such that for all $n \in \mathbb{N}$ $b_n = a_{\sigma(n)}$

Theorem 3.3.5.2

Suppose $(b_n)_{n \in \mathbb{N}}$ is a rearrangement of $(a_n)_{n \in \mathbb{N}}$ and $a_n > 0$ for all n and $\sum_{k=1}^{\infty} a_k$ converges. Then $\sum_{k=1}^{\infty} b_k$ converges and

$$\sum_{k=1}^{\infty} b_k = \sum_{k=1}^{\infty} a_k$$

Proof. Suppose that $s_n = \sum_{k=1}^{n} a_k$, $\sum_{k=1}^{\infty} a_k = s$ and $t_n = \sum_{k=1}^{n} b_k$. Suppose that $a_{\sigma(n)} = b_n$. Fix N and consider $M_N = \max\{\sigma(r) : r \leq N\}$. Then the first a_{M_n} terms contains the first b_n terms. Since $a_n, b_n > 0$ for all n, we have $t_n < s_{M_N} < s$. Thus t_n is increasing and bounded and converges by the completeness axiom to, say t with $t \leq s$. Reverse the roles of a_n and b_n now given that t_n converges, and we get $s \leq t$. Thus $s = t$. □

Theorem 3.3.5.3

Suppose $(b_n)_{n\in\mathbb{N}}$ is a rearrangement of $(a_n)_{n\in\mathbb{N}}$ and $\sum_{k=1}^{\infty} a_k$ converges absolutely. Then $\sum_{k=1}^{\infty} b_k$ converges and $\sum_{k=1}^{\infty} b_k = \sum_{k=1}^{\infty} a_k$.

Proof. Let $\sum_{k=1}^{\infty} a_k$ converges absolutely and $\sum_{k=1}^{\infty} b_k$ a rearrangement of the series. Consider $u_n = \frac{1}{2}(|a_n| + a_n)$, $v_n = \frac{1}{2}(|a_n| - a_n)$, $x_n = \frac{1}{2}(|b_n| + b_n)$, $y_n = \frac{1}{2}(|b_n| - b_n)$. We have that

$$u_n = \begin{cases} a_n & a_n \geq 0 \\ 0 & a_n \leq 0 \end{cases}$$

$$v_n = \begin{cases} 0 & a_n \geq 0 \\ a_n & a_n \leq 0 \end{cases}$$

$$x_n = \begin{cases} b_n & b_n \geq 0 \\ 0 & b_n \leq 0 \end{cases}$$

$$y_n = \begin{cases} 0 & b_n \geq 0 \\ b_n & b_n \leq 0 \end{cases}$$

We have $\sum_{k=1}^{\infty} u_k \leq \sum_{k=1}^{\infty} |a_k|$ thus $\sum_{k=1}^{\infty} u_k$ is convergent by completeness axiom and

$$\sum_{k=1}^{\infty} -v_k \leq \sum_{k=1}^{\infty} |a_k|$$

thus $\sum_{k=1}^{\infty} -v_k$ converges by completeness axiom and

$$\sum_{k=1}^{\infty} -v_k = -\sum_{k=1}^{\infty} v_k$$

converges. We have that $\sum_{k=1}^{\infty} u_k = \sum_{k=1}^{\infty} x_k$ and $\sum_{k=1}^{\infty} v_k = \sum_{k=1}^{\infty} y_k$ by construction. Thus

$$\sum_{k=1}^{\infty} a_k = \sum_{k=1}^{\infty} (u_k - v_k) = \sum_{k=1}^{\infty} (x_k - y_k) = \sum_{k=1}^{\infty} b_k$$

is convergent. □

Definition 3.3.5.4: Conditionally Convergent

A series $\sum_{k=1}^{\infty} a_k$ is said to be conditionally convergent if $\sum_{k=1}^{\infty} a_k$ is convergent but

$$\sum_{k=1}^{\infty} |a_k|$$

diverges.

Theorem 3.3.5.5

Suppose $\sum_{k=1}^{\infty} a_k$ is conditionally convergent. Consider $\sum_{k=1}^{\infty} u_k$ and $\sum_{k=1}^{\infty} v_k$ where

$$u_n = \begin{cases} a_n & \text{if } a_n \geq 0 \\ 0 & \text{if } a_n < 0 \end{cases}$$

and

$$v_n = \begin{cases} 0 & \text{if } a_n \geq 0 \\ a_n & \text{if } a_n < 0 \end{cases}$$

Then $\sum_{k=1}^{\infty} u_k = +\infty$ and $\sum_{k=1}^{\infty} v_k = -\infty$.

Proof. Note that $a_n = u_n - v_n$ and $|a_n| = u_n + v_n$. Suppose for contradiction that $\sum u_n$ converges. Then $|a_n| = 2u_n - a_n$ and thus $\sum |a_n|$ converges, a contradiction. Suppose also that $\sum v_n$ converges. Then again $|a_n| = 2v_n + a_n$ and thus $\sum |a_n|$ converges, a contradiction. Thus we have $\sum u_n$ and $\sum v_n$ diverges. Since $u_n \geq 0$ and $v_n \leq 0$ for all n, we have that $\sum u_n = \infty$ and $\sum v_n = -\infty$. □

Theorem 3.3.5.6

Suppose $\sum_{k=1}^{\infty} a_k$ is conditionally convergent. Then for every $l \in \mathbb{R}$ there exists a rearrangement (b_n) of (a_n) such that $\sum_{k=1}^{\infty} b_k = l$.

Proof. Let (p_n) be the subsequence of all positive terms of (a_n) and (q_n) all its negative terms. We have that $(p_n) \to +\infty$ and $(q_n) \to -\infty$. Consider $l \geq 0$. There exists N_1 such that $S_1 = \sum_{k=1}^{N_1} a_k < l$ and $S_1 + a_{N_1+1} > l$. Then since $(q_n) \to -\infty$. There exists M_1 such that $T_1 = S_1 + \sum_{k=1}^{M_1} q_k < l$ but $S_1 + \sum_{k=1}^{M_1-1} q_k > l$. Repeat the process to find the rearrangement of our sequence

$$p_1, \dots, p_{N_1}, q_1, \dots, q_{M_1}, p_{N_1+1}, \dots, p_{N_2}, q_{M_1+1}, \dots, q_{M_2}, \dots$$

Its partial sums converges since $|S_i - l| \leq p_{N_i}$ and $|T_i - l| \leq -q_{M_i}$ for all i and p_{N_i} and q_{M_i} tends to 0 since a_n is null. Swap the roles of q_n and p_n for the case $l < 0$. □

3.4 Limits and Continuity

3.4.1 Limits

Definition 3.4.1.1: Limits

Let $f : (a, b) \to \mathbb{R}$ a real valued function defined except possibly on $c \in (a, b)$. We say that

$$\lim_{x\to c} f(x) = L$$

if for every $\epsilon > 0$ there exists a number $\delta > 0$ such that

$$0 < |x - c| < \delta \implies |f(x) - L| < \epsilon$$

Definition 3.4.1.2: Convergence of Functions

Let $c, L \in \mathbb{R}$ and let $f : \mathbb{R} \to \mathbb{R}$.

- $f(x)$ is said to converge to L as $x \to \infty$ if and only if for every $\epsilon > 0$ there exists a $M > 0$ such that
$$x > M \implies |f(x) - L| < \epsilon$$
We write $\lim_{x\to\infty} f(x) = L$ in this case.
- $f(x)$ is said to converge to L as $x \to -\infty$ if and only if for every $\epsilon > 0$ there exists a $M < 0$ such that
$$x < M \implies |f(x) - L| < \epsilon$$
We write $\lim_{x\to-\infty} f(x) = L$ in this case.
- $f(x)$ is said to converge to ∞ as $x \to c$ if for every $M > 0$ there exists $\delta > 0$ such that
$$|x - c| < \delta \implies f(x) > M$$
We write $\lim_{x\to c} f(x) = \infty$ in this case.
- $f(x)$ is said to converge to $-\infty$ as $x \to c$ if for every $M < 0$ there exists $\delta > 0$ such that
$$|x - c| < \delta \implies f(x) < M$$
We write $\lim_{x\to c} f(x) = -\infty$ in this case.

Definition 3.4.1.3: One Sided Limits

Let $f : (a, b) \to \mathbb{R}$ a real valued function defined except possibly on $c \in (a, b)$.

- We say that $\lim_{x\to c^+} f(x) = L$ if for every $\epsilon > 0$ there exists a number $\delta > 0$ such that if
$$c < x < c + \delta \implies |f(x) - L| < \epsilon$$
- We say that $\lim_{x\to c^-} f(x) = L$ if for every $\epsilon > 0$ there exists a number $\delta > 0$ such that if
$$c - \delta < x < c \implies |f(x) - L| < \epsilon$$

Lemma 3.4.1.4

Let $L \in \mathbb{R}$. Then

$$\lim_{x\to c} f(x) = L \iff \lim_{x\to c^+} f(x) = \lim_{x\to c^-} f(x) = L$$

Proof. An easy manipulation of the definition of limits. □

Theorem 3.4.1.5: Sequential Limits

Let $f : (a, b) \to \mathbb{R}$ a real valued function defined except possibly on $c \in (a, b)$. Then

$$\lim_{x\to c} f(x) = L$$

if and only if for every sequence (x_n) in $I \setminus \{c\}$ with $(x_n) \to c$ we have $(f(x_n)) \to L$.

Proof. Suppose that $\lim_{x\to c} f(x) = L$. Then for every $\epsilon > 0$ there exists a number $\delta > 0$ such that if $|x - c| < \delta$ then $|f(x) - L| < \epsilon$. Consider a sequence with $(x_n) \to c$. Then for every $\delta > 0$ we have that $|x_n - c| < \delta$ for all n larger than some N. However this implies that $|f(x_n) - L| < \epsilon$ for all n larger than N. Thus we have $(f(x_n)) \to L$.

Now suppose that $(x_n) \to c \implies (f(x)) \to L$. This means that for all $\delta > 0$, there exists N such that $|x_n - c| < \delta$ for all $n > N$. Consider any x in the interval I. Then there is a sequence (x_n) that contains x. That sequence has the properties that $|x_n - c| < \delta$. We thus have $|f(x) - L| < \epsilon$ by assumption. □

Proposition 3.4.1.6: Algebra of Limits

Let $f, g : (a, b) \to \mathbb{R}$ a real valued function defined except possibly on $c \in (a, b)$ and $\lim_{x\to c} f(x) = L$ and $\lim_{x\to c} g(x) = M$, then

- $\lim_{x\to c} (f(x) + g(x)) = L + M$
- $\lim_{x\to c} (f(x)g(x)) = LM$
- $\lim_{x\to c} \left(\frac{f(x)}{g(x)}\right) = \frac{L}{M}$ provided that $M \neq 0$

Proof. Convert them to sequential definition of limits and apply algebra of sequences. □

Theorem 3.4.1.7: Sandwich Theorem for Limits

Let $f, g, h : (a, b) \to \mathbb{R}$ a real valued function defined except possibly on $c \in (a, b)$.

- If $f(x) \leq h(x) \leq g(x)$ for all $x \in I \setminus \{c\}$, and
$$\lim_{x\to c} f(x) = \lim_{x\to c} g(x) = L$$
then the limit of $h(x)$ as $x \to c$ exists, and
$$\lim_{x\to c} h(x) = L$$
- If $g(x) \leq M$ for all $x \in I \setminus \{c\}$ and $f(x_n) \to 0$ as $(x_n) \to c$ then $\lim_{x\to c} f(x)g(x) = 0$

Proof. Convert them to sequential definition and apply the sandwich theorem of sequences. For the second item, first note that for all $\frac{\epsilon}{M} > 0$ there exists some N such that

$$-\frac{\epsilon}{M} < f(x_n) < \frac{\epsilon}{M}$$

for all $n > N$. Note also that

$$-M \leq g(x_n) \leq M$$

for all n. Thus we have that

$$|f(x_n)g(x_n)| < \epsilon$$

for all $n > N$ □

3.4.2 Continuity

Definition 3.4.2.1: Continuity

A function $f : I \subseteq \mathbb{R} \to \mathbb{R}$ is said to be continuous at $c \in I$ if and only if for every $\epsilon > 0$ there exists $\delta > 0$ such that for every $x \in I$,

$$|x - c| < \delta \implies |f(x) - f(c)| < \epsilon$$

Theorem 3.4.2.2: Sequential Continuity

A function $f : I \subseteq \mathbb{R} \to \mathbb{R}$ is said to be continuous at $c \in I$ if and only if for every sequence (x_n) of points in I which converges to c, $f(x_n) \to f(c)$.

Proof. We first prove the forward implication. Suppose that f is continuous at c and $(x_n) \to c$. Then for every $\epsilon > 0$, there exists $\delta > 0$ such that for all $x \in I$,
$|x - c| < \delta \implies |f(x) - f(c)| < \epsilon$. Now, choose N so that if $n > N$, then $|x_n - c| < \delta$. Then if $n > N$ we have $|f(x_n) - f(c)| < \epsilon$. Thus $(f(x_n)) \to c$.

Now suppose f is not continuous at c. Then there are some $\epsilon > 0$ such that there exists some x with $|x - c| < \delta \implies |f(x) - f(c)| \geq \epsilon$. We now construct a sequence as follows. For each n, choose x_n such that $|x_n - c| < \delta = \frac{1}{n}$ but $|f(x_n) - f(c)| \geq \epsilon$. Then the sequence constructed tends to c since $\left(\frac{1}{n}\right) \to 0$ but $f(x_n)$ does not converge to $f(c)$. □

Theorem 3.4.2.3: Limits and Continuity

If $f : (a, b) \to \mathbb{R}$ is a function and $c \in (a, b)$, f is continuous at c if and only if $\lim_{x\to c} f(x) = f(c)$.

Proof. Simple proof involving replacing L to $f(c)$ in the definition of limits. □

Proposition 3.4.2.4

Let $f, g : I \subseteq \mathbb{R} \to \mathbb{R}$ and continuous at $c \in I$. Then

- $f + g$ is continuous at c
- $f \cdot g$ is continuous at c
- If $g(c) \neq 0$ then $\frac{f}{g}$ is continuous at c

Proof. The three proofs are simple. Using sequential continuity, since we have the algebra of sequences, we also conclude the algebra of continuous functions. □

Proposition 3.4.2.5

Let $f : I \subseteq \mathbb{R} \to \mathbb{R}$ and $g : J \subseteq \mathbb{R} \to I$. If f is continuous at $g(c)$ and g is continuous at c, then $f \circ g$ is continuous at c.

Proof. Let $x_n \to c$, then $g(x_n) \to g(c)$ thus $f(g(x_n)) \to f(g(c))$. □

Theorem 3.4.2.6: Intermediate Value Theorem

Let $f : [a, b] \to \mathbb{R}$ be continuous and suppose that $f(a) < u < f(b)$. Then

$$f(c) = u$$

for some $c \in (a, b)$.

Proof. Consider the set

$$A = \{x \in [a, b] : f(x) \leq u\}$$

Since $a \in A$, A is non-empty. Let $s = \sup(A)$. Suppose that $f(s) < u$. Since f is continuous, we can choose $\epsilon = u - f(s)$. Then for some $\delta > 0$, $|f(x) - f(s)| < \epsilon$ as long as $|x - s| < \delta$. In particular, consider $x = s + \frac{\delta}{2}$. Then we have

$$f\left(s + \frac{\delta}{2}\right) < f(s) + \epsilon = u$$

Thus $s + \frac{\delta}{2} \in A$, a contradiction.

Now suppose that $f(s) > u$. Since f is continuous, we can choose $\epsilon = f(s) - u$. Then for some $\delta > 0$, all x that satisfies $|x - s| < \delta$ also satisfy $|f(x) - f(s)| < \epsilon$. In particular, consider $x = s - \frac{\delta}{2}$. Then we have

$$f\left(s - \frac{\delta}{2}\right) > f(s) - \epsilon = u$$

Hence $s - \frac{\delta}{2}$ is an upper bound smaller than s, a contradiction. □

Proposition 3.4.2.7

If $f : I \subseteq \mathbb{R} \to \mathbb{R}$ is continuous on the interval I then its range is an interval.

Proof. Suppose that $a, b \in I$. Then $f(a), f(b) \in f(I)$ and by the IVT, for every y between $f(a)$ and $f(b)$, you can find a corresponding x such that $f(x) = y$. □

Proposition 3.4.2.8: Inverse Function Theorem

Let $f : [a, b] \to \mathbb{R}$ be continuous and strictly increasing. Then f has an inverse defined on its range and f^{-1} is continuous.

Proof. Firstly, since f is increasing, all of the values of f lies between $f(a) = c$ and $f(b) = d$. Thus the range of f is exactly $[c, d]$. By the IVT, for each $y \in [c, d]$ there is an unique number x such that $f(x) = y$. Define $g = f^{-1}$ by setting $g(y) = x$ for every pair of (x, y). By construction, g is then increasing. Since f is increasing, we have that $f(x - \epsilon) < y = f(x) < f(x + \epsilon)$ with $\epsilon > 0$. Since f is continuous, we have that

$$f(x - \epsilon) < y - \delta < y < y + \delta < f(x + \epsilon)$$

for some $\delta > 0$. Thus we have that for every y between $y - \delta$ and $y + \delta$, $x - \epsilon < g(y) < x + \epsilon$. Thus the definition of continuity is satisfied. □

Theorem 3.4.2.9: Boundedness of Continuous Functions

Let $f : [a, b] \to \mathbb{R}$ be continuous. Then f is bounded.

Proof. Suppose that f is unbounded. For each n, choose $x_n \in [a, b]$ such that $|f(x_n)| \geq n$. Since f is continuous, we can choose a subsequence (x_{n_k}) such that it converges to x. Since the interval is closed we must have $x \in [a, b]$. Then $f(x_{n_k}) \to f(x)$. But this is not possible since $f(x_{n_k})$ is becoming arbitrarily large. □

Theorem 3.4.2.10: Extreme Value Theorem

Let $f : [a, b] \to \mathbb{R}$ be continuous. Then f has a maximum and a minimum attained in the interval.

Proof. Let M be the supremum of $\{f(x) : x \in [a, b]\}$. Suppose that no point in the interval such that M is attained. Then the function $g(x) = M - f(x)$ is strictly positive and continuous on the interval. By the algebra of continuous functions, $\frac{1}{M-f(x)}$ is continuous and therefore bounded by, say R. Then $\frac{1}{R} \leq M - f(x)$ and hence $f(x) \leq M - \frac{1}{R}$. This shows that M is not the supremum, a contradiction. □

3.4.3 Uniform Continuity

Definition 3.4.3.1: Uniform Continuity

Let $f : I \subseteq \to \mathbb{R}$. Then f is uniformly continuous on I if and only if for every $\epsilon > 0$ there exists $\delta > 0$ such that for all $x, y \in I$ and

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

Proposition 3.4.3.2

Every uniformly continuous function is continuous.

Proof. Fix y and allow x to vary. Then we have the definition of continuity at $(y, f(y))$. □

Theorem 3.4.3.3

If f is continuous on a closed interval $[a, b]$, then f is uniformly continuous on $[a, b]$.

Proof. Suppose that f is not uniformly continuous. Then there exists $\epsilon > 0$ such that for all $\delta > 0$ there are points $x, y \in [a, b]$ such that

$$|x - y| < \delta \implies |f(x) - f(y)| \geq \epsilon$$

Now for every $k \in \mathbb{N}$, choose $x_k, y_k \in [a, b]$ such that $|x_k - y_k| < \frac{1}{k}$ and

$$|f(x_k) - f(y_k)| \geq \epsilon$$

We have that (x_k) is bounded by $[a, b]$, thus by the Bolzano-Weierstrass theorem, there exists a convergent subsequence (x_{k_i}) such that it has a limit $x_0 \in [a, b]$. Mirror for y_k. Now we have

$$\begin{aligned} |x_0 - y_{k_i}| &\leq |x_0 - x_{k_i}| + |x_{k_1} - y_{k_i}| \\ &\leq |x_0 - x_{k_j}| + \frac{1}{k_j} \end{aligned}$$

Thus we also have $y_{k_j} \to x_0$. But since f is continuous at x_0, we must have

$$\left|f(x_{k_j}) - f(x_0)\right| < \frac{\epsilon}{2}$$

and

$$\left|f(y_{k_j}) - f(x_0)\right| < \frac{\epsilon}{2}$$

Thus

$$\begin{aligned}\left|f(x_{k_j}) - f(y_{k_j})\right| &\leq \left|f(x_{k_j}) - f(x_0)\right| + \left|f(y_{k_j}) - f(x_0)\right| \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2} \\ &= \epsilon\end{aligned}$$

This is a contradiction since we assumed that $|f(x) - f(y)| \geq \epsilon$. □

Theorem 3.4.3.4

If f is uniformly continuous on a set I and $(s_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in I, then $(f(s_n))_{n\in\mathbb{N}}$ is a Cauchy sequence.

Proof. Since s_n is Cauchy in S, then fix $\delta > 0$. There exists N such that $|s_n - s_m| < \delta$ for all $m, n > N$. Since f is uniformly continuous, for every $\epsilon > 0$, there exists $\delta > 0$ such that

$$|s_n - s_m| < \delta \implies |f(s_n) - f(s_m)| < \epsilon$$

for all $m, n > N$. Thus $f(s_n)$ is Cauchy. □

3.4.4 Power Series

Definition 3.4.4.1: Power Series

Suppose that $a_n \in \mathbb{R}$ for all $n \in \mathbb{N}$. Then the function

$$f(x) = \sum_{k=0}^{\infty} a_k x^k$$

is said to be a power series in $\mathbb{R}$.

Theorem 3.4.4.2

Let $\sum_{k=0}^{\infty} a_k x^k$ be a power series with $\sum_{k=0}^{\infty} a_k t^k$ convergent for some $t \in \mathbb{R}$. Then $\sum_{k=0}^{\infty} a_k x^k$ converges absolutely for all x with $|x| < |t|$.

Proof. Note that the convergence of $\sum_{k=0}^{\infty} a_k t^k$ implies $(a_n t^n) \to 0$. This implies that $a_n t^n$ is bounded by, say M.

$$\begin{aligned}\sum_{k=0}^{\infty}\left|a_k x^k\right| &= \sum_{k=0}^{\infty}\left|a_k t^k\right|\left|\frac{x^k}{t^k}\right| \\ &< \sum_{k=0}^{\infty} M\left|\frac{x^k}{t^k}\right| \\ &= \frac{M}{1-\left|\frac{x}{t}\right|}\end{aligned}$$

Since it is increasing and bounded, the infinity sum converges absolutely thus the sum converges. □

Theorem 3.4.4.3

For any power series $\sum_{k=0}^{\infty} a_k x^k$, the radius of convergence is

$$R = \frac{1}{\limsup_{n\to\infty}|a_n|^{\frac{1}{n}}}$$

The power series converges for $|x| < R$. The power series diverges for $|x| > R$.

Proof. Application of the root test. □

Proposition 3.4.4.4

Let $\sum_{k=0}^{\infty} a_k x^k$ with radius of convergence R. Then $\sum_{k=0}^{\infty} |a_k| x^k$ also has radius of convergence R.

Proof. We have that

$$\left|\sum_{k=0}^{\infty}|a_k|x^k\right| \leq \sum_{k=0}^{\infty}|a_k|\left|x^k\right|$$

which converges between $(-R, R)$. □

Theorem 3.4.4.5: Continuity of Power Series

Let $\sum_{k=0}^{\infty} a_k x^k$ with radius of convergence R. Then the function $f(x) = \sum_{k=0}^{\infty} a_k x^k$ is continuous on the interval $(-R, R)$.

Proof. Suppose $|x| < R$. We show that the function is continuous at x. Let T such that $|x| < T < R$. Then $\sum_{k=0}^{\infty}|a_k|T^k$ converges, so for every $\epsilon > 0$ there is some number N such that

$$\sum_{k=N+1}^{\infty}|a_k|T^k < \frac{\epsilon}{3}$$

Let y such that $|y - x| < T - |x|$. Then we will have $|y| < T$ and $|x| < T$. Hence

$$\sum_{k=N+1}^{\infty}|a_k||x|^k < \frac{\epsilon}{3}$$

and

$$\sum_{k=N+1}^{\infty}|a_k||y|^k < \frac{\epsilon}{3}$$

The partial sum

$$\sum_{k=0}^{N}a_ky^k$$

is a polynomial thus is continuous. There exists some $\delta_0 > 0$ such that $|y - x| < \delta_0$ implies

$$\left|\sum_{k=0}^{N}a_ky^k - \sum_{k=0}^{N}a_kx^k\right| < \frac{\epsilon}{3}$$

Choose $\delta = \min(\delta_0, T - |x|)$. Then we have that $|y - x| < \delta$ we get

$$\begin{aligned}\left|\sum_{k=0}^{\infty}a_ky^k - \sum_{k=0}^{\infty}a_kx^k\right| &\leq \left|\sum_{k=N+1}^{\infty}a_ky^k\right| + \left|\sum_{k=0}^{N}a_ky^k - \sum_{k=0}^{N}a_kx^k\right| + \left|\sum_{k=N+1}^{\infty}a_kx^k\right|\\ &\leq \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3}\\ &= \epsilon\end{aligned}$$

□

3.5 Differentiation

3.5.1 Properties of the Derivative

Definition 3.5.1.1: Derivative

Let $f : (a,b) \to \mathbb{R}$ be a function and $c \in (a,b)$. We say f is differentiable at c or f has a derivative at c if the limit

$$\lim_{x\to c} \frac{f(x)-f(c)}{x-c}$$

exists and is finite. We write $f'(c)$ or $\frac{df}{dx}|_{x=c}$ as the derivative of f at c.

Proposition 3.5.1.2

Let $f : (a,b) \to \mathbb{R}$ be a function and $c \in (a,b)$. Then f is differentiable at c if and only if

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

exists and is finite. In this case this limit is equal to $f'(c)$.

Proof. Simple change of variables where $x = c+h$. □

Proposition 3.5.1.3

If f is differentiable at c then f is continuous at c.

Proof. Suppose that f is differentiable at c.

$$\begin{aligned}\lim_{x\to c} f(x) - f(c) &= \lim_{x\to c} \frac{(f(x)-f(c))(x-c)}{x-c} \\ &= f'(c) \lim_{x\to c}(x-c) \\ &= 0\end{aligned}$$

□

Theorem 3.5.1.4: Algebra of Derivatives

Let f and g are differentiable at the point c, then

- $(f+g)'(c) = f'(c) + g'(c)$
- $(cf)'(c) = cf'(c)$
- $(fg)'(c) = f'(c)g(c) + g'(c)f(c)$
- $\left(\frac{f}{g}\right)'(c) = \frac{f'(c)g(c)-g'(c)f(c)}{(g(c))^2}$ as long as $g(c) \neq 0$.

Proof. Use the limit definition of the derivatives. □

Theorem 3.5.1.5: Chain Rule

Suppose that $g : (a,b) \to \mathbb{R}$ and $f : g((a,b)) \to \mathbb{R}$ such that g is differentiable at $c \in (a,b)$ and f is differentiable at $g(c) \in g((a,b))$. Then $f \circ g$ is differentiable at c and $(f \circ g)'(c) = f'(g(c))g'(c)$.

Proof.

$$\begin{aligned}\lim_{x\to c}\frac{f(g(x))-f(g(c))}{x-c} &= \lim_{x\to c}\frac{f(g(x))-f(g(c))}{g(x)-g(c)}\frac{g(x)-g(c)}{x-c}\\ &= \lim_{x\to c}\frac{f(g(x))-f(g(c))}{g(x)-g(c)}\lim_{x\to c}\frac{g(x)-g(c)}{x-c}\\ &= f'(g(c))g'(c)\end{aligned}$$

□

3.5.2 Inverse of the Derivative

Theorem 3.5.2.1

If $f : I \subseteq \mathbb{R} \to \mathbb{R}$ is continuous and injective then f^{-1} is either increasing or decreasing on the interval.

Proof. Suppose that f^{-1} is neither increasing or decreasing on the interval I. Then f^{-1} is decreasing on some interval $(a,b) \in I$ and not decreasing on end points. This means that $(c,a) \in I$ is not decreasing thus there is some interval $(c',a) \in (c,a)$ that is increasing, Then for every $y \in (\min f(c), f(d), f(a))$, the function is not injective. □

Theorem 3.5.2.2

Let f be a continuous injective function defined on an interval and suppose that f is differentiable at $f^{-1}(b)$, with derivative $f'\left(f^{-1}(b)\right) \neq 0$. Then f^{-1} is differentiable at b and

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

Proof. Since f is continuous and injective, the derivative exists. Suppose that $g = f^{-1}$. We have that $f(g(x)) = x$ for all $x \in \operatorname{dom}(g)$. Then we have $f'(g(x))g'(x) = 1$ and $g'(x) = \frac{1}{f'(g(x))}$. □

3.5.3 Mean Value Theorem

Definition 3.5.3.1: Local Maximum and Minimums

Let $f : (a,b) \to \mathbb{R}$ be a function and $c \in (a,b)$.

- f is said to have a local maximum at c if $f(x) \leq f(c)$ for all $x \in (a,b)$
- f is said to have a local minimum at c if $f(x) \geq f(c)$ for all $x \in (a,b)$

Both local maximums and local minimums are called local extremums.

Theorem 3.5.3.2

If f is defined on an open interval containing c, if f assumes its maximum or minimum at c, and if f is differentiable at c, then

$$f'(c) = 0$$

Proof. Suppose that f attains maximum at c. If $x > c$, then $\frac{f(x)-f(c)}{x-c} \leq 0$ Thus

$$\lim_{x\to c} \frac{f(x)-f(c)}{x-c} \leq 0$$

If $x < c$, then $\frac{f(x)-f(c)}{x-c} \geq 0$. Thus

$$\lim_{x\to c} \frac{f(x)-f(c)}{x-c} \geq 0$$

Thus $f'(c) = 0$. The proof for minimum is similar. □

Theorem 3.5.3.3: Rolle's Theorem

Suppose that $f : [a, b] \to \mathbb{R}$ is continuous on the closed interval $[a, b]$ and differentiable at the open interval (a, b) and that $f(a) = f(b)$. Then there is a point c in the open interval where

$$f'(c) = 0$$

Proof. If f is constant, then its derivative is 0 everywhere. If f is non-constant, then by the theorem 4.2.9, f is bounded and by theorem 4.2.10, f attains a maximum and minimum in the interval. Those two points have their derivatives equal to 0. □

Theorem 3.5.3.4: The Mean Value Theorem

Suppose that $f : [a, b] \to \mathbb{R}$ is continuous on the closed interval $[a, b]$ and differentiable on the open interval (a, b). Then there is a point $c \in (a, b)$ such that

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

Proof. Consider the function $g(x) = f(x) - x\frac{f(b)-f(a)}{b-a}$. We have that

$$\begin{aligned} g(b) - g(a) &= f(b) - f(a) - (b-a)\frac{f(b)-f(a)}{b-a} \\ g(b) - g(a) &= 0 \\ g(b) &= g(a) \end{aligned}$$

Thus by Rolle's Theorem there exists a point c such that $g'(c) = 0$.

$$\begin{aligned} g'(c) &= 0 \\ f'(c) - \frac{f(b)-f(a)}{b-a} &= 0 \\ f'(c) &= \frac{f(b)-f(a)}{b-a} \end{aligned}$$

□

Theorem 3.5.3.5: Generalized Mean Value Theorem

Let f and g be continuous functions on $[a, b]$ that are differentiable on (a, b) Then there exists at least one x in (a, b) such that

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

Proof. Consider $h(x) = f(x)(g(b) - g(a)) - (f(b) - f(a))g(x)$. We have that $h(a) = h(b) = f(a)g(b) - g(a)f(b)$. By Rolle's Theorem there is a point between a and b where $h'(c) = 0$. Thus we have

$$\begin{aligned} h'(c) &= 0 \\ f'(c)(g(b) - g(a)) - g'(c)(f(b) - f(a)) &= 0 \end{aligned}$$

□

Theorem 3.5.3.6: Constant Function

If $f : (a, b) \to \mathbb{R}$ is differentiable and $f'(x) = 0$ for all $x \in (a, b)$ then f is constant.

Proof. Suppose that $h, k \in (a, b)$. By the Mean Value Theorem, there exists $c \in (h, k)$ such that

$$f'(c) = \frac{f(k) - f(h)}{k - h}$$

Since $f'(c) = 0$ we have that $f(h) = f(k)$ Since this is true for every $h, k \in (a, b)$, we have that

$$f(x) = m$$

for some $m \in \mathbb{R}$ for all $x \in (a, b)$. □

Theorem 3.5.3.7

Let $f : (a, b) \to \mathbb{R}$ be differentiable on (a, b).

- If $\forall x \in I$, $f'(x) > 0$ then f is strictly increasing on the interval
- If $\forall x \in I$, $f'(x) < 0$ then f is strictly decreasing on the interval
- If $\forall x \in I$, $f'(x) \leq 0$ then f is increasing on the interval
- If $\forall x \in I$, $f'(x) \geq 0$ then f is decreasing on the interval

Proof. We demonstrate the proof only of the first item. $f'(x) > 0$ implies $\lim_{x\to c} \frac{f(x)-f(c)}{x-c} > 0$. If $x > c$, we must have $f(x) > f(c)$. If $x < c$, we must have $f(x) < f(c)$. Thus $f(x)$ is strictly increasing. □

Theorem 3.5.3.8: L'Hopital's Rule

Suppose that $f, g : (a, b) \to \mathbb{R}$ are differentiable functions and $c \in (a, b)$. Suppose that

$$\lim_{x\to c} \frac{f'(x)}{g'(x)} = L$$

exists. Then if $\lim_{x\to c} f(x) = \lim_{x\to c} g(x) = 0$ then

$$\lim_{x\to c} \frac{f(x)}{g(x)} = L$$

Proof. Suppose that $\lim_{x\to c} \frac{f'(x)}{g'(x)}$ exists. Then $g'(x_n) \neq 0$ for any sequence of x_n that converges to c. Thus we can apply Cauchy's MVT. Since $f(c) = g(c) = 0$,

$$\lim_{x\to c} \frac{f(x)}{g(x)} = \lim_{x\to c} \frac{f(x) - f(c)}{g(x) - g(c)}$$

By Cauchy's MVT, there exists t between x and c such that $\frac{f(x)-f(c)}{g(x)-g(c)} = \frac{f'(t)}{g'(t)}$. As $x \to c$, $t \to c$. Thus we have that

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

□

3.5.4 More on Power Series

Theorem 3.5.4.1

Let $\sum_{k=0}^{\infty} a_k x^k$ be a power series with radius of convergence R. Then f is differentiable on $(-R, R)$ and

$$f'(x) = \sum_{k=1}^{\infty} k a_k x^{k-1}$$

Proof. We first show that $\sum_{k=1}^{\infty} k a_k x^{k-1}$ has the same radius of convergence. We have that $\sum_{k=0}^{\infty} |a_k| x^k$ has the same radius of convergence. Choose x and y such that $x < y < R$. Then $\sum_{k=0}^{\infty} a_k x^k$ and $\sum_{k=0}^{\infty} a_k y^k$ converges.

$$\begin{aligned}\sum_{k=1}^{\infty} k a_k x^{k-1} &< \sum_{k=1}^{\infty} |a_k|(y^{k-1} + y^{k-2}x + \cdots + x^{k-1}) \\ &= \sum_{k=0}^{\infty} |a_k| \frac{y^k - x^k}{y - x}\end{aligned}$$

Since $\sum_{k=0}^{\infty} |a_k| \frac{y^k - x^k}{y-x}$ converges, we complete our lemma.

Now for the main proof. Choose T such that $|x| < T < R$. We have proved that $\sum_{k=1}^{\infty} k a_k x^{k-1}$ converges so given $\epsilon > 0$ there is a number N so that

$$\sum_{k=N+1}^{\infty} k|a_k| T^{k-1} < \frac{\epsilon}{3}$$

Now if $0 < |y - x| < T - |x|$, we have $|y| < T$ and $|x| < T$. Thus we have that

$$\left|\sum_{k=N+1}^{\infty} k a_k x^{k-1} \leq \sum_{k=N+1}^{\infty} k|a_k||x|^{k-1} < \frac{\epsilon}{3}\right|$$

and also

$$\begin{aligned}\left|\sum_{k=N+1}^{\infty} a_k \frac{y^k - x^k}{y - x}\right| &= \left|\sum_{k=N+1}^{\infty} a_k(y^{k-1} + y^{k-2}x + \cdots + x^{k-1})\right| \\ &\le \sum_{k=N+1}^{\infty} |a_k|(|y|^{k-1} + \cdots + |x|^{k-1}) \\ &\le \sum_{k=N+1}^{\infty} k|a_k|T^{k-1} \\ &\le \frac{\epsilon}{3}\end{aligned}$$

The sum

$$\sum_{k=1}^{N} a_k(y^{k-1} + y^{k-2}x + \cdots + x^{k-1})$$

is a polynomial in y whose value at x is $\sum_{k=1}^{N} ka_kx^{k-1}$ so there exists $\delta_0 > 0$ such that $|y - x| < \delta_0$ implies

$$\begin{aligned}\left|\sum_{k=1}^{N} a_k \frac{y^k - x^k}{y - x} - \sum_{k=1}^{N}\right| ka_kx^{k-1} &= \left|\sum_{k=1}^{N} a_k(y^{k-1} + y^{k-2}x + \cdots + x^{k-1}) - \sum_{k=1}^{N} ka_kx^{k-1}\right| \\ &\le \frac{\epsilon}{3}\end{aligned}$$

Finally, choose $\delta = \min(\delta_0, T - |x|)$ then $|y - x| < \delta$ implies

$$\begin{aligned}\left|\sum_{k=1}^{\infty} a_k \frac{y^k - x^k}{y - x} - \sum_{k=1}^{\infty} ka_kx^{k-1}\right| &\le \left|\sum_{k=N+1}^{\infty} a_k \frac{y^k - x^k}{y - x}\right| + \left|\sum_{k=1}^{N} a_k \frac{y^k - x^k}{y - x} - \sum_{k=1}^{N} ka_kx^{k-1}\right| \\ &+ \left|\sum_{k=N+1}^{\infty} ka_kx^{k-1}\right| \\ &\le \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} \\ &= \epsilon\end{aligned}$$

□

3.5.5 Taylor Series

Definition 3.5.5.1: Taylor Series

Let $f : (a, b) \to \mathbb{R}$ and $c \in (a, b)$. If f possesses derivatives of all orders at c, then the series

$$\sum_{k=0}^{\infty} \frac{f^{(k)}(c)}{k!}(x - c)^k$$

is called the Taylor Series for f at c.

Definition 3.5.5.2: Remainder

Let $f : (a, b) \to \mathbb{R}$ be n times differentiable and $c, x \in (a, b)$. Then define the remainder to be

$$R_n(x) = f(x) - \sum_{k=0}^{n-1} \frac{f^{(k)}(c)}{k!}(x - c)^{n-1}$$

Theorem 3.5.5.3

Let $f : (a, b) \to \mathbb{R}$ possess derivatives of all order at $c \in (a, b)$. Then

$$f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(c)}{k!}(x-c)^k$$

if and only if

$$\lim_{n\to\infty} R_n(x) = 0$$

Proof. Suppose that $f(x)$ is equal to its taylor series at c. Since the taylor series converges to $f(x)$, $R_n(x)$ converges to 0 by definition.

Suppose that $R_n(x)$ converges to 0. Then by the definition of $R_n(x)$ we must have $f(x)$ equal to its taylor and we are done. □

Theorem 3.5.5.4: Lagrange Remainder

If $f : (c, d) \to \mathbb{R}$ is n times differentiable and $a, b \in (c, d)$, then

$$R_n(b) = \frac{f^{(n)}(t)}{n!}(b-a)^n$$

for some t between a and b.

Proof. Define

$$g(x) = f(x) - \left(f(a) + f'(a)(x-a) + \cdots + \frac{f^{(n-1)}(a)}{(n-1)!}(x-a)^{n-1}\right)$$

$g(x)$ satisfies $g(a) = g'(a) = \cdots = g^{(n-1)}(a) = 0$. It also satisfies $g^{(n)}(x) = f^{(n)}(x)$ for all x. Define $h(x) = g(x) - g(b)\frac{(x-a)^n}{(b-a)^n}$. Then $h(a) = h'(a) = \cdots = h^{(n-1)}(a) = 0$ and $h(b) = 0$. Now since $h(a) = h(b) = 0$, there exists $t_1 \in (a, b)$ where $h'(t_1) = 0$ by Rolle's Theorem. Since $h'(t_1) = h'(a) = 0$, there exists $t_2 \in (a, t_1)$ where $h''(t_2) = 0$. Continuing this way we get that $t = t_n$ where $h^{(n)}(t) = 0$. In terms of g, we have

$$g^{(n)}(t) = g(b)\frac{n!}{(b-a)^n}$$

Rewriting the equation we have that

$$g(b) = \frac{g^{(n)}(t)}{n!}(b-a)^n = \frac{f^{(n)}(t)}{n!}(b-a)^n$$

Substituting $x = b$ into the definition of $g(x)$, we have

$$\begin{aligned} g(b) &= f(b) - \left(f(a) + f'(a)(b-a) + \cdots + \frac{f^{(n-1)}(a)}{(n-1)!}(b-a)^{n-1}\right) \\ \frac{f^{(n)}(t)}{n!}(b-a)^n &= f(b) - \left(f(a) + f'(a)(b-a) + \cdots + \frac{f^{(n-1)}(a)}{(n-1)!}(b-a)^{n-1}\right) \\ &= f(b) - \sum_{k=0}^{n-1} \frac{f^{(k)(a)}}{(k-1)!}(b-a)^{k-1} \\ &= R_n(b) \end{aligned}$$

Thus we are done. □

3.5.6 The Exponential Series

Now that we have explored most of the theorems in power series, we will give a through investigation on the exponential function and the logarithm function.

We begin by recalling the definition of the exponential function.

Definition 3.5.6.1: The Exponential Function

Let $x \in \mathbb{R}$. The series

$$e^x = \sum_{k=0}^\infty \frac{x^k}{k!}$$

is called the exponential series

Theorem 3.5.6.2

$f(x) = e^x$ is continuous for $x \in R$.

Proof. Power series are continuous. □

Theorem 3.5.6.3: Property of the Exponential

Let $x, y \in \mathbb{R}$. Then

- $e^{x+y} = e^x e^y$
- $e^x \geq 1 + x$ for all x
- $e^x \leq \frac{1}{1-x}$ for all $x < 1$
- e^x is strictly increasing

Proof. Let $x, y \in \mathbb{R}$.

- We want to show that

$$\sum_{k=0}^{2m} \frac{(x+y)^k}{k!} - \left(\sum_{i=0}^m \frac{x^i}{i!}\right)\left(\sum_{j=0}^m \frac{x^j}{j!}\right) \to 0$$

This is equivalent to saying that $e^{x+y} - e^x e^y = 0$. From the binomial theorem, we have

$$(x+y)^k = \sum_{i=0}^k \frac{k!}{i!(k-i)!} x^i y^{k-i} = \sum_{i+j=k} k! \frac{x^i}{i!} \frac{y^j}{j!}$$

Hence

$$\begin{aligned}
&\sum_{k=0}^{2m}\frac{(x+y)^k}{k!}-\left(\sum_{i=0}^{m}\frac{x^i}{i!}\right)\left(\sum_{j=0}^{m}\frac{x^j}{j!}\right)\\
&=\sum_{i+j=k}k!\frac{x^i}{i!}\frac{y^j}{j!}-\left(\sum_{i=0}^{m}\frac{x^i}{i!}\right)\left(\sum_{j=0}^{m}\frac{x^j}{j!}\right)\\
&=\sum_{i\geq m+1,i+j\leq 2m}\frac{x^i}{i!}\frac{y^j}{j!}\\
&\leq\sum_{i\geq m+1,i+j\leq 2m}\frac{|x^i|}{i!}\frac{|y^j|}{j!}\\
&=\sum_{i\geq m+1,j\geq 0}\frac{|x^i|}{i!}\frac{|y^j|}{j!}\\
&=\left(\sum_{i\geq m+1}\frac{|x^i|}{i!}\right)\left(\sum_{j=0}^{\infty}\frac{|y^j|}{j!}\right)
\end{aligned}$$

The first sum tends to 0 as m tends to infinity.

- When $x\geq 0$, $e^x=1+x+\frac{x^2}{2!}+\frac{x^3}{3!}+\cdots\geq 1+x$
- When $0\leq x<1$, we have $e^x=1+x+\frac{x^2}{2!}+\frac{x^3}{3!}+\cdots\leq 1+x+x^2+x^3+\cdots=\frac{1}{1-x}$. Now suppose that $x=-u$ is negative, then

  $$e^u\geq 1+u$$

  thus $e^{-x}\geq 1-x$. This implies that $\frac{1}{1-x}\geq e^x$ thus this equality is true for all $x<1$. Now to prove the previous inequality, suppose $x\leq -1$, then $e^x>0>1+x$. Now suppose that $-1<x<0$. Let $x=-u$ then $0<u<1$ hence $e^u\leq\frac{1}{1-u}$. This implies that $e^{-x}\leq\frac{1}{1+x}$ thus

  $$1+x\leq e^x$$

- Suppose $x<y$. Then we have

  $$\begin{aligned}e^y&=e^{y-x}e^x\\&\geq(1+y-x)e^x\\&>e^x\end{aligned}$$

□

Theorem 3.5.6.4

$f(x)=e^x$ is differentiable for $x\in R$ and the derivative is

$$f'(x)=e^x$$

Proof. Power series are differentiable and differentiating the power series for e^x gives e^x. □

Theorem 3.5.6.5

The inverse of e^x exists.

Proof. It suffices to show that e^x is injective. Since it is strictly increasing, it must be increasing. Since the range of e^x is $(0, \infty)$, the domain of the inverse is $(0, \infty)$ □

This theorem allows us to define properly the inverse of the exponetial function, namely the logarithm. We will also show that the logarithm is also a power series.

Definition 3.5.6.6: The Logarithm

Define the logarithm function as the inverse of e^x as $\ln(x)$.

Theorem 3.5.6.7

$f(x) = \ln(x)$ is continuous and differentiable for all $x \in (0, \infty)$ and the derivative is

$$f'(x) = \frac{1}{x}$$

Proof. Since e^x is continuous then $\ln(x)$ is also continuous. By theorem 5.2.2, $\ln(x)$ is differentiable. Also by theorem 5.2.2 we have $f'(x) = \frac{1}{x}$. □

Theorem 3.5.6.8

The Taylor Series for the $\ln(1 - x)$ is

$$\ln(1 - x) = -\sum_{k=1}^{\infty} \frac{x^k}{k}$$

for $-1 \leq x < 1$

Be cautious that the taylor series is only defined for $x \in [-1, 1)$. Outside of the interval, the power series will diverge. With that in mind, we will complete the chapter by defining real exponents. This is defined through both the exponential function and the logarithm.

Definition 3.5.6.9: Real Exponents

If $x > 0$ and $p \in \mathbb{R}$ define

$$x^p = e^{p \ln(x)}$$

Theorem 3.5.6.10: Law of Exponents

Let $x > 0$ and $p, q \in \mathbb{R}$.

- $x^{p+q} = x^p x^q$
- $(x^p)^q = x^{pq}$

Proof. Simply revert the real exponents in terms of e. □

3.6 Integration

Students often misunderstand the importance of integration. It originated not from begin an inverse operation of differentiation, but as a tool to find the area under the function. It just so happens that they two serve as an inverse to each other. This point will be made very clear once you encounter differentiation and integration of multiple dimensions.

Integration has two underlying theories, that can be shown equivalent. These are the Darboux Integral and the Riemann Integral. We start our rigorous study on integration with the Darboux Integral.

3.6.1 Darboux Integral

The area under the function will be approximated by small rectangles, their height will be exactly the function while the collection of their various widths, will be called a partition. Once we establish most of the proporties of paritions we will then take limits.

Definition 3.6.1.1: Partition

Let $a < b$. A partition P of the interval $[a, b]$ is a finite collection of points in $[a, b]$, one of which is a and one of which is b. We write $P = \{a = t_0 < t_1 < t_2 < \cdots < t_n = b\}$ and $I_k = [t_{k-1}, t_k]$.

Do note that a partition does not have to partition the x-axis equally. This means that rectangles can have different lengths.

For the heights, we obtain two approximations for the area, one that uses the shortest rectangle, and the other that uses the longest rectangles.

Definition 3.6.1.2: Lower and Upper Sum

Suppose f is bounded on $[a, b]$ and $P = \{t_0, \dots, t_n\}$ is a partition of $[a, b]$. Let

$$m_k = \inf\{f(x) : t_{k-1} \leq x \leq t_k\}$$

$$M_k = \sup\{f(x) : t_{k-1} \leq x \leq t_k\}$$

The lower sum of f for P, denoted by $L(f, P)$, is defined as

$$L(f, P) = \sum_{k=1}^{n} m_k(t_k - t_{k-1}) = \sum_{k=1}^{n} m_k|I_k|$$

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

$$U(f, P) = \sum_{k=1}^{n} M_k(t_k - t_{k-1}) = \sum_{k=1}^{n} M_k|I_k|$$

As indicated by the notation, the lower and upper sum not only depends on the function, but it also depends on the partition one takes since the supremum and infinum of each particular rectangle would be different if two different partitions are chosen.

We now give an obvious proposition.

Proposition 3.6.1.3

Suppose that f is a function and P is a partition. Then

$$L(f, P) \leq U(f, P)$$

Proof. For all $k \in \{1, \ldots, n\}$, we have $m_k \leq M_k$. Thus

$$\sum_{k=1}^{n} m_k(t_k - t_{k-1}) \leq \sum_{k=1}^{n} M_k(t_k - t_{k-1})$$

and we are done. □

The above proposition states universally that the lower sum must be less than the upper sum. We will later see that the choice of parition need not matter, the lower sum will always be less than the upper sum. To formally prove that, we need the notion of a "better" partition.

Definition 3.6.1.4: Refinement

A partition $Q = \{a = s_0 < s_1 < s_2 < \cdots < s_m = b\}$ is a refinement of $P = \{a = t_0 < t_1 < t_2 < \cdots < t_n = b\}$ if every interval in P is the union of one or more interval in Q. We write $P \subseteq Q$ in this case.

Be careful with the notion of refinement. It requires the intervals in P to be unions of intervals of Q. This means that there can be two partitions that are not refinements of one and another. One way to think of refinements easier is that every point in P must be contained in Q in order for intervals in P to be union of intervals in Q, which is why we use a subset notation to indicate refinements.

The above proposition is also obvious. As we make better approximations, the lower and upper sums get closer, which as you can guess, once the lower sum and upper sum are equal, that number will be the area under the function.

Proposition 3.6.1.5

Suppose that P, Q are partitions. If $P \subseteq Q$, then

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

Proof. Let $P = \{t_0, \ldots, t_n\}$. Let Q be a refinement of P. Then there exists some interval in P, say $[t_{i-1}, t_i]$ such that it is equal to $[s_{j-1}, s_j] \cup [s_j, s_{j+1}]$ for some j. Then $m_i \leq m_j, m_{j+1}$. Thus $m_i(t_i - t_{i-1}) \leq m_j(t_i - t_{i-1}), m_{j+1}(t_i - t_{i-1})$. Summing all these m_i and m_j we have that $L(f, P) \leq L(f, Q)$. The proof is mirrored for $U(f, Q) \leq U(f, P)$ □

We will need a lemma in order to prove the required result.

Lemma 3.6.1.6

For any two partitions P_1, P_2, there exists a partition P such that $P_1 \subseteq P$ and $P_2 \subseteq P$.

Proof. Let $P_1 = \{t_0, \ldots, t_m\}$ and $P_2 = \{s_0, \ldots, s_n\}$. Then define P by ordering all the elements of P_1 and P_2 from smallest to largest. Then $P_1 \subseteq P$ and $P_2 \subseteq P$. □

Below is the proposition we need to prove a lot of things. The trick is to make the common refinement so that comparison can be made between them.

Proposition 3.6.1.7

Suppose that P_1 and P_2 are any two partitions of $[a, b]$. Then

$$L(f, P_1) \leq U(f, P_2)$$

Proof. Let Q be the common refinement of P_1 and P_2. Then

$$L(f,P_1)\leq L(f,Q)\leq U(f,Q)\leq U(f,P_2)$$

□

Finally, we can develop limits with better approximations to reach our goal. In particular, we have the following fact.

Lemma 3.6.1.8

Suppose that f is a function and P any partition of an interval. Then

$$\sup_P\{L(f,P)\}\leq\inf_P\{U(f,P)\}$$

Proof. Suppose that $T=\sup_P\{L(f,P)\}$. By lemma 6.1.5, we have that all upper sums are upper bounds of all low sums. Thus $T\leq U(f,P)$ for any P. This means that T is a lower bound for $U(f,P)$ thus $\sup_P\{L(f,P)\}=T\leq\inf_P\{U(f,P)\}$ □

We now give the proper notion of integrability with Darboux Integrals.

Definition 3.6.1.9: Darboux Integrable Functions

A function f which is bounded on $[a,b]$ is Darboux integrable on $[a,b]$ if $\sup_P\{L(f,P)\}=\inf_P\{U(f,P)\}$. In this case, the common number is called the integral of f on $[a,b]$ and is denoted by

$$\int_a^b f$$

Notice that the variable x is not present in the integral representation of the equivalent supremum and infinum to indicate the number is the same regardless of what we take for dx, which is the length of the triangle.

We then have an equivalent characterization of intergability to work on.

Theorem 3.6.1.10

If f is bounded on $[a,b]$, then f is integrable on $[a,b]$ if and only if for every $\epsilon>0$ there exists a partition P of $[a,b]$ such that $U(f,P)-L(f,P)<\epsilon$.

Proof. Suppose that f is integrable. Let $T=\sup_P\{L(f,P)\}=\inf_P\{U(f,P)\}$. By the approximation property, fix $\frac{\epsilon}{2}>0$. There exists P_1,P_2 such that

$$T-\frac{\epsilon}{2}<L(f,P_1)\leq T$$

and

$$T\leq U(f,P_2)<T+\frac{\epsilon}{2}$$

Define a new partition P_3 such that $P_1\subseteq P_3$ and $P_2\subseteq P_3$, then

$$T-\frac{\epsilon}{2}<L(f,P_1)\leq L(f,P_3)\leq T$$

and

$$T\leq U(f,P_3)\leq U(f,P_2)<T+\frac{\epsilon}{2}$$

Then we have $T - L(f, P_3) < \frac{\epsilon}{2}$ and $U(f, P_3) - T < \frac{\epsilon}{2}$ and

$$U(f, P_3) - L(f, P_3) < \epsilon$$

Now suppose that for every $\epsilon > 0$, there exists P such that $U(f, P) - L(f, P) < \epsilon$. Fix $\epsilon > 0$. Then $U(f, P) - L(f, P) < \epsilon$ implies $\inf_P\{U(f, P)\} - \sup_P\{L(f, P)\} < \epsilon$. Thus for all $\epsilon > 0$, $\inf_P\{U(f, P)\} - \sup_P\{L(f, P)\} < \epsilon$ Thus we must have $\inf_P\{U(f, P)\} - \sup_P\{L(f, P)\}$ less than every positive number and larger than every negative number. Thus $\sup_P\{L(f, P)\} = \inf_P\{U(f, P)\}$. □

This epsilon definition makes it easier to prove things since we are very familiar with closeness and limits. Often the supremum and infinum will be hard to compute which is why we will often resort to this theorem.

To end this section, we will discover another useful characterization of integrability. We start with a new definition on partitions.

Definition 3.6.1.11: Mesh

The mesh of a partition P is the maximum length of the subintervals of p. This means that $\text{mesh}(P) = \max\{t_k - t_{k-1} : k = 1, 2, \dots, n\}$

Now we have the following equivalent characterization oif Darboux Integrability.

Theorem 3.6.1.12

A bounded function f on $[a, b]$ is integrable if and only if for each $\epsilon > 0$ there exists a $\delta > 0$ such that

$$\text{mesh}(P) < \delta \implies U(f, P) - L(f, P) < \epsilon$$

for all partitions P of $[a, b]$.

Proof. The converse is trivial. Since there exists a $\delta > 0$ such that the condition hold, choose one of such P will complete the proof using the epsilon delta definition of Darboux Integrability.

Suppose that f is Darboux Integrable. Let $\epsilon > 0$, select a partition P_0 such that

$$U(f, P_0) - L(f, P_0) < \frac{\epsilon}{2}$$

Since f is bounded, there exists $B > 0$ such that $|f(x)| < B$ for all x. Let $\delta = \frac{\epsilon}{8mB}$, where m is the number of partitions of P_0. Consider any partition P. We want to prove that $U(f, P) - L(f, P) < \epsilon$. Let Q be the common refinement of P and P_0. I claim that

$$L(f, Q) - L(f, P) \le 2mB \cdot \text{mesh}(P) < 2mB\delta = \frac{\epsilon}{4}$$

When $m = 1$, Suppose that the refinement of P by P_0 is located at $[t_{k-1}, t_k]$. Then

$$\begin{aligned} L(f, Q) - L(f, P) &= \inf\{f(t) : t_{k-1} \le t \le u\}(u - t_{k-1}) + \inf\{f(t) : u \le t \le t_k\}(t_k - u) \\ &\quad - m_k(t_k - t_{k-1}) \\ &\le B \cdot \text{mesh}(P) - (-B) \cdot \text{mesh}(P) && (-B \text{ is a minimum of } f(x)) \\ &< 2B\text{mesh}(P) \end{aligned}$$

For the case that Q has the maximum m elements not in P,

$$\begin{aligned} L(f, Q) - L(f, P) &\le 2mB \cdot \text{mesh}(P) && (\text{Same argument as the case for } m = 1) \\ &< 2mB\delta \\ &= \frac{\epsilon}{4} && (\text{constructed } \delta) \end{aligned}$$

Since Q is a refinement of P_0, we have $L(f, P_0) \leq L(f, Q)$ thus

$$L(f, P_0) - L(f, P) < \frac{\epsilon}{4}$$

By a similar argument,

$$U(f, P) - U(f, P_0) < \frac{\epsilon}{4}$$

Thus

$$U(f, P) - L(f, P) < U(f, P_0) - L(f, P_0) + \frac{\epsilon}{2}$$

From the start of the proof, we have $U(f, P_0) - L(f, P_0) < \frac{\epsilon}{2}$, thus we now have $U(f, P) - L(f, P) < \epsilon$ as desired. □

This gives some more restriction on the partition with a delta as well, and allows us to choose our partition better. These different notions of integrability will prove to be useful in different scenarious. Often in proving fundamental functions that cannot be decomposed with algebra of integrability, we will resort to these definitions. This is why it also crucial to learn how to properly find the required partitions for the epsilon, and epsilon-delta definitions.

3.6.2 Riemann Integral

The Riemann integral is more abstract thus standard lecture notes in developing the notion of integrability will most likely use the Darboux integral. Treat this section as a bonus.

Definition 3.6.2.1: Tagged Partition

A tagged partition is a partition $P = \{t_0, \dots, t_n\}$ such that every $[t_{k-1}, t_k]$ is associated with a $x_k \in [t_{k-1}, t_k]$.

Definition 3.6.2.2: Riemann Sum

Let f be a bounded function on $[a, b]$ and let $P = \{a = t_0 < t_1 < \cdots < t_n = b\}$. A Riemann sum of f associated with the tagged partition P is a sum of the form

$$S(f, P) = \sum_{k=1}^{n} f(x_k)(t_k - t_{k-1}) = \sum_{k=1}^{n} x_k |I_k|$$

Definition 3.6.2.3: Riemann Integrable Functions

The function f is Riemann Integrable on $[a, b]$ if there exists a number L such that for every $\epsilon > 0$, there exists δ such that $\text{mesh}(P) < \delta$ implies $|S(f, P) - L| < \epsilon$ where P is a tagged partition.

As one can see, riemann integrability is more abstract in the sense that at least the Darboux integrability explicitly gives what the limit is. But here not only is the limit made implicitly, even the choice of the tag in the tagged partition is arbitrary.

Theorem 3.6.2.4

A bounded function f on $[a, b]$ is Riemann Integrable if and only if it is Darboux Integrable. In this case, both methods produce the same sum.

Proof. Suppose that f is Darboux Integrable. Then for any tagged partition P',

$$L(f, P) \leq S(f, P') \leq U(f, P)$$

Since it is Darboux Integrable,

$$U(f,P) < L(f,P) + \epsilon \leq \sup_P\{L(f,P)\} + \epsilon = \int_a^b f + \epsilon$$

and

$$L(f,P) > U(f,P) - \epsilon \geq \inf_P\{U(f,P)\} - \epsilon = \int_a^b f - \epsilon$$

Thus we have

$$\left|S(f,P') - \int_a^b f\right| < \epsilon$$

Suppose that f is a bounded function that is Riemann Integrable and that converges to L. Let P be any partition of $[a,b]$. Then for every $\epsilon > 0$, there exists a δ such that $\text{mesh}(P) < \delta$ implies $|S(f,P) - L| < \epsilon$. In particular, choose a partition P such that $\text{mesh}(P) < \delta$. For each interval in P, choose the tag x_k such that

$$f(x_k) < m_k + \epsilon$$

Thus we have

$$S(f,P) \leq L(f,P) + \epsilon(b-a)$$

and

$$|S(f,P) - L| < \epsilon$$

by definition of Riemann Integrability. Thus

$$\begin{aligned}\sup_P\{L(f,P)\} &\geq L(f,P)\\ &\geq S(f,P) - \epsilon(b-a)\\ &> L - \epsilon - \epsilon(b-a)\end{aligned}$$

Thus we have $L \leq \sup_P\{L(f,P)\}$. Similarly we have $L \geq \inf_P\{U(f,P)\}$. Thus we have

$$\sup_P\{L(f,P)\} = L = \inf_P\{U(f,P)\}$$

□

3.6.3 Properties of the Riemann Integral

This section gives us instant knowledge on whether a function is integrable, once we also have integrability of the basic functions.

We start the section with two sufficient and powerful conditions for integrability.

Theorem 3.6.3.1

Every monotonic function f on $[a,b]$ is integrable.

Proof. Consider a uniform partition P. We have

$$\begin{aligned}
U(f,P)-L(f,P) &= \sum_{k=1}^{n} M_k(t_k-t_{k-1}) - \sum_{k=1}^{n} m_k(t_k-t_{k-1}) \\
&\leq \frac{b-a}{n}\sum_{k=1}^{n}\left(f\left(a+\frac{k}{n}(b-a)\right)-f\left(a+\frac{k-1}{n}(b-a)\right)\right) \\
&= \frac{b-a}{n}(f(b)-f(a))
\end{aligned}$$

Thus given ϵ we can choose n large enough such that $\frac{(b-a)(f(b)-f(a))}{n}<\epsilon$. □

Theorem 3.6.3.2

Every continuous function $f:[a,b]\to\mathbb{R}$ is integrable.

Proof. Suppose that f is continuous. Then f is uniformly continuous in $[a,b]$. Pick a partition P such that $\text{mesh}(P)<\delta$, where δ is chosen from an ϵ from uniform continuity. Since f is continuous is in a closed interval, f is bounded and a maximum and minimum is achieved. For any M_k and m_k,

$$M_k - m_k = f(x_k)-f(y_k) < \frac{\epsilon}{b-a}$$

by uniform continuity, where $x_k, y_k \in [t_{k-1}, t_k]$. Hence

$$\begin{aligned}
U(f,P)-L(f,P) &= \sum_{k=1}^{n} M_k(t_k-t_{k-1}) - \sum_{k=1}^{n} m_k(t_k-t_{k-1}) \\
&= \sum_{k=1}^{n}(M_k-m_k)(t_k-t_{k-1}) \\
&< \sum_{k=1}^{n}\frac{\epsilon}{b-a}(t_k-t_{k-1}) \\
&= \epsilon
\end{aligned}$$

□

The above two theorems already provide a rich foundation of integrable functions. These include the typical trigonometric functions, exponential functions, polynomials and more. However do be careful that being integrable does not means that you can alawys find a nice expression for the result, or even an existence of an antiderivative, as we will see in other sections.

We also have the basic addition and scalar multiplication of integrable functions.

Theorem 3.6.3.3: Algebra of Integrals

Let f,g be integrable functions on $[a,b]$, and let $c\in\mathbb{R}$. Then

- cf is integrable and $\int_a^b cf = c\int_a^b f$
- $f+g$ is integrable and $\int_a^b (f+g) = \int_a^b f + \int_a^b g$

Proof. We first note that

$$\sup_P\{L(cf,P)\} = \sup_P\{c\cdot L(f,P)\} = c\sup_P\{L(f,P)\}$$

and

$$\inf_P\{(U(cf,P)\} = \inf_P\{c \cdot U(f,P)\} = c\inf_P\{U(f,P)\}$$

if $c > 0$. The first equality is achieved since $\sum_{k=1}^n cm_k(t_k - t_{k-1}) = c\sum_{k=1}^n m_k(t_k - t_{k-1})$. The second equality is just a property of the supremum and infinum. This results in $\sup_P\{L(cf,P)\} = \inf_P\{(U(cf,P)\}$ by integrability of f. Thus cf is integrable. From those equality we can also deduce that

$$\int_a^b cf = \inf\{(U(cf,P)\} = c\inf\{U(f,P)\} = c\int_a^b f$$

The negative version can be proven just by setting $c = -1$. Note that $\sup_P(-f) = -\inf_P(f)$ and $\inf_P(-f) = -\sup_P(f)$. Thus $U(-f,P) = -L(f,P)$ and $L(-f,P) = -U(f,P)$ and

$$\inf\{U(-f,P)\} = -\sup\{L(f,P)\} = -\inf\{U(f,P)\} = \sup\{L(-f,P)\}$$

and we are done.

For the sum rule, note that $\sup(f+g) \leq \sup(f) + \sup(g)$ and thus

$$U(f+g,P) \leq U(f,P) + U(g,P)$$

and similarly

$$L(f,P) + L(g,P) \leq L(f+g,P)$$

Also since $U(f,P) - L(f,P) < \frac{\epsilon}{2}$ and $U(g,P) - L(g,P) < \frac{\epsilon}{2}$ for some P, we have $U(f+g,P) - L(f+g,P) < \epsilon$. Thus we have $f+g$ is integrable.

Now

$$L(f,P) + L(g,P) \leq L(f+g,P) \leq U(f+g,P) \leq U(f,P) + U(g,P)$$

thus $\int_a^b f + g = \int_a^b f + \int_a^b g$ □

Similar to sequences, we also have inequalities between values of the integral and the value of the function.

Theorem 3.6.3.4: Monotinicity of the Integral

If f, g are intergrable on $[a,b]$ and $f(x) \leq g(x)$ for all $x \in [a,b]$. Then $\int_a^b f \leq \int_a^b g$.

Proof. We note that since $g - f \geq 0$, $U(g-f,P) > \inf\{U(g-f,P)\} > 0$ for any partition P. By the algebra of integrals, $g - f$ is integrable and $\inf\{U(g-f,P)\} = \int_a^b g - f \geq 0$ which implies

$$\int_a^b f \leq \int_a^b g$$

□

With the above theorem in play, we can give the integral version of intermediate value theorem. But before that, we need a proposition.

Proposition 3.6.3.5

Let $f : [a,b] \to \mathbb{R}$ be integrable. Let $m = \inf_{x\in[a,b]}\{f(x)\}$ and $M = \sup_{x\in[a,b]}\{f(x)\}$. Then

$$m(b-a) \leq \int_a^b f \leq M(b-a)$$

Proof. Let $g = \sup\{f\}$ and $h(x) = \inf\{f\}$ be constant functions. Then $h(x) \leq f(x) \leq g(x)$ for all $x \in [a,b]$. By the above theorem, we have our desired inequality. □

This can be made clear with a graph, the proposition simply states that the smallest rectangle above the function and the largest rectangle under the function bounds the area of the integral. Now we can give the intermediate value theorem.

Theorem 3.6.3.6: Intermediate Value Theorem for Integration

Let $f : [a,b] \to \mathbb{R}$ be a continuous function. Then there exists $c \in [a,b]$ such that

$$f(c) = \frac{1}{b-a}\int_a^b f$$

Proof. By the above proposition, we have

$$m \leq \frac{1}{b-a}\int_a^b f \leq M$$

Since f is continuous in the bounded interval $[a,b]$, f attains its maximum and minimum. Thus by the IVT, f must attain $\frac{1}{b-a}\int_a^b f$ for some $c \in (a,b)$. □

Theorem 3.6.3.7

If f is integrable on $[a,b]$ then $|f|$ is integrable on $[a,b]$ and

$$\left|\int_a^b f\right| \leq \int_a^b |f|$$

Proof. Since we have the inequality $\sup\{|f|\} - \inf\{|f|\} \leq \sup\{f\} - \inf\{f\}$, we have

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

and thus $|f|$ is integrable. Now by monotinicity, $-|f| \leq f \leq |f|$ and thus

$$\left|\int_a^b f\right| \leq \int_a^b |f|$$

□

Theorem 3.6.3.8

Let f be a function defined on $[a,b]$. f is integrable on $[a,c]$ and $[c,b]$ with $c \in (a,b)$ if and only if f is integrable on $[a,b]$ and

$$\int_a^b f = \int_a^c f + \int_c^b f$$

Proof. First suppose that $f : [a,b] \to \mathbb{R}$ is integrable. Then for every $\epsilon > 0$ there exists a partition P such that $U(f,P) - L(f,P) < \epsilon$. Let P_c be the refinement of P that contains c. Let

Q be the partition of $[a, c]$ induced by P_c and R be the partition at $[c, b]$. Then we have

$$U(f, P_c) = U(f, Q) + U(f, R)$$

and

$$L(f, P_c) = L(f, Q) + L(f, R)$$

which means

$$U(f, Q) - L(f, Q) + U(f, R) - L(f, R) = U(f, Q) - L(f, Q) < \epsilon$$

Note that the first two terms combined on the left hand side are positive and the last two terms as well thus they each must be less than ϵ and thus is integrable.

Now let $f : [a, c] \to \mathbb{R}$ and $f : [c, b] \to \mathbb{R}$ be integrable. For every $\frac{\epsilon}{2}$ there exists partitions Q and R such that

$$U(f, Q) - L(f, Q) < \frac{\epsilon}{2}$$

and

$$U(f, R) - L(f, R) < \frac{\epsilon}{2}$$

Take P to be the common refinement of Q and R. Then

$$U(f, Q) - L(f, Q) = U(f, Q) - L(f, Q) + U(f, R) - L(f, R) < \epsilon$$

thus f is integrable on $[a, b]$.

Note that we have

$$\begin{aligned}\int_a^b f &\leq U(f, P)\\ &= U(f, Q) + U(f, R)\\ &\leq L(f, Q) + L(f, R) + \epsilon\\ &\leq \int_a^c f + \int_c^b f + \epsilon\end{aligned}$$

and

$$\begin{aligned}\int_a^b &\geq L(f, P)\\ &= L(f, Q) + L(f, R)\\ &\geq U(f, Q) + U(f, R) - \epsilon\\ &\geq \int_a^c f + \int_c^b f - \epsilon\end{aligned}$$

Thus we have

$$\left(\int_a^c f + \int_c^b f\right) - \epsilon \leq \int_a^b f \leq \left(\int_a^c f + \int_c^b f\right) + \epsilon$$

□

The following theorem shows that composing a Riemann integrable function with a continuous function gives another Riemann integrable function. Mind the differences of this theorem and the theorem stated for the substitution rule in the later chapters.

Theorem 3.6.3.9

Let $f : [a,b] \to \mathbb{R}$ be a Riemann Integrable function and $g : \mathbb{R} \to \mathbb{R}$ is a continuous function. Then $g \circ f$ is Riemann Integrable.

Proof. Since f is integrable, f is bounded, thus we only need to consided $g : [-M, M] \to \mathbb{R}$. On this bounded interval, g is uniformly continuous and bounded, say $|g(x)| \leq K$.

Let $\epsilon > 0$, set $\epsilon' = \frac{\epsilon}{2(b-a)}$. By uniform continuity, there exists $\delta > 0$ such that $|x - y| < \delta$ implies $|g(x) - g(y)| < \epsilon'$. Choose $\nu = \frac{\delta}{4K}\epsilon$, by integrability of f, there exists Q such that $U(f,Q) - L(f,Q) < \nu$. Let $h = g \circ f$ for convenience. Now consider $U(h,Q) - L(h,Q)$

$$\begin{aligned}
U(h,Q) - L(h,Q) &= \sum_{k=1}^{n} (\sup_{I_k}(h) - \inf_{I_k}(h))|I_k| \\
&= \sum_{\substack{\sup_{I_k}(f) \\ -\inf_{I_k}(f) < \delta}} \left(\sup_{I_k}(h) - \inf_{I_k}(h)\right)|I_k| + \sum_{\substack{\sup_{I_k}(f) \\ -\inf_{I_k}(f) \geq \delta}} \left(\sup_{I_k}(h) - \inf_{I_k}(h)\right)|I_k| \\
&\leq \sum_{\substack{\sup_{I_k}(f) \\ -\inf_{I_k}(f) < \delta}} \epsilon'|I_k| + \sum_{\substack{\sup_{I_k}(f) \\ -\inf_{I_k}(f) \geq \delta}} (\sup_{I_k}(h) - \inf_{I_k}(h))|I_k| \\
&\qquad\qquad \text{(By the } \epsilon' \text{ from uniform continuity)} \\
&\leq \sum_{\substack{\sup_{I_k}(f) \\ -\inf_{I_k}(f) < \delta}} \epsilon'|I_k| + \sum_{\substack{\sup_{I_k}(f) \\ -\inf_{I_k}(f) \geq \delta}} 2K|I_k| \qquad \text{(By boundedness of } g\text{)} \\
&\leq \epsilon' \sum_{k=1}^{n} |I_k| + 2K \sum_{\substack{\sup_{I_k}(f) \\ -\inf_{I_k}(f) \geq \delta}} |I_k| \\
&= \epsilon'(b-a) + 2K \sum_{\substack{\sup_{I_k}(f) \\ -\inf_{I_k}(f) \geq \delta}} |I_k|
\end{aligned}$$

Now as a side note, observe that

$$\begin{aligned}
\sum_{\substack{\sup_{I_k}(f) \\ -\inf_{I_k}(f) \geq \delta}} |I_k| &= \frac{1}{\delta} \sum_{\substack{\sup_{I_k}(f) \\ -\inf_{I_k}(f) \geq \delta}} \delta|I_k| \\
&< \frac{1}{\delta} \sum_{\substack{\sup_{I_k}(f) \\ -\inf_{I_k}(f) \geq \delta}} (\sup_{I_k}(f) - \inf_{I_k}(f))|I_k| \\
&\leq \frac{1}{\delta} \sum_{k=1}^{n} (\sup_{I_k}(f) - \inf_{I_k}(f))|I_k| \\
&= \frac{1}{\delta} \sum_{k=1}^{n} (M_k - m_k)|I_k| \\
&= \frac{1}{\delta}(U(f,Q) - L(f,Q)) \\
&< \frac{\nu}{\delta}
\end{aligned}$$

Thus continuing from the main calculation,

$$\begin{aligned}\epsilon'(b-a) + 2K \sum_{\substack{\sup_{I_k}(f) \\ -\inf_{I_k}(f) \geq \delta}} |I_k| &< \epsilon'(b-a) + \frac{2K\nu}{\delta} \\ &= \frac{\epsilon}{2} + \frac{\epsilon}{2} \\ &= \epsilon\end{aligned}$$

Thus we are done. □

Theorem 3.6.3.10

Let $f, g : [a, b] \to \mathbb{R}$ be Riemann Integrable functions. Then fg and $\frac{f}{g}$ are both Riemann Integrable, with the second one provided that $\frac{1}{g}$ is bounded.

Proof. Note that if f is integrable, then f^2 is integrable by the above theorem, and choosing $g(x) = x^2$. Now we have that

$$fg = \frac{1}{2}((f+g)^2 - f^2 - g^2)$$

by the algebra of integrable functions, fg is integrable. For the quotient rule, note that if g is bounded, then there exists $\epsilon > 0$ such that $\epsilon < |g|$. Now consider

$$h(x) = \begin{cases} \frac{1}{x} & \text{if } |x| > \epsilon \\ \frac{x}{\epsilon^2} & \text{if } |x| \leq \epsilon \end{cases}$$

Notice that $h \circ g = \frac{1}{g}$. Since h is continuous and g is integrable, $\frac{1}{g}$ is integrable by the above theorem. □

The comoposition rule, although lengthy in proof, is extremely powerful as once can see. We are unable to prove the product and quotient rule without the fundamental theorem of calculus and so we will begin with it on the next section.

3.6.4 Fundamental Theorem of Calculus

Theorem 3.6.4.1: Fundamental Theorem of Calculus I

Suppose that $f : [a, b] \to \mathbb{R}$ is an integrable function and define the function $F : [a, b] \to \mathbb{R}$ by

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

Then F is continuous at $[a, b]$. Also if f is continuous at $c \in [a, b]$ then $F'(c) = f(c)$.

Proof. Note that since f is integrable, f is bounded by say M. Then

$$F(x+h) - F(x) = \int_x^{x+h} f(t)\, dt$$

by linearity. Considering the horizontal length and vertical length, we have

$$\begin{aligned}|F(x+h) - F(x)| &= \left| \int_x^{x+h} f(t)\, dt \right| \\ &\leq M|h|\end{aligned}$$

Thus F is continuous by choosing $\delta = \epsilon$. Now

$$\lim_{h\to 0} \frac{1}{h} \int_x^{x+h} f(t)\,dt - f(x) = \lim_{h\to 0} \frac{1}{h} \int_x^{x+h} (f(t) - f(x))\,dt$$

We need to show that this is equal to 0. Now since f is continuous, given $\epsilon > 0$, there exists $\delta > 0$ such that $|x - y| < \delta$ implies $|f(x) - f(y)| < \epsilon$. Therefore, choose $y = x + h$ and x to be x. For $|x + h - x| = |h| < \delta$,

$$\begin{aligned} \left|\frac{1}{h} \int_x^{x+h} (f(t) - f(x))\,dt\right| &\leq \left|\frac{1}{h} \int_x^{x+h} |f(t) - f(x)|\,dt\right| && \text{(By 6.3.7)} \\ &\leq \left|\frac{1}{h} \int_x^{x+h} \epsilon\,dt\right| && \text{(By Uniform Continuity)} \\ &= \epsilon \end{aligned}$$

Thus we have

$$\lim_{h\to 0} \frac{1}{h} \int_x^{x+h} (f(t) - f(x))\,dt = 0$$

and

$$\lim_{h\to 0} \frac{1}{h} \int_x^{x+h} f(t)\,dt = f(x)$$

Thus $F'(x) = f(x)$. □

Theorem 3.6.4.2: Fundamental Theorem of Calculus II

Let $F : [a, b] \to \mathbb{R}$ be a continuous function that is differentiable on (a, b) with $F' = f$. Suppose that $f : [a, b] \to \mathbb{R}$ is an integrable function. Then

$$\int_a^b f(x)dx = F(b) - F(a)$$

Proof. Consider any partition P of the interval $[a, b]$. For every interval in P,

$$\inf_{I_k}(f(x))(t_k - t_{k-1}) \leq f(c_k)(t_k - t_{k-1}) \leq \sup_{I_k}(f(x))(t_k - t_{k-1})$$

for every $c_k \in (t_{k-1}, t_k)$. Since F is continuous on $[t_{k-1}, t_k]$ and differentiable on (t_{k-1}, t_k), by the mean value theorem, there exists c_k such that $F(t_k) - F(t_{k-1}) = f(c_k)(t_k - t_{k-1})$. Thus

$$\inf_{I_k}(f(x))(t_k - t_{k-1}) \leq F(t_k) - F(t_{k-1}) \leq \sup_{I_k}(f(x))(t_k - t_{k-1})$$

and

$$L(f, P) \leq \sum_{k=1}^{n} (F(t_k) - F(t_{k-1}) \leq U(f, P)$$

and

$$L(f, P) \leq F(b) - F(a) \leq U(f, P)$$

This is true for every P thus

$$\sup_P(L(f, P)) \leq F(b) - F(a) \leq \inf_P(U(f, P))$$

and thus

$$\int_a^b f(x)\,dx = F(b) - F(a)$$

□

For the first time, we are given an explicit way of calculating the integral, given that the antiderivative exists. This is another reason why the fundamental theorem of calculus is so important. Without it, we will need to resort to the basic definitions of partitions.

Theorem 3.6.4.3

Let $f : [a, b] \to \mathbb{R}$ be an integrable function on $[a, b]$ and continuous at a. Let $x \in I_h$ and $|I_h| \to 0$. Then

$$\lim_{h\to 0} \frac{1}{|I_h|} \int_{I_h} f(t)\, dt = f(x)$$

Proof. From the last part of the proof of the FTC I, this is already proven. □

Theorem 3.6.4.4: Product Rule of Integrals

Let $f, g : [a, b] \to \mathbb{R}$ be continuous functions on $[a, b]$ that are differentiable on (a, b) and such that f', g' are integrable on $[a, b]$. Then

$$\int_a^b f(x)g'(x)\, dx = f(b)g(b) - f(a)g(a) - \int_a^b f'(x)g(x)\, dx$$

Proof. Note that by the algebra of integrable function, $f'g, fg', (fg)'$ are all integrable. Also by the FTC, we have

$$\int_a^b (fg)' = f(b)g(b) - f(a)g(a)$$

and

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

thus we have the required result. □

Theorem 3.6.4.5: Chain Rule of Integrals

Let $f : [a, b] \to \mathbb{R}$ be a differentiable function such that f' is integrable on $[a, b]$. Let g be a continuous function on $f([a, b])$. Then

$$\int_a^b g\left(f(x)\right) f'(x)\, dx = \int_{f(a)}^{f(b)} g(t)\, dt$$

Proof. Define $G(x) = \int_{f(a)}^{x} g(t)\, dt$. By the FTC, $G'(x) = g(x)$. By the chain rule of differentiation, we obtain $G'(f(x)) = g(f(x))f'(x)$. Since g is continuous and f is integrable, $g \circ f$ is integrable. Also f' is integrable. By the algebra of integrable functions, $G'(f(x))$ is integrable. Thus we have

$$\begin{aligned} \int_a^b g(f(x))f'(x)\, dx &= \int_a^b G'(f(x))\, dx \\ &= G(f(b)) - G(f(a)) && \text{(FTC)} \\ &= \int_{f(a)}^{f(b)} g(t)\, dt && \text{(FTC)} \end{aligned}$$

□

A neat trick to apply the FTC on non-standard integral bounds is to use the following formula:

$$\frac{d}{dx}\int_{a(x)}^{b(x)} f(t)\,dt = b'(x)f(b(x)) - a'(x)f(a(x))$$

This involves using the chain rule. The proof is simple with the given conditions for both the FTC and the chain rule.

3.6.5 Improper Integrals

Definition 3.6.5.1: Improper Integral

Let $f : [a,b] \to \mathbb{R}$ be Riemann Integrable for every $[c,b]$ with $a < c$. Then the improper integral of f on $[a,b]$ is defined as

$$\int_a^b f(x)\,dx = \lim_{\epsilon\to 0^+}\int_{a+\epsilon}^b f(x)\,dx$$

Similarly, if $f : [a,b] \to \mathbb{R}$ is Riemann Integrable for every $[a,c]$ with $c < b$, then the improper integral of f on $[a,b]$ is define as

$$\int_a^b f(x)\,dx = \lim_{\epsilon\to 0^+}\int_a^{b-\epsilon} f(x)\,dx$$

Definition 3.6.5.2: Improper Integrals on Interior Points

Let $f : [a,b] \to \mathbb{R}$ be a function that is integrable on $[a, c-\epsilon]$ and $[c+\delta, b]$ for all $\epsilon, \delta > 0$ where the interval makes sense. Then define the integral of f on $[a,b]$ to be

$$\int_a^b f(x)\,dx = \lim_{\epsilon\to 0^+}\int_a^{c-\epsilon} f(x)\,dx + \lim_{\delta\to 0^+}\int_{c+\delta}^b f(x)\,dx$$

Definition 3.6.5.3: Improper Integrals with Infinity

let $f : [a,\infty) \to \mathbb{R}$ be a function that is integrable for every interval $[a,c]$ with $a < c < \infty$. We define the improper integral of f on $[a,\infty]$ by

$$\int_a^\infty f(x)\,dx = \lim_{c\to\infty}\int_a^c f(x)\,dx$$

Similarly if $g : (-\infty, b] \to \mathbb{R}$ is an integrable function for every integral $[c,b]$ with $-\infty < c < b$ then we define the improper integral of g on $(-\infty, b]$ by

$$\int_{-\infty}^b g(x)\,dx = \lim_{c\to-\infty}\int_c^b g(x)\,dx$$

Finally, if $h : \mathbb{R} \to \mathbb{R}$ is a function that is integrable on every bounded interval $[a,b]$, then define the improper integral

$$\int_{-\infty}^\infty f(x)\,dx = \lim_{a\to-\infty}\int_a^c f(x)\,dx + \lim_{b\to\infty}\int_c^b f(x)\,dx$$

where c is any point in $\mathbb{R}$.

Theorem 3.6.5.4: Absolute Comparison Test

Let $f : [a,\infty) \to \mathbb{R}$ be integrable on $[a,b]$. For every $b > a$. If $\int_a^\infty |f| < \infty$, then $\int_a^\infty f$ converges. Moreover, if $g : [a,\infty) \to [0,\infty)$ is such that $|f| \leq g$ and $\int_a^\infty g < \infty$, then $\int_a^\infty f$ is absolutely convergent.

Proof. Recall from Cauchy Criterion that $\lim_{t\to\infty}\int_a^t |f|$ is finite if and only if for every $\epsilon > 0$, there exists N such that $R_1, R_2 > N$ implies

$$\left|\int_a^{R_2} |f| - \int_a^{R_1} |f|\right| = \left|\int_{R_1}^{R_2} |f|\right| < \epsilon$$

Thus we have that

$$\left|\int_{R_1}^{R_2} f\right| \leq \int_{R_1}^{R_2} |f| < \epsilon$$

So $\int_a^\infty$ converges by the Cauchy Criterion.

Similarly, we have that for every $\epsilon > 0$, there exists N such that $R_1, R_2 > N$ implies

$$\left|\int_a^{R_2} g - \int_a^{R_1} g\right| = \left|\int_{R_1}^{R_2} g\right| < \epsilon$$

Therefore

$$\left|\int_{R_1}^{R_2} |f|\right| \leq \left|\int_{R_1}^{R_2} g\right| < \epsilon$$

□

3.7 Sequences and Series of Functions

3.7.1 Uniform Convergence

Definition 3.7.1.1: Pointwise Convergence

Suppose $(f_n)_{n\in\mathbb{N}}$ is a sequence of functions defined on an interval I. Suppose that the sequence of numbers $(f_n(x))_{n\in\mathbb{N}}$ converges for every $x \in I$. We can define

$$f(x) = \lim_{n\to\infty} f_n(x)$$

for all $x \in I$. We say that $(f_n)_{n\in\mathbb{N}}$ converges to f.

Definition 3.7.1.2: Uniform Convergence

We say that $(f_n)_{n\in\mathbb{N}}$ converges uniformly on I to a function f if for every $\epsilon > 0$ there is an integer N such that $n > N$ implies

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

for all $x \in I$.

Corollary 3.7.1.3

Uniform convergence implies pointwise convergence.

Proof. Suppose that $(f_n)_{n\in\mathbb{N}}$ is uniformly convergent to f. Then in particular fix any x in its domain, its epsilon-delta definition of convergence on x is precisely the definition of pointwise convergence. □

Theorem 3.7.1.4: Cauchy Criterion

The sequence of functions $(f_n)_{n\in\mathbb{N}}$ defined on I converges uniformly if and only if for every $\epsilon > 0$ there eixsts N such that $m, n > N$ implies

$$|f_m(x) - f_n(x)| < \epsilon$$

for all x in the domain.

Proof. Suppose that $(f_n)_{n\in\mathbb{N}}$ converges uniformly to f. Then fix $\frac{\epsilon}{2} > 0$, there exists $N \in \mathbb{N}$ such that $n > N$ implies $|f_n - f| < \frac{\epsilon}{2}$ and $|f_m - f| < \frac{\epsilon}{2}$ if $m > N$. Then

$$|f_m(x) - f_n(x)| \leq |f_m(x) - f(x)| + |f_n(x) - f(x)| < \epsilon$$

Now suppose that the cauchy criterion is satisfied. Then for every fixed x, $f_n(x)$ is Cauchy in $\mathbb{R}$ and thus is convergent. This means there exists f on its domain such that at least f_n is pointwise convergent to f. From the cauchy criterion, we have that

$$f_m(x) - \epsilon < f_n(x) < f_m(x) + \epsilon$$

for all x and $n, m > N$. Since the bounds hold for all $m > N$, and we have that $f_m(x) \to f$ pointwise, we have that

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

and thus

$$|f(x) - f_n(x)| < 2\epsilon$$

for all $n > N$. □

Side note: Since we want a universal N such that it works for every $x \in I$, we can simply consider the maximum of the difference between $f_m(x)$ and $f_n(x)$, which leads to some notes developing the norm of a functionn $\|f\|_\infty = \sup_{x\in I} |f(x)|$. They are essentially the same even when substitued in the definition.

Proposition 3.7.1.5

Suppose $\lim_{n\to\infty} f_n(x) = f(x)$. Let

$$\sup_{x\in I} |f_n(x) - f(x)| = \|f_n - f\|_\infty$$

Then $(f_n)_{n\in\mathbb{N}}$ converges to f uniformly if and only if

$$\lim_{n\to\infty} \|f_n - f\|_\infty = 0$$

Proof. Suppose that $(f_n)_{n\in\mathbb{N}}$ converges to f uniformly. Then trivially $\sup_{x\in I} |f_n - f| < \epsilon$. Thus we are done.

Suppose that $\|f_n - f\| \to 0$ as $n \to \infty$. Then since $|f_n - f| \leq \|f_n - f\|_\infty < \epsilon$ in the ϵ definition, we must have $(f_n)_{n\in\mathbb{N}}$ converging to f uniformly. □

3.7.2 Uniform Convergence and Continuity

Theorem 3.7.2.1

Let $(f_n)_{n\in\mathbb{N}}$ be a sequence of continuous functions on I, and if f_n converges to f uniformly on I, then f is continuous on I.

Proof. By uniform convergence, we know that if $\frac{\epsilon}{3} > 0$, there exists $N \in \mathbb{N}$ such that $|f_n(x) - f(x)| < \frac{\epsilon}{3}$ for all $x \in I$ and $n > N$. Fix n such that $n > N$ now. I will show that f is continuous at x_0. Since $f_n(x)$ is continuous at x_0, we know that there exists δ such that $x \in (x_0 - \delta, x_0 + \delta) \cap I$ implies $|f_n(x) - f_n(x_0)| < \frac{\epsilon}{3}$. Now

$$\begin{aligned} |f(x) - f(x_0)| &\leq |f(x) - f_n(x)| + |f_n(x) - f_n(x_0)| + |f_n(x_0) - f(x_0)| \\ &< \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} \\ &= \epsilon \end{aligned}$$

Thus f is continuous at x_0. □

Proposition 3.7.2.2

Let $(f_n)_{n\in\mathbb{N}}$ be a sequence of continuous and bounded functions on $I \subseteq \mathbb{R}$ that is cauchy. Then it converges to a continuous and bounded function.

Proof. We alredy showed that uniform convergence implies continuity on its limit, we just have to show that it is bounded. Note that for all $x \in I$, the domain of f,

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

By uniformly convergence, there exists $N \in \mathbb{N}$ such that $n > N$ implies $|f_n(x) - f(x)| < 1$. For that n, since f_n is bounded, we have $|f_n| < M$ is f_n is bounded. Thus we have that $|f(x)| \leq M + 1$ for all $x \in I$. □

Definition 3.7.2.3: Continuously differentiable

Let $I \subseteq \mathbb{R}$. We use $C^k(I)$ to denote the collection of all functions that are differentiable k times and has a continuous kth derivative and has domain I.

3.7.3 Uniform Convergence and Integrability

Theorem 3.7.3.1

Suppose $(f_n)_{n\in\mathbb{N}}$ is a sequence of functions that is Riemann Integrable on $[a, b]$ that converges uniformly on $[a, b]$ to a function f. Then f is Riemann Integrable and

$$\lim_{n\to\infty} \int_a^b f_n(x)\,dx = \int_a^b \lim_{n\to\infty} f_n(x)\,dx$$

Proof. We first show that f is integrable. Fix $\frac{\epsilon}{4(b-a)} > 0$. Then there exists $N \in \mathbb{N}$ such that $n > N$ implies

$$|f_n(x) - f(x)| < \frac{\epsilon}{4(b-a)}$$

for all x. Fix $n > N$. Since f_n is integrable, given $\frac{\epsilon}{2} > 0$ there exists a partition such that $U(f_n, P) - L(f_n, P) < \frac{\epsilon}{2}$. For that P,

$$\begin{aligned}
U(f, P) - L(f, P) &= \sum_{k=1}^{n} (\sup_{I_k} f - \inf_{I_k} f)|I_k| \\
&= \sum_{k=1}^{n} (\sup_{I_k}(f - f_n + f_n) - \inf_{I_k}(f - f_n + f_n))|I_k| \\
&\leq \sum_{k=1}^{n} \left(\sup_{I_k} |f - f_n| + \sup_{I_k} f_n + \inf_{I_k} |f - f_n| - \inf_{I_k} f_n\right)|I_k| \\
&= 2\sum_{k=1}^{n} \|f_n - f\|_\infty |I_k| + \sum_{k=1}^{n} (\sup_{I_k} f_n - \inf_{I_k} f_n)|I_k| \\
&\leq 2\|f_n - f\|_\infty (b - a) + U(f_n, P) - L(f_n, P) \\
&\leq \frac{\epsilon}{2} + \frac{\epsilon}{2} \\
&= \epsilon
\end{aligned}$$

Thus we have that f is integrable.

Now note that

$$\begin{aligned}
\left|\int_a^b f_n - \int_a^b f\right| &= \left|\int_a^b (f_n - f)\right| \\
&\leq \int_a^b |f_n - f| \\
&\leq \int_a^b \|f - f_n\|_\infty \\
&= \|f_n - f\|_\infty (b - a)
\end{aligned}$$

which goes to 0 as $n \to \infty$ by the uniform convergence of f_n to f. □

3.7.4 Uniform Convergence and Differentiability

Theorem 3.7.4.1

Suppose $(f_n)_{n\in\mathbb{N}} \subset C^1([a,b])$ and it is pointwise convergent to f. If $(f_n')_{n\in\mathbb{N}}$ converges uniformly on $[a,b]$, then $f \in C^1([a,b])$ and $(f_n')_{n\in\mathbb{N}}$ in fact converges uniformly to f'. In other words, we have

$$\lim_{n\to\infty}\left(\frac{d}{dx}f_n(x)\right) = \frac{d}{dx}\left(\lim_{n\to\infty} f_n(x)\right)$$

Proof. Suppose that f_n' converges uniformly to g. By 7.3.1, we have that

$$\int_a^x g(t)\,dt = \int_a^x \lim_{n\to\infty} f_n'(t)\,dt = \lim_{n\to\infty}\int_a^x f_n'(t)\,dt$$

By the FTC, this gives

$$\int_a^x g(t)\,dt = \lim_{n\to\infty}(f_n(x) - f_n(a)) = f(x) - f(a)$$

Now since $\{f_n\}$ is continuous thus g is continuous. By the above, we also have that f is continuous. Thus the FTC implies that $\frac{d}{dx}\int_a^x g(t)\,dt = g(x)$. Thus we have that $g = f'$. □

An easy way to remember this is $(f_n)_{n\in\mathbb{N}}$ pointwise convergent, $(f_n')_{n\in\mathbb{N}}$ are all continuous and converges uniformly, we can exchange the order of differentiation and limit.

3.7.5 Series of Functions

Definition 3.7.5.1: Pointwise Convergence of Series of Functions

Let $(f_k)_{k\in\mathbb{N}}$ be a sequence of functions $f_k : [a,b] \to \mathbb{R}$. Define

$$S_n(x) = \sum_{k=1}^n f_k(x)$$

Then the series converges pointwise to $S : [a,b] \to \mathbb{R}$ if $S_n \to S$ pointwise.

Definition 3.7.5.2: Uniform Convergence of Series of Functions

Let $(f_k)_{k\in\mathbb{N}}$ be a sequence of functions $f_k : [a,b] \to \mathbb{R}$. Define

$$S_n(x) = \sum_{k=1}^n f_k(x)$$

Then the series converges uniformly to $S : [a,b] \to \mathbb{R}$ if S_n converges to S uniformly.

Theorem 3.7.5.3

Let $(f_n)_{n\in\mathbb{N}}$ such that $f_n : [a,b] \to \mathbb{R}$ is a sequence of Riemann Integrable functions. If $S_n = \sum_{k=1}^n f_k$ converges uniformly, then $\sum_{k=1}^\infty f_k$ is Riemann Integrable and

$$\int_a^b \sum_{k=1}^\infty f_k(x)\,dx = \sum_{k=1}^\infty \int_a^b f_k(x)\,dx$$

Proof. Since S_n is a finite sum of integrable functions, S_n is also integrable by additivity. By theorem 7.3.1, S is also integrable and we have

$$\lim_{n\to\infty}\int S_n = \int \lim_{n\to\infty} S_n$$

Substituting $S_n = \sum_{k=1}^{n} f_k$ and we get the result. □

Theorem 3.7.5.4: Term by Term Differentiation

Let $(f_n)_{n\in\mathbb{N}}$ such that $f_n : [a,b] \to \mathbb{R}$ is a sequence of C^1 functions. If $S_n = \sum_{k=1}^{n} f_k$ converges pointwise, and $\sum_{k=1}^{n} f_k'$ converges uniformly, then

$$\left(\sum_{k=1}^{\infty} f_k(x)\right)' = \sum_{k=1}^{\infty} f_k'(x)$$

Proof. Similar to the proof above, except that it utilizes theorem 7.4.1 instead. □

Theorem 3.7.5.5: Weierstrass M-test

Let $\{f_n\}$ be a sequence of functions $f_n : [a,b] \to \mathbb{R}$, and assume that for every n there exists $M_n > 0$ such that $|f_n(x)| \leq M_n$ for every $x \in [a,b]$ and $\sum_{k=1}^{\infty} M_k$ is finite. Then

$$S_n = \sum_{k=1}^{n} f_k$$

converges uniformly on $[a,b]$ to the function $\sum_{k=1}^{\infty} f_k$.

Proof. Since $\sum_{k=1}^{n} M_k < \infty$, given $\epsilon > 0$ there exists N such that

$$\sum_{k=m+1}^{n} M_k < \epsilon$$

for all $m, n > N$ by cauchy condition of infinite sums. Now note that

$$\begin{aligned}|S_n(x) - S_m(x)| &= \left|\sum_{k=m+1}^{n} f_k(x)\right| \\ &\leq \sum_{k=m+1}^{n} |f_k(x)| \\ &\leq \sum_{k=m+1}^{n} M_k \\ &< \epsilon\end{aligned}$$

This proves that S_n is uniformly cauchy, which implies that S_n is uniformly convergent. □

3.8 Examples

3.8.1 Integrable Functions

Example 3.8.1.1

Define

$$f(x) = \begin{cases} \frac{1}{q} & \text{if } x = \frac{p}{q} \text{ where } p, q \in \mathbb{N} \text{ and } \gcd(p, q) = 1 \\ 0 & \text{otherwise} \end{cases}$$

for $x \in [0, 1]$. Then f is integrable. This example show that non-continuous functions can be integrable.

Example 3.8.1.2

Define

$$f(x) = \begin{cases} \frac{1}{q} & \text{if } x = \frac{p}{q} \text{ where } p, q \in \mathbb{N} \text{ and } \gcd(p, q) = 1 \\ 0 & \text{otherwise} \end{cases}$$

for $x \in [0, 1]$. Also define $g(x) = 1 - f(x)$. Then $f(x) + g(x)$ is integrable and

$$\inf_P\{U(f + g, P)\} \leq \inf_P\{U(f, P)\} + \inf_P\{U(g, P)\}$$

and

$$\sup_P\{L(f + g, P)\} \geq \sup_P\{L(f, P)\} + \sup_P\{L(g, P)\}$$

This example shows the upper and lower riemann integrals of functions are not necessarily linear.

3.8.2 Uniform Convergence

Example 3.8.2.1

Define

$$f_n(x) = \frac{x}{n}$$

for $n \in \mathbb{N}$. The sequence $(f_n)_{n\in\mathbb{N}}$ is not uniformly convergent.

Example 3.8.2.2

Define

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

for $n \in \mathbb{N}$. The sequence $(f_n)_{n\in\mathbb{N}}$ is not uniformly convergent.

Example 3.8.2.3

Define

$$f_n(x) = \begin{cases} n & \text{if } 0 \leq x \leq \frac{1}{n} \\ 0 & \text{otherwise} \end{cases}$$

for $n \in \mathbb{N}$. The sequence $(f_n)_{n\in\mathbb{N}}$ is not uniformly convergent.

This example show that without uniform convergence, the limit and the integration opeartor cannot be interchanged.

Example 3.8.2.4

Define

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

for $n \in \mathbb{N}$. The sequence $(f_n)_{n\in\mathbb{N}}$ is continuously differentiable. It also converges uniformly to $f(x) = |x|$.

This example show that without uniform convergence of the derivatives of $f_n(x)$, the limit and the differentiation opeartor cannot be interchanged.

Example 3.8.2.5

Define

$$f_n(x) = \frac{1}{n}\sin\left(n^2 x\right)$$

for $n \in \mathbb{N}$. The sequence $(f_n)_{n\in\mathbb{N}}$ is continuously differentiable. It also converges uniformly.

This example show that without uniform convergence of the derivatives of $f_n(x)$, the limit and the differentiation opeartor cannot be interchanged.

Chapter 4

Metric Spaces

4.1 Metric Spaces

4.1.1 Basic Definitions

A lot of the times we would like to add a structure of a metric to space so that analysis such as continuity and integration can be performed on it.

Definition 4.1.1.1: Metric

Let X be a set. Let $x, y, z \in X$. A metric is a function $d : X \times X \to \mathbb{R}$ satisfying the following.

- $d(x, y) \geq 0$ with equality if and only if $x = y$
- $d(x, y) = d(y, x)$
- $d(x, y) \leq d(x, z) + d(z, y)$

Definition 4.1.1.2: Metric Space

A metric space is an oredered pair (X, d) where X is a set and d is a metric on X.

Definition 4.1.1.3: Open Balls

Let X be a metric space. Let $a \in X$. Define the open ball of radius r around a to be

$$B_r(a) = \{x \in X | d(x, a) < r\}$$

Lemma 4.1.1.4: Metric Subspace

Let (X, d) be a metric space. Let $A \subseteq X$, then $(A, d|_A)$ is also a metric space.

Proof. $d|_A$ inherits the metric properties of X while being restricted to A. □

Proposition 4.1.1.5: Metric Space Product

Let (X_1, d_1) and (X_2, d_2) be metric spaces. Let $x_1, y_1 \in X_1$ and $x_2, y_2 \in X_2$. Then for $1 \leq p < \infty$,

$$d_p((x_1, x_2), (y_1, y_2)) = (d_1(x_1, y_1)^p + d_2(x_2, y_2)^p)^{1/p}$$

defines a metric on $X_1 \times X_2$.

Proof. We prove the triangle inequality here, the others are easy. We have

$$\begin{aligned} d_p((x_1,x_2),(y_1,y_2))^p &= d_1(x_1,y_1)^p + d_2(x_2,y_2)^p \\ &\leq (d_1(x_1,z_1)+d(z_1,y_1))^p + (d_2(x_2,z_2)+d_2(z_2,y_2))^p \end{aligned}$$

□

4.1.2 Sets in a Metric Space

Definition 4.1.2.1: Open Sets

Let M be a metric space. Let $U \subset M$. We say that U is open if for every $a \in U$, there exists r such that

$$B_r(a) \subseteq U$$

Definition 4.1.2.2: Closed Sets

Let M be a metric space. Let $U \subset M$. We say that U is closed if $M \setminus U$ is open.

Lemma 4.1.2.3

Open balls are open.

Proof. Let $B_r(a)$ be our open ball. Let $x \in B_r(a)$. Then

$$B_{(r-d(x,a))/2}(x) \subseteq B_r(a)$$

thus we are done. □

Proposition 4.1.2.4

Countable union of open sets is open and countable intersections of closed sets is closed.

Proof. Let $U_1, U_2, \dots$ be a sequence of open sets. Let $U = \bigcup_{n\in\mathbb{N}} U_n$. Let $x \in U$. Then there exists $k \in \mathbb{N}$ such that $x \in U_k$. Since U_k is open, there exists $r \in \mathbb{R}^+$ such that

$$B_r(x) \subseteq U_k \subseteq U$$

and we are done.

Observe that

$$X \setminus \bigcup_{n\in\mathbb{N}} U_n = \bigcap_{n\in\mathbb{N}} (X \setminus U_n)$$

By definition of closed sets, $X \setminus U$ is closed and we are done. □

Proposition 4.1.2.5

Finite intersection of open sets is open and finite union of closed sets is closed.

Proof. Let $U_1, \dots, U_n$ be opens sets. Then let $x \in \bigcap_{k=1}^n U_k$. Then $x \in U_k$ for all $k \in \{1, \dots, n\}$ and there exists $r_k > 0$ such that $B_{r_k}(x) \subseteq U_k$ for each k. Take $r = \min\{r_1, \dots, r_n\}$. Then

$$B_r(x) \subseteq B_{r_k}(x) \subseteq U_k$$

for each k and thus $B_r(x) \subseteq \bigcap_{k=1}^n U_k$ and we are done.

Observe that

$$X \setminus \bigcap_{k=1}^n U_k = \bigcup_{k=1}^n (X \setminus U_k)$$

and by definition of closed sets, $X \setminus \bigcap_{k=1}^n U_k$ is closed and we are done. □

4.1.3 Points in a Subset

Definition 4.1.3.1: Interior Points

Let M be a metric space. Let $x \in U \subset M$. We say that x is an interior point of U if there exists r such that

$$B_r(x) \subset U$$

Denote the set of all interior points by U°.

Definition 4.1.3.2: Exterior Points

Let M be a metric space. Let $x \in U \subset M$. We say that x is an exterior point of U if there exists r such that

$$B_r(x) \subset M \setminus U$$

Denote the set of all interior points by $\text{Ext}(U)$.

Definition 4.1.3.3: Boundary

Let M be a metric space. Let $x \in U \subset M$. We say that x is a boundary point of U if for every r,

$$B_r(x) \cap U \neq \emptyset \text{ and } B_r(x) \cap M \setminus U \neq \emptyset$$

Denote the set of all boundary points by ∂U.

Definition 4.1.3.4: Closure

Let M be a metric space. Let $U \subset M$. Define the closure of U to be

$$\overline{U} = U \cup \partial U$$

Proposition 4.1.3.5

Let M be a metric space. Let $U \subset M$. Then U is open if and only

$$U \cap \partial U = \emptyset$$

Proof. Suppose that U is open. Let $x \in U \cap \partial U$. This means that $x \in \partial U$ and $B_r(x) \cap M \setminus U \neq \emptyset$ for all r. But this means that $B_r(x)$ cannot be a subset of U is it always contains point outside U, thus $x \notin U$ and thus $U \cap \partial U = \emptyset$.

Let $U \cap \partial U = \emptyset$. Let $x \in U$. Then $x \notin \partial U$. Thus by negation of the definition of boundary, there exists $r > 0$ such that $B_r(x) \cap M \setminus U = \emptyset$. Thus $B_r(x) \subseteq U$ and we are done. □

Proposition 4.1.3.6

Let M be a metric space. Let $U \subset M$. Then U is closed if and only

$$\overline{U} = U$$

4.1.4 Sequences, Limits and Continuity

Definition 4.1.4.1: Sequences

Let X be a metric space. A sequence in X is an ordered set of numbers $x_0, x_1, x_2, \dots$ such that they all are in X. We denote this sequence by $(x_n)_{n\in\mathbb{N}}$.

Definition 4.1.4.2: Convergence

A sequence $(x_n)_{n\in\mathbb{N}} \subset X$ a metric space is said to converge to $x \in X$ if for every $\epsilon > 0$ there exists N such that $n > N$ implies

$$d(x_n, x) < \epsilon$$

Proposition 4.1.4.3: Uniqueness of Limit

If a sequence converges, then its limit is unique.

Proposition 4.1.4.4

Let X be a metrix space. $U \subseteq X$ is closed if and only if for every sequence $(x_n)_{n\in\mathbb{N}} \subseteq U$ that converges, it converges to some $x \in U$.

Proof. Suppose that U is closed. Then $X \setminus U$ is open by definition. Let $\{x_n\} \subset U$ converge to $x \notin U$. Then $x \in X \setminus U$. By definition of convergence, for every $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that $x_n \in B_\epsilon(x)$ for $n > N$. But since $X \setminus U$ is open, there should be some ϵ such that $B_\epsilon(x) \subset X \setminus U$. In this case, we would have $x_n \in B_r(x) \subset X \setminus U$ which is a contradiction.

Suppose that the right hand side is true. Suppose for a contradiction that $X \setminus U$ is not open. Then for every $\epsilon > 0$, $B_\epsilon(x)$ is not a subset of $X \setminus U$ for some $x \in X \setminus U$. Let $y_k \in B_{1/k}(x)$ but $y_k \notin X \setminus U$. Then $y_k \in U$ and $y_k \to x \in X \setminus U$, a contradiction. □

Definition 4.1.4.5: Continuity

Let (U, d_1), (V, d_2) be metric spaces. $f : U \to V$ is said to be continuous at $p \in U$ if for every $\epsilon > 0$, there exists $\delta > 0$ such that

$$x \in B_\delta(p) \implies f(x) \in B_\epsilon(f(p))$$

Or equivalently,

$$f(B_\delta(p)) \subset B_\epsilon(f(p))$$

Proposition 4.1.4.6

Let $f : X \to Y$ be a funciton between metric spaces. Then f is continuous at a if and only if for every sequence x_n such that $\lim_{n\to\infty} x_n \to a$, we have

$$\lim_{n\to\infty} f(x_n) = f(a)$$

Theorem 4.1.4.7

Let U, V be metric spaces. Let $f : U \to V$ be a function. Then f is continuous if and only if for every open subsets $\Omega \subset V$, $f^{-1}(\Omega)$ is open.

Proof. Suppose that f is continuous. Let $\Omega \subset V$ such that Ω is open. Then for every $p \in f^{-1}(V)$, there exists $\epsilon > 0$ such that $B_\epsilon(f(p)) \subset V$. By continuity, there exists $\delta > 0$ such that $f(B_{\delta(p)}) \subset B_\epsilon(f(p))$. This implies $B_\delta(p) \subset f^{-1}(B_\epsilon(f(p)))$. But also since $B_\epsilon(f(p)) \subset V$, we have

$$B_\delta(p) \subset f^{-1}(B_\epsilon(f(p))) \subset f^{-1}(V)$$

Since this is true for every p, $f^{-1}(V)$ must be open.

Now suppose that $\Omega \subset V$ is open imply $f^{-1}(\Omega)$ is open. Let $p \in \Omega$. Then there exists $\epsilon > 0$ such that $B_\epsilon(f(p)) \subset \Omega$. By assumption, we must have $f^{-1}(B_\epsilon(f(p)))$ is open. The fact that this is open means there exists $\delta > 0$ such that $B_\delta(p) \subset f^{-1}(B_\epsilon(f(p)))$. Then we have

$$f(B_\delta(p)) \subset B_\epsilon(f(p))$$

and we are done. □

4.1.5 Equivalent Metrics

Theorem 4.1.5.1

Let d_1, d_2 be two metrics on X. Then the following statements are equivalent.

- The open sets in (X, d_1) and (X, d_2) coincide
- For any metric space (Y, d_Y), a function $g : X \to Y$ is continuous from (X, d_1) to (Y, d_Y) if and only if g is continuous from (X, d_2) to (X, d_1)
- For any metric sapce (Y, d_Y), a function $f : Y \to X$ is continuous from (Y, d_Y) to (X, d_1) if and only if f is continuous from (Y, d_Y) to (X, d_2)

Definition 4.1.5.2: Topologically Equivalent Metrics

Two metrics d_1, d_2 on X are said to be topologically equivalent if the above statements are true.

Definition 4.1.5.3: Lipschitz Equivalent Metrics

Two metrics d_1, d_2 on X are said to be Lipschitz equivalent if there exists $0 < c_1 \leq c_2 < \infty$ such that

$$c_1 d_1(x, y) \leq d_2(x, y) \leq c_2 d_1(x, y)$$

for all $x, y \in X$.

Lemma 4.1.5.4

Lipschitz equivalence implies topologically equivalence on metrics.

Definition 4.1.5.5: Equivalent Norms

Two norms $\|\cdot\|_1$ and $\|\cdot\|_2$ for a vector space V over a field $F = \mathbb{R}$ or $\mathbb{C}$ are said to be equivalent if there exists $c_1, c_2 \in F$ such that for every $x \in V$,

$$c_1\|x\|_1 \leq \|x\|_2 \leq c_2\|x\|_1$$

Proposition 4.1.5.6

The equivalence on norms is an equivalent relation.

Proposition 4.1.5.7

Suppose that two norms are equivalent on a normed vector space, then they induce topologically equivalent metrics.

Proof. Suppose that $\|\cdot\|_1$ and $\|\cdot\|_2$ are equivalent. Then define their corresponding metrics by $d_1(x,y) = \|x-y\|_1$ and $d_2(x,y) = \|x-y\|_2$ for x, y in a normed vector space X. We show that the open sets coincide.

Suppose that $U \subseteq (X, d_1)$ is open. Then for every $x \in U$, there exists $r > 0$ such that $B_r(x) \subset U$. From the equivalent norms, we have that there exists c such that $\|x-y\|_2 \leq c\|x-y\|_1$ and thus

$$\left\{x \in X \middle| \|x-y\|_2 < \frac{r}{c}\right\} \subseteq \{x \in X | \|x-y\|_1 < r\}$$

Thus $B_{\frac{r}{c}}(x)$ in the d_2 metric is a subset of $B_r(x)$ in the d_1 metric. This means that we have constructed an open ball in (X, d_2) so that it is contained in U. Thus U is also open in (X, d_2).

Mirror this to show that the open sets of (X, d_2) must also be open sets of (X, d_1) using the fact that there exists c such that $\|x-y\|_1 \leq c\|x-y\|_2$ and we are done. □

Lemma 4.1.5.8

If X is a vector space and two norms induce topologically equivalent metrics, then the norms are equivalent.

4.2 Connectedness

4.2.1 Definitions and Properties

Definition 4.2.1.1: Connectedness

We say that a metric space is disconnected if we can write it as the disjoint union of two nonempty open sets. Otherewise it is connected.

Notice that the definition of connectedness is implicit from the definition of disconnectedness. We give an explicit criteria to prove connectedness.

Proposition 4.2.1.2

Let X be a metric space. Then the following are equivalent.

- X is connected
- If $f : X \to \{0, 1\}$ is a continuous function then f is constant.
- The only subsets of X which are both open and closed are X and $\emptyset$.

Proof.

- (1) $\iff$ (2): We prove the contrapositive. Namely X is disconnected if and only if there exists a continuous function $f; X \to \{0, 1\}$ that is non-constant. Suppose that X is disconnected. Then there exists $A, B \subset X$ that are open such that $A \cap B = \emptyset$ and $A \cup B = X$. Define f by
$$f(x) = \begin{cases} 0 & \text{if } x \in A \\ 1 & \text{if } x \in B \end{cases}$$
This function is continuous since every open set in $\{0, 1\}$ is mapped to an open set in X. It clearly is non-constant thus we are done.

  Now suppose that $f : X \to \{0, 1\}$ is non-constant continuous function. Then by defining $A = f^{-1}(0)$ and $B = f^{-1}(1)$, we are done.
- (1) $\iff$ (3): Suppose that X is connected but there exists non-empty $A \subset X$ such that it is both open and closed. Then $X \setminus A$ is open and is disjoint with A and $A \cup X \setminus A = X$. This is a contradiction to X being connected.

  Now suppose that the only subsets which are both open and closed are X and $\emptyset$. Suppose for a contradiction that X is disconnected. Then there exists open sets $A, B \subset X$ such that $A \cap B = \emptyset$ and $A \cup B = X$. Then clearly $B = X \setminus A$ is open, but $X \setminus A$ is the set difference of an open set thus it should be closed. Then B is both open and closed and we have a contradiction.

□

These two criteria will prove themselves to be particularly useful in proving theorems related to connectedness as well as begin able to identify concrete examples on connectedness.

Proposition 4.2.1.3

Let X be a metric space. Let $S \subset X$ be a metric subspace. Then S is connected if and only if the following is true. If U, V are open subsets of X and $U \cap V \cap S = \emptyset$ and $S \subseteq U \cup V$ implies $S \subseteq U$ or $S \subseteq V$.

Proof. □

Lemma 4.2.1.4

If $C \subset (X, d)$ is connected then so is any set S satisfying $C \subset S \subset \overline{C}$.

Lemma 4.2.1.5

Let X be a metric space. The countable union of connected subsets of X such that they have a nonempty intersection is connected.

Proof. Suppose that $\{A_i | i \in I\}$ are all connected and has a nonempty intersection $x \in X$. Suppose that $f : X \to \{0, 1\}$ is a continuous function such that $f(x) = 0$. For every A_i, $f|_{A_i}$ is a constant function since f is continuous. This means that $f|_{A_i}(x) = 0$ for all $x \in A_i$. Then f when only restricted to the countable union of A_i, it will be identically zero. Thus we are done. □

Proposition 4.2.1.6

Continuity preserves connectedness. That is, if $f : X \to Y$ is a continuous function between metric spaces and X is connected, then $f(X)$ is conneted.

Proof. Suppose that $f(X)$ is disconnected. Then there exists a non-empty $A \subset f(X)$ that is both open and closed. By continuity, $f^{-1}(A)$ is also both open and closed, which is a contradiction since X is connected. □

Proposition 4.2.1.7

The product of two connected spaces is connected.

Notice that none of the above propositions involve any notion of distance. This is baecause these are topological properties rather than metric properties, which will be discussed more on a topology course.

4.2.2 Path-Connectedness

Definition 4.2.2.1: Path-Connected Metric Space

Let X be a metric space. Then we say that X is path-connected if the following are true. For any $a, b \in X$, there exists a continuous map $\gamma : [0, 1] \to X$ with $\gamma(0) = a$ and $\gamma(1) = b$. γ is called a path.

Lemma 4.2.2.2

Let X be a metric space. Define a relation $\sim$ on X as $a \sim b$ if and only if there exists a path $\gamma : [0, 1] \to X$ with $\gamma(0) = a$ and $\gamma(1) = b$. Then $\sim$ is an equivalent relation.

Proposition 4.2.2.3

Every path-connected metric space is connected.

Proposition 4.2.2.4

A connected open subset of a normed space is path-connected.

4.2.3 Connectedness on $\mathbb{R}^n$

Theorem 4.2.3.1

A subset of $\mathbb{R}$ is connected if and only if it is an interval.

Below is a partial converse of path connectedness implying connectedness over $\mathbb{R}^n$.

Theorem 4.2.3.2

Connected open subsets of $\mathbb{R}^n$ are path connected.

Theorem 4.2.3.3

Open subsets of $\mathbb{R}^n$ have open connected components.

Theorem 4.2.3.4

A subset U of $\mathbb{R}$ is open if and only if it is the disjoint union of countably many open intervals.

4.3 Compactness

4.3.1 Compactness and Sequential Compactness

Definition 4.3.1.1: Open Cover

An open cover of a metric space X is a collection $\mathcal{U}$ of open subsets of X such that E$X = \bigcup_{U\in\mathcal{U}} U$E

Definition 4.3.1.2: Compact Metric Spaces

Let X be a metric space. Let $K \subseteq X$. K is said to be compact if every open cover of K contains a finite subcover.

Definition 4.3.1.3: Lebesgue Number

Let $\mathcal{U}$ be an open cover of a metric space X. A number $\delta > 0$ is called a Lebesgue number for $\mathcal{U}$ if for any $x \in X$ there exists $U \in \mathcal{U}$ such that $B_\delta(x) \subset U$.

Lemma 4.3.1.4

Every open cover $\mathcal{U}$ of a compact metric space X has a Lebesgue number.

Definition 4.3.1.5: Sequential Compactness

Let X be a metric space. Then X is said to be sequentially compact if any sequence of elements in X has a convergent subsequence.

Lemma 4.3.1.6

If X is sequentially compact that any open cover of X has a Lebesgue number.

Proposition 4.3.1.7

Let (X, d) be a metric space. Then the following are equivalent.

- X is compact
- X is sequentially compact
- X is closed and totally bounded

4.3.2 Properties of Compactness

Proposition 4.3.2.1

A compact subset of a metric space is closed.

Proof. Let $K \subset X$ be compact. Let $a \in X \setminus K$. For every $x \in K$, $B_{d(a,x)/2}(a)$ and $B_{d(a,x)/2}(x)$ are disjoint open balls. Then $\{B_{d(a,x)/2}(x) | x \in K\}$ is an open cover of K. Since K is compact, it has a finite subcover $B_{d(a,x_1)/2}(x_1), \dots, B_{d(a,x_n)/2}(x_n)$. But

$$K \cap \bigcap_{k=1}^{n} B_{d(a,x_k)/2}(a) = \emptyset$$

since the two types of balls are disjoint. Thus K is closed. □

Proposition 4.3.2.2

A compact subset of a metric space is bounded.

Proof. Let $a \in X$. Let $x \in K$. Then $x \in B_r(a)$ for all $r > d(a, x)$. Thus K is covered by the collection of open balls $B_r(a)$. Thus it has a finite subcover $B_{r_1}(a), \dots, B_{r_n}(a)$. Thus

$$K \subset \bigcup_{k=1}^{n} B_{r_k}(a) = B_{\max\{r_1,\dots,r_n\}}(a)$$

and we are done. □

Proposition 4.3.2.3

Let X be a compact metric space. Let $C \subseteq X$ be a closed subset. Then C is compact.

Proof. Let U be a cover of C by open subsets of X. Then $U \cup X \setminus C$ is an open cover of X, thus has a finite subcover. This provides an open subcover of C since $X \setminus C$ is open and you can remove this element fromt the subcover. □

4.3.3 Compactness and Continuity

Theorem 4.3.3.1

Continuity preserves compactness.

Proof. Let $f : X \to Y$ be a continuous function between metric spaces. Suppose that X is compact. □

Lemma 4.3.3.2

Let X, Y be metric spaces. A sequence $\{(x_n, y_n)\} \subset X \times Y$ converges if and only if $\{x_n\} \subset X$ converges in X and $\{y_n\} \subset Y$ converges in Y.

Proposition 4.3.3.3

The product of two compact metric spaces is compact.

Theorem 4.3.3.4: Heine-Borel Theorem

A subset of $\mathbb{R}^n$ is compact if and only if it is closed and bounded.

Proof. Let K be a compact subset of $\mathbb{R}^n$. K is closed by proposition 3.2.1 and K is bounded by proposition 3.2.2.

Let K be a closed and bounded subset of $\mathbb{R}^n$. If K is bounded then $K \subset [-r, r]^n$ for some $r > 0$. I claim that $[-r, r]^n$ is compact. Once it is compact, applying 3.2.3 to the closed subset K of $[-r, r]^n$, we have that K is compact.

Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $[-r, r]$ by bolzano weierstrass it has a convergent subsequence. Thus $[-r, r]$ is sequentially compact and thus compact. Using the productivity of compact metric spaces, we have that $[-r, r]^n$ is compact thus we are done. □

4.3.4 Uniform Continuity

Definition 4.3.4.1: Uniformly Continuous

A map $f : X \to Y$ is uniformaly continuous if for every $\epsilon > 0$, there exists $\delta > 0$ such that

$$d_X(x, y) < \delta \implies d_Y(f(x), f(y)) < \epsilon$$

for any $x, y \in X$.

Theorem 4.3.4.2

A continuous map from a compact metric into a metric space is uniformly continuous.

4.4 Completeness

4.4.1 Motivation and Definitions

Definition 4.4.1.1: Cauchy Sequence

We say that $\{x_n\} \subset (X, d)$ is a Cauchy sequence if for every $\epsilon > 0$, there exists some N such that $d(x_n, x_m) < \epsilon$ for all $n, m > \epsilon$.

Proposition 4.4.1.2

Every convergent sequence is Cauchy.

Proof. Let $(x_n)_{n\in\mathbb{N}}$ be a convergent sequence in a metric space X. Let $\epsilon > 0$, then from convergence we have that for $d(x_n, x) < \frac{\epsilon}{2}$ for all $n > N$ for some $N \in \mathbb{N}$. Then choosing the same N, we have that

$$d(x_n, x_m) \leq d(x_n, x) + d(x, x_m) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

thus we are done. □

We now give the definition of a complete space in terms of Cauchy sequences.

Definition 4.4.1.3: Complete Spaces

A metric space (X, d) is complete if any Cauchy sequence in X converges.

Proposition 4.4.1.4

Every compact metric space is complete.

Proof. Suppose that $(x_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in a compact metric space X. Then X being sequentially compact means that there exists a subsequence of $(x_n)_{n\in\mathbb{N}}$ such that it converges in X. But then clearly

$$d(x_n, x) \leq d(x_n, x_{n_k}) + d(x_{n_j}, x)$$

implies that $x_n \to x$ since in the inequality, the first part of the sum corresponds to the sequence being Cauchy and thus tends to 0, while the latter part correponds to the subsequence being convergent and thus tends to 0. □

4.4.2 Properties of Complete Spaces

Proposition 4.4.2.1

A subspace of a metric space is complete if and only if it is closed under a complete metric space.

Proof. Suppose that X is a metric space and $U \subset X$ is a complete metric space. Let $(x_n)_{n\in\mathbb{N}} \subset U$ and that $x_n \to x \in X$. Then $(x_n)_{n\in\mathbb{N}}$ is Cauchy thus it convergence to some $y \in U$. We will show that in fact $x = y$. This is true from the fact that

$$d|_U(x_n, y) = d(x_n, y)$$

Thus $(x_n)_{n\in\mathbb{N}}$ is in fact a sequence that converges in U. This shows that U is closed.

Now suppose that U is closed under a complete metric space X. Let $(x_n)_{n\in\mathbb{N}}$ be a Cauchy sequence in U. Then trivially it is also a Cauchy sequence in X and thus is convergent. Since U is closed, the limit is necessarily in U and thus U is complete. □

Theorem 4.4.2.2: Cantor's Intersection Theorem

Let X be a complete metric space. Let $S_1 \supseteq S_2 \supseteq \dots$ form a nested sequence of non-empty closed sets in X with the property that $\operatorname{diam}(S_n) \to 0$ as $n \to \infty$. Then

$$\bigcap_{n=1}^{\infty} S_n \neq \emptyset$$

Proof. For each $N \in \mathbb{N}$, choose $x_N \in S_N$. Then for all $n > N$, $x_n \in S_N$. Thus for $n, m > N$, we have that $d(x_n, x_m) \leq \operatorname{diam}(S_n)$. It follows that $(x_n)_{n\in\mathbb{N}}$ is Cauchy. Thus $x_n \to x$ for some $x \in X$. Since each S_n is closed and $x_n \in S_N$ for all $n > N$, we must have that $x \in S_n$ for each n. Thus $x \in \bigcap_{k=1}^{\infty} S_n$ is nonempty. □

Below are a few examples of complete spaces.

Proposition 4.4.2.3

$\mathbb{R}^n$ and $\mathbb{C}$ are both complete.

Proof. Let $(x_k)_{k\in\mathbb{N}}$ be a Cauchy sequence in $\mathbb{R}^n$. Denote the ith component of x_k by $x_{k,i}$. Then for every $\epsilon > 0$, there exists N such that

$$\|x_k - x_m\| = \left(\sum_{i=1}^{n} |x_{k,i} - x_{m,i}|^2\right)^{\frac{1}{2}} < \epsilon$$

for $k, m > N$. In particular, we have that each individual

$$|x_{k,i} - x_{m,i}| < \epsilon$$

for $m, n > N$. Thus $(x_{k,i})_{k\in\mathbb{N}}$ is a Cauchy sequence in $\mathbb{R}$. But we know that Cauchy sequences in $\mathbb{R}$ converges, thus $(x_{k,i})_{k\in\mathbb{N}}$ converges to $x_i \in \mathbb{R}$. Now define $x = (x_1, \dots, x_n)$, then

$$\|x_k - x\| = \left(|x_{k,i} - x_i|^2\right)^{\frac{1}{2}} < n\epsilon$$

by convergence of each individual component. Thus $(x_n)_{n\in\mathbb{N}}$ is a convergent sequence.

The proof for $\mathbb{C}$ is the same in considering $\mathbb{R}^2$. □

Proposition 4.4.2.4

Every normed vector space is complete.

4.4.3 Completion

The goal of this section is to attempt to complete a metric space by adding in the missing limits of a metric space.

Definition 4.4.3.1: Space of Bounded Real Functions

Denote $B(X)$ the space of all bounded real valued functions on a metric (topological) space X. This means that

$$B(X) = \{f : X \to \mathbb{R} | |f| \leq M \text{ for some } M \in \mathbb{R}\}$$

Proposition 4.4.3.2

The metric space with distance induced by the supremum norm

$$\|f\|_\infty = \sup_{x \in X} |f(X)|$$

for $f \in B(X)$ is complete.

Proof. Let $(f_n)_{n\in\mathbb{N}}$ be a Cauchy sequence in $B(X)$. Then for every $\epsilon > 0$, there exists N such that

$$\|f_n - f_m\|_\infty = \sup_{x \in X} |f_n(x) - f_m(x)| < \epsilon$$

for all $n, m > N$. In particular, for each $x \in X$, the property of supremum implies that $|f_n(x) - f_m(x)| < \epsilon$ for $n, m > N$. Thus $(f_n(x))_{n\in\mathbb{N}} \subset \mathbb{R}$ is Cauchy for each x. Since $\mathbb{R}$ is complete, $(f_n(x))_{n\in\mathbb{N}}$ converges for each $x \in X$.

Now define the function $f : X \to \mathbb{R}$ by

$$f(x) = \lim_{n\to\infty} f_n(x)$$

Then fix $\epsilon > 0$, we have that

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

for all $n > N$ by letting $m \to \infty$ from the fact that $|f_n(x) - f_m(x)| < \epsilon$. This N does not depend on x. Fix $\epsilon = 1$, then there exists $N_1 \in \mathbb{N}$ such that

$$\begin{aligned} |f(x) - f_n(x)| &\leq |f(x) - f_{N_1}(x)| \\ &\leq 1 + |f_{N_1}(x)| \end{aligned}$$

for all $x \in X$ and $n > N_1$ thus f is bounded. This means that $f \in B(X)$ and that $\|f_n - f\|_\infty < \epsilon$ for all $n > N$. □

Proposition 4.4.3.3

Any metric space X can be isometrically embedded into the complete metric space $B(X)$.

4.4.4 Compactness, Completeness and Totally Bounded

Definition 4.4.4.1: Totally Bounded

A metric space X is totally bounded if for any $\epsilon > 0$, there exists $B_\epsilon(p_k)$ for $k \in \{1, \dots, n\}$ such that

$$X \subseteq \bigcup_{k=1}^{n} B_\epsilon(p_k)$$

Theorem 4.4.4.2

A subspace Y of a metric space X that is complete is compact if and only if it is closed and totally bounded.

Theorem 4.4.4.3

A subspace Y of a complete metric space is totally bounded if and only if its closure is compact.

4.4.5 Contraction Mapping and Completion

Definition 4.4.5.1: Lipschitz Continuous

Let (X, d_X) and (Y, d_Y) be metric spaces and suppose that $f : X \to Y$. We say that f is a Lipschitz map if there is a constant $K \geq 0$ such that

$$d_Y(f(x), f(y)) \leq Kd(x, y)$$

for all x, y in X.

If $Y = X$ and $K \in [0, 1)$ then f is a contraction mapping.

Lemma 4.4.5.2

If $f : X \to Y$ is Lipschitz continuous then it is continuous.

Theorem 4.4.5.3: Contraction Mapping Theorem

Let X be a nonempty complete metric space and suppose that $f : X \to X$ is a contraction. Then f has a unique fixed point, meaning there is a unique $x \in X$ such that $f(x) = x$.

Proof. Let $x_0 \in X$ and define a sequence by $x_{n+1} = f(x_n)$ for $n \in \mathbb{N}$. Then we have that

$$d(x_{n+1}, x_n) \leq Kd(x_n, x_{n-1}) \leq \cdots \leq K^n d(x_1, x_0)$$

Then for any $k > n$, we have that

$$\begin{aligned} d(x_k, x_n) &\leq \sum_{i=n}^{k-1} d(x_{i+1}, x_i) \\ &\leq \sum_{i=n}^{k-1} K^i d(x_1, x_0) \\ &\leq \frac{K^i}{1-K} d(x_1, x_0) \end{aligned}$$

This is Cauchy since we can choose $\epsilon > 0$ such that $\frac{K^i}{1-K} < \epsilon$. Since X is complete, we have that $x_n \to x$ for some $x \in X$. Since f is continuous we have that $f(x_n) \to f(x)$. Now taking limits on

$$x_{n+1} = f(x_n)$$

we have that $x = f(x)$.

To prove uniqueness, note that if $f(x) = x$ and $f(y) = y$, then

$$d(x, y) = d(f(x), f(y)) \leq Kd(x, y)$$

which implies that $(1 - K)d(x, y) = 0$. Thus $x = y$. □

Another name for this theorem would be Banach's Fixed Point Theorem.

Theorem 4.4.5.4: Picard-Lindelof Theorem

Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be Lipschitz continuous with

$$|f(x) - f(y)| \leq L|x - y|$$

where $x, y \in \mathbb{R}^n$. Then for any $x_0 \in \mathbb{R}^n$, the differential equation

$$\frac{dx}{dt} = f(x)$$

with initial condition $x(0) = x_0$ has a unique solution on $[-t, t]$ for any $Lt < 1$.

4.4.6 Cantor's Theorem

Theorem 4.4.6.1

If X is a complete metric space and $\{F_n | n \in \mathbb{N}\}$ is a collection of open dense subsets of X, then

$$F = \bigcap_{k=1}^{\infty} F_n t$$

is dense in X. Equivalently, if $\{G_n | n \in \mathbb{N}\}$ is a collection of nowhere dense subsets of a nonempty complete metric space X, then

$$\bigcup_{k=1}^{\infty} F_k \neq X$$

Lemma 4.4.6.2

The Cantor set is uncountable.

4.5 Notable Metric Spaces

4.5.1 $\mathbb{R}^n$ on Different Metrics

Theorem 4.5.1.1

Let $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ and similarly for $y \in \mathbb{R}^n$. The following are all metrics of $\mathbb{R}^n$.

- l_p metric:
$$d_p(x, y) = \left(\sum_{k=1}^{n} (x_k - y_k)^p\right)^{1/p}$$
for $1 \leq p < \infty$
- l_∞ metric:
$$d_\infty(x, y) = \max_{k \in \{1,\dots,n\}} \{|x_k - y_k|\}$$
- Jungle river metric on $\mathbb{R}^2$:
$$d_{\text{Jr}}(x, y) = \begin{cases} |x_2 - y_2| & \text{if } x_1 = y_1 \\ |y_2| + |x_2| + |x_1 - y_1| & \text{if } x_1 \neq y_1 \end{cases}$$
- French Railway Metric (Sunflower metric) on $\mathbb{R}^2$:
$$d_{\text{Fr}}(x, y) = \begin{cases} |x - y| & \text{if there exists } \lambda \in \mathbb{R} \text{ such that } y = \lambda x \\ |x| + |y| & \text{otherwise} \end{cases}$$
- Discrete Metric:
$$d_{\text{Discrete}}(x, y) = \begin{cases} 0 & \text{if } x = y \\ 1 & \text{if } x \neq y \end{cases}$$
- British Railway Metric on $\mathbb{R}^2$:
$$d(x, y) = \begin{cases} 0 & \text{if } x = y \\ |x| + |y| & \text{if } x \neq y \end{cases}$$

Do try and draw at least the unit ball for each of these metrics and see what happens (at least for $\mathbb{R}^2$).

Proposition 4.5.1.2

All l_p metrics are topologically equivalent.

Proof. The metric are all induced by the l_p norms and we know that they are equivalent. Equivalent norms induce topologically equivalent metrics and we are done. □

Proposition 4.5.1.3

Let (X, d) be a metric space. Then the function
$$d_{\text{B}}(x, y) = \min\{d(x, y), 1\}$$
for any $x, y \in X$ is a metric on X.

4.5.2 The Space of Continuous Functions

Definition 4.5.2.1

We denote $C([a,b])$ the space of real valued continuous functions whose domain is $[a,b]$.

Proposition 4.5.2.2

Let $f \in C([a,b])$. Define the supremum norm of f to be

$$\|f\|_\infty = \sup_{x\in[a,b]}]|f(x)|$$

Then the supremum norm is a norm on $C([a,b])$.

Proposition 4.5.2.3

Let $f \in C([a,b])$. Define the L^p norm of f to be

$$\|f\|_{L^p} = \left(\int_a^b |f(x)|^p \, dx\right)^{\frac{1}{p}}$$

for $p \in [1,\infty)$. Then the supremum norm is a norm on $C([a,b])$.

4.5.3 Sequence Space

Definition 4.5.3.1: Sequence Space

The sequence space l^p for $1 \leq p < \infty$ consists of all sequences $\{x_n\}$ such that

$$\sum_{k=1}^\infty |x_k|^p < \infty$$

If $p=\infty$ then l^∞ is the space of all bound sequences.

Proposition 4.5.3.2

The function

$$\|x\|_{l^p} = \left(\sum_{k=1}^\infty |x_k|^p\right)^{\frac{1}{p}}$$

on l^p space defines a norm on it.

If $p=\infty$ then $\|x\|_{l^\infty} = \sup_{k\in\mathbb{N}} |x_k|$ defines a norm on l^∞.

Proposition 4.5.3.3

l^p is a complete metric space with metric

$$d(\{x_n\},\{y_n\}) = \|x-y\|_{l^p}$$

Chapter 5

Multivariable Calculus

5.1 Analysis on $\mathbb{R}^n$ and $M_{m\times n}(\mathbb{R})$

5.1.1 Sequences on $\mathbb{R}^n$

Definition 5.1.1.1: Euclidean Metric

The Euclidean Metric of $x, y \in \mathbb{R}^n$ is defined to be

$$d(x,y) = \left(\sum_{k=1}^{n}(x_k - y_k)^2\right)^{\frac{1}{2}}$$

We denote it as $|x-y|$.

Proposition 5.1.1.2

$\mathbb{R}^n$ and the Euclidean Metric forms a metric space.

Proof. We show that the Euclidean Metric is indeed a metric on $\mathbb{R}^n$. Let $x, y, z \in \mathbb{R}^n$.

- $(x_k - y_k)^2 \geq 0$ for all $k \in \{1, \dots, n\}$ with equality if and only if $x_k = y_k$
- $(x_k - y_k)^2 = (y_k - x_k)^2$ for all $k \in \{1, \dots, n\}$
- $d(x,y) \leq d(x,z) + d(y,z)$ for any $z \in \mathbb{R}^n$

□

Definition 5.1.1.3: Convergence in $\mathbb{R}^n$

A sequence $\{x_n\} \subseteq \mathbb{R}^n$ is said to converge to $x \in \mathbb{R}^n$ if for every $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that $d(x_n, x) < \epsilon$ for all $n > N$.

Proposition 5.1.1.4: Componentwise Convergence

$x_k \in \mathbb{R}^n$ converges if and only if $x_{i,k}$ converges for every $i \in \{1, \dots, n\}$.

Proof. Suppose that $|x_k - x| < \epsilon$. Then in particular $|x_{i,k} - x_i| < |x_k - x| < \epsilon$ and thus is convergent componentwise. Suppose that $\{x_k\}$ is convergent componentwise. Then there exists $N_1, \dots, N_n$ such that $n_j > N_j$ implies $|x_{j,k} - x_j| < \epsilon$ for $j \in \{1, \dots, n\}$. Take the max of all

$N_1, \dots, N_n$. Then

$$\begin{aligned} |x_k - x| &= \left(\sum_{j=1}^{n} (x_j - x_{j,k})\right)^{\frac{1}{2}} \\ &< \sqrt{n}\epsilon \end{aligned}$$

□

Theorem 5.1.1.5: Bolzano-Weierstrass Theorem

Any bounded sequence in $\mathbb{R}^n$ has a convergent subsequence.

Proof. Let $\{x_n\} \subset \mathbb{R}^n$ be bounded. By Bolzano-Weierstrass Theorem on $\mathbb{R}$, the sequence $\{x_{1,m}\} \subset \mathbb{R}$ has a convergent subsequence $\{x_{1,m_k}\}$. Apply Bolzano-Weierstrass Theorem to $\{x_{2,m_k}\}$ and keep going until you reach the nth component. We will end up with a subsequence that converges for all components and thus is convergent in $\mathbb{R}^n$. □

Proposition 5.1.1.6: Sum Rule of Sequences

Let $\{x_n\} \subset \mathbb{R}^n$ converge to $x \in \mathbb{R}^n$ and $\{y_n\} \subset \mathbb{R}^n$ converge to $y \in \mathbb{R}^n$. Let $a, b \in \mathbb{R}$. Then

$$ax_n + by_n \to ax + by$$

Proof. Trivial using the characterization of componentwise convergence and real analysis. □

5.1.2 Continuity on $\mathbb{R}^n$

Definition 5.1.2.1: Limits

Let $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$. Let $a \in \mathbb{R}^n$ be a limit point of U. Let $b \in \mathbb{R}^m$. We say that

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

if for every $\epsilon > 0$ there is some $\delta > 0$ such that for every $\epsilon > 0$ there exists some $\delta > 0$ such that for all $x \in U$,

$$\|x - a\| < \delta \implies \|f(x) - b\| < \epsilon$$

Definition 5.1.2.2: Continuity

Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$. We say that f is continuous at $a \in U$ if

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

Theorem 5.1.2.3: Componentwise Continuity

Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ be a function. Then f is continuous if and only if each of its components $f_1, \dots, f_n : U \to \mathbb{R}$ is continuous.

Proof. Suppose that f is continuous at $a \in \mathbb{R}^n$. Let $\epsilon > 0$. Then there exists $\delta > 0$ from continuity such that $\|x-a\|<\delta$ implies $\|f(x)-f(a)\|<\epsilon$. Then $\|x-a\|<\delta$ implies

$$\|f_i(x)-f_i(a)\|\leq\|f(x)-f(a)\|<\epsilon$$

Thus each component is continuous.

Now let f be component wise continuous at $a\in\mathbb{R}^n$. Let $\epsilon>0$. For each component $f_i:\mathbb{R}^n\to\mathbb{R}$, using continuity with $\frac{\epsilon}{\sqrt{n}}>0$there exists $\delta_i>0$ such that $\|x-a\|<\delta$ implies $\|f_i(x)-f_i(a)\|<\frac{\epsilon}{\sqrt{n}}$. Choose $\delta=\max\{\delta_1,\dots,\delta_n\}$. Then $\|x-a\|<\delta$ implies that

$$\begin{aligned}\|f(x)-f(a)\|^2&=\sum_{k=1}^n\|f_k(x)-f_k(a)\|^2\\&\leq n\max_{k\in\{1,\dots,n\}}\|f_k(x)-f_k(a)\|^2\\\|f(x)-f(a)\|&\leq\sqrt{n}\max_{k\in\{1,\dots,n\}}\|f_k(x)-f_k(a)\|\\&\leq\sqrt{n}\frac{\epsilon}{\sqrt{n}}\\&=\epsilon\end{aligned}$$

Thus f is continuous. □

Theorem 5.1.2.4: Sequential Continuity

Let $f:U\subseteq\mathbb{R}^n\to\mathbb{R}^m$. Then f is continuous at $c\in U$ if and only if for every sequence $\{x_n|n\in\mathbb{N}\}\subset U$ that $x_n\to c\in U$ it has the property that

$$f(x_n)\to f(c)$$

Proof. Exactly the same proof as that of real analysis. □

Proposition 5.1.2.5: Sum Rule

Let $f,g:U\subseteq\mathbb{R}^n\to\mathbb{R}^m$ be continuous at $p\in\mathbb{R}^n$. Then for any $a,b\in\mathbb{R}^m$, $af(x)+bg(x)$ is continuous at p.

Proof. Simple proof involving componentwise continuity and real analysis. □

Proposition 5.1.2.6: Product and Quotient Rule

Let $f,g:U\subseteq\mathbb{R}^n\to\mathbb{R}$ be continuous at $p\in\mathbb{R}^n$. Then

- $f(x)g(x)$ is continuous at p
- $\frac{f(x)}{g(x)}$ is continuous at p given that $g(x)\neq 0$ for all $x\in U$

Proof. Also a simple proof involving componentwise continuity and real analysis. □

Proposition 5.1.2.7: Composition Rule

Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ and $g : V \subseteq \mathbb{R}^m \to \mathbb{R}^k$ be continuous at $p \in \mathbb{R}^n$ and $f(p) \in V$ respectively. Then $g(f(x))$ is continuous at p.

Proof. Another simple proof involving componentwise continuity and real analysis. □

Definition 5.1.2.8: Linear Continuity

A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is linearly continuous if given any line $g(t) = x_0 + tv$ in $\mathbb{R}^n$ where $x_0, v \in \mathbb{R}^n$, we have that

$$\lim_{t\to 0} f(x_0 + tv) = f(x_0)$$

Definition 5.1.2.9: Separate Continuity

A function $f : \mathbb{R}^2 \to \mathbb{R}$ is separately continuous at (x_0, y_0) if $g(x) = f(x, y_0)$ and $h(y) = f(x_0, y)$ are both continuous.

Proposition 5.1.2.10

Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be continuous. If f is continuous, then f is linearly continuous. If f is linearly continuous, then f is seapartely continuous.

Beware that none of the inverse implications hold. But we can use the contrapositive to prove that functions are not continuous. Namely if we can prove that f is not separately continuous, then clearly f will not be continuous. This is precisely why we made this definitions in the first place.

Proposition 5.1.2.11

Let $g : U \subset \mathbb{R} \to \mathbb{R}$. Define $U_i = \{x \in \mathbb{R}^n | x_i \in U\}$. This U_i is the set of all $\mathbb{R}^n$ such that the ith element is in U. Define $f : U_i \subset \mathbb{R}^n \to \mathbb{R}$ by $f(x) = g(x_i)$. If g is continuous at $a \in U$ then f is continuous at $\{x \in \mathbb{R}^n | x_i = a\}$.

5.1.3 Topological Properties

We will only develop sufficient topological properties so that we can apply it to our proofs. The complete development of topology is in another set of notes.

Definition 5.1.3.1: Open Ball

Let $p \in \mathbb{R}^n$ and $r > 0$. The open ball at p with radius r is defined to be

$$B_r(p) = \{x \in \mathbb{R}^n || x - p| < r\}$$

Definition 5.1.3.2: Open and Closed Sets

We say that a set $U \subset \mathbb{R}^n$ is

- open if for all $p \in U$ there exists $r > 0$ such that $B_r(p) \subseteq U$
- closed if $\mathbb{R}^n \setminus U$ is open

Proposition 5.1.3.3

Let $U \subseteq \mathbb{R}^n$ is closed if and only if for every sequence $(x_n)_{n\in\mathbb{N}} \subset U$ that converges to $x \in \mathbb{R}^n$, $x \in U$.

Proof. Suppose that U is closed. Suppose that $(x_n)_{n\in\mathbb{N}} \subset U$ is a sequence such that it converges to some $x \in \mathbb{R}^n$. Suppose for a contradiction that $x \in \mathbb{R}^n \setminus U$. Then $\mathbb{R}^n \setminus U$ being open means that there exists some $r > 0$ such that $B_r(x) \subseteq \mathbb{R}^m \setminus U$. By definition of convergent, there exists $N \in \mathbb{N}$ such that $x_n \in B_r(x)$ for all $n > N$. But this contradicts the fact that $x_n \in U$. Thus we must have $x \in U$.

Now suppose that every sequence $(x_n)_{n\in\mathbb{N}} \subset U$ has limit $x \in U$. Suppose for a contradiction that U is not closed. Then $\mathbb{R}^n \setminus U$ is not open and that there exists $x \in \mathbb{R}^n \setminus U$ such that for all $r > 0$, $B_r(x)$ is not a subset of $\mathbb{R}^n \setminus U$. Let $(y_n)_{n\in\mathbb{N}}$ be a sequence with the property that $y_n \in B_{1/k}(x)$ but $y_n \notin \mathbb{R}^n \setminus U$. Then $y_n \in U$ and $y_n \to x \in \mathbb{R}^n \setminus U$, a contradiction. □

Definition 5.1.3.4: Bounded Sets

We say that a set $U \subset \mathbb{R}^n$ is bounded if there exists an open ball $B_r(p)$ such that $U \subset B_r(p)$.

Theorem 5.1.3.5

Let $f : \mathbb{R}^n \to \mathbb{R}^k$ be a function. Then the following are equivalent.

- f is continuous
- $V \subseteq \mathbb{R}^k$ is open implies $f^{-1}(V)$ is open
- $U \subseteq \mathbb{R}^k$ is closed implies $f^{-1}(U)$ is closed

Proof. Suppose that f is continuous. Then V being open means that if $f(x) \in V$, then $B_\epsilon(f(x)) \subseteq V$ for some $\epsilon > 0$. For this ϵ, apply continuity. Then there exists $\delta > 0$ such that $f(B_\delta(x)) \subseteq B_\epsilon(f(x))$. Then this means that

$$B_\delta(x) \subseteq f^{-1}(B_\epsilon(f(x))) \subseteq f^{-1}(V)$$

and we are done.

Now suppose that f has the second property. Let $f(x) \in V \subseteq \mathbb{R}^k$ be open. Then there exists $\epsilon > 0$ such that $B_\epsilon(f(x)) \subseteq V$. Thus $f^{-1}(B_\epsilon(f(x)))$ is also open. This being open means that there exists $\delta > 0$ such that $B_\delta(x) \subseteq f^{-1}(B_\epsilon(f(x)))$ and

$$f(B_\delta(x)) \subseteq B_\epsilon(f(x))$$

and we are done.

For the closed property, simply take the complements of the entire proof. □

Definition 5.1.3.6: Sequential Compactness

Let $K \subset \mathbb{R}^n$ be a set. Then K is sequentially compact if every sequence $(x_n)_{n\in\mathbb{N}} \subset K$ has a convergent subsequence $(x_{n_j})_{j\in\mathbb{N}}$ that converges to some $k \in K$.

Proposition 5.1.3.7: Heine-Borel Theorem

$K \subset \mathbb{R}^n$ is sequentially compact if and only if K is closed and bounded.

Proof. Let K be compact. Let $(x_n)_{n\in\mathbb{N}} \subset K$ be a convergent sequence. By sequential compactness, $(x_{n_k})_{k\in\mathbb{N}}$ is a subsequence that converges to $x \in K$. But $(x_n)_{n\in\mathbb{N}}$ has the same limit as its subsequence thus $x_n \to x$. Now suppose for a contradiction that K is unbounded. Then there exists a sequence $(x_n)_{n\in\mathbb{N}}$ such that $|x_n| \geq n$ for all $n \in \mathbb{N}$. By sequential compactness, there is a subsequence $(x_{n_k})_{k\in\mathbb{N}}$ such that its limit is $x \in K$. This means that $(x_{n_k})_{k\in\mathbb{N}}$ is bounded, a contradiction.

Now suppose that K is closed and bounded. Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in K. Then K being bounded means that $(x_n)_{n\in\mathbb{N}}$ is bounded. By the Bolzano-Weierstrass theorem, it has a convergent subsequence $(x_{n_k})_{k\in\mathbb{N}}$ with limit in K since K is closed. □

Theorem 5.1.3.8

Let $f : K \subset \mathbb{R}^n \to \mathbb{R}^m$ be continuous and K sequentially compact. Then $f(K)$ is sequentially compact.

Proof. Let $(y_n)_{n\in\mathbb{N}}$ be a sequence in $f(K)$. Then there exists a sequence $(x_n)_{n\in\mathbb{N}}$ such that $y_n = f(x_n)$ for all $n \in \mathbb{N}$. By sequential compactness, there exists a subsequence $(x_{n_k})_{k\in\mathbb{N}}$ that is convergent to $x \in K$. By sequential continuity of f at x, $y_{n_k} \to f(x)$ and we are done. □

Theorem 5.1.3.9: Extreme Value Theorem

Let $f : K \subset \mathbb{R}^n \to \mathbb{R}$ be continuous and K sequentially compact. Then there exists $a, b \in K$ such that

$$f(a) \leq f(x) \leq f(b)$$

for all $x \in K$.

Proof. We know that $f(K)$ is closed and bounded. Thus it must have a supremum M and infinum m that are finite. Then we must have $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ such that $a_n \to m$ and $b_n \to M$ by definition of supremum and infinum. Since $f(K)$ is closed, $m, M \in f(K)$ and we are done. □

5.1.4 The Space of Matrices

Definition 5.1.4.1: Frobenius Norm

Let $A \in M_{k\times n}(\mathbb{R})$. Define the Frobenius norm to be

$$\|A\|_F = \left(\sum_{j=1}^{n}\sum_{i=1}^{k} a_{ij}^2\right)^{\frac{1}{2}}$$

Definition 5.1.4.2: Operator Norm

Let $T \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^k)$. Define the operator norm to be

$$\|T\| = \sup_{x\in\mathbb{R}^n\setminus\{0\}} \frac{|T(x)|}{|x|} = \sup_{\substack{x\in\mathbb{R}^n \\ |x|=1}} |T(x)|$$

The definition of the opaertor norm given on S^{n-1} is particularly useful because S^{n-1} is sequentially compact. Since $\|A\|$ is a supermum, there must exist a sequence $(x_n)_{n\in\mathbb{N}}$ such that $|Ax_n| \to \|A\|$.

Proposition 5.1.4.3

Let $T, S \in L(\mathbb{R}^n, \mathbb{R}^k)$. The operator norm satisfies the following and thus is a norm on $L(\mathbb{R}^n, \mathbb{R}^k)$

- $\|T\| = 0$ if and only if T is the zero mapping
- $\|\lambda T\| = |\lambda|\|T\|$ for any $\lambda \in \mathbb{R}$
- $\|T + S\| \leq \|T\| + \|S\|$

We want to compare the two norms and show that they are equivalent. But they are in two different vector spaces. We need to associate the two vector space into the same one first.

Proposition 5.1.4.4

Both $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^k)$ and $M_{k\times n}(\mathbb{R})$ is a vector space over $\mathbb{R}$ where $n, k \in \mathbb{N} \setminus \{0\}$. Moreover, we have

$$\mathcal{L}(\mathbb{R}^n, \mathbb{R}^k) \cong M_{k\times n}(\mathbb{R}) \cong \mathbb{R}^{nk}$$

Proof. The fact that $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^k)$ and $M_{k\times n}(\mathbb{R})$ is a vector space is trivial. Since they have the same dimension and we know that all finite dimensional vector space is isomorphic to $\mathbb{R}^p$ for some $p \in \mathbb{N} \setminus \{0\}$, we get the desired result. □

Proposition 5.1.4.5

The frobenius norm and the operator norm are equivalent.

Proof. Given that the they actually act on the same vector space, we can associate every linear transformation T with the matrix A. Now we have

$$\begin{aligned}
\|A\|_F^2 &= \sum_{j=1}^{n}\sum_{i=1}^{k} a_{ij}^2 \\
&= \sum_{j=1}^{n} |T(e_j)|^2 \\
&= \sum_{j=1}^{n} |T(e_j)|^2 \\
&\leq \|T\|^2 \sum_{j=1}^{n} |e_j|^2 && \left(\frac{|T(x)|}{|x|} \leq \|T\|\right) \\
&= n\|T\|^2
\end{aligned}$$

Thus we have $\|A\|_F \leq \sqrt{n}\|T\|$.

Now for any $x \in \mathbb{R}^n$, we also have

$$\begin{aligned}
|T(x)|^2 &= |Ax|^2 \\
&= \sum_{i=1}^{k} \left(\sum_{j=1}^{n} a_{ij} x_j \right)^2 \\
&\leq \sum_{i=1}^{k} \left(\left(\sum_{j=1}^{n} a_{ij}^2 \right) \left(\sum_{j=1}^{n} x_j^2 \right) \right) && \text{(Cauchy-Schwarz Inequality)} \\
&= \left(\sum_{i=1}^{k} \sum_{j=1}^{n} a_{ij}^2 \right) |x|^2 \\
&= \|A\|_F^2 |x|^2
\end{aligned}$$

If $x \neq 0$, we can write it as

$$\begin{aligned}
\frac{|T(x)|^2}{|x|^2} \leq \sup_{x \in \mathbb{R}^n \setminus \{0\}} \frac{|T(x)|^2}{|x|} &\leq \|A\|_F^2 \\
\|T\|^2 &\leq \|A\|_F^2
\end{aligned}$$

Thus we now have

$$\|T\| \leq \|A\|_F \leq \sqrt{n} \|T\|$$

and we are done. □

Now we can make sense of using the operator norm for matrices, and the frobenius norm for linear maps.

Proposition 5.1.4.6

Let $T \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^k)$ and $S \in \mathcal{L}(\mathbb{R}^k, \mathbb{R}^m)$. Then

$$\|S \circ T\| \leq \|S\| \|T\|$$

Proof. We have that

$$\begin{aligned}
|S(T(x))| &\leq \|S\| |T(x)| \\
&\leq \|S\| \|T\| |x|
\end{aligned}$$

Thus we are done. □

Proposition 5.1.4.7

The function $\det(\cdot) : \mathbb{R}^{n \times n} \to \mathbb{R}$, which is the determinant, is continuous with respect to the elements of the matrix.

Proof. Note that $\mathbb{R}^{n \times n} \cong \mathbb{R}^{n^2}$ thus the determinant really is just a linear form of $\mathbb{R}^{n^2}$ since the determinant is defined by a polynomial. Polynomials are clearly continuous thus we are done. □

Proposition 5.1.4.8

Let $n \in \mathbb{N} \setminus \{0\}$. Then

$$GL(n, \mathbb{R}) \subset L(\mathbb{R}^n)$$

is open.

Proof. Consider the function $\det(\cdot) : \mathbb{R}^{n \times n} \to \mathbb{R}$. We have that $\det(GL(n, \mathbb{R})) = \mathbb{R} \setminus \{0\}$ and image is clearly open in $\mathbb{R}$. Thus by continuity with open sets, $GL(n, \mathbb{R})$ is open. □

Proposition 5.1.4.9

Let $A \in GL(n, \mathbb{R})$. If $B \in M_{n \times n}(\mathbb{R})$ and $\|B - A\| < \frac{1}{\|A^{-1}\|}$, then B is invertible. This means that

$$B_{1/\|A^{-1}\|}(A) \subset GL(n, \mathbb{R})$$

is open in $M_{n \times n}(\mathbb{R})$. Furthermore, we must have

$$\|B^{-1}\| \leq \frac{1}{\frac{1}{\|A^{-1}\|} - \|B - A\|}$$

Proposition 5.1.4.10

Let $f : \mathbb{R} \to M_{k \times n}(\mathbb{R})$ be a function. Then f is continuous if and only if every item in the matrix is a continuous function in the sense of a real function.

Proof. This is clear since $M_{k \times n}(\mathbb{R}) \cong \mathbb{R}^{kn}$ and we have proved this in the above section. □

5.2 Differentiation

5.2.1 Frechet Derivative

Definition 5.2.1.1: Frechet Derivatives

Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ with U open and $x \in U$. We say that f is differentiable at $x \in U$ if there exists a linear map $T \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ such that

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

If T has a matrix representation A we write $Df(x) = A$ as the derivative of f at x.

It is important that this definition uses a linear map as an approximation rather than a matrix so that in higher order derivatives, we can quit using matrices for the approximation and instead use bilinear forms and more.

Definition 5.2.1.2: Derivative Operator

Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ be differentiable for all $x \in U$. We define the operator D on f to be the function that takes x to $Df(x)$. This means that $Df : U \subseteq \mathbb{R}^n \to M_{m\times n}(\mathbb{R})$.

Note the notational differences. Df is a function that takes in x and spits out $Df(x)$. $Df(x)$ itself is a linear map $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$. It can also be a function $Df(x) : \mathbb{R}^n \to \mathbb{R}^m$ that takes a direction $v \in \mathbb{R}^n$ and spits out the directional derivative $Df(x)v \in \mathbb{R}^m$.

We then have the extension of theorems as in the one-dimensional case.

Theorem 5.2.1.3

The linear transformation representing the derivative is unique if it exists.

Proof. Suppose that A, B both represent the derivative of $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$. Then fix $\epsilon > 0$. There exists δ_1, δ_2 such that $|h| < \min\{\delta_1, \delta_2\}$ implies

$$\frac{|f(x+h) - f(x) - Ah|}{|h|} < \epsilon$$

and

$$\frac{|f(x+h) - f(x) - Bh|}{|h|} < \epsilon$$

Then we have

$$\begin{aligned} |(A-B)h| &= |f(x+h) - f(x) - Bh - f(x+h) - f(x) - Ah| \\ &\leq |f(x+h) - f(x) - Bh| + |f(x+h) - f(x) - Ah| \\ &\leq \epsilon|h| + \epsilon|h| \\ &= 2\epsilon|h| \end{aligned}$$

Thus $A = B$ by definition of limit and we are done. □

Proposition 5.2.1.4

If $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $x \in U$, then f is continuous at x.

Proof. Since f is continuous at x, there exists $A \in \mathbb{R}^{m\times n}$ such that

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

Fix $\epsilon > 0$, there exists $\delta_1 > 0$ such that $|h| < \delta_1$

$$\begin{aligned} |f(x+h)-f(x)-Ah| &\leq \epsilon|h| \\ |f(x+h)-f(x)| &\leq |Ah|+\epsilon|h| \\ &< (\|A\|+\epsilon)|h| \end{aligned}$$

Set $\delta_2 = \min\left\{\delta_1, \frac{\epsilon}{\|A\|+\epsilon}\right\}$. Then $|h| < \delta_2$ implies

$$|f(x+h)-f(x)| < (\|A\|+\epsilon)\delta_2 < \epsilon$$

and we are done. □

Similar to continuity, differentiability in higher dimensions can be broken down by its individual components. This way we only need to show individual differentiability to save trouble.

Proposition 5.2.1.5: Componentwise Differentiability

Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^k$, where

$$f(x) = \begin{pmatrix} f_1(x) \\ \vdots \\ f_k(x) \end{pmatrix}$$

Then f is differentiable at $x \in U$ if and only if for each $i \in \{1, \dots, k\}$, $f_i : U \to \mathbb{R}$ is differentiable at x.

The rest of this section are also extensions of one-dimensional case.

Proposition 5.2.1.6

If $f : \mathbb{R}^n \to \mathbb{R}^m$ is a constant function then $Df(x) = 0$.

Proof. Suppose that $f(x) = k$ for some $k \in \mathbb{R}^m$. Then I claim that $Df(x) = 0$. Indeed since

$$\lim_{h\to 0}\frac{|0+0-0h|}{|h|}=0$$

thus the definition of differentiability is satisfied. □

Proposition 5.2.1.7

If $A \in \mathbb{R}^{m\times n}$ and $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is defined by $f(x) = Ax$, then $Df(x) = A$.

Proof. Clearly $Df(x) = A$ satisfies the definition of differentiability since

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

□

5.2.2 Jacobian Matrix and Directional Derivatives

Definition 5.2.2.1: Directional Derivative

Let $v \in \mathbb{R}^n$. Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^k$ be a function. Define the directional derivative along v passing through x to be

$$\partial_v f(x) = \lim_{t\to 0} \frac{f(x+tv)-f(x)}{t} = \frac{d}{dt} f(x+tv)\bigg|_{t=0}$$

if the limit exists.

Proposition 5.2.2.2: I

$f : U \subseteq \mathbb{R}^n \to \mathbb{R}^k$ has a directional derivative along $v \in \mathbb{R}^n$, then f is linearly continuous along v.

Proposition 5.2.2.3

Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^k$ be a function. If Df exists then $\partial_v f(x) = Df(x)v$. Moreover, $\partial_v f(x)$ is linear. Meaning

$$\partial_{av+bw} f(x) = a\partial_v f(x) + b\partial_w f(x)$$

for all $a, b \in \mathbb{R}$ and for all $v, w \in \mathbb{R}^n$.

Proof. Suppose that Df exists. Then using the definition of differentiability, we have that

$$\begin{aligned}
\lim_{t\to 0} \frac{f(x+tv)-f(x)-Df(x)(tv)}{t|v|} &= 0 \\
\lim_{t\to 0} \frac{f(x+tv)-f(x)-tDf(x)v}{t} &= 0 \\
\lim_{t\to 0} \frac{f(x+tv)-f(x)}{t} &= \lim_{t\to 0} \frac{tDf(x)v}{t} \\
\lim_{t\to 0} \frac{f(x+tv)-f(x)}{t} &= Df(x)v \\
\partial_v f(x) &= Df(x)v
\end{aligned}$$

Linearity follows since $Df(x)$ is matrix and matrix multiplication on a vector is linear. □

We should make absolutely clear distinction with the existence of frechet derivatives and the existence of directional derivatives. Even if the directional derivatives exists, as long as they are not linear, the matrix for the frechet derivative will not exist. In particular, it is possible to write out the Jacobian matrix as we see below, but this Jacobian matrix will not be equal to the matrix in the frechet derivative.

The above theorem is also useful for detecting non-differentiability. By verifying the directional derivatives are non-linear, one can use the contrapositive to prove that f would not be differentiable.

Definition 5.2.2.4: Partial Derivatives

Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^k$. Define

$$\partial_j f(x) = \lim_{t\to 0} \frac{f(x+t\mathbf{e}_j)-f(x)}{t}$$

which is the directional derivatives with standard basis as the direction. In particular,

$$\partial_j f_i(x) = \lim_{t\to 0} \frac{f_i(x+t\mathbf{e}_j)-f_i(x)}{t}$$

is just the one dimensional differentiation in real analysis and

$$\partial_j f(x) = \begin{pmatrix} \partial_j f_1(x) \\ \vdots \\ \partial_j f_n(x) \end{pmatrix}$$

The partial derivatives are simply the special case of directional derivatives, when taken their direction to be unit vectors. Therefore we also have the following lemma.

Proposition 5.2.2.5: I

the partial derivatives of $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^k$ exists, then f is separately continuous along v.

Now we have the six notions, namely continuity, linear continuity, separate continuity, differentiable, directional derivatives and partial derivatives interconnected with each other. Namely each type of differentiability implies its own continuity.

Definition 5.2.2.6: Jacobian Matrix

Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^k$ be a function and let $x \in U$. Define the Jacobian matrix at x to be

$$\partial f(x) = \begin{pmatrix} \partial_1 f_1(x) & \cdots & \partial_n f_1(x) \\ \vdots & & \vdots \\ \partial_1 f_k(x) & \cdots & \partial_n f_k(x) \end{pmatrix}$$

Clearly the existence of the Jacobian matrix simply relies on the existence of partial derivatives. Note that in order for Df to exist, we require the directional derivatives to be linear.

Theorem 5.2.2.7

If $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^k$ is differentiable at $x \in U$ and $v \in \mathbb{R}^n$, then

$$Df(x)v = \partial f(x)v$$

Proof. Note that we have

$$\begin{aligned} Df(x)v &= Df(x) \sum_{k=1}^n v_i e_i \\ &= \sum_{k=1}^n v_i Df(x) e_i \\ &= \sum_{k=1}^n v_i \partial_i f(x) \\ &= \partial f(x) v \end{aligned}$$

□

The above theorem shows us that as long as f is differentiable, the Jacobian and the matrix given in the frechet derivative will be equal and interchangable.

Lemma 5.2.2.8

If $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $x \in U$ and $v \in \mathbb{R}^n$, then

$$Df(x)v = \partial f(x)v = \partial_v f(x)$$

Proof. This immediately follows from previous theorems. □

This theorem also shows that if f is differentiable, then f must have directional derivatives.

5.2.3 Properties of the Derivative

Theorem 5.2.3.1: Algebra of Differentiable functions

If $f, g : U \to \mathbb{R}^m$ are differentiable functions at x and $\lambda, \mu \in \mathbb{R}$ are constants, then

$$D(\lambda f + \mu g)(x) = \lambda Df(x) + \mu Dg(x)$$

Lemma 5.2.3.2

Let $f : U \subset \mathbb{R}^n \to \mathbb{R}^k$, $x \in U$ and $r > 0$ such that $B_r(x) \subset U$ and $A \in L(\mathbb{R}^n, \mathbb{R}^k)$. Define $\Delta_{x,A} f : B_r(0) \to \mathbb{R}^k$ by

$$\Delta_{x,A} f(h) = \begin{cases} \frac{f(x+h)-f(x)-Ah}{|h|} & \text{if } h \neq 0 \\ 0 & \text{if } h = 0 \end{cases}$$

Then f is differentiable at x with $Df(x) = A$ if and only if $\Delta_{x,A}$ is continuous at 0.

Proof. Let $\Delta_{x,A}$ be continuous at 0. Then

$$\lim_{h \to 0} |\Delta_{x,A} f(h)| = |\Delta_{x,A} f(0)| = 0$$

Thus by definition of differentiability f is differentiable at x with $Df(x) = A$.

Now let f be differentiable at x and set $A = Df(x)$. Then by definition of differentiability we have that

$$\lim_{h \to 0} \left| \frac{f(x+h) - f(x) - Ah}{|h|} \right| = \lim_{h \to 0} |\Delta_{x,A} f(h)| = 0 = \Delta_{x,A} f(0)$$

Thus $\Delta_{x,A} f(x)$ is continuous at 0. □

Lemma 5.2.3.3

Let $\tau > 0$. Let $\xi : B_r(0) \to \mathbb{R}$ be bounded and $\nu : B_r(0) \to \mathbb{R}^k$ be continuous at $0 \in B_r(0)$ and $\nu(0) = 0$. Then

$$\delta(h) = \xi(h)\nu(h)$$

where $0 < |h| < \tau$ and $\delta(0) = 0$ is continuous at $0 \in B_r(0)$.

Proof. By continuity of ν at 0, let $\epsilon > 0$. Then there exists $\delta \in (0, \tau)$ such that $|h| < \tau$ implies $|\nu(h)| < \epsilon$. By boundedness of ξ, there exists $M > 0$ such that $|\xi| < M$ for all $h \in B_r(0) \setminus \{0\}$. Thus $0 < |h| < \delta$ implies $|\delta(h)| < M\epsilon$, meaning $\lim_{h \to 0} \delta(h) = 0 = \delta(0)$ thus we are done. □

Proposition 5.2.3.4: Chain Rule

Let $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$ and $g : V \subset \mathbb{R}^m \to \mathbb{R}^k$ be two differentiable functions. Let $x \in U$. Then $g \circ f$ is differentiable and

$$D(g \circ f)(x) = (Dg)(f(x)) \cdot Df(x)$$

Proof. Let

$$\Delta_x f(h) = \begin{cases} \frac{f(x+h)-f(x)-Df(x)h}{|h|} & \text{if } h \neq 0 \\ 0 & \text{if } h = 0 \end{cases}$$

and

$$\Delta_{f(x)} g(k) \begin{cases} \frac{g(f(x)+k)-g(f(x))-D_g(f(x))k}{|k|} & \text{if } k \neq 0 \\ 0 & \text{if } k = 0 \end{cases}$$

where both functions are continuous by lemma 2.3.2. Then we have that

$$f(x+h) = f(x) + Df(x)h + \Delta_x f(h)|h|$$

and

$$g(f(x)+k) = g(f(x)) + D_g(f(x))k + \Delta_{f(x)}g(k)|k|$$

Let $k(h) = Df(x)h + \Delta_x f(h)|h|$. Then by linearity of $D_g(f(x))$,

$$g(f(x+h)) = g(f(x)) + D_g(f(x))Df(x)h + D_g(f(x))\Delta_x f(h)|h| + \Delta_{f(x)}g(k(h))|k(h)|$$

Let

$$\delta_1(h) = D_g(f(x))(\Delta_x f(h))$$

and

$$\delta_2(h) = \begin{cases} \frac{|k(h)|}{|h|}\Delta_{f(x)}g(k(h)) & \text{if } h \neq 0 \\ 0 & \text{if } h = 0 \end{cases}$$

Then we have

$$g(f(x+h)) - g(f(x)) - D_g(f(x)) \circ Df(x)h = |h|(\delta_1(h) + \delta_2(h))$$

We now show that $\lim_{h\to 0} |\delta_1(h)| = 0$ and $\lim_{h\to 0} |\delta_2(h)| = 0$. Note that

$$|\delta_1(h)| \leq \|D_g(f(x))\||\Delta_x f(h)|$$

Since $\lim_{h\to 0} |\Delta_x f(h)| = 0$ by construction, we are done. Now let

$$\xi(h) = \frac{|k(h)|}{|h|} \leq \frac{|Df(x)h|}{|h|} + |\Delta_x f(h)| \leq \|Df(x)\| + |\Delta_x f(h)|$$

Since $\Delta_x f$ is continuous, ξ is bounded on $B_r(0) \setminus \{0\}$ for some $\tau > 0$. Now set $\nu(h) = \Delta_{f(x)}g(k(h))$. Since k is contuinuous at $k(0) = 0$ and g is differentiable at $f(x)$, we have that $\nu(h)$ is continuous and $\nu(0) = 0$, Thus we can apply lemma 2.3.3 and

$$\lim_{h\to 0} |\delta_2(h)| = 0$$

Thus we must have

$$g(f(x+h)) - g(f(x)) - D_g f(x) \circ Df(x)h = |h|(\delta_1(h) + \delta_2(h)) \to 0$$

and we are done. □

5.2.4 Mean Value Inequality

Theorem 5.2.4.1: Mean Value Theorem

Let $x, y \in \mathbb{R}^n$. Let $r : [a,b] \to \mathbb{R}^n$ be continuously differentiable. Let $f : C^1(r([a,b]), \mathbb{R}^k)$ and there exists some M such that $|\partial f(x)| \leq M$ for all $x \in U$. Then

$$|f(r(b)) - f(r(a))| \leq M \int_a^b |r'(t)|\,dt$$

Proof. We have that

$$\begin{aligned} |f(r(b)) - f(r(a))| &= \left| \int_a^b \frac{d}{dt} f(r(t))\,dt \right| && \text{(By the FTC)} \\ &= \left| \int_a^b \partial f(r(t)) r'(t)\,dt \right| \\ &\leq \int_a^b |\partial f(r(t)) r'(t)|\,dt \\ &\leq \int_a^b |\partial f(r(t))||r'(t)|\,dt \\ &\leq \int_a^b M|r'(t)|\,dt \end{aligned}$$

Thus we are done. □

Notive that $\int_a^b |r'(t)|\,dt$ is exactly the length of the path r from a to b.

Proposition 5.2.4.2

Suppose that $U \subseteq \mathbb{R}^n$ is path-connected and every path is differentiable. Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ be differentiable and that $\partial f(x) = 0$ for all $x \in U$. Then f is constant on U.

Proof. Let $x, y \in U$, then any path $r : [a, b] \to U$ with $r(a) = x$ and $r(b) = y$ is differentiable. Then

$$\begin{aligned} f(y) - f(x) &= f(r(b)) - f(r(a)) \\ &= \int_a^b \frac{d}{dt} f(r(t))\,dt && \text{(By the FTC)} \\ &= \int_a^b \partial f(r(t)) r'(t)\,dt \\ &= 0 \end{aligned}$$

Thus $f(x) = f(y)$ for any $x, y \in U$. □

5.2.5 Conditions for Differentiability

Theorem 5.2.5.1

Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^k$ and suppose that $B_r(x_0) \subset U$ for some $r > 0$. Suppose that the Jacobian $\partial f(x)$ exists and is continuous for all $x \in B_r(x_0)$. Then f is differentiable.

Proof. We show this for the case that $k = 1$ since $f : \mathbb{R}^n \to \mathbb{R}^k$ is differentiable if and only if each component $f_1, \dots, f_k$ is differentiable. Thus now $f : \mathbb{R}^n \to \mathbb{R}$. Suppose that ∂f is continuous in $B_r(x)$ for fixed x. This means that $\partial_{x_j} f_i$ is continuous for any $j \in \{1, \dots, n\}$ and

$i \in \{1, \dots, k\}$. Take a point $x + h \in B_r(x)$. Define $h = \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix}$ and

$$h'_k = \begin{pmatrix} h_1 \\ \vdots \\ h_k \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Trivially define $h'_0 = 0$ and $h_0 = 0$ for convenience. Now $f(x + h'_k) - f(x + h'_{k-1})$ is just a one variable function on the kth slot, thus we can apply the mean value theorem in real analysis and conclude that there exists $\theta_k \in (0, 1)$ such that
$f(x + h'_k) - f(x + h'_{k-1}) = \partial_{x_k} f(x + h'_{k-1} + (\theta_k h_k)e_k)h_k$.

Summing all these up, we have that

$$\begin{aligned} \sum_{k=1}^{n} f(x + h'_k) - f(x + h'_{k-1}) &= \sum_{k=1}^{n} \partial_{x_k} f(x + h'_{k-1} + (\theta_k h_k)e_k)h_k \\ f(x + h) - f(x) &= \sum_{k=1}^{n} \partial_{x_k} f(x + h'_{k-1} + (\theta_k h_k)e_k)h_k \end{aligned}$$

We now use the continuity of the partial derivatives. Given $\epsilon > 0$, there $\delta_k > 0$ such that $|h| < \delta_k$ implies $|\partial_{x_k} f(x + h) - \partial_{x_k} f(x)| < \epsilon$. Choose $\delta = \min\{\delta_1, \dots, \delta_n\}$. Since $\theta_k \in (0, 1)$ for all $k \in \{1, \dots, n\}$, we must have $\left|h'_{k-1} + (\theta_k h_k)e_k\right| < |h| < \delta$. Thus

$$\begin{aligned} \left| f(x + h) - f(x) - \sum_{k=1}^{n} \partial_{x_k} f(x)h_k \right| &= \left| \sum_{k=1}^{n} \partial_{x_k} f(x + h'_{k-1} + (\theta_k h_k)e_k)h_k - \sum_{k=1}^{n} \partial_{x_k} f(x)h_k \right| \\ &< \epsilon \sum_{k=1}^{n} |h_k| \\ &< \epsilon\sqrt{n}|h| \end{aligned}$$

if $|h| < \delta$. The second inequality is due to the fact that $\sum_{k=1}^{n} h_k \leq \sqrt{n}|h|$.

Now define $A \in \mathcal{L}(\mathbb{R}^n, \mathbb{R})$ by $A(h) = (\partial f)h$. This makes sense in matrix multiplication since $\partial f \in M_{1\times n}(\mathbb{R})$. Then $0 < |h| < \delta$ implies

$$\begin{aligned} \frac{|f(x + h) - f(x) - Ah|}{|h|} &\leq \epsilon\sqrt{n}\frac{|h|}{|h|} \\ &= \epsilon\sqrt{n} \end{aligned}$$

since $A(h) = \sum_{k=1}^{n} \partial_{x_k} f(x)h_k$. Since this is true for all ϵ, the condition for differentiability is satisfied and we are done. □

Notice that in the above proof, in order to show that f is differentiable at x, we need to know the partial derivatives of its neighbours.

The below theorem acts as a summary or recollection of the relationships between differentiability and continuous Jacobian Matrices.

Theorem 5.2.5.2

Let U be an open subset of $\mathbb{R}^n$. Then $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^k$ is continuously differentiable on U if and only if $\partial f : U \subseteq \mathbb{R}^n \to \mathbb{R}^{k\times n}$ is continuous on U.

Proof. Suppose that f is differentiable at x and its derivative $Df(x)$ is continuous. Then $Df(x) = \partial f(x)$ thus $\partial f(x)$ is also continuous.

Now suppose that $\partial f(x)$ is continuous in U. Then by the above theorem f is differentiable and thus we have $Df(x) = \partial f(x)$. Hence $Df(x)$ is also continuous. □

5.3 More Properties of Differentiability

5.3.1 Inverse Function Theorem

The inverse function theorem is a powerful multidimensional analog of finding the derivative of an inverse function. The function has a few conditions to satisfy in order for it to be applied. Since the proof is exceptionally long, I will split it to a number of theorems and prepositions.

Proposition 5.3.1.1

Suppose that $\Psi : U \to V$ is a bijection which is differentiable at $x \in U$. Suppose that Ψ^{-1} is differentiable at $y = \Psi(x) \in V$. Then $D\Psi(x)$ and $D\Psi^{-1}(y)$ are both invertible and

$$(D\Psi^{-1})(y) = (D\Psi(\Psi^{-1}(y)))^{-1}$$

Proof. Differentiating the relation $\Psi(\Psi^{-1}(y)) = y$ gives

$$D\Psi(\Psi^{-1}(y)) \circ D\Psi^{-1}(y) = I_n$$

which is the identity transformation thus we are done. □

Lemma 5.3.1.2

$T \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^k)$ is injective if and only if there exists $a > 0$ such that $|T(x)| \geq a|x|$ for all $x \in \mathbb{R}^n$.

Proof. Firstly suppose that there exists $a > 0$ such that $|T(x)| \geq a|x|$ for all $x \in \mathbb{R}^n$. Let $T(x) = T(y)$. Then

$$\begin{aligned} |T(x-y)| &\geq a|x-y| \\ |T(x) - T(y)| &\geq a|x-y| \\ 0 &\geq a|x-y| \end{aligned}$$

Since $a > 0$ we must have $x = y$.

Now suppose that for all $a > 0$, $|T(x)| < a|x|$ for all $x \in \mathbb{R}^n$. This means that $\frac{|T(x)|}{|x|} < a$ for all $a > 0$. This is precisely the definition of a limit. We can find a sequence $(x_j)_{j\in\mathbb{N}}$ such that

$$\frac{|T(x_j)|}{|x_j|} \to 0$$

Now consider a new sequence defined by $y_j = \frac{x_j}{|x_j|}$. Clearly $(y_j)_{j\in\mathbb{N}} \subset S^{n-1}$. But S^{n-1} is compact by the Heine-Borel theorem. Thus there exists $(y_{j_m})_{m\in\mathbb{N}}$ such that it has its limit in S^{n-1}. But then

$$T(y_j) = \frac{T(x_j)}{|x_j|}$$

thus $|T(y_j)| \to 0$. All the y_j are clearly nonzero, this means that T has a non-trivial kernel and thus T is not injective. □

We split the inverse function theorem into its injective and surjective part for better readability.

Proposition 5.3.1.3: Injective Part of Inverse Function Theorem

Let U be an open subset of $\mathbb{R}^n$ and suppose that $\Psi : \mathbb{R}^n \to \mathbb{R}^m$ such that it is differentiable and its derivative is continuous in U. Assume that $D\Psi(p)$ is injective at a point $p \in U$. Then there exists $\delta > 0$ such that $B_\delta(p) \subset U$ and such that f is injective on $B_\delta(p)$.

Proof. From the above lemma, $D\Psi(p)$ being injective means that there exists $a > 0$ such that

$$|Df(p)h| \geq \epsilon|h|$$

for all $h \in \mathbb{R}^n$. Since $Df : U \to \mathcal{L}(\mathbb{R}^n, \mathbb{R}^k)$ is continuous, there exists $\delta > 0$ such that $x \in B_\delta(p) \subset U$ implies

$$\|Df(p) - Df(x)\| < \frac{1}{2}\epsilon$$

Define a new function $F : U \to \mathbb{R}^k$ by $F(x) = f(x) - Df(p)x$. Then F is differentiable since $Df(p)x$ is just a linear transformation and we have

$$DF(x) = Df(x) - Df(p)$$

Thus we now have

$$\|DF(x)\| = \|Df(x) - Df(p)\| < \frac{1}{2}\epsilon$$

We can now apply the mean value inequality to get

$$|F(z) - F(x)| \leq \frac{1}{2}\epsilon|z - x|$$

for all $x, z \in B_\delta(p)$. Finally we have

$$\begin{aligned} |f(x) - f(z)| &= |Df(p)(x - z) - (F(z) - F(x))| \\ &\geq \epsilon|x - z| - \frac{1}{2}\epsilon|x - z| \\ &= \frac{1}{2}\epsilon|x - z| \end{aligned}$$

This means that $x \neq z$ implies $f(x) \neq f(z)$ and we are done. □

Proposition 5.3.1.4: Surjective Part of Inverse Function Theorem

Let U be an open subset of $\mathbb{R}^n$ and suppose that $\Psi : \mathbb{R}^n \to \mathbb{R}^n$ such that it is differentiable and its derivative is continuous in U. Assume that $D\Psi(p)$ is surjective at a point $p \in U$. Then there exists $\rho > 0$ such that $B_\rho(\Psi(p)) \subset \Psi(U)$.

Proof. By the rank nullity theorem, $D\Psi(p)$ is injective (surjectivity implies bijectivity in linear maps). By the above proposition, there exists $\epsilon > 0$ such that

$$|D\Psi(p)h| \geq \epsilon|h|$$

for all $h \in \mathbb{R}^n$. Again define $F : U \to \mathbb{R}^n$ by $F(x) = \Psi(x) - Df(p)x$. Then exactly the same as the above proof, we must have $|F(x) - F(z)| \leq \frac{1}{2}\epsilon|x - z|$ and $|\Psi(x) - \Psi(z)| \geq \frac{1}{2}\epsilon|x - z|$. Set

$$K = \overline{B_{\frac{1}{2}\delta}(p)} = \{x \in \mathbb{R}^n||x - p| \leq \frac{1}{2}\delta\}$$

and $\partial K = \{x \in \mathbb{R}^n||x - p| = \frac{1}{2}\delta\}$. Now we have

$$\begin{aligned} |\Psi(x) - \Psi(p)| &\geq \frac{1}{2}\epsilon|x - p| \\ &= \frac{1}{4}\epsilon\delta \end{aligned}$$

For all $x \in \partial K$. Set $\rho = \frac{1}{8}\epsilon\delta$ and fix $y \in B_\rho(\Psi(p))$. We will show that

$$y \in \Psi(B_{\frac{1}{2}\delta}(p))$$

to finish the proof. Define a new function $\phi : K \to \mathbb{R}$ by $\phi(x) = |\Psi(x) - y|$. Then ϕ is continuous by composition of continuous functions. By the extreme value theorem, there exists $c \in K$ such that $\phi(c) \leq \phi(x)$ for all $x \in K$.

Now I will show that $c \in B_{\frac{1}{2}\delta}(p)$ by showing that $c \notin \partial K$ and using the fact that $B_{\frac{1}{2}\delta}(p) = K \setminus \partial K$. $c \notin \partial K$ can be proved by $\phi(x) > \phi(p)$ for all $x \in \partial K$. Now if $x \in \partial K$, we have

$$\begin{aligned} \phi(x) &= |\Psi(x) - y| \\ &\geq |\Psi(x) - \Psi(p)| - |y - \Psi(p)| \\ &\geq \frac{1}{4}\epsilon\delta - \frac{1}{8}\epsilon\delta \\ &= \rho \\ &> |y - \Psi(p)| \\ &= \phi(p) \end{aligned}$$

Thus we are done with this part.

Now since $D\Psi(p)$ is surjective, there exists $h \in \mathbb{R}^n$ such that $D\Psi(p)h = y - \Psi(c)$ (by nuisances). Since $c \in B_{\frac{1}{2}\delta}(p)$, there exists $\nu > 0$ such that $|t| < \nu$ implies $|c + th - p| < \frac{1}{2}\delta$. Thus we have

$$\begin{aligned} \Psi(c + th) - y &= \Psi(c + th) - \Psi(c) - tD\Psi(p)h + (\Psi(c) - y + tD\Psi(p)h) \\ &= F(c + th) - F(c) + (1 - t)(\Psi(c) - y) \end{aligned}$$

for $|t| < \nu$. Now, we have that

$$\begin{aligned} \phi(c) &\leq \phi(c + th) \\ &= |\Psi(c + th) - y| \\ &\leq \frac{1}{2}t|D\Psi(p)h| + (1 - t)|\Psi(c) - y| \\ &= \left(1 - \frac{1}{2}t\right)(\Psi(c) - y) \end{aligned}$$

Since this holds for all $|t| < \nu$, we must have that $\phi(c) = 0$, which implies that $y = \Psi(c)$. Thus we have shown that

$$B_\rho(\Psi(p)) \subset \Psi(B_{\frac{1}{2}\delta}(p)) \subset \Psi(U)$$

□

Corollary 5.3.1.5

Let $U \subset \mathbb{R}^n$ be open. Let $\Psi \in \mathcal{C}^1(U, \mathbb{R}^n)$ and suppose that $D\Psi(p)$ is invertible for all points $p \in U$, then Ψ maps open subsets of U to open subsets of $\mathbb{R}^n$.

Proof. Let $V \subset U$ be open. By the above proposition applied to $\Psi|_V$, we have for all $p \in V$, there exists $\rho > 0$ such that

$$B_\rho(\Psi(p)) \subset \Psi(V)$$

, thus $\Psi(V)$ is open. □

Theorem 5.3.1.6: Inverse Function Theorem

Let U be an open subset of $\mathbb{R}^n$ and suppose that $\Psi \in \mathcal{C}^1(U, \mathbb{R}^n)$. Assume that $D\Psi(p)$ is invertible at a point $p \in U$. Let $q = \Psi(p)$. Then there exists a neighbourhood N_p and N_q such that

$\Psi : N_p \to N_q$ is a bijection and $\Psi^{-1} : N_q \to N_p$ is continuously differentiable and

$$(D\Psi^{-1})(y) = (D\Psi(\Psi^{-1}(y)))^{-1}$$

for all $y \in N_q$.

Proof. Clearly Ψ satisfies the above two propositions. Thus this means that there exists $\epsilon > 0$ such that $|Df(p)h| \geq \epsilon|h|$ for all $h \in \mathbb{R}^n$ and that there exists $\delta > 0$ such that $x \in B_\delta(p) \subset U$ implies

$$\|Df(p) - Df(x)\| < \frac{1}{2}\epsilon$$

Using these two facts and the rank nullity theorem we deduce that $D\Psi(x)$ is invertible for all $x \in B_\delta(p)$. The above corollary implies that $\Psi(B_\delta(p))$ is open. Thus $\Psi|_{B_\delta(p)} : B_\delta(p) \to \Psi(B_\delta(p))$ is a bijection.

Now we show that Ψ^{-1} is continuously differentiable. □

Note that there could be multiple p such that Ψ maps p to q. The inverse function theorem guarantees that as long as $D\Psi(p)$ is invertible, then they will have bijective neighbourhoods. This is why we can have multiple branches for the inverse. A good example would be $y = x^2$. It has two inverses, namely $\sqrt{x}$ and $-\sqrt{x}$ precisely because of this reason. In this case the two neighbourhoods would be $\mathbb{R}^+$ and $\mathbb{R}^-$ respectively.

Furthermore, even if there are multiple p mapping to q, not every p could have a neighbourhood such that they are bijective because we must require the fact that $D\Psi(p) \neq 0$.

5.3.2 Implicit Function Theorem

Theorem 5.3.2.1

Let U be an open subset of $\mathbb{R}^{n+l}$ and $c \in \mathbb{R}^l$. Suppose that $F \in \mathcal{C}^1(U, \mathbb{R}^l)$ and that the equation $F(x, y) = c$ has a solution $(x_0, y_0) \in U$ such that $\det(\partial_y F(x_0, y_0)) \neq 0$. Then there exists an open set $x_0 \in N_{x_0} \subset \mathbb{R}^n$ and $g \in \mathcal{C}^1(N_{x_0}, \mathbb{R}^l)$ such that

- $g(x_0) = y_0$, $\{(x, g(x)) : x \in N_{x_0}\} \subset U$ and $F(x, g(x)) = c$ for all $x \in N_{x_0}$
- $\partial_y F(x, g(x))$ is invertible for all $x \in N_{x_0}$ and

$$\partial g(x) = -(\partial_y F(x, g(x)))^{-1} \cdot \partial_x F(x, g(x))$$

  for all $x \in N_{x_0}$

Beware that more often we seen that the invertible matrix is not necessarily on the right hand side of ∂F. It could consist of multiple columns of ∂F simply because of the ordering of the variables which is completely by the writer's choice. Moreover, there may be more than one g to convert the implicit function into an explicit one if you consider different variables for the domain and the codomain.

5.3.3 Higher Order Derivatives

From here onwards, our function $f : \mathbb{R}^n \to \mathbb{R}$ will be a scalar function else the Hessian Matrix cannot be defined. Do note that general second order differential operators for $f : \mathbb{R}^n \to \mathbb{R}^m$ do exists. It is just that we will not discuss it here.

Definition 5.3.3.1: Second Order Partial Derivatives

Suppose that $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ is differentiable with partial derivative operator $\partial_j f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ for $1 \leq j \leq n$. Define the second order partial derivative at $x_0 \in U$ to be

$$\partial_{ij} f(x_0) = \frac{\partial}{\partial x_i} \partial_j f(x) \bigg|_{x=x_0} = \frac{\partial^2}{\partial x_i \partial x_j} f(x) \bigg|_{x=x_0}$$

for $1 \leq i \leq n$ if it exists. This is done by treating $\partial_j f$ as a function from $\mathbb{R}^n$ to $\mathbb{R}$ and taking the partial derivative of it.

Definition 5.3.3.2: Hessian Matrix

Suppose that all second order partial derivatives of $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ exists. Define the Hessian Matrix f to be

$$H_f(x) = \partial^2 f(x) = \begin{pmatrix} \partial_{11} f(x) & \cdots & \partial_{1n} f(x) \\ \vdots & \ddots & \vdots \\ \partial_{n1} f(x) & \cdots & \partial_{nn} f(x) \end{pmatrix}$$

With $f : \mathbb{R}^n \to \mathbb{R}$, to say that Df is differentiable means that there exists a linear operator T such that

$$\lim_{h \to 0} \frac{|Df(x+h) - Df(x) - T(h)|}{|h|} = 0$$

as defined by the definition of differentiability. Notice that $Df(x) \in \mathcal{L}(\mathbb{R}^n, \mathbb{R})$. This means that T must take the form $T : \mathbb{R}^n \to \mathcal{L}(\mathbb{R}^n, \mathbb{R})$ so that $T(h) \in \mathcal{L}(\mathbb{R}^n, \mathbb{R})$ makes sense. Thus $T \in \mathcal{L}(\mathbb{R}^n, \mathcal{L}(\mathbb{R}^n, \mathbb{R}))$. This leads to the following identification:

Proposition 5.3.3.3

Let $n \in \mathbb{N} \setminus \{0\}$. Then

$$\mathcal{L}(\mathbb{R}^n, \mathcal{L}(\mathbb{R}^n, \mathbb{R})) \cong \mathcal{L}^2(\mathbb{R}^n, \mathbb{R})$$

where $\mathcal{L}^2(\mathbb{R}^n, \mathbb{R})$ is the space of all bilinear forms $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$.

Proof. Let $T \in \mathcal{L}(\mathbb{R}^n, \mathcal{L}(\mathbb{R}^n, \mathbb{R}))$. Then for every $x \in \mathbb{R}^n$, $T(x) : \mathbb{R}^n \to \mathbb{R}$. Thus $T(x)(y) \in \mathbb{R}$ with $x, y \in \mathbb{R}^n$. Obviously we can write it into $T(x)(y) = B(x, y)$. Conversely any bilinear form can be written into the above form. In particular they both are vector spaces of dimension n^2.

From linear algebra we know that $B(x, y) = x^T A y$ for some matrix A. If $T(x) = Cx$ with C representing T, then

$$B(x, y) = (T(x)) \cdot y = (Cx)^T y = x^T C^T y$$

□

Proposition 5.3.3.4

Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ be differentiable for all $x \in U$ and $Df : U \subseteq \mathbb{R}^n \to \mathcal{L}(\mathbb{R}^n, \mathbb{R})$ be differentiable. Then $D(Df)(x) = H_f(x)$. Moreover, $H_f(x)$ is symmetric and all second order partial derivatives commute.

Lemma 5.3.3.5

If all second order partial derivatives of $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ at x is continuous for $x \in U$, then

$$\frac{\partial^2}{\partial x_i \partial x_j} f(x) = \frac{\partial^2}{\partial x_j \partial x_i} f(x)$$

Definition 5.3.3.6: Twice differentiable and Continuous derivatives

Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ where $f(x) = \begin{pmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{pmatrix}$ with $f_k(x) : \mathbb{R}^n \to \mathbb{R}$ for $k \in \{1, \dots, m\}$. Define

$$\mathcal{C}^2(U, \mathbb{R}^m) = \{f : U \to \mathbb{R}^m | \partial^2 f_k(x) : U \subseteq \mathbb{R}^n \to \mathbb{R}^n \text{ is continuous for } k \in \{1, \dots, m\}\}$$

Theorem 5.3.3.7

[Second Order Taylor Expansion] Let $U \subset \mathbb{R}^n$ be convex and $x, x + h \in U$. If $f \in \mathcal{C}^2(U)$ then

$$f(x+h) = f(x) + \sum_{k=1}^n h_i \frac{\partial f(x)}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^n h_i h_j \frac{\partial^2 f(x)}{\partial x_i \partial x_j} + R(h)$$

where

$$\lim_{h \to 0} \frac{|R(h)|}{|h|^2} = 0$$

5.3.4 Second Order Derivative Test

Definition 5.3.4.1: Critical Point

We say that $p \in U$ is a critical point of $f \in \mathcal{C}^1(U)$ if $\nabla f(p) = 0$.

Proposition 5.3.4.2

If f has a local minimum ot maximum at p then $\nabla f(p) = 0$.

Theorem 5.3.4.3: Second Order Derivative Test

Suppose that $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ such that $H_f(x)$ exists and is continuous ($f \in \mathcal{C}^2(U)$). Suppose that $\nabla f(p) = 0$ for some $p \in U$ (All first order derivatives are 0 at p or p is a critical point).

- If $x^T H_f(p) x > 0$ for all $x \in \mathbb{R}^n \setminus \{0\}$ then f has a strict local minimum at p
- If $x^T H_f(p) x < 0$ for all $x \in \mathbb{R}^n \setminus \{0\}$ then f has a strict local maximum at p
- If there exists $x, y \in \mathbb{R}^n \setminus \{0\}$ such that $x^T H_f(p) x > 0$ and $y^T H_f(p) y < 0$ then p is a saddle point
- The test is inconclusive otherwise.

5.4 Vector Calculus

5.4.1 Regions

Before we define integration of functions, we need to make sense of the domain of our integral.

Definition 5.4.1.1: Regions

A region in $\mathbb{R}^n$ is a bounded open subset Ω of $\mathbb{R}^n$ such that there exists a function $f : \mathbb{R}^n \to \mathbb{R}$ with the property that

- all partial derivatives of f are continuous
- $\Omega = \{x \in \mathbb{R}^n | f(x) < 0\}$
- $\nabla f(p) \neq 0$ for all $p \in \mathbb{R}^n$ such that $f(p) = 0$.

Also define the boundary of a region, $\partial\Omega$ to be $\{x \in \mathbb{R}^n | f(x) = 0\}$.

Proposition 5.4.1.2

Let $\Omega \subset \mathbb{R}^n$ be a region with boundary $f(x) = 0$. Then there exists $0 < l < n$ and $r : \mathbb{R}^l \to \mathbb{R}^n$ such that $f(r(x)) = 0$. We say that r is the parametrization of $\partial\Omega$.

Proof. Since $\nabla f(p) \neq 0$ for all $p \in \mathbb{R}^n$, we can apply the implicit function theorem to find function $g : \mathbb{R}^l \to \mathbb{R}^{n-l}$ such that $f(x, g(x)) = 0$ for any $p \in f^{-1}(0)$. Now define $r : \mathbb{R}^l \to \mathbb{R}^n$ by $r(x) = (x, g(x))$. Then clearly $f(r(x)) = 0$. □

In particular, if $l = 1$, then we call r a curve. If $l = 2$, then r would be a surface. This is precisely the parametrization of a boundary, be it a curve or a surface.

Definition 5.4.1.3: Tangent Space

Let $\Omega \subset \mathbb{R}^n$ be a region with boundary $f(x) = 0$ and parametrization $r : \mathbb{R}^l \to \mathbb{R}^n$. Define the tangent space at $p \in \partial\Omega$ to be

$$T_p(\partial\Omega) = \text{span}\{\partial_{x_1} r, \dots, \partial_{x_l} r\}$$

Tangent spaces will be given a more rigorous definition in differential geometry. But the notion does coincide.

Proposition 5.4.1.4

Let $\Omega \subset \mathbb{R}^n$ be a region with boundary $f(x) = 0$ and parametrization $r : \mathbb{R}^l \to \mathbb{R}^n$. Let $p \in \partial\Omega$. Then

$$T_p(\partial\Omega) = \ker(\partial f)$$

Proof. Notice that from the equation $f(r(x)) = 0$, we have that by the chain rule,

$$(\partial f)(\partial r) = 0$$

But ∂r is precisely $\left(\partial_{x_1} r, \dots, \partial_{x_l} r\right)$. Thus we are done. □

5.4.2 Differentiation and Integration of Curves

This section is dedicated to vector functions, which are functions with domain $\mathbb{R}$ and codomain $\mathbb{R}^n$.

Definition 5.4.2.1: Curves

Let $r : I \subset \mathbb{R} \to \mathbb{R}^n$. Then

$$C = \{\mathbf{r}(t) | t \in I\}$$

is called a curve in $\mathbb{R}^n$. r is a parametrization of C, and is called a vector function.

Definition 5.4.2.2: Classifying Curves

Let $r : I \subseteq \mathbb{R} \to \mathbb{R}^3$ induce a curve C.

- We say $r \in \mathcal{C}^k$ if it can be differentiated k times. It is smooth if $r \in \mathcal{C}^\infty$
- C is simple if it does not intersect itself
- C is closed if $I = [a, b]$ and $r(a) = \mathbf{r}(b)$
- C is regular if $r'(t) \neq 0$ for all $t \in I$

From regions, curves arises. On the other hand, we can construct regions from sufficiently nice curves.

Proposition 5.4.2.3

A curve $r : I \to \mathbb{R}^n$ gives rise to a region Ω if r is in $\mathcal{C}^1$, r is simple, closed and regular.

Definition 5.4.2.4: Arc Length Function and Parametrization

Let $r : I \to \mathbb{R}^n$ be a vector function. Define the Arc Length Function by

$$s(t) = \int_{t_0}^{t} |\mathbf{r}'(u)| du$$

Define the Arc Length Parametrization by $\mathbf{r}(s)$.

Lemma 5.4.2.5

$\mathbf{r}(s)$ has unit speed.

Proof. We show that $|\mathbf{r}'(s)| = 1$. By the chain rule,

$$\begin{aligned} \left| \frac{d}{ds} \mathbf{r}(s(t)) \right| &= \left| \frac{d\mathbf{r}}{dt} \cdot \frac{dt}{ds} \right| \\ &= \frac{|\mathbf{r}'(t)|}{|\mathbf{r}'(t)|} && \text{(By the FTC)} \\ &= 1 \end{aligned}$$

□

Definition 5.4.2.6: Line Integrals on Scalar Functions

Let $f : \mathbb{R}^n \to \mathbb{R}$. Let $r : [a, b] \to \mathbb{R}^n$ a smooth curve with image C. Define the line integral along a function $f : \mathbb{R}^n \to \mathbb{R}$ to be

$$\int_C f \, dr = \int_a^b f(r(t)) |r'(t)| \, dt$$

If r is a closed curve then we write the integral as $\oint_C f$.

dr in the integral is justified by noting that if r is a function of t, then $dr = r'(t)dt$. But since dr is a positive line segment, we take the absolute value, giving us

$$dr = |dr| = |r'(t)||dt| = dt$$

There is no proper infinite sum that we can justify for $\int_C f$ because it is simply a notation short hand for the integral on the right hand side. Notice that the parameterization r does not appear on this notation because there can be mutiple parametrizations with the same image.

5.4.3 Vector Fields

Definition 5.4.3.1: Vector Fields

We say that a function $v : U \subset \mathbb{R}^n \to \mathbb{R}^n$ is a vector field. For every $x \in \mathbb{R}^n$ it assigns a vector $v(x) \in \mathbb{R}^n$ to it. If each $v_k \in \mathcal{C}^1$ then we say that $v \in \mathcal{C}^1$.

Definition 5.4.3.2: Conservative Vector Fields

We say that a vector field $v : U \subset \mathbb{R}^n \to \mathbb{R}^n$ is conservative if there exsits some scalar function $\phi : \mathbb{R}^n \to \mathbb{R}$ such that

$$v = \nabla\phi$$

Definition 5.4.3.3: Divergence

Let $f : U \subset \mathbb{R}^n \to \mathbb{R}^n$ be a vector field. Define the divergence of f to be

$$\nabla \cdot f = \begin{pmatrix} \frac{\partial}{\partial x_1} & \cdots & \frac{\partial}{\partial x_n} \end{pmatrix} \begin{pmatrix} f_1 \\ \vdots \\ f_n \end{pmatrix} = \sum_{k=1}^{n} \frac{\partial f_k}{\partial x_k}$$

Definition 5.4.3.4: Line Integrals on Vector Fields

Let $v : U \subset \mathbb{R}^n \to \mathbb{R}^n$ be a vector field. Let $r : [a, b] \to \mathbb{R}^n$ be a curve with image $C \in U$. Define the integral of v along the curve C to be

$$\int_C v \cdot\, dr = \int_a^b v(r(t)) \cdot r'(t)\, dt$$

If r is a closed curve then we write the integral as $\oint_C v \cdot\, dr$.

Proposition 5.4.3.5

Let $v : U \subset \mathbb{R}^n \to \mathbb{R}^n$ be a vector field. Let $r : [a, b] \to \mathbb{R}^n$ be a curve with image $C \in U$. Then

$$\int_C v \cdot\, dr = \sum_{k=1}^{n} \int_C v_k\, dx_k$$

Proposition 5.4.3.6

If $v : U \subset \mathbb{R}^n \to \mathbb{R}^n$ is conservative with $v = \nabla f$ and $r : [a, b] \to \mathbb{R}^n$ is a curve with image C then

$$\int_C v \cdot dr = f(r(b)) - f(r(a))$$

Theorem 5.4.3.7: Equivalent Characterization of Conservative Vector Fields

Let $\gamma : [a, b] \to \mathbb{R}^n$ be a path. Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be a vector field. Then the following are equivalent.

- F is conservative
- $\int_\gamma F(r) \cdot dr$ is independent of the choice of γ
- $\int_\gamma F(r) \cdot dr = 0$ if γ is closed

5.5 2 Dimensional Calculus

5.5.1 Double Integrals

Definition 5.5.1.1: Positively Oriented Tangents and Outward Normals

Let Ω be a region in $\mathbb{R}^2$ with boundary $f(x) = 0$ and parametrization $r : \mathbb{R} \to \mathbb{R}^2$. We say that r is positively oriented if as the domain of r increases, the tangent vector $r'(t)$ moves anti-clockwise. In this case, we say that $\begin{pmatrix} -r_2(p) \\ r_1(p) \end{pmatrix}$ for $p \in \partial\Omega$ is the outward normal vector which is perpendicular to the tangent space $T_p(\partial\Omega)$.

Definition 5.5.1.2: Double Integrals

Let $f : \mathbb{R}^2 \to \mathbb{R}$. Define the double integral over f on a region Ω to be

$$\iint_\Omega f(x, y)\,dx\,dy$$

Proposition 5.5.1.3

Let $f, g : \mathbb{R}^2 \to \mathbb{R}$ be integrable on a region Ω and $a, b \in \mathbb{R}$. Then the following are true for double integrals.

- $\iint_\Omega (af(x, y) + bg(x, y))\,dA = a\iint_\Omega f(x, y)\,dA + b\iint_\Omega g(x, y)\,dA$
- If $f(x, y) \geq g(x, y)$ for all $x, y \in \Omega$, then $\iint_\Omega f(x, y)\,dA \geq \iint_\Omega g(x, y)\,dA$
- If $\Omega = D_1 \cup D_2$ and $D_1 \cap D_2 = \emptyset$, then $\iint_\Omega f(x, y)\,dA = \iint_{D_1} f(x, y)\,dA + \iint_{D_2} f(x, y)\,dA$

Proposition 5.5.1.4: Fubini's Theorem

If $f : \mathbb{R}^2 \to \mathbb{R}$ is continuous and bounded on the rectangle $R = \{(x, y) \in \mathbb{R}^2 | a \leq x \leq b, c \leq y \leq d\}$. Then

$$\int_a^b \int_c^d f(x, y)\,dy\,dx = \int_c^d \int_a^b f(x, y)\,dx\,dy$$

Proposition 5.5.1.5

If $f : \mathbb{R}^2 \to \mathbb{R}$ is a function such that $f(x, y) = g(x)h(y)$ where $g, h : \mathbb{R} \to \mathbb{R}$ and $R = \{(x, y) \in \mathbb{R}^2 | a \leq x \leq b, c \leq y \leq d\}$, then

$$\iint_R f(x, y)\,dA = \int_a^b g(x)\,dx \int_c^d f(y)\,dy$$

Theorem 5.5.1.6

Let R be a region. Let $f : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $f(x, y) = \begin{pmatrix} u(x, y) \\ v(x, y) \end{pmatrix}$ such that $f|_R$ is a bijection and $\partial f \in \mathcal{C}^1(R)$. Let $g : \mathbb{R}^2 \to \mathbb{R}$. Then

$$\iint_S g(u, v)\,du\,dv = \iint_R g(x, y)|\partial f|\,dx\,dy$$

The conditions for f being a local bijection is given by the inverse function theorem.

Theorem 5.5.1.7: Moments

Let $f : \mathbb{R}^2 \to \mathbb{R}$. Let $\rho(x, y)$ give the moment at (x, y). The moment about the x-axis on a region $D \subset \mathbb{R}^2$ is given by

$$M_x = \iint_D y\rho(x, y)\, dA$$

and the moment about the y-axis is given by

$$M_y = \iint_D x\rho(x, y)\, dA$$

Theorem 5.5.1.8: Mass

Let $f : \mathbb{R}^2 \to \mathbb{R}$. Let $\rho(x, y)$ give the moment at (x, y). Let $D \subset \mathbb{R}^2$ be a region. Then the mass is given by

$$m = \iint_D \rho(x, y)\, dA$$

Theorem 5.5.1.9: Center of Mass

Let $f : \mathbb{R}^2 \to \mathbb{R}$. Let $\rho(x, y)$ give the moment at (x, y). Let $D \subset \mathbb{R}^2$ be a region. The center of mass is given by

$$(\overline{x}, \overline{y}) = \left(\frac{M_y}{m}, \frac{M_x}{m}\right)$$

5.5.2 Green's Theorem for a Planar Region

Definition 5.5.2.1: Curl

Let $v : U \subseteq \mathbb{R}^2 \to \mathbb{R}^2$ be a continuously differentiable planar vector field given by $v(x, y) = (a(x, y), b(x, y))$. Define the curl of v to be

$$\operatorname{curl}(v) = \frac{\partial b}{\partial x} - \frac{\partial a}{\partial y}$$

Theorem 5.5.2.2: Green's Theorem for a Rectangular Region

Let $v : U \subseteq \mathbb{R}^2 \to \mathbb{R}^2$ be a continuously differentiable planar vector field where $v(x, y) = \begin{pmatrix} v_1(x, y) \\ v_2(x, y) \end{pmatrix}$. Let $\Omega = \{(x, y) \in \mathbb{R}^2 | a \leq x \leq b, c \leq y \leq d\}$ such that $\partial\Omega \subset U$. Then

$$\iint_\Omega \operatorname{curl}(v)\, dA = \int_{\partial\Omega} v \cdot dr$$

where r is positively oriented.

Proof. Clearly, we have

$$\iint_\Omega \operatorname{curl}(v)\, dA = \int_a^b v_1(x, c)\, dx + \int_b^c v_2(d, y)\, dy + \int_c^a v_1(x, d)\, dx + \int_d^b v_2(a, y)\, dy$$

□

Theorem 5.5.2.3: Green's Theorem for a Planar Region

Let $v : U \subseteq \mathbb{R}^2 \to \mathbb{R}^2$ be a continuously differentiable planar vector field. Let Ω be a region such that $\Omega \cup \partial\Omega \subset U$ and r a parametrization of $\partial\Omega$. Then

$$\iint_\Omega \text{curl}(v)\, dA = \oint_{\partial\Omega} v \cdot \, dr$$

where r is positively oriented.

Theorem 5.5.2.4: Conservative Implies Zero Curl

If $v : U \subseteq \mathbb{R}^2 \to \mathbb{R}^2$ is a conservative vector field with continuous partial derivatives on U, then

$$\text{curl}(v) = 0$$

5.5.3 Divergence Theorem for a Planar Region

Definition 5.5.3.1: Flux across a Curve

Let $v : \mathbb{R}^2 \to \mathbb{R}^2$ be a planar vector field. Let $\Omega \subset \mathbb{R}^2$ be a region with $\partial\Omega$ parametrized by r. Define the flux of v across Ω to be

$$\text{flux of } v = \int_{\partial\Omega} v \cdot n \, dr$$

wwhere r is positively oriented such that n is the ouwards unit normal.

Theorem 5.5.3.2: Divergence Theorem for a Planar Region

Let $v : \mathbb{R}^2 \to \mathbb{R}^2$ be a continuously differentiable planar vector field. Let Ω be a region such that $\Omega \cup \partial\Omega \subset U$ and r a parametrization of $\partial\Omega$. Then

$$\iint_\Omega \nabla \cdot v \, dA = \int_{\partial\Omega} v \cdot n \, dr$$

where r is positively oriented such that n is the ouwards unit normal.

5.6 3 Dimensional Calculus

5.6.1 Triple Integrals

Definition 5.6.1.1: Triple Integral

Let $f : \mathbb{R}^3 \to \mathbb{R}$ be a function. Define the triple integral of f in a region Ω to be

$$\iiint_\Omega f(x, y, z)\, dx\, dy\, dz$$

Theorem 5.6.1.2

Let $R \subset \mathbb{R}^3$ be a region. Let $f : \mathbb{R}^3 \to \mathbb{R}^3$ defined by $f(x, y, z) = \begin{pmatrix} u(x, y, z) \\ v(x, y, z) \\ w(x, y, z) \end{pmatrix}$ such that $f|_R$ is a bijection and $\partial f \in \mathcal{C}^1(R)$. Let $g : \mathbb{R}^3 \to \mathbb{R}$. Then

$$\iiint_S g(u, v, w)\, du\, dv\, dw = \iiint_R g(x, y, z)|\partial f|\, dx\, dy\, dz$$

Theorem 5.6.1.3: Cylindrical Coordinates

Let $f : \mathbb{R}^3 \to \mathbb{R}$. Let $E \subset \mathbb{R}^3$ be a region. Then

$$\iiint_R f(x, y, z)\, dV = \iiint_R f(r\cos\theta, r\sin\theta, z)r\, drd\theta dz$$

with $x = r\cos(\theta)$ and $y = r\sin(\theta)$ and $z = z$

Theorem 5.6.1.4: Spherical Coordinates

Let $f : \mathbb{R}^3 \to \mathbb{R}$. Let $E \subset \mathbb{R}^3$ be a region. Then

$$\iiint_R f(x, y, z)\, dV = \iiint_R f(r\cos\theta\sin\phi, r\sin\theta\sin\phi, r\cos\phi)r^2\sin(\phi)\, drd\theta d\phi$$

with $x = r\sin(\phi)\cos(\theta)$ and $y = r\sin(\phi)\sin(\theta)$ and $z = r\cos(\phi)$

Theorem 5.6.1.5: Mass

Consider a solid occupying a region $\Omega \subset \mathbb{R}^3$ with density function $\rho(x, y, z)$ Then its total mass is

$$M = \iiint_\Omega \rho(x, y, z)\, dV$$

Theorem 5.6.1.6: Center of Mass

Consider a solid occupying a region $\Omega \subset \mathbb{R}^3$ with density function $\rho(x, y, z)$ and total mass M. Then

$$(\bar{x}, \bar{y}, \bar{z}) = \left(\frac{1}{M}\iiint_\Omega \rho x\, dV, \frac{1}{M}\iiint_\Omega \rho y\, dV, \frac{1}{M}\iiint_\Omega \rho z\, dV\right)$$

5.6.2 Surface Integrals

Theorem 5.6.2.1

Let S be a surface parametrized by $\mathbf{r}(u,v)$. The surface area is given by

$$A = \iint_{\Omega} |\mathbf{r}_u \times \mathbf{r}_v| \, dudv$$

where Ω is the parameter domain.

Lemma 5.6.2.2

The formula of the area of a surface given by $z = f(x,y)$ where $(x,y) \in R \subset \mathbb{R}^2$ is given by

$$A = \iint_R \sqrt{1 + \left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2} \, dxdy$$

5.6.3 Divergence Theorem

Definition 5.6.3.1: Curl

Let $f : U \subset \mathbb{R}^3 \to \mathbb{R}^3$ be a vector field. Then the curl of f is defined as

$$\nabla \times f = \begin{vmatrix} i & j & k \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ f_1 & f_2 & f_3 \end{vmatrix}$$

Proposition 5.6.3.2

If $v : U \subseteq \mathbb{R}^3 \to \mathbb{R}^3$ is a conservative vector field with continuous partial derivatives on U, then

$$\text{curl}(v) = 0$$

Definition 5.6.3.3: Flux of a Surface

Let $v : \mathbb{R}^3 \to \mathbb{R}^3$ be a vector field. Let $\Omega \subset \mathbb{R}^3$ be a volume. Define the flux of v across Ω to be

$$\text{flux of } v = \iint_{\partial\Omega} v \cdot n \, dA$$

where n is the outward pointing normal.

Lemma 5.6.3.4: Practical Definition of Flux

Let $\mathbf{v} : \mathbb{R}^3 \to \mathbb{R}^3$. Let $\Omega \subset \mathbb{R}^3$. Let $r : \mathbb{R}^2 \to \mathbb{R}^3$ parameterize the surface $\partial\Omega$. Then

$$\iint_{\partial\Omega} v \cdot n \, dA = \iint_{\partial\Omega} \mathbf{v}(r(u,v)) \cdot \left(\frac{\partial r}{\partial u} \times \frac{\partial r}{\partial v}\right) dudv$$

Theorem 5.6.3.5: The Divergence Theorem

Let $v : U \subset \mathbb{R}^3 \to \mathbb{R}^3$ be a $\mathcal{C}^1$ vector field. Let Ω be a region in $\mathbb{R}^3$ with $\partial\Omega$ parametrized by r such that $\Omega \cup \partial\Omega \subset U$. Then

$$\iiint_{\Omega} \nabla \cdot v \, dV = \iint_{\partial\Omega} v \cdot n \, dA$$

where r is positively oriented such that n is the ouwards unit normal.

5.6.4 Stoke's Theorem

Definition 5.6.4.1: Pointwise Circulation

Let $\mathbf{F}(x, y, z)$ give the velocity of a fluid. The pointwise circulation of $\mathbf{F}$ is given by

$$\nabla \times \mathbf{F} \cdot \hat{n}$$

Theorem 5.6.4.2: Net Circulation

The net circulation over a surface S is given by

$$\iint_S \nabla \times \mathbf{F} \cdot \hat{n} \, dS$$

Theorem 5.6.4.3: Stoke's Theorem

Let $\mathbf{F}(x, y, z)$ be a vector field. Let S be a surface with unit normal $\hat{n}$ and boundary curve C, oriented according to the right hand rule. Then

$$\iint_S \nabla \times \mathbf{F} \cdot \hat{n} \, dS = \int_C \mathbf{F} \cdot \, d\mathbf{r}$$

5.7 Integration

5.7.1 Sigma Fields

Definition 5.7.1.1: Sigma Fields

A σ-field on a non-empty set S is a collection $\mathcal{F}$ of subsets of S such that

- $S, \emptyset \in \mathcal{F}$
- $A \in \mathcal{F}$ implies $A^c \in \mathcal{F}$
- $A_k \in \mathcal{F}$, $k \in \mathbb{N}$ implies $\bigcup_{k=1}^{\infty} A_k \in \mathcal{F}$

Definition 5.7.1.2

[Measure] A measure on a σ-field $\mathcal{F}$ of subsets of S is a function $\mu : \mathcal{F} \to [0, +\infty]$ such that

- $\mu(\emptyset) = 0$
- $\mu\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \mu\left(A_k\right)$ where $A_k \in \mathcal{F}$ are pairwise disjoint.

Proposition 5.7.1.3

Let μ be a measure on a σ-field $\mathcal{F}$ and $A_1, A_2, \cdots \in \mathcal{F}$

- If $A_1 \subseteq A_2$, then $\mu(A_1) \leq \mu(A_2)$
- $\mu(\bigcup_{k=1}^{\infty} A_k) \leq \sum_{k=1}^{\infty} \mu(A_k)$
- $\mu(A_1) + \mu(A_2) = \mu(A_1 \cup A_2) + \mu(A_1 \cap A_2)$

5.7.2 Lesbegue Outer Measure

Definition 5.7.2.1

[Intervals] An interval I is a subset of $\mathbb{R}$ of the form

- $[a, b] = \{x \in \mathbb{R} | a \leq x \leq b\}$
- $(a, b] = \{x \in \mathbb{R} | a < x \leq b\}$
- $[a, b) = \{x \in \mathbb{R} | a \leq x < b\}$
- $(a, b) = \{x \in \mathbb{R} | a < x < b\}$

Define the measure of an interval to be its length, $|I| = b - a$

Definition 5.7.2.2

[Boxes] A box in $\mathbb{R}^n$ is a cartesian product

$$B = I_1 \times I_2 \times \cdots \times I_n$$

of n intervals. Define the measure of a box to be its "volume",

$$|B| = |I_1| \cdots |I_n|$$

Definition 5.7.2.3

[Lebesgue Measure] Let $E \subset \mathbb{R}^n$ be a bounded set. Define the Lebesgue Outer Measure as

$$m^*(E) = \inf_{E \subset \bigcup_{n=1}^{\infty} A_n, A_n \in \mathcal{M}_E} \sum_{n=1}^{\infty} |A_n|$$

Proposition 5.7.2.4

Let A, B be a subset of $\mathbb{R}^n$

- $0 \leq m^*(A) \leq +\infty$
- $m^*(\emptyset) = 0$
- $m^*(I) = |I|$ if I is a box
- $A \subseteq B \implies m^*(A) \leq m^*(B)$
- $m^*\left(\bigcup_{k=1}^{\infty} A_k\right) \leq \sum_{k=1}^{\infty} m^*(A_k)$

Lemma 5.7.2.5

For every $E, C \subseteq \mathbb{R}^n$,

$$m^*(C) \leq m^*(C \cap E) + m^*(C \cap E^c)$$

Definition 5.7.2.6

$E \subseteq \mathbb{R}^n$ is said to be Lesbegue Measurable if

$$m^*(C) \geq m^*(C \cap E) + m^*(C \cap E^c)$$

Denote the set of all Lesbegue measurable sets $\mathcal{M}$. In case where E is Lesbegue measurable we denote its measure with $m(E)$.

Proposition 5.7.2.7

$\mathcal{M}$ is a sigma field and m is a measure on M.

5.7.3 Measurable Functions

Definition 5.7.3.1

[Measurable Functions] A function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be Lesbegue measurable if for every $t \in \mathbb{R}$,

$$\{x | f(x) > t\} \in \mathcal{M}$$

Proposition 5.7.3.2

Let f, g be Lesbegue measurable functions. Let $\alpha \in \mathbb{R}$.

- $f + g$ is measurable
- αf is measurable
- f^2 is measurable
- fg is measurable

Proposition 5.7.3.3

Let f_k be a sequence of functions that are Lesbegue measurable. Then

- $\sup_k f_k$ is measurable
- $\int_k f_k$ is measurable
- $\limsup_k f_k$ is measurable
- If $f_k \to f$, then f is measurable

Definition 5.7.3.4

[Indicator Function] The indicator function of a set $A \subseteq \mathbb{R}^n$ is the function 1_A defined on $\mathbb{R}^n$ such that

$$1_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$$

Definition 5.7.3.5

A function $f : \mathbb{R}^n \to \mathbb{R}$ with finite range is called a simple function. The collection of all non-negative Lesbegue measurable simple functions is denoted by $\mathcal{S}_+(\mathcal{M})$

5.7.4 Lesbegue Integration

Definition 5.7.4.1

Let $f \in \mathcal{S}_+(\mathcal{M})$ such that

$$f = \sum_{k=1}^{r} a_k 1_{A_k}$$

where $A_k = \{x : f(x) = a_k\}$ and $\{A_1, \dots, A_r\}$ is a measurable partition of $\mathbb{R}^n$. Define the lesbegue integral of f to be

$$\int f \, dm = \sum_{k=1}^{r} a_k m(A_k)$$

Definition 5.7.4.2

Let $f : \mathbb{R}^n \to \mathbb{R}$ be Lesbegue measurable. If $f \geq 0$, define

$$\int f \, dm = \sup \left\{ \int f_s \, dm \middle| f_s \leq f, f_s \in \mathcal{S}_+(\mathcal{M}) \right\}$$

Definition 5.7.4.3

In general define the Lesbegue integral on $\mathbb{R}^n$ by

$$\int f \, dm = \int f^+ \, dm - \int f^- \, dm$$

provided that at least one of the terms on the right is finite. For $E \in \mathcal{M}$ define the Lesbegue integral on E by

$$\int_E f \, dm = \int f \cdot 1_E \, dm$$

5.8 Integration of Differential Forms

5.8.1 Basic Definitions

Definition 5.8.1.1

[Multilinear Function] Let V be a vector space over $\mathbb{R}$. A function $f : V^k \to \mathbb{R}$ is k-linear if it is linear in each of its k arguments

$$f(v_1, \dots, av_i + bw_i, \dots, v_k) = af(v_1, \dots, v_i, \dots, v_k) + bf(v_1, \dots, w_i, \dots, v_k)$$

for $i \in \{1, \dots, k\}$ nd $a, b \in \mathbb{R}$. It is also called a k-tensor on V. Denote the set of all k-tensors on V by $L_k(V)$

Definition 5.8.1.2

[Symmetric] Let V be a vector space over $\mathbb{R}$. $f : V^k \to \mathbb{R}$ is symmetric if

$$f(v_{\sigma(1)}, \dots, v_{\sigma(k)}) = f(v_1, \dots, v_k)$$

for all $\sigma \in S_k$

Definition 5.8.1.3

[Alternating] Let V be a vector space over $\mathbb{R}$. $f : V^k \to \mathbb{R}$ is alternating if

$$f(v_{\sigma(1)}, \dots, v_{\sigma(k)}) = \text{sign}(\sigma) f(v_1, \dots, v_k)$$

for all $\sigma \in S_k$. Alternating k-tensors are also called k-covectors. Denote the set of all k-covectors $\Lambda_k(V)$

Definition 5.8.1.4

Let $f : V^k \to \mathbb{R}$ be a k-linear function. Define

$$(Sf)(v_1, \dots, v_k) = \sum_{\sigma \in S_k} \sigma(f)$$

Define

$$(Af)(v_1, \dots, v_k) = \sum_{\sigma \in S_k} \text{sign}(\sigma)\sigma(f)$$

Proposition 5.8.1.5

Let $f : V^k \to \mathbb{R}$ be a k-linear function. Then Sf is symmetric and Af is alternating.

Lemma 5.8.1.6

If f is an alternating k-linear function on a vector space V, then $Af = (k!)f$.

5.8.2 Tensor Product and Wedge Product

Definition 5.8.2.1

[Tensor Product] Let f be k-linear on V and g be l linear on V. Their tensor product is defined to be the $k+l$ linear function

$$(f \otimes g)(v_1, \dots, v_{k+l}) = f(v_1, \dots, v_k)g(v_{k+1}, \dots, v_{k+l})$$

Proposition 5.8.2.2

Let f, g, h be multilinear functions on V. Then

$$f \otimes (g \otimes h) = (f \otimes g) \otimes h$$

Definition 5.8.2.3

[Wedge Product] Let f be k-linear on V and g be l linear on V. Their wedge product is defined to be the $k+l$ linear function

$$f \wedge g = \frac{1}{k!l!}A(f \otimes g)$$

Proposition 5.8.2.4

Let $f \in \Lambda_k(V)$ and $g \in \Lambda_l(V)$. Then

$$f \wedge g = (-1)^{k+l} g \wedge f$$

Corollary 5.8.2.5

Let $f \in \Lambda_k(V)$ and k is odd. Then $f \wedge f = 0$

Proposition 5.8.2.6

Let f, g, h be multilinear functions on V. Then

$$f \wedge (g \wedge h) = (f \wedge g) \wedge h$$

Proposition 5.8.2.7

Let $f_k \in \Lambda_{d_k}(V)$ for $k \in \{1, \dots, n\}$. Then

$$f_1 \wedge \cdots \wedge f_n = \frac{1}{(d_1)! \cdots (d_n)!}A(f_1 \otimes \cdots \otimes f_n)$$

Definition 5.8.2.8

[Multi-index Notation] Suppose that V is a vector space and $\alpha^1, \dots, \alpha^n$ the dual basis of V. Define $I = (i_1, \dots, i_k)$ and write α^I for $\alpha^{i_1} \wedge \cdots \wedge \alpha^{i_k}$. We usually want $i_1 < \cdots < i_k$.

Lemma 5.8.2.9

Let $e_1, \dots, e_n$ be a basis for V and $\alpha^1, \dots, \alpha^n$ be the dual basis of V. Then

$$\alpha^I(e_J) = \delta^I_J \begin{cases} 1 & \text{if } I = J \\ 0 & \text{if } I \neq J \end{cases}$$

Proposition 5.8.2.10

The set of all α^I where $I = (i_1 < \cdots < i_k)$ form a basis for the space $\Lambda_k(V)$. The dimension of $\Lambda_k(V)$ is $\binom{n}{k}$

Corollary 5.8.2.11

If $k > \dim(V)$, then $\Lambda_k(V) = 0$

5.8.3 Tangent Space

Definition 5.8.3.1

[Tangent Space] The set of all vectors with tail at $p \in \mathbb{R}^n$ is denoted $T_p(\mathbb{R}^n)$. We write a point in $\mathbb{R}^n$ as $p = (p_1, \dots, p_n)$ and a vector v in $T_p(\mathbb{R}^n)$ as $\langle v_1, \dots, v_n \rangle$

Definition 5.8.3.2

[Line Through a Point] The line through a point $p \in \mathbb{R}^n$ with direction v has parametrization

$$c(t) = (p_1 + tv_1, \dots, p_n + tv_n)$$

with its i-component $c_i(t) = p_i + tv_i$

Definition 5.8.3.3

[Directional Derivative] Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ be $\mathcal{C}^\infty$. Let $v \in T_p(\mathbb{R}^n)$. The directional derivative of f in the direction v at p is defined to be

$$D_v(f) = \lim_{t \to 0} \frac{f(c(t)) - f(p)}{t} = \frac{d}{dt}\bigg|_{t=0} f(c(t))$$

Proposition 5.8.3.4

Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ be $\mathcal{C}^\infty$. Then

$$D_v(f) = \sum_{k=1}^n v_k \frac{\partial f}{\partial x_k}\bigg|_p$$

and D_v is a map from $\mathcal{C}_p^\infty(\mathbb{R}^n) \to \mathbb{R}$

Proposition 5.8.3.5

The map $\phi : T_p(\mathbb{R}^n) \to \mathcal{D}_p(\mathbb{R}^n)$ given by $\phi(v) = D_v$ is an isomorphism of vector spaces.

Proposition 5.8.3.6

The standard basis of $T_p(\mathbb{R}^n)$ corresponds to

$$\left\{ \frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_n} \right\}$$

Definition 5.8.3.7

[Vector Fields] A vector field X on an open subset U of $\mathbb{R}^n$ is a function that assigns to each point p in U a tangent vector denoted $X_p \in T_p(R^n)$. This means that $X : \mathbb{R}^n \to T_p(\mathbb{R}^n)$

Proposition 5.8.3.8

For every vector field X,

$$X_p = \sum_{k=1}^{n} a_k(p) \frac{\partial}{\partial x_k}\bigg|_p$$

where $a_k(p) \in \mathbb{R}$

5.8.4 Differential 1-forms

Definition 5.8.4.1

[Cotangent Space] Define the cotangent space to $\mathbb{R}^n$ at p to be $T_p^*(\mathbb{R}^n)$, the dual space of $T_p(\mathbb{R}^n)$.

Definition 5.8.4.2

[Differential 1-form] A differential 1-form is a function $\omega : U \subseteq \mathbb{R}^n \to \bigcup_{p\in U} T_p^*(\mathbb{R}^n)$ from $p \in \mathbb{R}^n$ to $\omega_p \in T_p^*(\mathbb{R}^n)$

Proposition 5.8.4.3

Fix $f \in \mathcal{C}^\infty(\mathbb{R}^n)$. Define $df_p : T_p(\mathbb{R}^n) \to \mathbb{R}$ by

$$(df)_p(X_p) = X_p(f)$$

Then the mapping $(df)(p) = (df)_p$ from p to $(df)_p$ is a differential 1-form.

Proposition 5.8.4.4

Suppose that $x_1, \dots, x_n$ are the standard coordinate for $\mathbb{R}^n$. Then for each point $p \in \mathbb{R}^n$,

$$\{(dx_1)_p, \dots, (dx_n)_p\}$$

is the basis for $T_p^*(\mathbb{R}^n)$ dual to

$$\left\{\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_n}\right\}$$

in $T_p(\mathbb{R}^n)$

Proposition 5.8.4.5

If $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ is $\mathcal{C}^\infty$, then

$$df = \sum_{k=1}^{n} \frac{\partial f}{\partial x_k} dx_k$$

5.8.5 Differential k-forms

Definition 5.8.5.1

[Differential k-forms] A differential k-form ω on $U \subseteq \mathbb{R}^n$ is a function that assigns to each point $p \in U$ an alternating k-linear function. This means $\omega : \mathbb{R}^n \to \Lambda_k(T_p(\mathbb{R}^n))$ Denote $\Omega^k(U)$ the vector space of $\mathcal{C}^\infty$ k-forms on U.

Proposition 5.8.5.2

A basis of $\Lambda_k(T_p(\mathbb{R}^n))$ is given by $dx_p^I = dx_p^{i_1} \wedge \cdots \wedge dx_p^{i_k}$ for $1 \leq i_1 < \cdots < i_k \leq n$.

Proposition 5.8.5.3

A differential k-form ω is of the form

$$\omega = \sum_I \alpha_I dx^I$$

with $a_I : U \subseteq \mathbb{R}^n \to \mathbb{R}$

5.8.6 Exterior Derivative

Definition 5.8.6.1

[Exterior Derivative of 0-forms] Let $f \in \mathcal{C}^\infty(U)$. Then f is a 0-form. Define its exterior derivative to be its differential $df \in \Omega^1(U)$.

Definition 5.8.6.2

[Exterior Derivative of k-forms] Let $\omega = \sum_I \alpha_I dx^I \in \Omega^k(U)$. Define

$$d\omega = \sum_I d\alpha_I \wedge dx^I = \sum_I \left(\sum_j \frac{\partial \alpha_I}{\partial x_j} dx_j \right) \wedge dx^I \in \Omega^{k+1}(U)$$

Proposition 5.8.6.3

Let $\omega \in \Omega^k(\mathbb{R}^n)$. Then $d^2\omega = 0$

Definition 5.8.6.4

[Closed Forms] A k-form ω on U is closed if $d\omega = 0$

Definition 5.8.6.5

[Exact Forms] A k-form ω on U is exact if there exists a $k-1$ form τ such that $\omega = d\tau$.

5.8.7 Integration of Differential Forms

Definition 5.8.7.1

Two basis of $\mathbb{R}^n$ are said to have the same orientation if their determinants have the same sign.

Definition 5.8.7.2

Let $\omega = f(x)dx_1 \wedge \cdots \wedge dx_n$ be a smooth n-form on an open subset $U \subset \mathbb{R}^n$. Its integral over a subset $A \subset U$ is defined to be the Riemann integral

$$\int_A \omega = \int_A f(x)\, dx_1 \cdots dx_n$$

Part III

The Fundamentals of Geometry and Topology

Chapter 6

Linear Algebra 1

6.1 Vector Spaces

6.1.1 Introduction to Vector Spaces

Although the complete development of fields is given in an abstract algebra course, we give the definition of a field here for completeness.

Definition 6.1.1.1: Fields

A field $(\mathbb{F}, +, \cdot)$ is a triple where $+ : \mathbb{F} \times \mathbb{F} \to \mathbb{F}$ and $\cdot : \mathbb{F} \times \mathbb{F} \to \mathbb{F}$ such that they satisfy the following rules if $a, b, c \in \mathbb{F}$:
$(\mathbb{F}, +)$ is an abelian group.

- $a + (b + c) = (a + b) + c$
- There exists $0 \in \mathbb{F}$ such that $a + 0 = 0 + a = a$
- There exists $-a \in \mathbb{F}$ such that $a + (-a) = (-a) + a = 0$
- $a + b = b + a$

$(\mathbb{F} \setminus \{0\}, \ \cdot\ )$ is an abelian group.

- $a \cdot (b \cdot c) = (a \cdot b) \cdot c$
- There exists $1 \in \mathbb{F}$ such that $a \cdot 1 = 1 \cdot a = a$
- There exists $a^{-1} \in \mathbb{F}$ such that $a \cdot a^{-1} = a^{-1} \cdot a = 1$ if $a \neq 0$
- $a \cdot b = b \cdot a$

Distributive law.

- $a \cdot (b + c) = (a \cdot b) + (a \cdot c)$

Definition 6.1.1.2: Vector Space

A vector space V over a field $\mathbb{F}$ is a set of elements V together with $0 \in V$ and two binary opeartions $+ : V \times V \to V$ and $\cdot : F \times V \to V$, vector addition and scalar multiplcation respectively, satisfying the following with $a, b \in \mathbb{F}$ and $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{V}$.
$(V, +)$ is an abelian group.

- $\mathbf{u} + \mathbf{v} \in V$
- $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$
- There exists a vector $\mathbf{0}_V$ such that $\mathbf{0}_V + \mathbf{u} = \mathbf{u}$

- There exists an additive inverse $-\mathbf{u}$ such that $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$
- $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$

$\mathbb{F}$ acts on V as a group action, with an identity in V.

- $a \cdot \mathbf{u} \in V$
- $a(b\mathbf{u}) = (ab)\mathbf{u}$
- There exists a vector 1_V such that $1_v \cdot \mathbf{u} = \mathbf{u}$

Distributive laws.

- $a(\mathbf{u} + \mathbf{v}) = a\mathbf{u} + a\mathbf{v}$
- $(a + b)\mathbf{u} = a\mathbf{u} + b\mathbf{u}$

Proposition 6.1.1.3

Let $a \in \mathbb{F}$ and $u \in V$ be a vector space over $\mathbb{F}$.

- $a \cdot \mathbf{0}_V = \mathbf{0}_V$
- $0 \cdot \mathbf{u} = \mathbf{0}_V$
- $(-a)\mathbf{v} = -(a\mathbf{v}) = a(-\mathbf{v})$
- $a\mathbf{v} = \mathbf{0}_V \implies a = 0$ or $\mathbf{v} = \mathbf{0}_V$

Proof.

- $a(\mathbf{0}_V) = a(\mathbf{0}_V + \mathbf{0}_V) = a\mathbf{0}_V + a\mathbf{0}_V$. Adding the additive inverse of $a\mathbf{0}_V$ on both sides gives our result.
- $0\mathbf{u} = (0 + 0)\mathbf{u} = 0\mathbf{u} + 0\mathbf{u}$. Adding the additive inverse of $0\mathbf{u}$ on both sides gives our result.
- Naturally $-(a\mathbf{v})$ is the inverse of $a\mathbf{v}$. Consider $a\mathbf{v} + a(-\mathbf{v})$.
  $a\mathbf{v} + a(-\mathbf{v}) = a(\mathbf{v} - \mathbf{v}) = a\mathbf{0}_V = 0_V$. Thus $a(-\mathbf{v})$ is also the inverse of $a\mathbf{v}$ and $-(a\mathbf{v}) = a(-\mathbf{v})$. The same could be done to the third item with the other distributive law.
- Suppose that $a \neq 0$. Then $\mathbf{v} = (a^{-1}a)\mathbf{v} = a^{-1}(a\mathbf{v}) = 0$.

□

Proposition 6.1.1.4

The additive identity, multiplicative identity, additive inverse of a vector space is unique.

Proof. Suppose that $\mathbf{e}$ and $\mathbf{f}$ are additive identities. Then $\mathbf{e} + \mathbf{f} = \mathbf{e}$ and $\mathbf{e} + \mathbf{f} = \mathbf{f}$. Thus $\mathbf{e} = \mathbf{f}$. Suppose that $\mathbf{e}$ and $\mathbf{f}$ are multiplicative identities. Then $\mathbf{ef} = \mathbf{e}$ and $\mathbf{ef} = \mathbf{f}$ and $\mathbf{e} = \mathbf{f}$. Let $a \in V$. Suppose that $b, c \in V$ are additive inverses of a. Then

$$\begin{aligned} a + b = a + c &\implies b + a + b = b + a + c \\ &\implies (b + a) + b = (b + a) + c \\ &\implies b = c \end{aligned}$$

□

6.1.2 Basis and Dimension

Definition 6.1.2.1: Linearly Independent

We say that a set of vectors $\{v_1, \dots, v_n\}$ of a vector space V over $\mathbb{F}$ are linearly independent if

$$\sum_{k=1}^{n} a_k v_k = 0$$

for $a_1, \dots, a_n \in \mathbb{F}$ implies $a_1 = \cdots = a_n = 0$.

Definition 6.1.2.2: Span

We say that a set of vectors $\{v_1, \dots, v_n\}$ of a vector space V over $\mathbb{F}$ spans V if for all $v \in V$, there exists $a_1, \dots, a_n \in \mathbb{F}$ such that

$$v = \sum_{k=1}^{n} a_k v_k$$

Definition 6.1.2.3: Basis

We say that a set of vectors $\{v_1, \dots, v_n\}$ of a vector space forms a basis for V if they are linearly independent and spans V.

Definition 6.1.2.4: Dimension

We say that the dimension of a vector space V is the number of elements in a basis of V. If a basis has n elements, then we say that $\dim(V) = n$.

If n is a finite number, then we say that V is finite dimensional.

We have yet to shown that the dimension of a vector space is well defined since we do not know whether the cardinality of any two bases are the same. Therefore we have the following important theorem for finite dimensional vector space.

Theorem 6.1.2.5: Steinitz Exchange Lemma

Let U, W be finite subsets of a finite dimensional vector space V. If U is a set of linearly independent vectors and W spans V, then

- $|U| \leq |W|$
- There exists a set $W' \subset W$ with $|W'| = |W| - |U|$ such that $U \cup W'$ spans V.

Proof. Take $U = \{u_1, \dots, u_m\}$ and $W = \{w_1, \dots, w_n\}$. We will show that after reordering elements of W, we will have a set $\{u_1, \dots, u_m, w_m + 1, \dots, w_n\}$ that it spans V. We proceed by induction on m. Suppose that $m = 0$. In this case, $|U| \leq |W|$ necesssarily holds and by construction, W already spans V.

Now suppose that the proposition is true for $m - 1$. By the induction hypothesis, we may reorder elements of W so that $\{u_1, \dots, u_{m-1}, w_m, \dots, w_n\}$ spans V. Since $u_m \in V$, there exists $a_1, \dots, a_n$ such that

$$u_m = \sum_{k=1}^{m-1} a_k u_k + \sum_{k=m}^{n} a_k w_k$$

At least one of $a_m, \dots, a_n$ must be nonzero else the equality will contradict the linear independence of $u_1, \dots, u_m$. This must mean that $m \leq n$.

Now by reordering $a_m w_m, \dots, a_n w_n$, we may assume that $a_m \neq 0$. Thus we have that

$$w_m = \frac{1}{a_m}\left(u_m - \sum_{k=1}^{m-1} a_k u_k - \sum_{k=m+1}^{n} a_k w_k\right)$$

This means that w_m lies in the span of $\{u_1, \dots, u_m, w_{m+1}, \dots, w_n\}$. Since this span contains each of the vectors $u_1, \dots, u_{m-1}, w_m, \dots, w_n$, by the inductive hypothesis it spans V. □

Clearly this implies that linearly independent sets of vectors must have cardinality less than sets of vectors that span V. By taking the highest cardinality of such linearly indendent set of vectors, and the lowest cardinality of such sets of vectors that span V, we necessarily have that they are equal and thus is exactly the dimension of V.

We will discuss about dimensions and infinite dimensional vector spaces more in functional analysis. For the rest of the notes we will mostly go with finite dimensional vector spaces.

We now give a criterion with matrices to find whether a set of vectors span V or whether they are linearly independent.

Theorem 6.1.2.6

Let V be a vector space of dimension n and $S = \{v_1, \dots, v_n\} \subset V$. Then

- Elements of S are linearly independent if and only if the row echelon form of $\begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix}$ has a leading one in every column
- Elements of S span V if and only if the row echelon form of $\begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix}$ has no zero rows
- S is a basis of V if and only if the row echelon form of $\begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix}$ is equal to the identity

6.1.3 Vector Subspaces

Definition 6.1.3.1: Vector Subspaces

A subset U of a vector space V is called a subspace of V if U is also a vector space.

Proposition 6.1.3.2: Subspace Criterion

U is a subspace of V if and only if U is closed under vector addition and scalar multiplication and contains the zero vector.

Proof. Suppose that U is a subspace of V. Then necessarily U is closed under vector addition and scalar multiplication and contains the zero vector.

Now suppose that the latter conditions are fulfilled by a subset U of V. Then it is easy to see that U satisfies all the criteria for being a vector space. □

Proposition 6.1.3.3

If U_1 and U_2 are subspaces of V then $U_1 \cap U_2$ is also a subspace.

Proof. Suppose that $\mathbf{v}, \mathbf{w} \in U_1 \cap U_2$. Then $\mathbf{v}, \mathbf{w} \in U_1$ and U_2. Since U_1, U_2 are subspaces, $\mathbf{v} + \mathbf{w} \in U_1$ and U_2 thus $\mathbf{v} + \mathbf{w} \in U_1 \cap U_2$. The proof is similar for scalar multiplication. □

Definition 6.1.3.4: Sum of Subspaces

Let U, W be subspaces of the vector space V. Then define

$$U + W = \{\mathbf{u} + \mathbf{w} : \mathbf{u} \in U \text{ and } \mathbf{w} \in W\}$$

Proposition 6.1.3.5

Let U, W be subspaces of a vector space V. Then $U + W$ is the smallest subspace of V containing U and W.

Proof. We first show that $U + W$ is indeed a subspace of V. Suppose that $v \in U + W$. Then there exists $u \in U$ and $w \in W$ such that $v = u + w$. Then since U and W are closed individually under vector addition and scalar multiplication, any product and addition in $U + W$ can be decomposed into a sum of vectors in U and W and thus the new vector will also be able to be decomposed into U and W and thus lie in $U + W$.

Now suppose that S is a subspace of V containing U and W. This means that any linear combination of elements of U and W are contained in S thus $U + W \subseteq S$. This means that if any subspace containing U and W must also contain $U + W$, which means that $U + W$ is the smallest subspace containing U and W. □

Definition 6.1.3.6: Independent Subspaces

Let $W_1, \dots, W_n$ be subspaces of a vector space V. We say that $W_1, \dots, W_n$ are independent if no vector of W_i is a linear combination of the remaining subspaces for every $i \in \{1, \dots, n\}$

Definition 6.1.3.7: Direct Sum

A vector space is the direct sum of its subspaces

$$V = W_1 \oplus \cdots \oplus W_n$$

if $W_1, \dots, W_n$ are independent and $V = W_1 + \cdots + W_n$.

Corollary 6.1.3.8

If $V = W_1 \oplus \cdots \oplus W_n$ then

$$\dim(V) = \sum_{k=1}^{n} \dim(W_k)$$

Proof. Each basis of W_k are not contained in any other linear combination of all the basis of $W_1, \dots, W_{k-1}, W_{k+1}, \dots, W_n$. This means that the set of all the basis of $W_1, \dots, W_n$ are linearly independent. Since they each span W_k independently, the set of all the basis of $W_1, \dots, W_n$ will span $W_1 \oplus \cdots \oplus W_n$ and thus is a basis of V. Thus we are done. □

6.1.4 Row and Column Ranks

The final section is devoted to matrices as we will soon see that matrices are particularly useful in a lot of things.

Definition 6.1.4.1: Row Space

Let $A_{m\times n}$ be a matrix. The row space of A is the subspace of $\mathbb{F}^m$,

$$\text{span}\{r_1, \dots, r_m\}$$

where r_i are the rows of A. The row rank of A is defined to be the dimension of the row space of A.

Definition 6.1.4.2: Column Space

Let $A_{m\times n}$ be a matrix. The column space of A is the subspace of $\mathbb{F}^n$,

$$\text{span}\{c_1, \dots, c_m\}$$

where c_i are the columns of A. The column rank of A is defined to be the dimension of the column space of A.

Lemma 6.1.4.3

Applying row operations does not change the row space, row rank of a matrix and column rank of a matrix.

Theorem 6.1.4.4

The row rank of a matrix is equal to the column rank.

We can now define the rank of a matrix without problem.

Definition 6.1.4.5: Rank of a Matrix

Define the rank of a matrix to be its row rank or column rank.

Proposition 6.1.4.6

Let A be a $n \times n$ matrix. Then the following are equivalent.

- The rank of A is n
- A is invertible
- The rows of A form a linearly independent set
- The columns of A form a linearly independent set

6.2 Linear Maps

6.2.1 Properties of Linear Maps

Definition 6.2.1.1: Linear Transformation

Let V, W be vector spaces over $\mathbb{F}$. A linear transformation or linear map T from V to W is a function $T : V \to W$ such that

- $T(v_1 + v_2) = T(v_1) + T(v_2)$ for all $v_1, v_2 \in V$
- $T(kv) = kT(v)$ for all $k \in \mathbb{F}$, $v \in V$

Lemma 6.2.1.2

Let $T : V \to W$ be a linear map.

- $T(0_v) = 0_w$
- $T(-v) = -T(v)$ for all $v \in V$

Proof. Suppose that $v \in V$. Then $T(0 \cdot v) = 0 \cdot T(v) = 0$. Also we have that $T(0 - v) = T(0) - T(v) = -T(v)$. □

Proposition 6.2.1.3

If $T : U \to V$ and $S : V \to W$ are linear then $S \circ T : U \to W$ is also linear.

Proof. Let $au + bv \in U$.

$$\begin{aligned} S \circ T(au + bv) &= S(aT(u) + bT(v)) \\ &= a(S \circ T(u)) + b(S \circ T(v)) \end{aligned}$$

□

6.2.2 Isomorphisms

Definition 6.2.2.1: Isomorphic Linear Maps

A linear map $T : V \to W$ is said to be an isomorphism if T is bijective. In this case we also say that V and W are isomorphic.

Theorem 6.2.2.2

Let $T : V \to W$ be an isomorphism of vector spaces V, W over F. Then its inverse map $T^{-1} : W \to V$ is a linear map.

Theorem 6.2.2.3

Let $T : V \to W$ be a linear map. Then the following are equivalent.

- T is isomorphic
- If $v_1, \dots, v_n \in V$ is a basis of V then $T(v_1), \dots, T(v_n) \in W$ is a basis of W

Corollary 6.2.2.4

Every finite dimensional vector space is isomorphic to $\mathbb{R}^n$ for some $n \in \mathbb{N} \setminus \{0\}$.

Proof. Direct consequence from the above. □

This corollary is especially important since it tells use that we only really need to study all of $\mathbb{R}^n$ to study all of finite dimensional spaces. Once we have our results on $\mathbb{R}^n$, we can translate it via an isomorphism.

Proposition 6.2.2.5

Let V, W be vetor spaces. The set of all linear maps from V to W forms a vector space. Denote it as $\mathcal{L}(V, W)$

6.2.3 Kernels and Images

Definition 6.2.3.1: Images and Kernels

Let $T : U \to V$ be a linear map. The image of T is defined as

$$\text{im}(T) = \{\mathbf{v} \in V | T(\mathbf{u}) = \mathbf{v}, \forall \mathbf{u} \in U\}$$

The kernel of T is defined as

$$\ker(T) = \{\mathbf{u} \in U | T(\mathbf{u}) = \mathbf{0}_V\}$$

Theorem 6.2.3.2

Let $T : U \to V$ be a linear map. Then

- $\text{im}(T)$ is a subspace of V
- $\ker(T)$ is a subspace of U

Proof. Let $T : U \to V$ be a linear map.

- We prove that $\text{im}(T)$ is a subspace of V. Let $\mathbf{u}, \mathbf{v} \in \text{im}(T)$ and $a \in \mathbb{F}$. Since $\mathbf{u}, \mathbf{v} \in \text{im}(T)$, there exists $\mathbf{u}_0, \mathbf{v}_0 \in U$ such that $T(\mathbf{u}_0) = \mathbf{u}$ and $T(\mathbf{v}_0) = \mathbf{v}$. Note that $a\mathbf{u}_0 \in U$ and $\mathbf{u}_0 + \mathbf{v}_0 \in U$. Consider $T(a\mathbf{u}_0)$. We have $T(a\mathbf{u}_0) = aT(\mathbf{u}_0) = a\mathbf{u}$. Thus $a\mathbf{u} \in \text{im}(T)$. Similarly, $T(\mathbf{u}_0 + \mathbf{v}_0) = T(\mathbf{u}_0) + T(\mathbf{v}_0) = \mathbf{u} + \mathbf{v}$. Thus $\mathbf{u} + \mathbf{v} \in \text{im}(T)$. By the subspace criterion $\text{im}(T)$ is a subspace of V.
- We now prove that $\ker(T)$ is a subspace of U. Suppose that $u, v \in \ker(T)$ and $a, b \in \mathbb{F}$. Then $T(au + bv) = aT(u) + bT(v) = 0$. Thus $au + bv \in \ker(T)$.

□

Theorem 6.2.3.3

Let A be the matrix representing a linear transformation. Let B be the row reduced form of A. Then a basis of the image of the linear transformation is given by the columns in A that has leading one in columns in B.

Definition 6.2.3.4: Rank and Nullity

Let $T : U \to V$ be a linear map.

- $\dim(\text{im}(T))$ is said to be the rank of T.
- $\dim(\ker(T))$ is said to be the nullity of T.

Theorem 6.2.3.5: Rank Nullity Theorem

Let $T : U \to V$ be a linear map. Then

$$\text{rank}(T) + \text{nullity}(T) = \dim(U)$$

Theorem 6.2.3.6

Let $T : U \to V$ be a linear map, where $\dim(U) = n$, $\dim(V) = m$. Let $e_1, \dots, e_n$ be a basis of U. Then the rank of T is equal to the largest size of a linearly independent subset of $T(e_1), \dots, T(e_n)$.

6.2.4 Role of Matrices

Definition 6.2.4.1: Matrix of a Linear Map

Let $T : U \to V$ be a linear map where $\dim(U) = n$ and $\dim(V) = m$. Let $\mathbf{e}_1, \dots, \mathbf{e}_n$ be the standard basis of U and $\{\mathbf{f}_1, \dots, \mathbf{f}_m\}$ the standard basis of V. Let

$$T(\mathbf{e}_i) = \sum_{k=1}^{m} \alpha_{ki}\mathbf{f}_k$$

for $i \in \{1, \dots, n\}$ Define the matrix of this linear map to be

$$\begin{pmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{m1} & \alpha_{m2} & \cdots & \alpha_{mn} \end{pmatrix}$$

Theorem 6.2.4.2

Let $T : U \to V$ be a linear map. Let A be the matrix of a linear map. Let $\mathbf{v} \in U$. Then the coordinates of $T(\mathbf{v})$ are given by

$$T(\mathbf{v}) = A\mathbf{v}$$

Theorem 6.2.4.3

The rank of a matrix equals the rank of any map that it represents.

Theorem 6.2.4.4

The composition of linear maps is represented by the matrix product of its representatives.

Theorem 6.2.4.5

T is isomorphic if and only if its matrix is nonsingular.

6.3 Eigenspaces

Eigenspaces are invariants of a linear map. Every vector in the eigenspace will only be scaled while maintaining its direction.

6.3.1 Eigenvalues and Eigenvectors

Definition 6.3.1.1: Eigenvalues and Eigenvectors

Let $T : V \to V$ be a linear map, where V is a vector space over $\mathbb{F}$. Suppose that for some non-zero vector $v \in V$, and some scalar $\lambda \in \mathbb{F}$, we have $T(v) = \lambda v$. Then v is called an eigenvector of T, and λ is called the eigenvalue of T.

Notice that λ can in fact be 0. If this is the case, then the eigenvectors are just the vectors in the kernel.

Definition 6.3.1.2: Eigenspace

Let λ be an eigenvalue of a linear map T. The set of all eigenvectors belonging to λ is called an eigenspace of T with respect to λ, denoted E_λ.

Lemma 6.3.1.3

Let λ be an eigenvalue of A. Then

$$E_\lambda = \ker(A - \lambda I)$$

Proof. Clearly since $Av = \lambda v$ for any eigenvector v of λ, we also have that $(A - \lambda I)v = 0$ which means that $v \in \ker(A - \lambda I)$. □

Proposition 6.3.1.4

Let $\lambda_1, \dots, \lambda_r$ be distinct eigenvalues of $T : V \to V$, and let $v_1, \dots, v_r$ be the corresponding eigenvectors. Then $v_1, \dots, v_r$ are linearly independent.

As we can see, distinct eigenvalues are linearly independent. Considering the span of each eigenvectors, we can clearly see that each of their spans are independent.

Proposition 6.3.1.5

If $\lambda_1, \dots, \lambda_n$ are distinct eigenvalues of a matrix A, then $E_{\lambda_1}, \dots, E_{\lambda_n}$ are independent.

Proof. Clear from the fact that the basis of eigenspaces of different eigenvalues are linearly independent. □

Definition 6.3.1.6: Characteristic Polynomial

Let A be an $n \times n$ matrix.

$$c_A(x) = \det(A - xI_n)$$

is called the characteristic polynomial of A.

Proposition 6.3.1.7

Let A be an $n \times n$ matrix. Then λ is an eigenvalue of A if and only if

$$c_A(\lambda) = 0$$

Definition 6.3.1.8: Invariant Subspaces

Let $T : V \to V$ be a linear transformation. Let U be a subspace of V. We say that U is T-invariant if

$$v \in U \implies T(v) \in U$$

for all $v \in U$ or equivalently, $T(U) \subseteq U$.

Theorem 6.3.1.9

Eigenspaces is an invariant subspace under its linear transformation.

The main result of this subsection, stated that eigenspaces remain invariant under the linear transformation. Clearly this depends on the linear transformation. We will also show that this fact is also unchanged when considering different basis for the linear transformation.

6.3.2 Change of Basis

Definition 6.3.2.1: Change of Basis Matrix

Let V be a vector space and B, B' are two basis of V. A change of basis linear map is a linear map $T : V \to V$ such that $T : V_B \to V_{B'}$, meaning the old basis is mapped to the new basis.

Proposition 6.3.2.2

Let V be a vector space and $v_1, \dots, v_n$, $v_1', \dots, v_n'$ two distinct basis of V. Then

$$v_k = p_{k1}v_1' + \cdots + p_{kn}v_n'$$

for all $k \in \{1, \dots, n\}$ and for any vector x in the basis $v_1, \dots, v_n$, the vector in the other basis x' is given by $x' = Px$ with the invertible matrix P

$$\begin{pmatrix} p_{11} & \cdots & p_{1n} \\ \vdots & \ddots & \vdots \\ p_{n1} & \cdots & p_{nn} \end{pmatrix}$$

Theorem 6.3.2.3

Let V, W be vector spaces. Let V consists of two different basis B and B' with a map $P : V_B \to V_{B'}$. Similarly for W we have C and C' and $Q : W_C \to W_{C'}$. Suppose $A : V_B \to W_C$ is a linear map. Then $A' : V_{B'} \to W_{C'}$ is given by

$$A' = QAP^{-1}$$

Definition 6.3.2.4: Similar Matrices

We say that two matrices $A, B \in M_{n\times n}(\mathbb{R})$ are similar if there exists an invertible matrix $P \in M_{n\times n}(\mathbb{R})$ such that $B = PAP^{-1}$.

Lemma 6.3.2.5

The relation of similarity in matrices is an equivalent relation in $M_{n\times n}(\mathbb{R})$.

Similar matrices will play an important role. We will soon see that every matrix will be similar to relatively nice matrix so that their properties can be investigated, as well as making computations significantly easier.

6.3.3 Diagonalization

We now show a very nice kind of matrices, diagonal matrices that will come into play with linear maps. Our goal is to attempt to classify, by similarity, of every matrix into a diagonal one. We will soon see that this is not possible, and thus giving the last section of these notes meaning.

Definition 6.3.3.1: Diagonalizable Linear Maps

An linear map T is diagonalizable if there exists a basis the matrix representation of T is linear.

Proposition 6.3.3.2: Diagonalizable Matrices

A linear map T represented by A is diagonalizable if there exists an invertible P and a diagonal matrix D such that $P^{-1}AP = D$. In that case, P consists of eigenvectors of T and the diagonals of D are the eigenvalues of A.

Theorem 6.3.3.3

If the linear map $T : V \to V$ has n distinct eigenvalues where $\dim(V) = n$, then T is diagonalizable.

Although not stated in the theorem, this does not mean that linear maps without n distinct eigenvalues are not diagonalizable. However by taking the contrapositive, we see that not every linear map is diagonalizable because clearly, not every linear map has n distinct eigenvalues.

Theorem 6.3.3.4

Let $T \in \mathcal{L}(V)$. Let $\lambda_1, \dots, \lambda_m$ be the distinct eigenvalues of T. Then following are equivalent.

- T is diagonalizable
- V has a basis consisting of eigenvalues of T
- $V = E_{\lambda_1} + \cdots + E_{\lambda_m}$
- $\dim(V) = \dim(E_{\lambda_1}) + \cdots + \dim(E_{\lambda_m})$

6.4 The Jordan Canonical Form

In the last section, we looked into what kinds of matrices can have ”nice” looking matrix under some basis. We now provide a less ”nice” looking form of a similar matrix. However, every matrix can be reduced to this relatively ”nice” looking form, as long as the field is algebraically closed. This form is called the Jordan Normal Form.

6.4.1 The Minimal Polynomial

Theorem 6.4.1.1

Let $\mathbb{F}$ be a field. Let A be a $n \times n$ matrix over $\mathbb{F}$. Then there is some non-zero polynomial $p \in \mathbb{F}[x]$ of degree at most n^2 such that $p(A) = \mathbf{0}_n$.

Proof. Note that $\{I, A, \dots, A^{n^2}\}$ is linerarly dependent in the vector space of $n \times n$ matrices. Thus there exists constant $c_0, \dots, c_{n^2}$ that are not all zero such that

$$c_0 I + \cdots + c_{n^2} A^{n^2} = \mathbf{0}_n$$

Thus $p(x) = c_0 + c_1 x + \cdots + c_{n^2} x^{n^2}$ is our desired polynomial. □

Theorem 6.4.1.2

Let $A_{n\times n}$ be a matrix over $\mathbb{F}$ representing the linear map $T : V \to V$. Then

- There is a unique monic non-zero polynomial $p(x)$ with minimal degree and coefficients in $\mathbb{F}$ such that $p(A) = \mathbf{0}_n$
- If $q(x)$ is any polynomial with $q(A) = \mathbf{0}_n$, then $p|q$

Proof. By the previous theorem, there exists a polynomial such that $p(A) = 0$. Divide the polynomial by c_{n^2} gives us the desired monic polynomial. Suppose that p_1, p_2 are distinct monic polynomials that are minimal such that $p_1(A) = 0$ and $p_2(A) = 0$, then $p = p_1 - p_2$ is a non zero polynomial with a smaller degree and $p(A) = 0$, contradicting the minimality of degree. Thus p is unique.

Let $p(x)$ be the minimal polynomial in the above proof. Let $q(A) = 0$. By division algorithm there exists some r with smaller degree than p such that $q = sp + r$. If r is non-zero, then $r(A) = q(A) - s(A)p(A) = 0$, contradiction of minimality, thus $r = 0$ and $p|q$. □

Definition 6.4.1.3: The Minimal Polynomial

The unique monic non-zero polynomial $\mu_A(x)$ of minimal degree with $\mu_A(A) = \mathbf{0}_n$ is called the minimal polynomial of A.

Proposition 6.4.1.4

Similar matrices have the same minimal polynomial.

Proof. Similar matrices represent the same linear map, thus both have their minimal polynomial same as T, the linear map. □

Proposition 6.4.1.5

Let D be a diagonal matrix with $\{d_1, \dots, d_r\}$ its unqiue diagonal entries, then

$$\mu_D(x) = (x - d_1) \cdots (x - d_r)$$

Proof. For any diagonal matrix,

$$p(D) = \begin{pmatrix} p(d_{11}) & 0 & \cdots & 0 \\ 0 & p(d_{22}) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & p(d_{nn}) \end{pmatrix}$$

Thus $p(D) = 0$ if and only if $p(d_{kk}) = 0$ for $k \in \{1, \dots, n\}$. Thus the smallest-degree monic polynomial vanishing at these points is clearly the polynomial above. □

Corollary 6.4.1.6

Every diagonalizable matrix has its minimal polynomial a product of distinct linear factors.

Proof. Since diagonalizable matrix is similar to some diagonal matrix and they both have the same minimal polynomial, by the above proposition it is a product of distinct linear factors. □

We will later see that in fact, the above criterion is a neccessary and sufficient condition: A is diagonalizable if and only if the minimal polynomial is a product of distinct linear factors.

6.4.2 Cayley-Hamilton Theorem

Theorem 6.4.2.1: Cayley-Hamilton

Let $c_A(x) = \det(A - xI)$ be the characteristic polynomial of the $n \times n$ matrix A over a field $\mathbb{F}$, then $c_A(A) = 0$.

Proof. Firstly note that if $P(x) = \sum_{i=1}^n P_i x^i$ and $Q(x) = \sum_{j=1}^m Q_j x^j$ are polynomials with matrix coefficients where the matrix is $n \times n$, and $R(x) = \sum_{k=1}^{n+m} R_k x^k$ is the product of the two polynomials with $R_k = \sum_{i+j=k} P_i Q_j$, then if M is a $n \times n$ matrix that commmutes with all of Q_j, then we have

$$R(M) = P(M)Q(M)$$

This can be seen by expanding the sums out.

Now take $Q(x) = A - xI$ and $P(x) = \text{adj}(Q)$. Then we have
$P(x)Q(x) = \det(A - xI)I = c_A(x)I$ by property of the adjoint. And since A commutes with all coefficients of the polynomial of Q, we have

$$c_A(A)I = P(A)Q(A) = P(A) \cdot 0 = 0$$

Thus $c_A(A) = 0$. □

Corollary 6.4.2.2

For any $A_{n\times n}$ over $\mathbb{F}$, we have $\mu_A | c_A$, and $\deg(\mu_A) \leq n$.

Proof. This is clear since $c_A(A) = 0$ and μ_A is the minimal polynomial such that $\mu_A(A) = 0$ by division with remainder. Since $\deg(c_A) = n$, $\deg(\mu_A) \leq n$. □

This lemma may help with finding out the minimal polynomial.

Lemma 6.4.2.3

Let λ be an eigenvalue of A. Then $\mu_A(\lambda) = 0$.

Proof. Let v be an eigenvector of the eigenvalue λ of A. Trivially $\mu_A(A)v = 0$. But also since

$$A^n v = \lambda^n v$$

we have $0 = \mu_A(A)v = \mu_A(\lambda)v$. Since v is nonzero we must have $\mu_A(\lambda) = 0$. □

In general, to deduce the formula for the minimal polynomial, we follow three steps.

Step 1: Find out the eigenvalues of the matrix.
Step 2: List out the possibilities of the minimal degree. This is done using the fact that $\mu_A(\lambda) = 0$ and $\deg(\mu_A) \leq n$.
Step 3: Plug in the matrix to find out which polynomial has its root at A.

There is another method to find out the formula using the following lemma.

Lemma 6.4.2.4

Let $T : V \to V$ be a linear map. Let

$$V = W_1 \oplus \cdots \oplus W_k$$

be the direct sum of invariant subspaces, meaning $W_1, \dots, W_k$ are invariant subspaces of T. Let $\mu_i(x)$ be the minimal polynomial of $T|_{W_i}$. Then

$$\mu_T(x) = \operatorname{lcm}(\mu_1, \dots, \mu_k)$$

Using this, we derive a better algorithm to find the minimal polynomial:

Step 1: Take $v \neq 0$ an eigenvector and set $W = \operatorname{span}\{v, T(v), T^2(v), \dots\}$. Then W is invariant under T. Let d be the minimal positive integer such that $v, T(v), \dots, T^d(v)$ are linearly dependent. Then $v, T(v), \dots, T^{d-1}(v)$ are linearly independent. Then we know that $\mu_T(x)$ has degree larger than d since else $\mu_T(x)v$ will never be 0. Then there is a nontrivial linear dependency relation of the form

$$T^d(v) + c_{d-1}T^{d-1}(v) + \cdots + c_1 T(v) + c_0 v = 0$$

Step 2: Consider the polynomial

$$x^d + c_{d-1}x^{d-1} + \cdots + c_1 x + c_0$$

Then this is precisely the minimal polynomial.

6.4.3 Generalized Eigenspace

Definition 6.4.3.1: Generalized Eigenvector

Let $T : V \to V$. Fix $k \in \mathbb{N}$. A non zero vector v such that

$$(T - \lambda I)^k v = 0$$

is called a generalized eigenvector of T with respect to the eigenvalue λ. Also we define

$$N_k(T,\lambda)=\{v\in V|(T-\lambda I)^kv=0\}=\ker((T-\lambda I)^k)$$

to be the generalized eigenspace of index k of T with respect to λ. The set of all generalized eigenvector regardless of the index, is defined to be

$$G(T,\lambda)=\{v\in V|(T-\lambda I)^kv=0\text{ for some }k\in\mathbb{N}\}=\bigcup_{k=1}^{\infty}N_k(T,\lambda)$$

Proposition 6.4.3.2

The dimensions of corresponding generalized eigenspaces of similar matrices are the same.

Proof. This is true since generalized eigenspaces are defined without explicitly defining a basis for the linear map. Thus similar matrices that induce the same linear map will have the same dimensions for generalized eigenspaces. □

Proposition 6.4.3.3

Let $T:V\to V$ be a linear map with eigenvalue λ. Then

$$N_1(T,\lambda)\subseteq N_2(T,\lambda)\subseteq N_3(T,\lambda)\subseteq\dots$$

Proof. Trivially if $v\in\ker(A-\lambda I)^i$. This means that $(A-\lambda I)^iv=0$ and $(A-\lambda I)^{i+1}v=0$. Thus $N_i(T,\lambda)\subseteq N_{i+1}(T,\lambda)$ for any i and we are done. □

Proposition 6.4.3.4

Let λ be an eigenvalue of $T:V\to V$. There exists some $n\in N$ such that

$$N_n(T-\lambda I)=N_{n+1}(T-\lambda I)=\dots$$

Denote $d(\lambda)$ the smallest of such n.

Proof. $d(\lambda)\leq\dim(V)$. □

Proposition 6.4.3.5

Let λ be an eigenvalue of $T:V\to V$. Then

$$G(T,\lambda)=N_{\dim(V)}(T,\lambda)$$

Proposition 6.4.3.6

Let $T:V\to V$ with eigenvalues $\lambda_1,\dots,\lambda_m$. Let $v_i\in G(T,\lambda_i)$ for $i\in\{1,\dots,m\}$. Then $v_1,\dots,v_m$ are linearly independent.

Theorem 6.4.3.7

Let V be a vector space of an algebraically closed field F. Let $T:V\to V$. Let $\lambda_1,\dots,\lambda_m$ be distinct eigenvalues of T. Then

- $V = G(T, \lambda_1) \oplus \cdots \oplus G(T, \lambda_m)$
- $G(T, \lambda_j)$ is invariant under T.

6.4.4 Jordan Canonical Form

Definition 6.4.4.1: Jordan Chain

A Jordan Chain of length k is a sequence of nonzero vectors $v_1, \dots, v_k$ such that

$$\begin{aligned} Av_1 &= \lambda v_1 \\ Av_2 &= \lambda v_2 + v_1 \\ &\vdots \\ Av_k &= \lambda v_k + v_{k-1} \end{aligned}$$

for some eigenvalue λ of A.

Corollary 6.4.4.2

Let $v_1, \dots, v_k$ be a Jordan Chain of λ. Then $v_i \in N_i(A, \lambda)$ for $i \in \{1, \dots, k\}$ and $(T - \lambda I)(v_i) = v_{i-1}$ except for $i = 1$.

Proof. The result is immediate from substitution in the Jordan Chains. □

Proposition 6.4.4.3

The vectors in a Jordan chain are linearly independent.

Proposition 6.4.4.4

The subspace spanned by a Jordan Chain is invariant under its linear map.

Proof. Note that we just have to find out where $v_1, \dots, v_k$ are mapped to since they are a basis of our subspace. But $T(v_i) = \lambda v_i + v_{i-1}$ for $i \in \{2, \dots, k\}$ which is a linear combination of our basis, we must have that the subspace is invariant. □

Definition 6.4.4.5: Jordan Block of Degree k

Define the Jordan block of degree k to be the $k \times k$ matrix

$$\gamma_{ij} = \begin{cases} \lambda & \text{if } j = i \\ 1 & \text{if } j = i + 1 \\ 0 & \text{otherwise} \end{cases}$$

This means the diagonal of the matrix is λ and the super diagonal is 1. It is denoted as $J_{\lambda,k}$

Corollary 6.4.4.6

The matrix of T with respect to the basis $v_1, \dots, v_n$ is a Jordan Block if and only if $v_1, \dots, v_n$ is a Jordan Chain.

Proof. Let $v_1, \dots, v_k$ be a Jordan Chain. Our matrix should be in the form

$$\begin{pmatrix} T(v_1) & T(v_2) & \cdots & T(v_k) \end{pmatrix}$$

Calculating each column gives

$$\begin{pmatrix} \lambda & 1 & 0 & 0 & 0 \\ 0 & \lambda & 1 & 0 & 0 \\ 0 & 0 & \ddots & \ddots & 0 \\ 0 & 0 & 0 & \lambda & 1 \\ 0 & 0 & 0 & 0 & \lambda \end{pmatrix}$$

For the other side, it is easy to simply compute $v_1, \dots, v_k$ out with the matrix. □

Definition 6.4.4.7: Jordan Basis

A Jordan basis is a basis consisting of one or more Jordan chains strung together.

Lemma 6.4.4.8

A Jordan Basis is indeed a basis.

Proof. We naturally assume the string of Jordan Chains consists of n vectors in total, corresponding to $\dim(V) = n$. Thus we just have to show linear independence. But this is also trivial. We have shown that vectors in the same Jordan Chain are independent, and vectors in different $G(T, \lambda)$ are proven to be linearly independent. □

Definition 6.4.4.9: Direct Sum

Let $A \in F^{n\times n}$ and $B \in F^{m\times m}$. Define the direct sum to be

$$A \oplus B = \begin{pmatrix} A & 0_{n\times m} \\ 0_{m\times n} & B \end{pmatrix}$$

Lemma 6.4.4.10

Let B, C be square matrices.

$$(B \oplus C)^n = B^n \oplus C^n$$

Corollary 6.4.4.11

The matrix of T with respect to a Jordan Basis is the direct sum

$$J_{\lambda_1,k_1} \oplus \cdots \oplus J_{\lambda_s,k_s}$$

of Jordan Blocks.

Theorem 6.4.4.12

Let A be a matrix over an algebraically closed field. Then there exists a Jordan basis for A. Moreover, A is similar to some J which is a direct sum of Jordan Blocks.

Proof. We will construct the basis with 3 methods. They each contribute to part of the Jordan Basis. Firstly, we will obtain a basis of a subspace. We induct on n. The case of $n = 1$ is trivial.

Now suppose $T : V \to V$ is a linear map with $\dim(V) = n$. We want to find a restriction of T that is an automorphism to apply the induction hypothesis. Fix λ to be one of the eigenvalues of T. This will be used throughout the entire proof. This λ is possible because the ground field is algebraically closed. Now I claim that $U = \text{im}(T - \lambda I)$ is invariant under T. If this is true, then $T|_U : U \to U$ can be used to apply induction hypothesis. So all we have to show is that U is T-invariant and that $\dim(U) < \dim(V)$. The second item must be true by the rank nullity theorem. There must be an eigenvector in $\ker(T - \lambda I)$. Thus $\ker(T - \lambda I) \geq 1$ which implies that $\dim(U) < n$ by the rank nullity theorem. Now we prove that U is invariant. Let $u \in U$, I show that $T(u) \in U$. If $u \in U$, then $u = (T - \lambda I)(v)$ for some $v \in V$, hence

$$T(u) = T(T - \lambda I)(v) = (T - \lambda I)(T(v)) \in \text{im}(T - \lambda I) = U$$

Thus we have proven that $T|_U : U \to U$. Apply induction hypothesis here to obtain a Jordan Basis for T_U. Call that Jordan Basis $e_1, \dots, e_m$.

We now construct our second set of vectors. Recall that a Jordan Basis is a string of l Jordan Chains. Let $v_1, \dots, v_k$ denote one of the Jordan Chains. We can extend this Jordan Chain one more by setting $(T - \lambda I)^{k+1}(v_{k+1}) = 0$. Do the same thing for everey Jordan Chain, and relabel them to $w_1, \dots, w_l$. As a side note, the two set of vectors we have now still form a Jordan Basis because we simply extended every Jordan Chain one more.

For the final set of vectors, observe that the first vector of each of the l Jordan Chains are eigenvectors of $T|_U$ with its eigenvalue being λ. This is because by definition of Jordan Chains, $T|_U(v_1) = \lambda v_1$. Also note that those l vectors are linearly independent. Thus the first vectors of each of the l Jordan Chains span an l dimensional subspace of the eigenspace of λ. Recall that the eigenspace of λ has dimension $\dim(V) - \dim(U) = \dim(\ker(T - \lambda I))$. To minimize notation let $m = \dim(U)$. Thus by extension theorem we can extend the basis of the l dimensional subspace to $\ker(T - \lambda I)$. Call the extension vectors $w_{l+1}, \dots, w_{n-m}$. As a side note, these $n - m - l$ vectors each Jordan Chains of length 1. Thus we have complete our last set of vectors.

We have n vectors

$$e_1, \dots, e_m, w_1, \dots, w_l, w_{l+1}, \dots, w_{n-m}$$

Thus we only need to prove that they are linearly independent. Let $x = \sum_{k=1}^m \beta_m e_m$. Let

$$\sum_{i=1}^{n-m} \alpha_i w_i + x = 0$$

Applying $T - \lambda I$ on both sides give

$$\sum_{i=1}^{l} \alpha_i (T - \lambda I) w_i + (T - \lambda I)(x) = 0$$

Since $w_{l+1}, \dots, w_{n-m}$ is in $\ker(T - \lambda I)$, they become 0. Now recall that our construction of $w_1, \dots, w_l$ is made by extending our Jordan Chains. So applying $(T - \lambda I)$ moves down our Jordan Chain. This means that $(T - \lambda I)x$ no longer contains the last term of each Jordan Chain and are linear combinations of $\{e_1, \dots, e_m\} \setminus \{\text{Last Term of each Jordan Chain}\}$, while all of the $(T - \lambda I)(w_1), \dots, (T - \lambda I)(w_l)$ are all last members of ecah Jordan Chain. From the fact that $e_1, \dots, e_m$ are a basis, we have $\alpha_1 = \cdots = \alpha_l = 0$.

Our sum now becomes $(T - \lambda I)x = 0$. Which means that $x \in \ker(T - \lambda I)$. Now our original sum becomes

$$\sum_{i=l+1}^{n-m} \alpha_i w_i + x = 0$$

By construction, $w_{l+1}, \dots, w_{n-m}$ extends a basis of the eigenspace of $T|_U$ for λ, thus $\alpha_{l+1} = \cdots = \alpha_{n-m} = 0$. Also since $e_1, \dots, e_m$ is a basis of U, we have $\beta_1 = \cdots = \beta_m = 0$. Finally by the above corollary, in a Jordan Basis, the matrix of T is a direct sum of Jordan Blocks. □

Lemma 6.4.4.13

If A, B are similar, then they have the same JCF up to reodering of the Jordan Blocks by direct sum.

Theorem 6.4.4.14

Let λ be an eigenvalue of a matrix A. Let J be the Jordan Canonical Form of A. Then

- The number of Jordan Blocks of J with eigenvalue λ is equal to $\dim(\ker(A - \lambda I))$
- Let $k > 0$. Then number of Jordan Blocks of J with eigenvalue λ of degree at least i is equal to $\dim(N_i(A, \lambda)) - \dim(N_{i-1}(A, \lambda))$

Proof. Since similar matrices have the same dimensions for their generalized eigenspaces corresponding to their eigenvalue, WLOG take $A = J = J_{\lambda_1,k_1} \oplus \cdots \oplus J_{\lambda_s,k_s}$. However, note that the dimension of $N_i(A \oplus B, \lambda)$ is equal to $\dim(N_i(A, \lambda)) + \dim(N_i(B, \lambda))$. So we just have to prove the theorem for a single Jordan Block.

Since $(J_{\lambda,k} - \lambda I)^i$ has a single diagonal line of ones i places above the diagonal for $i < k$, and is 0 for $i \geq k$, the dimension of its kernel is i for $0 \leq i \leq k$ and k for $i \geq k$. □

Corollary 6.4.4.15

The JCF of a matrix is unique up to a reordering of the Jordan Blocks.

Proof. The above theorem says that the number of Jordan Blocks associated with λ is determined by the nullity of A, and the size of every Jordan Block is determined by the dimension of the generalized eigenspaces. □

6.4.5 Results of the Jordan Normal Theorem

Lemma 6.4.5.1

Let $M = A \oplus B$. Then

$$c_M(x) = c_A(x)c_B(x)$$

and

$$\mu_M(x) = \gcd(\mu_A(x), \mu_B(x))$$

Proposition 6.4.5.2

Let A have JCF J. Let λ be an eigenvalue of A. Consider the Jordan Blocks in J related to λ. The string of Jordan Chains of these Jordan Blocks form a basis for $G(A, \lambda)$.

Theorem 6.4.5.3

Let $T : V \to V$ and $\lambda_1, \dots, \lambda_m$ be the set of eigenvalues of T. Then the characteristic polynomial of T is

$$c_A(x) = (-1)^n \prod_{k=1}^{m} (x - \lambda_k)^{a_k}$$

where a_k is the sum of the degrees of the Jordan Blocks of T of eigenvalue λ_k

Theorem 6.4.5.4

Let $T : V \to V$ and $\lambda_1, \ldots, \lambda_m$ be the set of eigenvalues of T. Then the minimal polynomial of T is

$$\mu_A(x) = \prod_{k=1}^{m} (x - \lambda_k)^{b_k}$$

where b_k is the largest among the degrees of the Jordan Blocks of T of eigenvalue of λ_k. Also, we have $d(\lambda_k) = b_k$

Theorem 6.4.5.5

Let $T : V \to V$ and $\lambda_1, \ldots, \lambda_m$ be the set of eigenvalues of T. Then T is diagonalizable if and only if $\mu_A(x)$ has no repeated factors.

Theorem 6.4.5.6

Let $A, B \in M_{n\times n}(\mathbb{C})$. Then A and B are similar if and only if the following two are true:

- A and B have the same set of eigenvalues
- $\dim(\ker(A - \lambda I)^i) = \dim(\ker(B - \lambda I)^i)$ for all i and eigenvalues λ

To finish this section, we show the process of determining the Jordan Canonical form of a matrix. The steps are usually as follows:

Step 1: Find out the nullity of $A - \lambda I$ as this gives us the number of Jordan Blocks with eigenvalue λ.
Step 2: To find out the number of Jordan blocks with eigenvalue λ and size at least i, we calculate $\dim(\ker(A - \lambda I_n)^i) - \dim(\ker(A - \lambda I)^{i-1})$.

To find out the change of basis matrix, meaning the Jordan Chains, we do the following step:

Step 3: Find out the last one in the chain v_k by solving $(A - \lambda I)^k v_k = 0$ while restricting v_k such that $(A - \lambda I)^{k-1} v_k \neq 0$, and then proceed to find out v_{k-1} by $(A - \lambda I)^{k-1} v_k = v_{k-1}$ and vice versa.

There are also extra information that we can use to determine the JCF:
The degree of $(x - \lambda)$ in μ_A indicates the maximum size of Jordan Blocks with eigenvalue λ.
The degree of $(x - \lambda)$ in c_A indicates the total size of used in the JCF of all Jordan Blocks with eigenvalue λ.

We give an example of finding the JCF of a matrix.

Example 6.4.5.7

Find the Jordan Canonical Form of

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 2 \end{pmatrix}$$

Proof. We begin by finding out the eigenvalues of A. We have that $c_A(x) = \det(A - xI) = (1 - x)^2(2 - x)$. This means that the eigenvalues are 1 and 2. Now we begin with step 1.

For eigenvalue 1, we have that row reduced form of $A - I$ is

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

Thus the nullity of $A - I$ is 1. This means that the number of Jordan blocks with eigenvalue 1 is 1. Using information from c_A, we know that the total size used for the eigenvalue 1 is 2. This means that there is exactly one Jordan block of size 2 in the JCF of A.

This leaves the fact that the remaining Jordan block of size 1 being the eigenvalue 2.

With this, we complete the JCF of A with

$$J = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

To compute the basis, or the change of basis matrix P, we use step 3. Since A has one Jordan block for eigenvalue 1, we need to find one string of Jordan chain. This chain needs to have length 2 since the size of the Jordan block is 2. (If there are multiple of Jordan blocks of the same eigenvalues, the end vector of the Jordan chains needs to be linearly indendent). We begin by finding the ending of the chain, v_2 by using the fact that $(A - I)^2 v_2 = 0$ and $(A - I)v_2 = v_1 \neq 0$. We have that

$$(A - I)^2 = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix}$$

We choose that $v_2 = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}$. Now we have

$$v_1 = (A - I)v_2 = \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix}$$

Finally, we choose $v_3 \in \ker(A - 2I)$. But row reducing $A - 2I$ gives

$$\begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

We can choose $v_3 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$. This means that

$$P = \begin{pmatrix} -1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix}$$

□

6.4.6 Functions of Matrices

We start of the last section with a formula for the nth power of a Jordan Block.

Lemma 6.4.6.1

Let $J_{\lambda,k}$ be a Jordan Block. Then

$$J_{\lambda,k}^n = \begin{pmatrix} \lambda^n & n\lambda^{n-1} & \cdots & \binom{n}{k-2}\lambda^{n-k+2} & \binom{n}{k-1}\lambda^{n-k+1} \\ 0 & \lambda^n & \cdots & \binom{n}{k-3}\lambda^{n-k+3} & \binom{n}{k-2}\lambda^{n-k+2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \lambda^n & n\lambda^{n-1} \\ 0 & 0 & \cdots & 0 & \lambda^n \end{pmatrix}$$

When used together with the fact that powers of matrices can be distributed through direct sums, we obtain the formula for finding powers of matrices. Namely if A has Jordan canonical form J and $A = PJP^{-1}$, then $A^n = PJ^nP^{-1}$.

We now define functions of matrices in terms of this decomposition.

Definition 6.4.6.2: Functions of Matrices

Let f be a function over $\mathbb{C}$. For every matrix $A \in M_{n\times n}(\mathbb{C})$, define $f(A)$ by

$$f(A) = Pf(J)P^{-1}$$

where $f(J) = f(J_{\lambda_1,k_1}) \oplus \cdots \oplus f(J_{\lambda_t,k_t})$ is the direct sum of Jordan blocks. And finally, for each Jordan block $J_{\lambda,k}$, define $f(J_{\lambda,k})$ by

$$f(J_{\lambda,k}) = \begin{pmatrix} f(\lambda) & f'(\lambda) & \cdots & \frac{1}{(k-2)!}f^{(k-2)}(\lambda) & \frac{1}{(k-1)!}f^{(k-1)}(\lambda) \\ 0 & f(\lambda) & \cdots & \frac{1}{(k-3)!}f^{(k-3)}(\lambda) & \frac{1}{(k-2)!}f^{(k-2)}(\lambda) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & f(\lambda) & f'(\lambda) \\ 0 & 0 & \cdots & 0 & f(\lambda) \end{pmatrix}$$

Note that taking $f(x) = x^n$, this coincides with the power of a Jordan block, which is presumably what motivated the this definition of matrix functions.

Chapter 7

Linear Algebra 2

7.1 Linear Forms and Quadratic Forms

7.1.1 Linear Forms

Definition 7.1.1.1: Linear Forms

A linear form on V is a linear map from V to $\mathbb{F}$.

Proposition 7.1.1.2: Dual Space

The set of all linear forms on V forms a vector space called the dual space V'.

Proof. Simply a check on the axioms of vector space. □

Lemma 7.1.1.3

Let V be a finite dimensional vector space. Then V' is also finite dimensional and $\dim(V') = \dim(V)$.

Definition 7.1.1.4: Dual Basis

Let $v_1, \ldots, v_n$ be a basis of V, then the dual basis of $v_1, \ldots, v_n$ is the list $\phi_1 \ldots, \phi_n$ of elements of V', where ϕ_k is a linear functional such that

$$\phi_k(v_i) = \begin{cases} 1 & \text{if } k = i \\ 0 & \text{if } k \neq i \end{cases}$$

Proposition 7.1.1.5

The dual basis of a basis of V is a basis of V'

Definition 7.1.1.6: Dual Map

Let $T \in \mathcal{L}(V, W)$. The dual map of T is the linear map $T' \in \mathcal{L}(W', V')$ defined by $T'(\phi) = \phi \circ T$ for $\phi \in W'$.

Proposition 7.1.1.7

Let $S, T \in \mathcal{L}(V, W)$ and $\lambda \in \mathbb{F}$.

- $(S + T)' = S' + T'$

- $(\lambda T)' = \lambda T'$
- $(ST)' = T'S'$.

7.1.2 Quadratic Forms

Definition 7.1.2.1: Quadratic Forms

A quadratic form in n variables $x_1, \ldots, x_n$ over a field K is a polynomial

$$q(x_1, \ldots, x_n) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j$$

Proposition 7.1.2.2

Every quadratic form $q(x_1, \ldots, x_n) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j$ can be represented by a matrix multiplication, namely

$$q(x_1, \ldots, x_n) = \begin{pmatrix} x_1 & \cdots & x_n \end{pmatrix} \begin{pmatrix} a_{11} & \frac{1}{2}a_{12} & \cdots & \frac{1}{2}a_{1n} \\ \frac{1}{2}a_{21} & a_{22} & \cdots & \frac{1}{2}a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{2}a_{n1} & \frac{1}{2}a_{n2} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

In particular, this matrix is symmetric with $a_{ij} = a_{ji}$ for $i, j \in \{1, \ldots, n\}$.

Proof. Multiplying out the entries of the matrix multiplication gives the original quadratic form. □

Proposition 7.1.2.3

A change of basis via the change of basis matrix P also changes the symmetric matrix of the quadratic form by $P^T AP$

Definition 7.1.2.4: Congruent Matrices

Two matrices A, B are said to be congruent if there exists some invertible matrix P such that $B = P^T AP$

Beware that congruences does not apply to only symmetric matrices. We will see more of it in action in bilinear forms.

Proposition 7.1.2.5

Two symmetric matrices are congruent if and only if they represent the same quadratic form with respect to different bases.

Theorem 7.1.2.6

Let $q(x_1, \ldots, x_n)$ be a quadratic form in n variables over a field K whose characteristic is not 2. Then there exists a basis such that $q(y_1, \ldots, y_n) = c_1 y_1^2 + \cdots + c_n y_n^2$ for some $c_1, \ldots, c_n \in K$.

Proof. There is a shorter proof for this theorem, but for the sake of the construction of $c_1, \ldots, c_n$, we will prove the theorem constructively. Suppose that q is represented by the symmetric matrix $A = (a_{ij})_{n\times n}$ with respect to the basis $b_1, \dots, b_n$. There are three steps in the construction. I use $b_1, \dots, b_n$ to indicate the old basis and $b'_1, \dots, b'_n$ to indicate the basis after the step.

Step 1: Arrange such that $q(b_1) \neq 0$. There are four cases here.

- If $a_{11} \neq 0$, then we are done.
- If $a_{11} = 0$ but $a_{kk} \neq 0$ for some $1 < k \leq n$. Then just set $b'_1 = b_k$ and $b'_k = b_1$. At the same time, the matrix for the quadratic form is changed by swapping rows r_1 and r_k, and then swapping the columns c_1 and c_k
- If $a_{kk} = 0$ for all $k \in \{1, \dots, n\}$, but there are some i, j such that $a_{ij} \neq 0$, then set $b'_i = b_i + b_j$ since $q(b_i + b_j) = 2a_{ij} \neq 0$ and so we reduced this case to the previous two cases.The matrix then becomes

$$\begin{pmatrix} 2a_{1k} & a_{12}+a_{k2} & \dots & a_{1k} & \cdots & a_{1n}+a_{kn} \\ a_{12}+a_{k2} & 0 & \cdots & a_{2k} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{1k} & a_{k2} & \cdots & 0 & \cdots & a_{kn} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{1n}+a_{kn} & a_{n2} & \cdots & a_{nk} & \cdots & 0 \end{pmatrix}$$

- If $a_{ij} = 0$ for all $i, j \in \{1, \dots, n\}$ then it is the zero function.

In this step the change of basis matrix is just the elementary matrices.

Step 2: Now we modify $b_2, \dots, b_n$ to make them orthogonal to b_1. Now set $b'_k = b_k - \frac{a_{1k}}{a_{11}} b_1$. This way, the matrix entry a_{1k} becomes zero. Now the change of basis matrix becomes

$$P = \begin{pmatrix} 1 & -\frac{a_{12}}{a_{11}} & \cdots & -\frac{a_{1n}}{a_{11}} \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}$$

After this step all the change of basis matrix should be compiled and calculated so that the new matrix for the quadratic form can be formed.

Step 3: Since the matrix for the quadratic form is now

$$\begin{pmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & ? & \cdots & ? \\ \vdots & \vdots & \ddots & \vdots \\ 0 & ? & \cdots & ? \end{pmatrix}$$

We can induct on n by repeating the process of step 1 with the entry a_{22} until we reach a_{nn}. □

This main theorem of quadratic forms shows that every quadratic form is congruent to a diagonal matrix. To illustrate the process of reduction, we look an example.

Example 7.1.2.7

Find a nice basis for the quadratic form $q\left(\begin{pmatrix} x \\ y \\ z \end{pmatrix}\right) = xy + 3yz - 5xz$.

Proof. Using the formula, we construct the matrix of q as

$$A = \begin{pmatrix} 0 & \frac{1}{2} & -\frac{5}{2} \\ \frac{1}{2} & 0 & \frac{3}{2} \\ -\frac{5}{2} & \frac{3}{2} & 0 \end{pmatrix}$$

We start the first change of basis. Notice that $a_{11} = a_{22} = a_{33} = 0$ in the matrix while a_{12} is not. Denote the standard basis by b_1, b_2, b_3. We perform the first basis change by $\{b_1' = b_1 + b_2, b_2' = b_2, b_3' = b_3\}$. This means that the change of basis matrix from new to old is

$$P' = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

To change A into the new matrix A', we simply replace r_1 by $r_1 + r_2$ and replace c_1 by $c_1 + c_2$. This gives

$$A' = \begin{pmatrix} 1 & \frac{1}{2} & -1 \\ \frac{1}{2} & 0 & \frac{3}{2} \\ -1 & \frac{3}{2} & 0 \end{pmatrix}$$

We keep track of changing the basis for clarity. We now have $A' = (P')^T AP'$.

The next step is to set the new basis to $b_2'' = b_2' - \frac{1}{2}b_1'$ and $b_3'' = b_3' + b_1'$. This means that the new basis is now $\{b_1 + b_2, \frac{1}{2}(b_2 - b_1), b_1 + b_2 + b_3\}$. The change of basis from this basis back to the old one is now

$$P'' = \begin{pmatrix} 1 & -\frac{1}{2} & 1 \\ 1 & \frac{1}{2} & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

Now the new matrix A'' is formed by replacing r_2 with r_2 by $r_2 - \frac{1}{2}r_1$ and r_3 with $r_3 + r_1$. Noticing that A'' must be symmetric, we need to take the new elements and replace the remanining lower traingualr elements so that A'' maintains symmetric. Also observe that the replacement of rows is exactly one changes into the new basis from the previous basis, where $b_2'' = b_2' - \frac{1}{2}b_1'$ etc. Now we have

$$A'' = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -\frac{1}{4} & 2 \\ 0 & 2 & -1 \end{pmatrix}$$

We now have $A'' = (P'')^T AP''$.

Now we perform the next change of basis. We set $b_3''' = b_3'' - \frac{2}{-\frac{1}{4}}b_2'' = b_3'' + 8b_2''$. Now the new basis is $\{b_1 + b_2, \frac{1}{2}(b_2 - b_1), -3b_1 + 5b_2 + b_3\}$. The change of basis matrix is now

$$P''' = \begin{pmatrix} 1 & -\frac{1}{2} & -3 \\ 1 & \frac{1}{2} & 5 \\ 0 & 0 & 1 \end{pmatrix}$$

Similar to the above, we replace r_3 with the same transformation and adjust A''' so that it remains symmetric. Thus now we have

$$A''' = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -\frac{1}{4} & 0 \\ 0 & 0 & 15 \end{pmatrix}$$

This means that we are done with $A''' = (P''')^T AP'''$. □

In general, this result of diagonalization is different from that of similar matrices. One should not be confusing reduction of congruent matrices into diagonal matrices and reduction of similar matrices into JCF as well as reduction of diagonalizable matrices into diagonal matrices.

7.1.3 Bilinear Forms

Definition 7.1.3.1: Bilinear Maps

Let V, W be vector spaces over a field $\mathbb{F}$. A bilinear map on V and W is a map $\tau : V \times W \to \mathbb{F}$ such that

- $\tau(a_1 + a_2v_2, w) = a_1\tau(v_1, w) + a_2\tau(v_2, w)$
- $\tau(v, b_1w_1 + b_2w_2) = b_1\tau(v, w_1) + b_2\tau(v, w_2)$

Theorem 7.1.3.2

Let V and W has basis $e_1, \dots e_n$ and $f_1, \dots, f_m$ respectively. Let $\tau(v, w)$ be a bilinear map. Then

$$\tau(v, w) = v^T \begin{pmatrix} \tau(e_1, f_1) & \cdots & \tau(e_1, f_m) \\ \vdots & \ddots & \vdots \\ \tau(e_n, f_1) & \cdots & \tau(e_n, f_m) \end{pmatrix} w$$

Proof. Simple to see by expanding the matrix multiplication. □

Proposition 7.1.3.3

Let A be the matrix of the bilinear map $\tau V \times W \to \mathbb{F}$ with respect to the basis $e_1, \dots, e_n$ and $f_1, \dots, f_m$ of V and W. Let B be the matrix with respect to the basis of $e_1', \dots, e_n'$ and $f_1', \dots, f_m'$ of V and W. Let P and Q be the change of basis matrix where $Pv' = v$ and $Qw' = w$ for $v \in V$ and $w \in W$. Then

$$B = P^T AQ$$

Definition 7.1.3.4: Bilinear Forms

A bilinear form is a bilinear map that maps $V \times V$ to $\mathbb{F}$.

Definition 7.1.3.5: Congruent Matrices

Two matrices are congruent if there exists an invertible matrix P such that $B = P^T AP$.

Note that this definition of congruence in matrices coincides with the definition given with symmetric matrices.

Definition 7.1.3.6: Types of Bilinear Forms

A bilinear form is said to be

- symmetric if $\tau(v, w) = \tau(w, v)$
- antisymmetric if $\tau(v, w) = -\tau(w, v)$
- positive definite if $\tau(v, v) > 0$ for all $v \in V$.

Proposition 7.1.3.7

Let $q : V \to K$ be a function. Then the following are equivalent.

- q is a quadratic form
- $q(cv) = c^2 v$ for $c \in K$ and $v \in V$ and $\tau(v, w) = \frac{1}{2}(q(v + w) - q(v) - q(w))$ is a bilinear form on V
- $q(v) = \tau(v, v)$ is a symmetric bilinear form on V

Proof. Let $q : V \to K$ be a function.

- (1) $\implies$ (2): Since every term in a quadratic form is quadratic, $q(cv) = c^2 q(v)$ is natural. The fact that $\tau(v, w)$ is bilinear is also easy to check.
- (2) $\implies$ (3): From (2) we know that $\tau(v, v) = q(v)$ by substituting v in the position of w and thus it clearly is a bilinear form. The position of w and v are also interchangable and thus it is symmetric.
- (3) $\implies$ (1): If $\tau(v, v)$ is a symmetric bilinear form then the matrix of τ is symmetric since $a_{ij} = \tau(e_i, f_j) = \tau(f_j, e_i) = a_{ji}$. Thus $q(v) = \tau(v, v)$ defines a quadratic form.

□

7.1.4 Sesquilinear Forms

7.2 Inner Product Spaces

7.2.1 Norms

Throught this section, $\mathbb{F}$ means either $\mathbb{R}$ or $\mathbb{C}$. In general normed vector spaces only perform well in these two fields.

Definition 7.2.1.1: Norm

Let V be a vector space. A norm on V is a function $\|\cdot\| : V \to \mathbb{F}$ such that

- $\|x\| \geq 0$ for all $x \in V$ with equality if and only if $x = 0$
- $\|\lambda x\| = |\lambda|\|x\|$ for all $x \in V$ and $\lambda \in \mathbb{F}$
- $\|x + y\| \leq \|x\| + \|y\|$ for all $x, y \in V$

Definition 7.2.1.2: Normed Vector Space

A normed vector space is a pair $(V, \|\cdot\|)$ where V is a vector space and $\|\cdot\|$ is a norm on V.

Proposition 7.2.1.3

Every normed vector space is a metric space.

Proof. Can easily be seen that setting $d(x, y) = \|x - y\|$ allows the norm to become a metric. □

Definition 7.2.1.4: Convex Set

Let V be a vector space. A subset K of V is convex if $x, y \in K$ implies

$$\lambda x + (1 - \lambda)y \in K$$

for $0 \leq \lambda \leq 1$.

Lemma 7.2.1.5

For every normed vector space, the unit ball $B_1(0) = \{v \in V | \|v\| \leq 1\}$ is convex.

Proposition 7.2.1.6

Let $N : V \to \mathbb{R}^+$ be a function that satisfies the first two requirements of a norm. If N also satisfies the fact that $\{x \in V | N(x) \leq 1\}$ is convex, then N is a norm.

Definition 7.2.1.7: Isometries

If (X, d_1) and (Y, d_2) are metric spaces, then a distancing preserving map between X and Y is a map

$$f : X \to Y$$

such that for any $P, Q \in X$, we have

$$d(f(P), f(Q)) = d(P, Q)$$

A bijective distancing preserving map is called an isometry.

7.2.2 Inner Products

Inner products are only properly defined for vector spaces over $\mathbb{R}$ and $\mathbb{C}$. From this point onwards we will limit our discussions with $V = \mathbb{R}^n$ or $\mathbb{C}^n$ and $K = \mathbb{R}$ or $\mathbb{C}$.

Definition 7.2.2.1: Inner Products

An inner product on V is a function that takes each ordered pair (x, y) of a vector space V to a number $\langle x, y\rangle \in K$ such that

- $\langle x, x\rangle \geq 0$ for all $x \in V$ with equality if and only if $x = 0$.
- $\langle x + z, y\rangle = \langle x, y\rangle + \langle z, y\rangle$ for all x for all $x, y, z \in V$
- $\langle \lambda x, y\rangle = \lambda\langle x, y\rangle$ for all $x, y \in V$ and $\lambda \in K$
- $\langle x, y\rangle = \overline{\langle y, x\rangle}$ for all $x, y \in V$

In this case V is called an inner product space.

Proposition 7.2.2.2

Let $u, v, w \in V$. Let $\lambda \in K$. Then the following are true.

- $\langle 0, u\rangle = \langle u, 0\rangle = 0$
- $\langle u, v + w\rangle = \langle u, v\rangle + \langle u, w\rangle$
- $\langle u, \lambda v\rangle = \overline{\lambda}\langle u, v\rangle$

Proof. Let $\langle \cdot, \cdot\rangle$ be an inner product over V.

- $\langle 0, u\rangle = \langle 0, u\rangle + \langle 0, u\rangle$ thus $\langle 0, u\rangle = 0$. $\langle u, 0\rangle = 0$ can be proven using the below property.
- Let $u, v, w \in V$. Then

$$\begin{aligned}
\langle u, v + w\rangle &= \overline{\overline{\langle u, v + w\rangle}} \\
&= \overline{\langle v + w, u\rangle} \\
&= \overline{\langle v, u\rangle + \langle w, u\rangle} \\
&= \overline{\langle v, u\rangle} + \overline{\langle w, u\rangle} \\
&= \langle u, v\rangle + \langle u, w\rangle
\end{aligned}$$

- Applying the same technique as above gives the desired result.

□

Proposition 7.2.2.3

Every inner product induces a norm.

Proof. Define the norm to be $\|x\| = \sqrt{\langle x, x\rangle}$. □

Proposition 7.2.2.4: Cauchy-Schwartz Inquality

For all $x, y \in V$,

$$|\langle x, y\rangle| \leq \|x\| \cdot \|y\|$$

with equality if and only if $y = \lambda x$ for some $\lambda \in \mathbb{F}$.

Proof. Let $z = x - \frac{|\langle x,y\rangle|}{\|y\|^2}y$. We have $\|z\| \geq 0$.

$$\begin{aligned}
\|z\|^2 &= \left\langle x - \frac{|\langle x,y\rangle|}{\|y\|^2}y, x - \frac{|\langle x,y\rangle|}{\|y\|^2}y \right\rangle \\
&= \langle x, x\rangle - 2\left\langle x, \frac{|\langle x,y\rangle|}{\|y\|^2}y \right\rangle + \left\langle \frac{|\langle x,y\rangle|}{\|y\|^2}y, \frac{|\langle x,y\rangle|}{\|y\|^2}y \right\rangle \\
&= \langle x, x\rangle - 2\frac{|\langle x,y\rangle|^2}{\|y\|^2} + \frac{|\langle x,y\rangle|^2}{\|y\|^4}\langle y, y\rangle \\
&= \langle x, x\rangle - 2\frac{|\langle x,y\rangle|^2}{\|y\|^2} + \frac{|\langle x,y\rangle|^2}{\|y\|^2} \\
&= \|x\|^2 - \frac{|\langle x,y\rangle|^2}{\|y\|^2}
\end{aligned}$$

Thus we have

$$\begin{aligned}
\|x\|^2 &\geq \frac{|\langle x,y\rangle|^2}{\|y\|^2} \\
\|x\| &\geq \frac{|\langle x,y\rangle|}{\|y\|} && \text{(Since they are all positive)} \\
\|x\| \cdot \|y\| &\geq |\langle x, y\rangle|
\end{aligned}$$

Note that we have equality if and only if $\|z\| = 0$, meaning $x = \frac{|\langle x,y\rangle|}{\|y\|^2}y$. We are done by taking $\lambda = \frac{\|y\|^2}{|\langle x,y\rangle|}$. □

7.3 Orthogonality

7.3.1 Orthogonal Vectors

Definition 7.3.1.1: Orthogonal Vectors

Two vectors $u, v \in V$ an inner product space are called orthogonal if $\langle u, v \rangle = 0$.

Corollary 7.3.1.2

Let V be an inner product space. Let $u, v \in V$. Then the following are true.

- 0 is orthogonal to every vector in V
- 0 is the only vector in V that is orthogonal to itself.

Proof. Easy check involving properties of the inner product. □

Theorem 7.3.1.3: Pythagorean Theorem

Suppose that $u, v \in V$ an inner product space and u, v are orthogonal. Then

$$\|u+v\|^2 = \|u\|^2 + \|v\|^2$$

Proof. The norm here is induced by the inner product and thus $\|x\| = \sqrt{\langle x, x \rangle}$. We have that

$$\begin{aligned}\|u+v\|^2 &= \langle u+v, u+v \rangle \\ &= \langle u, u \rangle + \langle u, v \rangle + \langle v, u \rangle + \langle v, v \rangle \\ &= \langle u, u \rangle + \langle v, v \rangle \\ &= \|u\|^2 + \|v\|^2\end{aligned}$$

□

Theorem 7.3.1.4: Orthogonal Decomposition

Let $u, v \in V$ and $v \neq 0$. Set $c = \frac{\langle u,v \rangle}{\|v\|^2}$ and $w = u - \frac{\langle u,v \rangle}{\|v\|^2} v$. Then $\langle w, v \rangle = 0$ and $u = cv + w$.

Proof. The fact that $u = cv + w$ is natural so we only have to prove that $\langle w, v \rangle = 0$. We have that

$$\begin{aligned}\langle w, v \rangle &= \left\langle u - \frac{\langle u, v \rangle}{\|v\|^2} v, v \right\rangle \\ &= \langle u, v \rangle - \left\langle \frac{\langle u, v \rangle}{\|v\|^2} v, v \right\rangle \\ &= \langle u, v \rangle - \frac{\langle u, v \rangle}{\|v\|^2} \langle v, v \rangle \\ &= \langle u, v \rangle - \frac{\langle u, v \rangle}{\|v\|^2} \|v\|^2 \\ &= 0\end{aligned}$$

Thus we are done. □

Proposition 7.3.1.5

Every orthonormal list of vectors are linearly independent.

Proof. Suppose that $v_1, \dots, v_n$ are orthnormal. We want to show that $v_n = \sum_{k=1}^{n-1} a_k v_k$ imnplies $a_1 = \dots = a_{n-1} = 0$. Then

$$\langle v_n, v_i \rangle = \sum_{k=1}^{n-1} a_k (v_k \cdot v_i) = a_i \|v_i\|^2$$

for $i \in \{1, \dots, n-1\}$ since $v_1, \dots, v_{n-1}$ are orthonormal. But since $\langle v_n, v_k \rangle = 0$ we must have $a_i = 0$. This means that $a_1 = \dots = a_{n-1} = 0$ and thus $v_1, \dots, v_n$ are linearly independent. □

7.3.2 Orthonormal Basis

Definition 7.3.2.1: Orthonormal Basis

A basis $v_1, \dots, v_n$ of an inner product space V with $\dim(V) = n$ is called orthonormal if

- $\|b_i\| = 1$ for $i \in \{1, \dots, n\}$
- $\langle b_i, b_j \rangle = \delta_{ij}$ for $i, j \in \{1, \dots, n\}$

Proposition 7.3.2.2

The orthonormal basis is indeed a basis for an inner product space V.

Proof. Since lists of orthonormal vectors are linearly independent and there are n vectors, they must also span V and thus is a basis. □

Theorem 7.3.2.3

Let $b_1, \dots, b_n$ be an orthonormal basis and $v = \sum_{k=1}^n a_k b_k$. Then

$$\|v\|^2 = \sum_{k=1}^n |a_k|^2$$

Proof. We have that

$$\begin{aligned}\|v\|^2 &= \left\langle \sum_{k=1}^n a_k b_k, \sum_{k=1}^n a_k b_k \right\rangle \\ &= \sum_{i=1}^n \sum_{j=1}^n a_i a_j (b_i \cdot b_j) \\ &= \sum_{i=1}^n \sum_{j=1}^n a_i a_j \delta_{ij} \\ &= \sum_{k=1}^n |a_k|^2\end{aligned}$$

and we are done. □

Proposition 7.3.2.4

Let $b_1, \dots, b_n$ be an orthonormal basis of V. Then

$$v = \sum_{k=1}^{n} \langle v, b_k \rangle b_k$$

Proof. Applying the inner product with b_i for each $i \in \{1, \dots, n\}$ gives $a_i = \langle v, b_i \rangle$ since $\langle b_i, b_k \rangle = 0$ for any $k \neq i$. Thus if $v = \sum_{k=1}^{n} a_k b_k$ then $v = \sum_{k=1}^{n} \langle v, b_k \rangle b_k$ and we are done. □

Theorem 7.3.2.5: Gram-Schmidt Procedure

Let $v_1, \dots v_m$ be a list of linearly independent vectors of V. Let $b_1 = \frac{v_1}{\|v_1\|}$. For $i = 2, \dots, m$. Define

$$b_i = \frac{v_i - \sum_{k=0}^{i-1} \langle v_i, b_k \rangle b_k}{\|v_i - \sum_{k=0}^{i-1} \langle v_i, b_k \rangle b_k\|}$$

Then $b_1, \dots, b_m$ are orthonormal and has the same span as $v_1, \dots, v_m$.

Theorem 7.3.2.6

Every finite dimensional inner product space has an orthonormal basis.

Proof. By the Gram-Schmidt procedure, every basis can be transformed into an orthonormal basis. □

7.3.3 Orthogonal Complements

Definition 7.3.3.1: Orthogonal Complement

Let U be a subset of an inner product space V. The orthogonal complement of U is defined as

$$U^\perp = \{v \in V | \langle v, u \rangle = 0 \text{ for all } u \in U\}$$

Proposition 7.3.3.2

Let U be a subset of V.

- U is a subspace of V if and only if $U^\perp$ is a subspace of V.
- $\{0\}^\perp = V$
- $V^\perp = \{0\}$
- $U \cap U^\perp = \{0\}$
- If $W \subseteq U$, then $U^\perp \subseteq W^\perp$

Proposition 7.3.3.3

Let U be a finite dimensional subspace of V. Then

$$U = (U^\perp)^\perp$$

Theorem 7.3.3.4

Suppose U is a finite dimensional subspace of V. Then

$$V = U \oplus U^{\perp}$$

and

$$\dim(U) + \dim(U^{\perp}) = \dim(V)$$

7.3.4 Orthogonal Maps

Definition 7.3.4.1: Orthgonal Maps

A linear map $T : V \to V$ is said to be orthogonal if

$$\langle T(v), T(w)\rangle = \langle v, w\rangle$$

for all $v, w \in V$.

One can think of orthgonal maps as orthgonality preserving maps. If $\langle v, w\rangle = 0$ then $\langle T(v), T(w)\rangle = 0$ which means orthogonality is preserved.

Proposition 7.3.4.2

Let $T : V \to V$ be a linear map over an inner product space V. Let A represent the linear map T. Then the following are equivalent.

- T is orthogonal
- A is orthogonal
- T maps orthonormal bases to orthonormal bases

Proof. Suppose that $T : V \to V$ is represented by A.

- (1) $\iff$ (2): We have that $\langle T(v), T(w)\rangle = v^T A^T A w$. Thus it is equal to $v^T w$ if and only if $A^T A = I$.
- (1) $\iff$ (3): Suppose that T is orthogonal. Suppose that $\{b_1, \dots, b_n\}$ is orthonormal. Then $\langle T(b_i), T(b_j)\rangle = \langle b_i, b_j\rangle = 0$ for $i, j \in \{1, \dots, n\}$. Thus $\{T(b_1), \dots, T(b_n)\}$ is orthogonal. But they are also orthonormal since $\|T(b_i)\|^2 = \langle T(b_i), T(b_i)\rangle = \langle b_i, b_i\rangle = \|b_i\|^2 = 1$. This means that $\|T(b_i)\| = 1$ for $i \in \{1, \dots, n\}$.

  Now suppose that T maps orthonormal bases to orthonormal bases. Then if $\{b_1, \dots, b_n\}$ is orthonormal, we have

$$\begin{aligned}
\langle T(v), T(w)\rangle &= \left\langle \left(\sum_{k=1}^n v_k T(b_k)\right), \left(\sum_{k=1}^n w_k T(b_k)\right)\right\rangle \\
&= \sum_{k=1}^n (v_k w_k)\langle T(b_k), T(b_k)\rangle && (\langle T(b_i), T(b_j)\rangle = 0 \text{ if } i \neq j) \\
&= \sum_{k=1}^n v_k w_k \\
&= \langle v, w\rangle
\end{aligned}$$

  Thus we are done.

□

7.4 Orthgonality in $\mathbb{R}^n$

7.4.1 Reduction of Quadratic Forms over $\mathbb{R}$

Orthogonality is interesting for real matrices because the notion of similarity and congruence coinincide under orthogonality. Notice that being similar and congruent to a diagonal matrix at the same time means that there exists an invertible P such that $P^TAP = P^{-1}AP = D$.

In the remaining sections we treat the adjugate in the case of $\mathbb{R}^n$ and save the case for $\mathbb{C}^n$ in another chapter. Then V in the remaining sections will only denote $\mathbb{R}^n$.

Definition 7.4.1.1: Euclidean Vector Space

An Euclidean vector space is $\mathbb{R}^n$ equipped with an inner product.

Proposition 7.4.1.2

A function $b : V \times V \to \mathbb{R}$ is an inner product over $\mathbb{R}$ if and only if b is bilinear and positive definite.

Lemma 7.4.1.3: Polarization Identity

For $x, y \in \mathbb{R}^n$,

$$\langle x, y\rangle = \frac{1}{4}\|x+y\|^2 - \frac{1}{4}\|x-y\|^2$$

Proof. Simple proof using the fact that $\|x\|^2 = \langle x, x\rangle$. □

Proposition 7.4.1.4: Sylvester's Law of Inertia

A quadratic form q over $\mathbb{R}$ has the form

$$q(x_1, \dots, x_n) = \sum_{k=1}^{t} x_k^2 - \sum_{k=1}^{u} x_k^2$$

where $t + u = \operatorname{rank}(q)$. The pair (t, u) is called the signature of q. This reduced quadratic form is also unique in the sense that the number of positives and number of negatives of any two reduced forms are the same.

Moreover, every symmetric matrix is congruent to a matrix of the form

$$\begin{pmatrix} I_t & 0 & 0 \\ 0 & -I_u & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

where (t, u) is the signature of the quadratic form.

Proof. We saw in theorem 1.2.7 that every quadratic form can be expressed as

$$q(y_1, \dots, y_n) = \sum_{k=1}^{n} c_k y_k^2$$

By doing a basis change with $b_k' = \frac{1}{\sqrt{c_k}} b_k$ whenever $c_k \neq 0$ will give us the above sum. For those that have $c_k = 0$, the terms vanish and are exactly in the kernel of the quadratic form thus $t + u = \operatorname{rank}(q)$.

The second part is direct from the fact that same quadratic forms with different matrix representations imply their representations are similar. □

7.4.2 Reduction of Inner Products

Definition 7.4.2.1: Dot Product

The dot product in $\mathbb{R}^n$ is defined to be the inner product given by

$$x \cdot y = x_1 y_1 + \cdots + x_n y_n$$

in standard basis.

Theorem 7.4.2.2

Let $\langle \cdot, \cdot \rangle$ be an inner product on a real vector space V. Then there exists an basis $b_1, \dots, b_n$ of V such that the inner product, when represented in the orthonormal basis, takes the form of exactly the dot product.

Proof. Let $\langle \cdot, \cdot \rangle$ be our inner product in question. Define a quadratic form by

$$q(x) = \langle x, x \rangle = \|x\|^2$$

We know that this quadratic form can be reduced to

$$q(x_1, \dots, x_n) = x_1^2 + \cdots + x_t^2 - x_{t+1}^2 - \dots x_{t+u}^2$$

Now we must have $u = 0$ since if $u > 0$, then the basis vector b_{t+1} satisfies $q(b_{t+1}) = -1$ and $q(b_{t+1}) = \langle b_{t+1}, b_{t+1} \rangle$ which is a contradiction since inner products are positive definite. Also $t = n$ since if $t < n$, then $\langle b_{t+1}, b_{t+1} \rangle = 0$ which is again a contradiction.

Using polarization, we see that $\langle x, y \rangle = x_1 y_1 + \cdots + x_n y_n$ in that basis and we are done. □

With this theorem, we know that any inner product can be expressed in the dot product as long as it is under a suitable basis. Thus we now reduce our discussion to only the dot product, as our standard inner product in $\mathbb{R}^n$.

The below theorem, while unrelated to the reduction of inner products, is a result of Gram-schmidt process that is only true for real matrices.

Theorem 7.4.2.3: QR Decomposition

Let A be an $n \times n$ matrix over $\mathbb{R}$. Then we can write $A = QR$ where Q is orthogonal and R is upper triangular.

Proof. We split the matrices into two cases. Firstly consider the case where A is invertible. We can treat A as a change of basis matrix from the basis $\{a_1, \dots, a_n\}$ where a_k is the column of A for $k \in \{1, \dots, n\}$. This change of basis matrix takes $\{a_1, \dots, a_n\}$ to $\{e_1, \dots, e_n\}$ which is the standard basis. Apply the Gram-schmidt process to $\{a_1, \dots, a_n\}$ to get $\{b_1, \dots, b_n\}$ which is an orthonormal basis. Let Q be the change of basis matrix from $\{b_1, \dots, b_n\}$ to $\{e_1, \dots, e_n\}$. Let R be the change of basis matrix from $\{a_1, \dots, a_n\}$ to $\{b_1, \dots, b_n\}$. Then clearly $A = QR$. We just have to show that Q is orthogonal and R is upper triangular.

Q being orthonormal is trivial since columns of Q are just $b_1, \dots, b_n$. Using the Gram-schimdt process, we can see that the change of basis from $\{b_1, \dots, b_n\}$ to $\{a_1, \dots, a_n\}$, each b_k is only affected by $a_1, \dots, a_k$ from the old basis. This means that the change of basis matrix must be upper triangular and its inverse must also be upper triangular.

Now we also have the case when A is not invertible. □

We now give an example of QR decomposition, in conjunction with the Gram-schmidt procedure.

Example 7.4.2.4

Find the QR decomposition of

$$A = \begin{pmatrix} -1 & 0 & -2 \\ 2 & 0 & -1 \\ 0 & -2 & -2 \end{pmatrix}$$

Proof. A quick check shows that A is invertible. Let a_1, a_2, a_3 be the columns of A. We start the Gram-schimdt process by taking the new basis $f_1 = \frac{a_1}{\|a_1\|} = \begin{pmatrix} -\frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \\ 0 \end{pmatrix}$. We also need to keep track on the change of basis matrix. We have that $a_1 = \sqrt{5} f_1$.

For the next step, we find that

$$\begin{aligned} f_2 &= \frac{a_2 - (a_2 \cdot f_1) f_1}{\|a_2 - (a_2 \cdot f_1) f_1\|} \\ &= \frac{a_2}{\|a_2\|} \\ &= \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix} \end{aligned}$$

This means that $a_2 = 2f_2$.

Finally, we have that

$$\begin{aligned} f_3 &= \frac{a_3 - (a_3 \cdot f_1) f_1 - (a_3 \cdot f_2) f_2}{\|a_3 - (a_3 \cdot f_1) f_1 - (a_3 \cdot f_2) f_2\|} \\ &= \frac{a_3 - 2f_2}{\|a_3 - 2f_2\|} \\ &= \frac{a_3 - 2f_2}{\sqrt{5}} \\ &= \begin{pmatrix} -\frac{2}{\sqrt{5}} \\ -\frac{1}{\sqrt{5}} \\ 0 \end{pmatrix} \end{aligned}$$

We have that $a_3 = 2f_2 + \sqrt{5} f_3$. Combining everything together, we have that

$$\begin{pmatrix} -1 & 0 & -2 \\ 2 & 0 & -1 \\ 0 & -2 & -2 \end{pmatrix} = \begin{pmatrix} -\frac{1}{\sqrt{5}} & 0 & -\frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & 0 & -\frac{1}{\sqrt{5}} \\ 0 & -1 & 0 \end{pmatrix} \begin{pmatrix} \sqrt{5} & 0 & 0 \\ 0 & 2 & 2 \\ 0 & 0 & \sqrt{5} \end{pmatrix}$$

□

7.4.3 Adjoints

Proposition 7.4.3.1

Let V be an inner product space and $T : V \to V$ be a linear map. Then there exists a unique linear map $T^* : V \to V$ such that

$$\langle T(v), w \rangle = \langle v, T^*(w) \rangle$$

for all $v, w \in V$.

Proof. Let T be a linear map. Then the function $\tau(v,w) = \langle T(v), w\rangle$ is a bilinear form since the inner product is bilinear. But we know that bilinear forms can be represented by a matrix multiplication, namely $\tau(v,w) = v^T Aw$ where A is defined as in theorem 1.3.2. Then treating Aw as the linear map $T^*(w)$ and since $v^T w = v \cdot w$, we have that $\tau(v,w) = v \cdot T^*(w)$ thus proving existence. Uniqueness follows naturally by construction of the matrix A. □

Definition 7.4.3.2: Adjoint of a Linear Map

T^* in the above case is called the adjoint of T.

Definition 7.4.3.3: Self-Adjoint

A linear map $T : V \to V$ is said to be self-adjoint if $T^* = T$

Proposition 7.4.3.4

Let T be a linear map represented by a matrix A. Then the following are true.

- T is self-adjoint if and only if A is symmetric.
- T is orthogonal if and only if $T^* = T^{-1}$.

Proof. Let T be self-adjoint. Then $Av \cdot w = v \cdot Aw$ for all $v, w \in V$. □

Proposition 7.4.3.5

Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be self-adjoint. Then every eigenvalues of T are real.

Proof. Suppose that $T(v) = Av$ where A is a representation of T. Suppose that λ is an eigenvalue of T. Then $Av = \lambda v$ for some eigenvector $v \in \mathbb{R}^n$. Then taking complex conjugates give

$$\overline{Av} = \overline{\lambda v}$$
$$A\overline{v} = \overline{\lambda}\overline{v}$$

Taking the transpose of $Av = \lambda v$ gives $v^T A^T = \lambda v^T$ and $v^T A = \lambda v^T$. Multiplying $\overline{v}$ on bothe sides give

$$v^T A\overline{v} = \lambda v^T \overline{v}$$
$$\overline{\lambda} v^T \overline{v} = \lambda v^T \overline{v}$$

But $v^T\overline{v} = v_1\overline{v_1} + \cdots + v_n\overline{v_n} = |v_1|^2 + \cdots + |v_n|^2$ which is 0 if and only if v is 0. Since eigenvectors are taken to be nonzero, we must have $\lambda = \overline{\lambda}$ and thus λ is real. □

Theorem 7.4.3.6

Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear map on the inner product space $\mathbb{R}^n$ that is seldf-adjoint. Then there exists an orthonormal basis consisting of eigenvectors of T.

Equivalently, for every quadratic form q on V, there exists an orthonormal basis $b_1, \dots, b_n$ such that

$$q(x_1, \ldots, x_n) = \sum_{k=1}^{n} c_k x_k^2$$

for some $c_1, \ldots, c_n \in \mathbb{R}$.

Proof. Notice that the two statements are exactly the same and I will ommit the reason.

We prove by induction on n. Suppose that the theorem holds for $n-1$. Let T be the linear map. By the above we know that T has at least one eigenvalue in $\mathbb{R}$, say λ_1. Let f_1 be the corresponding eigenvector with magnitude 1.

Consider the orthogonal complement $W = \{w \in V | w \cdot f_1 = 0\}$ of f_1. Since W is the kernel of the linear map $S : V \to \mathbb{R}$ defined by $S(v) = v \cdot f_1$, it is a subspace of V dimension $n-1$. I claim that $T(W) \subseteq W$.

Let $w \in W$. We have

$$T(w) \cdot f_1 = w \cdot T(f_1) = w \cdot \lambda_1 f_1 = 0$$

by self-adjoint. Thus we have shown that $T(W) \subseteq W$.

Applying the induction hypothesis on W, we have an orthonormal basis $f_2, \ldots, f_n$ of W consisting of eigenvectors of T. By definition, $f_1, \ldots, f_n$ is an orthonormal basis and we are done. □

Notice that the above two statements are also equivalent to saying that every real symmetric matrix is congruent and similar to a diagonal matrix.

Proposition 7.4.3.7

If $T : \mathbb{R}^n \to \mathbb{R}^n$ is self-adjoint, and λ, μ are distinct eigenvalues of T with eigenvectors v, w, then $v \cdot w = 0$.

Proof. We have that

$$\begin{aligned} v^T A w &= v \cdot Aw \\ &= v^T \mu w \\ &= \mu(v \cdot w) \end{aligned}$$

and

$$\begin{aligned} v^T A w &= v^T A^T w \\ &= (Av)^T w \\ &= (\lambda v)^T w \\ &= \lambda v^T w \\ &= \lambda(v \cdot w) \end{aligned}$$

Comparing the two results, we have that $(\mu - \lambda)(v \cdot w) = 0$ and thus $v \cdot w = 0$. □

The proposition will prove itself to be useful in finding an orthonormal basis for self-adjoint linear maps.

7.4.4 Singular Value Decomposition

Theorem 7.4.4.1: Singular Value Decomposition for Linear Maps

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear map of rank k. Then there eixsts unique positive numbers $\gamma_1 \geq \gamma_2 \geq \cdots \geq \gamma_k \geq 0$ and orthonormal bases of $\mathbb{R}^n$ and $\mathbb{R}^m$ such that the matrix of T with respect to these bases is

$$\begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix}$$

where $D = \text{diag}(\gamma_1, \dots, \gamma_k)$. In fact, the γ_i are exactly the nonzero eigenvalues of T^*T, each one appearing as many times as the dimension of the corresponding eigenspace.

Theorem 7.4.4.2: Singular Value Decomposition for Matrices

Let $A_{m\times n}$ be a matrix. There exists unique singular values $\gamma_1 \geq \gamma_2 \geq \dots \gamma_k \geq 0$ where $k =$rank(A), and orthogonal matrices P, Q such that

$$\begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} = P^T AQ$$

where $D = \text{diag}(\gamma_1, \dots, \gamma_k)$.

We present an example of singular value decomposition for illustration.

Example 7.4.4.3

Find the singular value decomposition of the matrix

$$A = \begin{pmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{pmatrix}$$

Proof. Step 1: We compute the singular values of A, which is just the squareroot of the eigenvalues of $A^T A$. Now

$$A^T A = \begin{pmatrix} 80 & 100 & 40 \\ 100 & 170 & 140 \\ 40 & 140 & 200 \end{pmatrix}$$

We have that $c_{A^T A}(x) = x(360 - x)(90 - x)$. This means that the singular values are $\gamma_1 = \sqrt{360} = 6\sqrt{10}$ and $\gamma_2 = \sqrt{90} = 3\sqrt{10}$. Now we want P and Q such that

$$P^T AQ = \begin{pmatrix} 6\sqrt{10} & 0 & 0 \\ 0 & 3\sqrt{10} & 0 \end{pmatrix}$$

Step 2: We find the orthonormal eigenvectors of $A^T A$ so that it forms the matrix Q. This gives

$$Q = \begin{pmatrix} \frac{1}{3} & -\frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & -\frac{1}{3} & -\frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} & \frac{1}{3} \end{pmatrix}$$

Step 3: We now calculate P by finding the image of the above basis under A, and dividing it with the nonzero singular values. This gives

$$\begin{aligned} P &= \begin{pmatrix} \frac{1}{6\sqrt{10}} Ab_1 & \frac{1}{3\sqrt{10}} Ab_2 \end{pmatrix} \\ &= \begin{pmatrix} \frac{3}{\sqrt{10}} & \frac{1}{\sqrt{10}} \\ \frac{1}{\sqrt{10}} & -\frac{3}{\sqrt{10}} \end{pmatrix} \end{aligned}$$

This means that we are done. □

7.5 Multilinear Algebra

7.5.1 Tensor Products

Definition 7.5.1.1: Bilinear Mappings

Let V_1, V_2, W be vector spaces over $\mathbb{F}$. Let $\phi : V_1 \times V_2 \to W$ be a mapping. Then ϕ is called bilinear if

- $\phi(\lambda x_1 + \mu x_2, y) = \lambda\phi(x_1, y) + \mu\phi(x_2, y)$ for all $x_1, x_2 \in V_1$, $y \in V_2$ and $\lambda, \mu \in \mathbb{F}$
- $\phi(x, \lambda y_1 + \mu y_2) = \lambda\phi(x, y_1) + \mu\phi(x, y_2)$ for all $x \in V_1$, $y_1, y_2 \in V_2$ and $\lambda, \mu \in \mathbb{F}$

If W is the ground field $\mathbb{F}$, we say that ϕ is a bilinear function.

The tensor product is defined through a universal property. The following universal property essentially determines the uniqueness for a tensor product, but not the existence part. To complete the definition and show that tensor products indeed exists, we explicitly define one such of tensor products (one such because it is defined up to canonical isomorphism).

Definition 7.5.1.2: Tensor Product and Universal Property

The tensor product of two vector spaces V_1, V_2 is a vector space denoted $V_1 \otimes V_2$, toegther with a bilinear map $\phi : V_1 \times V_2 \to V_1 \otimes V_2$ defined by $\phi(v_1, v_2) = v_1 \otimes v_2$ such that for every bilinear map $h : V_1 \times V_2 \to W$, there is a unique linear map $\overline{h} : V_1 \otimes V_2 \to W$ such that $h = \overline{h} \circ \phi$. In other words, the following diagram commutes:

$$\begin{tikzcd}
V_1 \times V_2 \arrow[r, "\phi"] \arrow[dr, "h"] & V_1 \otimes V_2 \arrow[d, "\overline{h}"] \\
& W
\end{tikzcd}$$

This universal property allows canonical isomorphism since if we have two tensor products $V_1 \otimes_1 V_2$ and $V_1 \otimes_2 V_2$, we can apply the universal property to the both of them to obtain an isomorphism between the two. One way of explicitly calculating the elements are as follows:

Proposition 7.5.1.3

Let V, W be vector spaces over a field $\mathbb{F}$. Define L to be the vector space that has $V \times W$ as a basis. Define R to be the linear subspace of L spanned by

$$\{(v_1+v_2, w)-(v_1, w)-(v_2, w), (v, w_1+w_2)-(v, w_1)-(v, w_2), (sv, w)-s(v, w), (v, sw)-s(v, w)\}$$

Constuct the quotient space $V \otimes W$ to be

$$V \otimes W = \frac{L}{R}$$

Then $V \otimes W$ is the tensor product of V and W.

This gives the existence of tensor products. In practise no one explicitly finds out the elements in this way, and would simply use $v \otimes w$ to denote the tensor product.

Proposition 7.5.1.4

Let U, V, W be vector spaces. Then the following are true for tensor products.

- Commutativity: There is a unique isomorphism from $V \otimes W$ to $W \otimes V$
- Associativity: There is a unique isomorphism from $(U \otimes V) \otimes W$ to $U \otimes (V \otimes W)$
- $\dim(V \otimes W) = \dim(V) \cdot \dim(W)$.

7.5.2 Tensor Algebra

Definition 7.5.2.1: kth Tensor Power

Let V be a vector space over a field $\mathbb{F}$. Let $k \in \mathbb{N}$. Define the kth tensor power of V to be the tensor product

$$V^{\otimes k} = V \otimes V \cdots \otimes V$$

where the tensor product over V is taken k times.

By convention, define $V^{\otimes k}$ to be $\mathbb{F}$.

Definition 7.5.2.2: Tensor Algebra

Let V be a vector space over $\mathbb{F}$. Define the tensor algebra over V to be the direct sum

$$T(V) = \bigoplus_{k=0}^{\infty} V^{\otimes k}$$

Define multiplication in $T(V)$ to be the determined by the canonical isomorphism $V^{\otimes k} \otimes V^{\otimes l} \to V^{\otimes k+l}$, which is extended by linearity to all of $T(V)$.

Proposition 7.5.2.3

Let V be a vector space over $\mathbb{F}$. Then $T(V)$ is a graded algebra with the above defined multiplication rule.

Proposition 7.5.2.4: Universal Property

7.5.3 Exterior Algebra

Definition 7.5.3.1: Exterior Algebra

Let V be a vector space over $\mathbb{F}$. Let I be the ideal generated by all elements of the form $v \otimes v$ for $v \in V$. Define the exterior algebra of V to be the quotien

$$\Lambda = T(V)/I$$

Elements of the form $v_1 \otimes v_2$ are written as $v_1 \wedge v_2$ by convention.

Lemma 7.5.3.2

Let V be a vector space over $\mathbb{F}$ where the characteristic of $\mathbb{F}$ is not equal to 2. Then the ideals

$$\{v \otimes v | v \in V\} = \{v \otimes w + w \otimes v\}$$

are equal and thus they both form the exterior algebra over V.

Proposition 7.5.3.3

Let V be a vector space over $\mathbb{F}$. Then the following are true for $\Lambda(V)$.

- $\Lambda(V)$ is a graded algebra
- $\Lambda(V)$ is alternating
- $\Lambda(V)$ is anticommutative, meaning that $x \wedge y = -y \wedge x$

Definition 7.5.3.4: kth Graded Component

Let $\Lambda(V)$ be an exterior algebra. Define the kth graded component of $\Lambda(V)$ to be the graded component $\Lambda^k(V)$.

Proposition 7.5.3.5: Universal Property

Proposition 7.5.3.6

Let $\{e_1, \dots, e_n\}$ be a basis of the vector space V. Then

$$\{e_{i_1} \wedge \cdots \wedge e_{i_r} | 1 \leq i_1 < \cdots < i_r \leq n\}$$

is a basis of $\Lambda^r(V)$ and

$$\dim(\Lambda^r(V)) = \binom{n}{r}$$

Lemma 7.5.3.7

Let $\Lambda(V)$ be an exterior algebra. Then the following are true.

- $v \wedge v = 0$ for any $v \in V$
- $v \wedge w = -w \wedge v$ for any $v, w \in V$
- $v \wedge w = (-1)^{rs} w \wedge v$ for any $v \in \Lambda^r(V)$ and $w \in \Lambda^s(V)$

7.5.4 Symmetric Algebra

Definition 7.5.4.1: Symmetric Algebra

Let V be a vector space over $\mathbb{F}$. Let J be the ideal generated by all elements of the form $v \otimes w - w \otimes v$ for $v, w \in V$. Define the symmetric algebra of V to be the quotien

$$\text{Sym} = T(V)/J$$

Elements of the form $v_1 \otimes v_2$ are written as $v_1 v_2$ by convention.

Again here we are quotienting out symmetric objectsn so that we can treat them as the same thing.

Chapter 8

Point Set Topology

8.1 Topological Spaces

8.1.1 Basic Definitions

We begin with our main object of study.

Definition 8.1.1.1: Topological Space

A topological space is a pair of objects $(X, \mathcal{T})$ where X is a set and $\mathcal{T}$ is a collection of subsets of X satisfying

- $\emptyset, X \in \mathcal{T}$
- $\{G_i : i \in I\} \subseteq \mathcal{T} \implies \bigcup_{i \in I} G_i \in \mathcal{T}$
- $G_1, \dots, G_n \in \mathcal{T} \implies \bigcap_{k=1}^n G_k \in \mathcal{T}$

The collection $\mathcal{T}$ is called the topology on X and sets in $\mathcal{T}$ are called open sets.

The definition of a topological space requires only set-theoretic language. This allows for a wide range of potentialy spaces that we want to encapsulate with this structure since its definition simply relies on sets and not other additional structures. Later on we will meet a lot of fancy names for some more sets in a topological space, below is one of them.

Definition 8.1.1.2: Closed Sets

A subset A of a topological space X is said to be closed if $X \setminus A$ is open.

Beware that this definition means that sets which are both open and closed could exist. They are not binary and should not be treated as such. As taking complements of sets allows De Morgans' law to be applied with unions and intersection, we have the following so called alternative definition of a topological space.

Proposition 8.1.1.3

Let $(X, \mathcal{T})$ be a topological space.

- X and $\emptyset$ is closed
- If $U_1, \dots, U_n$ are closed then $\bigcup_{k=1}^n U_k$ is closed
- If $\{U_i : i \in I\}$ is closed for all i then $\bigcap_{i \in I} U_i$ is closed

Proof. Let $(X, \mathcal{T})$ be a topological space.

- $X, \emptyset$ are open thus $\emptyset, X$ are closed
- $X \setminus U_1, \dots, X \setminus U_n$ are open thus $\bigcap_{k=1}^n (X \setminus U_k) = X \setminus \bigcup_{k=1}^n U_k$ is open and $\bigcup_{k=1}^n U_k$ is closed
- $\{X \setminus U_i : i \in I\}$ are open sets thus $\bigcup_{i \in I}(X \setminus U_i) = X \setminus \bigcap_{i \in I} U_i$ is open and $\bigcap_{i \in I} U_i$ is closed

□

It is an alternative definition in the sense that by defining closed sets in a topological space, its open sets becomes apparent by taking complements. Just like how specifying open sets automatically creates the collection of all closed sets in that particular space.
And then we have the problem of the amount of open sets.

Definition 8.1.1.4: Refinements

Let X be a set. Let $\mathcal{T}_1, \mathcal{T}_2$ be topologies on X. We say that $\mathcal{T}_1$ refines $\mathcal{T}_2$ if $\mathcal{T}_1 \supset \mathcal{T}_2$

We could define two different topologies on the same set so long as it satisfies the axioms of a toplogical space. This is why we may have the notion of coarse and fine topologies relative to each other.

Definition 8.1.1.5: Basis

Let X be a set. A collection of sets $\mathcal{B} \subseteq \mathcal{P}(X)$ is called a basis on X if

- For every $x \in X$, there exists $B \in \mathcal{B}$ such that $x \in B$, meaning $\bigcup_{B \in \mathcal{B}} B = X$
- For every $B_1, B_2 \in \mathcal{B}$, for every $x \in B_1 \cap B_2$, there exists $B \in \mathcal{B}$ such that $x \in B \subseteq B_1 \cap B_2$

Similar to a vector space, we allow a smaller collection of open sets to "generate" the entire topology. This allows us to simply define the topology based on some smaller collection of open sets instead of having to write out every single open set in the topology. We prove this fact with the below proposition.

Proposition 8.1.1.6

Let X be a set and $\mathcal{B}$ be a basis on X. Then

$$\mathcal{T}_{\mathcal{B}} = \left\{ \bigcup_{X \in \mathcal{C}} X : \mathcal{C} \subseteq \mathcal{B} \right\}$$

is a topology of X generated by $\mathcal{B}$.

Proof. We prove that $\mathcal{T}_{\mathcal{B}}$ satisfies the three definitions of a topology. Note that $\emptyset \subseteq \mathcal{B}$ and $\bigcup_{X \in \emptyset} X = \emptyset$ thus $\emptyset \in \mathcal{T}_{\mathcal{B}}$. Also $\bigcup_{B \in \mathcal{B}} B = X$. Now let $\{V_i : i \in I\} \subseteq \mathcal{T}_{\mathcal{B}}$. $V_i \in \mathcal{T}_{\mathcal{B}}$ means that there exists some $C_i \in \mathcal{B}$ such that $\bigcup_{X \in C_i} X = V_i$. Now consider $\bigcup_{i \in I} \bigcup_{X \in \mathcal{C}_i} X$.

$$\bigcup_{i \in I} \bigcup_{X \in \mathcal{C}_i} X = \bigcup_{X \in \bigcup_{i \in I} \mathcal{C}_i} X$$

However, $\bigcup_{i \in I} \mathcal{C}_i \subseteq \mathcal{B}$ thus we are done. Finally, we prove that for $U, V \in \mathcal{T}_{\mathcal{B}}$, $U \cap V \in \mathcal{T}_{\mathcal{B}}$. Let

$$U = \bigcup_{X \in \mathcal{A}} X$$

and

$$V = \bigcup_{X \in \mathcal{C}} X$$

We have that

$$U \cap V = \bigcup \{A \cap C : A \in \mathcal{A}, C \in \mathcal{C}\}$$

Now consider $A \cap C$ for any $A \in \mathcal{A}$ and $C \in \mathcal{C}$. Let $x \in A \cap C$. Since $\mathcal{B}$ is a basis, there exists $B_x \in \mathcal{B}$ such that $x \in B_x \subset A \cap C$. Since $x \in A \cap C$ implies $x \in B_x$,

$$A \cap C \subset \bigcup_{x \in A \cap C} B_x$$

But also, $B_x \subset A \cap C$ thus

$$\bigcup_{x \in A \cap C} B_x \subset A \cap C$$

Thus

$$\bigcup_{x \in A \cap C} B_x = A \cap C$$

Since $B_x \in \mathcal{B}$ for all x, we have $A \cap C \in \mathcal{T}_\mathcal{B}$ and since we prove property 2 of definition of a topology, we have $U \cap V = \bigcup \{A \cap C : A \in \mathcal{A}, C \in \mathcal{C}\} \subset \mathcal{T}_\mathcal{B}$ □

Aside from the concept of a basis which generates a topology, sub basis also generates a topology.

Definition 8.1.1.7: Sub Basis

A sub basis for a topology $\mathcal{T}$ on T is a collection $\mathcal{B} \subset \mathcal{T}$ such that every set $\mathcal{T}$ is a union of finite intersections of sets from $\mathcal{B}$.

Comparing the two notions of basis in a topology, a basis generates the topology by creating unions of sets while a sub basis generates a topology using finite intersections. In practice both are useful in their respectful places. Often sub basis consists of less sets, but requires you to think about finite intersections which may or may not be harder depending on the situation.

Proposition 8.1.1.8

If $\mathcal{B}$ is any collection of subsets of a set X, and $\bigcup_{B \in \mathcal{B}} B = X$ then there exists a unique topology $\mathcal{T}$ on X with sub basis $\mathcal{B}$. In particular,

$$\mathcal{T} = \{\text{Union of finite intersections of sets of } \mathcal{B}\}$$

Proof. Let $\mathcal{B}$ be a sub basis of a topology $\mathcal{T}$. Then by definition, since every $U \in \mathcal{T}$ is a finite intersection of sets from $\mathcal{B}$, the collection of finite intersections precisely form a basis for $\mathcal{T}$. □

8.1.2 Closure, Interior and Boundary

We give some more names to some more random sets.

Definition 8.1.2.1: Neighbourhood

A neighbourhood of $x \in X$ a topological space is a set $U \subset X$ such that $x \in B \subset U$ for some $B \in \mathcal{T}$.

The neighbourhood acts like the open ball in metric spaces. Since we have no open balls in a topological space, this acts like the open ball for more potential structures such as limits. In fact the open balls are stricter than open neighbourhoodds in a metric space, but they are mostly interchangable.

Also, the B seems quite random in the definition, but is in fact neccessary. There are some topological spaces where sets like $\{x\}$ does not exists. To talk about the surrounding of x, we must somehow quantify it. Therefore we first use some set that contains x, then surround it with another set.

In the remainder of the section, three operations will be given to sets in a topological space. They each have unique properties which will soon prove itself useful.

Definition 8.1.2.2: Closure

Let $(X, \mathcal{T})$ be a topological space. Let $A \subset X$.Define the closure of A, denoted $\overline{A}$ to be the intersection of all closed subsets of X that contain A, meaning

$$\overline{A} = \bigcap_{\substack{A \subseteq U \subseteq X \\ U \text{ is closed}}} U$$

You can think of closure as the boundary of a set plus the set itself. It sort of completes the set by closing it. Inherently it implies that $\overline{A}$ is the smallest closed set containing A. Indeed if there were an even smaller closed set F between A and $\overline{A}$, then it should be contained in the intersection as well, which means that $\overline{A} \subset F$, a contradiction of the construction of F. We dedicate the next few theorems in establishing properties of closure.

Lemma 8.1.2.3

Let X be a topological space. Let $A \subset X$. Then $\overline{A}$ is closed.

Proof. $\overline{A}$ is necessarily closed since it is the intersection of closed subsets. □

Closed sets are special in the sense that limits of convergence sequences will be contained in these sets (in particularly nice spaces), as we will see later, by taking the closure of a set, we are implicitly containing all all limits of sequences into the original set. And as you ahve already seem, closed intervals in real analysis give rise to many interesting theorems such as Bolzano-Weierstrass theorem as well as a bunch of continuity theorems that actually depend on the fact that its domain is a closed set .

Proposition 8.1.2.4

Let $(X, \mathcal{T})$ be a topological space. Let $A \subset X$. A is closed if and only if $\overline{A} = A$.

Proof. First let A be closed. Then A is the smallest subset of X that contains A. Thus the intersection of all closed subsets larger than A is equal to A. Thus $\overline{A} = A$.

Suppose that $\overline{A} = A$. Then $\overline{A}$ is proved to be closed already. □

That gives us a neat condition to show that whether a set is closed instead of having to show that its complement is open. However, this privilege must wait until we define the boundary of a set.

Proposition 8.1.2.5

Let $(X, \mathcal{T})$ be a topological space. Let $A, B \subset X$. Then

- $\overline{\overline{A}} = \overline{A}$
- $A \subseteq \overline{A}$
- $A \subseteq B \implies \overline{A} \subseteq \overline{B}$
- $\overline{A \cup B} = \overline{A} \cup \overline{B}$

Proof. Let X be a topological space and $A \subset X$.

- Immediate from the above two proposition.

- The closure is defined by taking intersection of subsets U of X where $A \subseteq U$. Thus necessarily A is also a subset of the intersection of those subsets.
- Suppose that $A \subseteq B$. Then by definition any closed set containing B must contain A thus

$$\bigcap_{\substack{A\subseteq U\subseteq X\\ U \text{ is closed}}} U \subseteq \bigcap_{\substack{B\subseteq U\subseteq X\\ U \text{ is closed}}} U$$

  and we are done.
- We first prove that $\overline{A\cup B} \subseteq \overline{A}\cup\overline{B}$. Notice that $\overline{A}\cup\overline{B}$ is closed and contains A and B. By definition, $\overline{A\cup B}$ is the smallest closed set containing A and B thus any closed set containing A and B must contain $\overline{A\cup B}$.

  Now we prove the opposite inclusion. Notice that since $A, B \subseteq A\cup B$, we have that $\overline{A}, \overline{B} \subseteq \overline{A\cup B}$ thus $\overline{A}\cup\overline{B} \subseteq \overline{A\cup B}$ and we are done.

□

Notice that $\overline{A\cap B} = \overline{A}\cap\overline{B}$ is generally wrong. A typical counter example would be two unit sqaures not including boundaries centered at $(0.5, 0)$ and $(-0.5, 0)$ respectively. Since they don't include their boudaries their intersection is empty, and taking closure of the mptyset is the emptyset. But if you take closure before you intersect them, their boundaries will meet and taking intersewctions would give the line segment between $(0, 0.5)$ and $(0, -0.5)$.

We finish the topic of closure with the following characterization of closure.

Theorem 8.1.2.6

Let $(X, \mathcal{T})$ be a topological space, let $A \subset X$. Then

$$\overline{A} = \{x \in X |\ \forall\, U \in \mathcal{T} \text{ with } x \in U, U \cap A \neq \emptyset\}$$

Proof. Let $x \in \overline{A}$. Suppose for a contradiction that there exists an open set U containing x such that $U \cap A = \emptyset$. Then $X \setminus U$ does not contain x and is closed. Then $\overline{A} \subseteq X \setminus U$ which is a contradiction since x in $\overline{A}$ but $x \notin X \setminus U$.

Now suppose that $x \in X$ such that for all open sets U, $U \cap A \neq \emptyset$. Suppose for a contradiction that $x \notin \overline{A}$. Then there exists a closed set F such that $x \notin F$ and $A \subseteq F$. Then $X \setminus F$ is open and contains x but $X \setminus F \cap A = \emptyset$ thus this is a contradiction. □

If we can "complete" a set into its closed form by closure, we could also dig out the contents and leave out its boundary (if it has one) with interior.

Definition 8.1.2.7: Interior

Let A be a subset of a topological space $(X, \mathcal{T})$. We say that $x \in A$ is an interior point of A if there exists an open set U such that $x \in U \subset A$. Denote the set of all interior points of A by A°. Then

$$A^\circ = \{x \in A | \exists U \in \mathcal{T}, x \in U \subset A\}$$

We have the following equivalent characterization of interiors.

Proposition 8.1.2.8

Let $(X, \mathcal{T})$ be a topological space and $A \subset X$. Then A° is the union of all open sets of X contained in A. In other words,

$$A^\circ = \bigcup_{\substack{U \subset A \\ U \text{ is open}}} U$$

Proof. Let $x \in \bigcup_{\substack{U \subset A \\ U \text{ is open}}} U$. Then $x \in U$ for some open set $U \subset A$. Thus $\bigcup_{\substack{U \subset A \\ U \text{ is open}}} U \subseteq A^\circ$.

Let $x \in A^\circ$. Then there exists $U \in \mathcal{T}$ such that $x \in U \subset A$ with U open. Thus $x \in \bigcup_{\substack{U \subset A \\ U \text{ is open}}} U$ and we are done. □

Similar to the observation in closure, from this definition we see that A° is the largest open set contained in A because if there were an even larger set, then taking union of that set will imply that $A^\circ \subset A^\circ \cup U \subset A$ which means that $A^\circ \cup U$ should be contained in A°, a contradiction.

Proposition 8.1.2.9

Let $(X, \mathcal{T})$ be a topological space and $A \subset X$. Then A is open if and only if $A = A^\circ$.

Proof. Let A be open. Since A is open and A is the largest subset of A, by the equivalent characterization $A^\circ = A$ thus we are done.

Let $A^\circ = A$. Then A is the union of open subsets thus A is also open. □

The following lemma is already implicitly used above but for completeness I will state it here.

Lemma 8.1.2.10

Let $(X, \mathcal{T})$ be a topological space and $A \subset X$. Then A° is open.

Proof. A° is a union of open sets. □

If taking closure has the useful property that all its limits will be included (in metric spaces), we will soon see, and as I have already hinted, that open sets provide a notion of how close two points are in a topological space. Open sets also serve as the core part of a notion in topological spaces such as continuity and connecredness as we will see later. It is therefore often useful to consider interiors of a set if wanted to do a prove with open sets.

Proposition 8.1.2.11

Let $(X, \mathcal{T})$ be a topological space. Let $A, B \subset X$. Then

- $(A^\circ)^\circ = A^\circ$
- $A^\circ \subseteq A$
- $A \subseteq B \implies A^\circ \subseteq B^\circ$
- $(A \cap B)^\circ = A^\circ \cap B^\circ$

Proof. Let $A, B \subset X$.

- Immediate from the above two propositions

- Direct from the definition
- Every open set that A contains B must also contain. Thus $x \in A^\circ$ implies there exists U open such that $x \in U \subset A \subset B$ which implies $x \in B^\circ$.
- We first prove that $(A \cap B)^\circ \subseteq A^\circ \cap B^\circ$. Note that $A \cap B \subseteq A, B$, taking the interior gives $(A \cap B)^\circ \subseteq A^\circ, B^\circ$. Thus $(A \cap B)^\circ \subseteq A^\circ \cap B^\circ$.

  Now we prove the opposite inclusion. Notice that $A^\circ \cap B^\circ$ is open and contains $A \cap B$. Since $(A \cap B)^\circ$ is the largest open set contained in $A \cap B$. It should also contain $A^\circ \cap B^\circ$ thus we are done.

□

In general, $(A \cup B)^\circ \neq A^\circ \cup B^\circ$. Try and consider the two unit squares again but this time include their boundaries.

Finally, we relate the two operations by the following rule.

Proposition 8.1.2.12

Let $(X, \mathcal{T})$ be a topological space. Let $A \subset X$. Then

$$A^\circ = X \setminus \overline{(X \setminus A)} \text{ and } \overline{A} = X \setminus (X \setminus A)^\circ$$

Proof. Let $(X, \mathcal{T})$ be a topological space and $A \subset X$.

- We have that
$$X \setminus \bigcup_{\substack{U \subset A \\ U \text{ is open}}} U = \bigcap_{\substack{U \subset A \\ U \text{ is open}}} X \setminus U$$

□

Closures and interiors are like opposite operations. However from the above we can see that they are not opposite in the sense of set complements.

Definition 8.1.2.13: Boundary

Let $(X, \mathcal{T})$ be a topological space and $A \subset X$. x is a boundary point of A if for every neighbourhood U of x, $U \cap A \neq \emptyset$ and $U \cap X \setminus A \neq \emptyset$. Denote the set of all boundary points by ∂A.

If you are able to concretely graspe the meaning of interiors and closures, you will see that they are very suitably names, especially when considering sets in $\mathbb{R}^n$. Boundary is no exception. These proposition will seem very trivial once the image is well produced in the back of your head.

Proposition 8.1.2.14

Let $(X, \mathcal{T})$ be a topological space and $A \subseteq X$. Then

- $\partial A = \overline{A} \cap \overline{X \setminus A}$
- $\overline{A} = A \cup \partial A$
- $\partial A = \overline{A} \setminus A^\circ$

Proof.

- Let $x \in \partial A$. Then for every open neighbourhood U of x, $U \cap A \neq \emptyset$ implies that $x \in \overline{A}$ and $U \cap (X \setminus A) \neq \emptyset$ implies that $x \in \overline{X \setminus A}$.

  The above statements can all be reversed thus we are done.

- Let $x \in \overline{A}$. Then for every open neighbourhood U of x, $U \cap A \neq \emptyset$. This has two cases, either $U \subseteq A$ or $U \nsubseteq A$. For the first case, $x \in U \subseteq A$ then we are done. For the latter case assume that $x \notin U \cap A$. Then $x \in U \cap X \setminus A \neq \emptyset$. This implies that $U \in \partial A$ thus we are done.

  Now suppose that $x \in A \cup \partial A$. If $x \in A$, we are done. Suppose that $x \in \partial A$. Then for every open set U containing x, $U \cap A \neq \emptyset$. But this is the definition of $x \in \overline{A}$ thus we are done.

- Let $x \in \partial A$, by the above, we know that $\partial A \subseteq \overline{A}$. We need to show that $x \notin A^\circ$. Suppose for a contradiction that $x \in A^\circ$. Then there exists an open set U containing x such that $x \in U \subseteq A$. But this is a contradiction of the definition of boundary since we require that $U \cap X \setminus A \neq \emptyset$ for every open set U.

  Now suppose that $x \in \overline{A} \setminus A^\circ$. Then since $x \in \overline{A}$, every open set U containing x has the property that $U \cap A \neq \emptyset$. Also $x \notin A^\circ$ implies that if U is open and contains x, then $U \nsubseteq A$ which implies that $U \cap X \setminus A \neq \emptyset$. Then this is precisely the definition of boundary thus $x \in \partial A$ thus we are done.

□

We now give a name to a special type of points, where will see the reasonning later.

Definition 8.1.2.15: Limit Points

Let $(X, \mathcal{T})$ be a topological space and $A \subset X$. x is a limit point of A if every neighbourhood of x intersects with $A \setminus \{x\}$. Meaning if U is a neighbourhood of x, then

$$U \cap A \setminus \{x\} \neq \emptyset$$

A point in A that is not a limit point is called an isolated point.

Note that limit points does not necessarily lie inside the set itself.

Proposition 8.1.2.16

Let $(X, \mathcal{T})$ be a topological space and $A \subset X$. Then

$$\overline{A} = A \cup \{\text{Limit points of } A\}$$

The final type of sets are called dense sets, which accurately as its name indicates, it dictates the sparseness of a subset with respect to its topological space.

Definition 8.1.2.17: Dense Sets

Let $(X, \mathcal{T})$ be a topological space. Let $D \subseteq X$. D is said to be dense if $\overline{D} = X$

Lemma 8.1.2.18

Every set is dense in and of itself.

Proof. Let X be a toplogical space. Then X is trivially closed thus $\overline{X} = X$. □

Unfortunately this lemma does not give much insight as to why this is the case. A good example would be that $\mathbb{Q}$ is dense in $\mathbb{Q}$ and also dense in $\mathbb{R}$. The meaning of dense here is something we have already seen in real analysis. We already know that rational numbers are dense in the sense that there exists a countably amount of rational numbers between any two. It is also dense in $\mathbb{R}$ because there are an countable number of real numbers between any two rational numbers. (In fact there should be an infinite amount instead of countably infinite but we have yet to seen this result).

Another way to think about this is that since I have secretly told you that taking closure will include the limit points of the original set, we know that every irrational number can be approximated by a sequence of rational numbers. Thus every real number is a limit of some sequence in $\mathbb{Q}$, which gives the name dense its meaning.

Proposition 8.1.2.19

Let $(X,\mathcal{T})$ be a topological space. D is dense if and only if for all non-empty $U \subset X$, $D \cap U \neq \emptyset$.

Definition 8.1.2.20: Separable Space

A topological space X is said to be separable if it has a countably dense subset.

8.1.3 Continuity and Homeomorphism

We would like to classify different types of spaces by similar properties they hold. One way to do this is to define homeomorphisms betwene spaces. Similar to how isomorphism will preserve structure between groups, hoemoemorphism preserves strucutres called topological invariants which we will soon see what properties fall under this category.

We first define the notion of continuity, slightly different from that of $\mathbb{R}$.

Definition 8.1.3.1: Continuity

Let $(X,\mathcal{T})$ and $(Y,\mathcal{U})$ be topological spaces. Let $f : X \to Y$ be a function. We say that f is continuous if $f^{-1}(U) \in \mathcal{T}$ for every $U \in \mathcal{U}$.

Proposition 8.1.3.2

Let X be a set and $\mathcal{T}_1, \mathcal{T}_2$ be two topologies on X. Then the identity function $id : X \to X$ is continuous if and only if $\mathcal{T}_1 \supseteq \mathcal{T}_2$

Proposition 8.1.3.3

Let $(X,\mathcal{T})$ and $(Y,\mathcal{U})$ be topological spaces. Let $f : X \to Y$ be a function. Let $\mathcal{B}$ be a basis of Y and $\mathcal{S}$ a subbasis on Y both generating $\mathcal{U}$. The following are equivalent.

- f is continuous
- $f^{-1}(U) \in \mathcal{T}$ for all $U \in \mathcal{B}$
- $f^{-1}(U) \in \mathcal{T}$ for all $U \in \mathcal{S}$
- For every closed set $C \subseteq Y$, $f^{-1}(C)$ is closed in X.
- For every $A \subseteq X$, $f(\overline{A}) \subseteq \overline{f(A)}$

Proposition 8.1.3.4

Let $(X, \mathcal{T}_1)$ and $(X, \mathcal{T}_2)$ be topological spaces. Then the following are equivalent.

- $\mathcal{T}_2$ refines $\mathcal{T}_1$
- For any toplogical space $(Y, \mathcal{U})$, if $f : Y \to (X, \mathcal{T}_2)$ is continuous then $f : Y \to (X, \mathcal{T}_1)$ is also continuous
- For any topological space $(Y, \mathcal{U})$, if $f : (X, \mathcal{T}_1) \to Y$ is continuous then $f : (X, \mathcal{T}_2) \to Y$ is also continuous.

Proposition 8.1.3.5

Let $(X_1, \mathcal{T}_1), (X_2, \mathcal{T}_2), (X_3, \mathcal{T}_3)$ be a topological spaces and $f : X_1 \to X_2$ and $g : X_2 \to X_3$ be continuous then $g \circ f$ is also continuous.

Proposition 8.1.3.6

Let $f, g : X \to \mathbb{R}$ be continuous functions from a toplogical space X. Then $f + g$, fg are continuous. f/g is also continuous on the set $\{x \in X | g(x) \neq 0\}$.

Definition 8.1.3.7: Homeomorphism

Let $(X, \mathcal{T})$ and $(Y, \mathcal{U})$ be topological spaces. Let $f : X \to Y$ be a bijective function. f is said to be homeomorphic and write $(X, \mathcal{T}) \cong (Y, \mathcal{U})$ if f is continuous and f^{-1} is continuous.

Proposition 8.1.3.8

Let $(X, \mathcal{T})$ and $(Y, \mathcal{U})$ be topological spaces. Let $f : X \to Y$ be a bijective function. The following are equivalent.

- f is a homeomorphism
- $U \subset X$ is open if and only if $f(U) \subset Y$ is open

Proof. Suppose that f is a homeomorphism. Then $U \subset X$ being open implies $f(U) \subset Y$ being open is clear by continuity of f^{-1}. $f(U) \subset Y$ being open implies $U \subset X$ being open is also clear from the continuity of f.

The only if part of the statement is also clear from the definition of continuity. □

Definition 8.1.3.9: Topological Invariants

A property ϕ of topological spaces is called topological invariant if whenever $(X, \mathcal{T})$ and $(Y, \mathcal{U})$ are homeomorphic topological space, one has property ϕ if and only if the other has it.

The following are some properties that we will see later which are indeed topological invariants: T_0, T_1, T_2 spaces and compactness and connectedness. These theorem will come up in the form of: Continuity preserves ϕ. This means that ϕ is a topological invariant.

8.2 New Topologies from Old

Similar to what we did with groups and subgroups as well as products of groups, we also have similar notion where we define subspaces, product spaces, and more. Naturally they will inherit some properties of the original topological space. We give special names as we will see later for properties that can be inherited by subspaces and product spaces respectively.

8.2.1 Subspace Topologies

We begin by considering subsets of a topological space.

Definition 8.2.1.1: Subspace Topology

Let $(X, \mathcal{T})$ be a topological space. Let $Y \subset X$. Define the subspace topology $\mathcal{T}_Y$ on Y by

$$\mathcal{T}_Y = \{U \cap Y | U \in \mathcal{T}\}$$

Proposition 8.2.1.2

In the above definition $\mathcal{T}_Y$ is a topology on Y.

Proof. Trivially, $\emptyset$ and Y are in $\mathcal{T}_Y$.

Now let $\{A_i | i \in I\} \subseteq \mathcal{T}_Y$. Then

$$\bigcup_{i\in I} A_i = \bigcup_{i\in I}(U_i \cap Y) = \left(\bigcup_{i\in I} U_i\right) \cap Y$$

and since X is a topological space, $V = \bigcup_{i\in I} U_i \in \mathcal{T}$ and thus $V \cap Y \in \mathcal{T}_Y$ and we have proved the second property.

Let $A, B \in \mathcal{T}_Y$. Then

$$A \cap B = (U_A \cap Y) \cap (U_B \cap Y) = (U_A \cap U_B) \cap Y \in \mathcal{T}_Y$$

thus we are done. □

We have seen that open sets are inherited from the original topological space. In fact, basis elements can also be inherited from the parent space.

Proposition 8.2.1.3

Let $(X, \mathcal{T})$ be a topological space and let $\mathcal{B}$ be a basis on X that generates $\mathcal{T}$. Let $Y \subset X$. The collection

$$\mathcal{B}_Y = \{B \cap Y | B \in \mathcal{B}\}$$

is a basis on Y that generates the subspace topology $\mathcal{T}_Y$ on Y.

Proof. Let $y \in Y$. Then there exists $B \in \mathcal{B}$ such that $y \in B$. Then $B \cap Y \in \mathcal{B}_Y$ and $y \in B \cap Y$ and the first property is proven.

Let $B_1, B_2 \in \mathcal{B}_Y$. Then there exists $A_1, A_2 \in \mathcal{B}$ such that $B_1 = A_1 \cap Y$ and $B_2 = A_2 \cap Y$. Let $y \in B_1 \cap B_2 = A_1 \cap A_2 \cap Y$. By definition of a basis there exists $A \in \mathcal{B}$ such that $y \in A \subset A_1 \cap A_2$. Then $y \in A \cap Y \subset B_1 \cap B_2$ and $A \cap Y \in \mathcal{B}$. Thus we are done. □

Proposition 8.2.1.4

Let $(X, \mathcal{T})$ be a topological space and let Y be a subspace of X. If U is an open subset of Y and Y is an open subset of X, then U is an open subset of X.

Proposition 8.2.1.5

Let $(X, \mathcal{T})$ be a topological space and let Y be a subspace of X. Let A be a subset of Y. The subspace topology inherits from Y is equal to the subspace topology inherits from X.

Proposition 8.2.1.6

Let $(X, \mathcal{T})$ be a topological space and let Y be a subspace of X. For any $Z \subset Y$, $\overline{Z}_Y = Y \cap \overline{Z}_X$, where $\overline{Z}_Y$ is the closure of Z in Y.

Proposition 8.2.1.7

Let $(X, \mathcal{T})$ be a topological space and let Y be a subspace of X. The inclusion map $i : Y \to X$ given by $i(x) = x$ is continuous.

Proof. Trivial by proposition 2.1.4. □

Proposition 8.2.1.8

Let $f : X \to Y$ be a continuous function and $A \subset X$ is a subspace. Then $f|_A : A \to Y$ is continuous.

Proof. Trivial by considering $f|_A : A \to f(A)$. □

Lemma 8.2.1.9: Pasting Lemma

Let $(X, \mathcal{T})$ and $(Y, \mathcal{U})$ be topological spaces. Let $A, B \subset X$, either both open or both closed, such that $X = A \cup B$ and both are subspaces of X. Let $f : A \to Y$ and $g : B \to Y$ be continuous functions that agree on $A \cap B$. Define $h : X \to Y$ by

$$h(x) = \begin{cases} f(x) & \text{if } x \in A \\ g(x) & \text{if } x \in B \end{cases}$$

Then h is continuous.

Proof. Problem arises when an open set that is part of both A and B is considered. Otherwise they are continuous in its own way. □

Definition 8.2.1.10: Hereditary Properties

A topological property ϕ is hereditary if every subspace of a topological space with ϕ has it.

Proposition 8.2.1.11

The following topological properties are hereditary.

- T_0, T_1, T_2
- Countable

- First Countable
- Second Countable

8.2.2 Product Topology

Definition 8.2.2.1: Product Topology

Let $(X,\mathcal{T})$ and $(Y,\mathcal{U})$ be topological spaces. The product topology on $X \times Y$ is the topology generated by the basis

$$\mathcal{B}_{X,Y} = \{U \times V : U \in \mathcal{T}, V \in \mathcal{U}\}$$

Lemma 8.2.2.2

$\mathcal{B}_{X,Y}$ is indeed a topology on $X \times Y$.

Proposition 8.2.2.3

Let $(X,\mathcal{T})$ and $(Y,\mathcal{U})$ be topological spaces. Let $\mathcal{B}_X$ and $\mathcal{B}_Y$ be bases on X and Y that generate $\mathcal{T}$ and $\mathcal{U}$, respectively. Then

$$\mathcal{B} = \{U \times V | U \in \mathcal{B}_X, V \in \mathcal{B}_Y\}$$

is a basis for the product topology on $X \times Y$.

Proposition 8.2.2.4

Let $(X,\mathcal{T})$ and $(Y,\mathcal{U})$ be topological spaces. Let $x \in X$. Define a map $f : Y \to X \times Y$ by $f(y) = (x,y)$ for all $y \in Y$. Then f is a homeomorphism.

Proof. The subspace $\{x\} \cup Y$ has open sets of the form $\emptyset \times U$ or $\{x\} \times U$ for $U \in \mathcal{U}$. Then f is continuous since $f^{-1}(\{x\} \cup U) = U$ and $f^{-1}(\emptyset \cup U) = \emptyset$. f is bijective since they have the same cardinality. The inverse map is also continuous since $f(U) = \{x\} \times U$ is open for open sets U. □

Definition 8.2.2.5: Productive Property

A property ϕ of topological spaces is said to be finitely productive if every finite product of topological spaces with ϕ has ϕ. It is said to be productive if every countable product of topological spaces with ϕ has ϕ.

Proposition 8.2.2.6

The following properties are finitely productive.

- T_0, T_1, T_2
- Finite
- Countable
- Separable
- First Countable
- Second Countable

Definition 8.2.2.7: Projection Maps

Let $(X_1, \mathcal{T}_1), \dots, (X_n, \mathcal{T}_n)$ be topological spaces. Define the projection maps

$$\pi_k : \prod_{i=1}^{n} X_i \to X_k$$

for $k \in \{1, \dots, n\}$ by

$$\pi_k(x_1, \dots, x_n) = x_k$$

Proposition 8.2.2.8

Let $A \subset X$ and $B \subset Y$. Then $\pi_1^{-1}(A) = A \times Y$ and $\pi_2^{-1}(B) = X \times B$. Moreover,

$$A \times B = \pi_1^{-1}(A) \cap \pi_2^{-1}(B)$$

Proposition 8.2.2.9

Let $(X, \mathcal{T})$ and $(Y, \mathcal{U})$ be topological spaces. Then the product topology on $X \times Y$ is the coarest topology on $X \times Y$ such that the projections π_1 and π_2 are continuous.

Proposition 8.2.2.10

Let $(X, \mathcal{T})$ and $(Y, \mathcal{U})$ be topological spaces. Then the set

$$\mathcal{S} = \{\pi_1^{-1}(U) | U \in \mathcal{T}\} \cup \{\pi_2^{-1}(V) | V \in \mathcal{U}\}$$

is a subbasis that generates the product topology on $X \times Y$.

Proposition 8.2.2.11

Let $(X, \mathcal{T})$ and $(Y_1, \mathcal{U}_1), (Y_2, \mathcal{U}_2)$ be topological spaces. Let $f : X \to Y_1 \times Y_2$ be a function. Then f is continuous if and only if $\pi_1 \circ f$ and $\pi_2 \circ f$ are continuous.

8.2.3 Quotient Topology

Definition 8.2.3.1: Quotient Topology

Let $\sim$ be an equivalence relation on a topological space X. Define the quotient topology on $X/\sim$ where open sets are sets of the form $U \subset X/\sim$ such that $q^{-1}(U) = \{x \in X | q(x) \in U\}$ is open. The map q is the quotient map $q : X \to X/\sim$ is the naturall induced map $q(x) = [x]$.

Proposition 8.2.3.2

The quotient topology for a topological space X is indeed a topology for $X/\sim$. Moreover, the map $q : X \to X/\sim$ is a continuous map.

8.3 Separation Axioms

8.3.1 Convergence and T_0, T_1, T_2 Spaces

Definition 8.3.1.1: Convergence

Let $(X,\mathcal{T})$ be a topological space. A sequence $\{x_n\}_{n=1}^\infty$ is said to converge to a point $x \in X$ if for every open set U containing x, there is an $N \in \mathbb{N}$ such that $x_n \in U$ for all $n > N$. In this case we write

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

Proposition 8.3.1.2

Let $(X,\mathcal{T})$ be a topological space. Let $A \subseteq X$. Let $\{a_n\} \subset A$ be a sequence. If $a_n \to a$ then $a \in \overline{A}$.

Definition 8.3.1.3: Kolmogorov T_0 Space

Let $(X,\mathcal{T})$ be a topological space. It is said to be T_0 if for any pair of distinct points $x, y \in X$, there exists an open set U that contains one of them and not the other.

Unfortunately, sequences in T_0 spaces are way too wild for us to perform analysis. They can virtually converge to any other point in the space. We need stronger separation axioms so that points in the space can be more distinctly presented. Even constant sequences can converge to different values!

Definition 8.3.1.4: Frechet T_1 Space

Let $(X,\mathcal{T})$ be a topological space. It is said to be T_1 if there exists open sets U, V such that U contains x but not y and V contains y but not x.

This is better. Constant sequences now have unique convergent values. But still, the same cannot be said to general sequences. We need yet another stronger separation axiom. But before we look at T_2, we have some equivalent characterizations of T_1 spaces.

Proposition 8.3.1.5

Let $(X,\mathcal{T})$ be a topological space. The following are equivalent.

- $(X,\mathcal{T})$ is T_1
- For every $x \in X$, $\{x\}$ is closed
- Every finite subset of X is closed.
- For every subset $A \subset X$, $A = \bigcap\{U \subset X : U \text{ is open and } A \subset U\}$

Now comes the main space of study.

Definition 8.3.1.6: Hausdorff T_2 Space

Let $(X,\mathcal{T})$ be a topological space. It is said to be T_2 if for every pair of distinct points $x, y \in X$, there exists open sets U, V such that $U \cap V = \emptyset$ and $x \in U$ and $y \in V$

Most analysis and well behaved spaces lie under this category. The reason is given by the following theorem.

Theorem 8.3.1.7

Let $(X, \mathcal{T})$ be a Hausdorff space. Then every sequence in X converges to at most one point.

Proof. Let $x_n \to x$ and $x_n \to y$ with $x \neq y$. Since X is Hausdorff there exists open sets U, V such that $x \in U$, $y \in V$ and $U \cap V = \emptyset$. By definition of convergence, there exists $N_1 \geq 1$ and $N_2 \geq 1$ such that $x_n \in U$ for all $n \geq N_1$ and $x_n \in V$ for all $n \geq n_2$. This is a contradiction since $x \in U \cup V$ when $n \geq \max(N_1, N_2)$ but $U \cap V = \emptyset$. □

Finally, sequences can converge uniquely. If you recall, notions such as continuity in real analysis are based on sequecnes. Therefore we really need sequences to be well behaved so that our foundations is solid. In fact, most of the more fun and useful spaces are Hausdorff. Therefore we paid less attention to T_0 and T_1 spaces. However, in order for sequences to capture key information of spaces, we need a bit more restrictions.

8.3.2 First Countability and Hausdorff

Definition 8.3.2.1: Local Basis

Let $(X, \mathcal{T})$ be a topological space. Let $x \in X$. A local basis at x is a collection of open sets $\mathcal{B}_x \subseteq \mathcal{T}$ following

- $x \in B$ for all $B \in \mathcal{B}_x$
- For any open set U containing x, there exists $B \in \mathcal{B}_x$ such that $B \subseteq U$

Definition 8.3.2.2: First Countable

Let $(X, \mathcal{T})$ be a topological space. It is first countable if every point in X has a countable local basis

Proposition 8.3.2.3

Let $(X, \mathcal{T})$ be a first countable topological space. Then for all $x \in X$, there exists a local basis $\mathcal{B}_x = \{B_n : n \in \mathbb{N}\}$ such that $B_1 \supseteq B_2 \supseteq B_3 \supseteq \cdots$

With the above proposition, we can reverse the previous proposition stating that convergent values lie inside the closure.

Proposition 8.3.2.4

Let $(X, \mathcal{T})$ be a first countable topological space. Let $A \subseteq X$. Then $x \in \overline{A}$ if and only if there is a sequence of elements of A converging to x.

The following is a partial converse to the unique limit point proposition that involves first countability.

Proposition 8.3.2.5

Let $(X, \mathcal{T})$ be a first countable topological space. If every convergent sequence has a unique limit point, then it is Hausdorff.

8.3.3 T_3 Spaces and Regularity

Definition 8.3.3.1: Regular Space

A topological space $(X, \mathcal{T})$ is said to be regular if for any $x \in X$ and any closed set C not containing x, there are disjoint open sets U, V such that $x \in U$ and $C \subset V$

Definition 8.3.3.2: T_3 Space

A topological space is said to be T_3 if it is T_1 and regular.

Proposition 8.3.3.3

Let $(X, \mathcal{T})$ be regular. Then $(X, \mathcal{T})$ is T_0 if and only if it is T_1 if and only if it is T_2.

Proposition 8.3.3.4

A topological space $(X, \mathcal{T})$ is regular if and only if for every point $x \in X$ and every open set U containing x, there is an open set V such that $x \in V \subseteq \overline{V} \subseteq U$

Proposition 8.3.3.5

Regularity and T_3 are topological invariants.

Proposition 8.3.3.6

Regularity and T_3 are hereditary.

Proposition 8.3.3.7

Regularity and T_3 are finitely productive.

Proposition 8.3.3.8

Let $(X, \mathcal{T})$ be a topological space that has a basis of clopen sets. Then $(X, \mathcal{T})$ is regular.

8.3.4 T_4 Spaces and Normality

Definition 8.3.4.1: Normal Spaces

A topological space $(X, \mathcal{T})$ is said to be normal if for any two disjoint, non-empty, closed subsets $C, D \subseteq X$, there are disjoint open sets U and V containing C and D respectively.

Definition 8.3.4.2: T_4 Spaces

A topological space $(X, \mathcal{T})$ is said to be T_4 if it is T_1 and normal.

Proposition 8.3.4.3

A topological space $(X, \mathcal{T})$ is normal if and only if for every open set U and every closed set $C \subset U$, there exists an open set V such that $C \subseteq V \subseteq \overline{V} \subseteq U$

Proposition 8.3.4.4

Normality and T_4 are topological invariants.

Proposition 8.3.4.5

Normality is not hereditary.

Proposition 8.3.4.6

Every closed subspace of a normal space is normal.

Proposition 8.3.4.7

Normality is not finitely productive.

Theorem 8.3.4.8

Every regular, second countable topological space is normal.

8.4 Metric Spaces

Metric spaces are more closely related to analysis in both its proofs and possible question types when compared to a more set theoretic approach for topology. However topology provides a more general context than metric spaces to discuss properties such as compactness and connectedness. Therefore I have decide to include metric spaces into a set of notes for point set topology. However be aware that metric spaces require a more analytical approach.

One should also be clear on what properties are unique to metric spaces. This is often reflected in the proofs. If the techniques of the proofs make use of sets rather than closeness of points for instance, then it may be a more topological property. However Hausdorff and metric spaces are also closely related which we will see below.

8.4.1 Basic Definitions

We begin with the definition of a metric.

Definition 8.4.1.1: Metric

Let X be a set. Let $x, y, z \in X$. A metric is a function $d : X \times X \to \mathbb{R}$ satisfying the following.

- $d(x, y) \geq 0$ with equality if and only if $x = y$
- $d(x, y) = d(y, x)$
- $d(x, y) \leq d(x, z) + d(z, y)$

Definition 8.4.1.2: Metric Space

A metric space is an oredered pair (X, d) where X is a set and d is a metric on X.

In topology, distance is loosely defined by separation axioms. Loosely speaking, the more separable they are through separation axioms, the more well defined distance is.

Definition 8.4.1.3: Open Balls

Let X be a metric space. The open ball centered at $p \in X$ of radius r is the set

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

As we will soon see that these open balls form a basis for a topology on X, one shoudl see that this is precisely why we call open sets of a topological space to be open sets. However this does not necesssarily mean that open sets in any topology on the same underlying set X coincide. We are very used to thinking that open sets in $\mathbb{R}$ are union of intervals. But this may not true in some other topology such as the Zariski topology, where the closed sets in the standard topology forms precisely the open sets of the Zariski topology.

Proposition 8.4.1.4

Let (X, d) be a metric space. The collection

$$\mathcal{B}_d = \{B_r(p) | p \in X, r > 0\}$$

is a basis on X.

Definition 8.4.1.5: Metric Topology

Let (X, d) be a metric space. The topology generated by the basis $\mathcal{B}_d$ on X is called the metric topology on X.

This means that every metric space is in fact a topological space with topology given by the metric topology. In fact, we can say more about the metric topology.

Proposition 8.4.1.6

Every metric space is Hausdorff.

Readers should stop and start to think about the notions we previously defined on topological spaces such as open sets and interiors, and think about what they mean in this more specific context.
The following notion is unique to metric spaces which is helpful in characterizing compactness in metric spaces.

Definition 8.4.1.7: Bounded Set

Let U be a subset of a metric spaces (X,d). We say that U is bounded if there exists $a \in X$ and $r>0$ such that

$$U \subseteq B_r(a)$$

8.4.2 New Metric Spaces from Old

Lemma 8.4.2.1: Metric Subspace

Let (X,d) be a metric space. Let $A \subseteq X$, then $(A, d|_A)$ is also a metric space.

Proof. $d|_A$ inherits the metric properties of X while being restricted to A. □

Proposition 8.4.2.2: Metric Space Product

Let (X_1,d_1) and (X_2,d_2) be metric spaces. Let $x_1,y_1 \in X_1$ and $x_2,y_2 \in X_2$. Then for $1 \leq p < \infty$,

$$d_p((x_1,x_2),(y_1,y_2)) = (d_1(x_1,y_1)^p + d_2(x_2,y_2)^p)^{1/p}$$

defines a metric on $X_1 \times X_2$.

Proof. We prove the triangle inequality here, the others are easy. We have

$$\begin{aligned} d_p((x_1,x_2),(y_1,y_2))^p &= d_1(x_1,y_1)^p + d_2(x_2,y_2)^p \\ &\leq (d_1(x_1,z_1) + d(z_1,y_1))^p + (d_2(x_2,z_2) + d_2(z_2,y_2))^p \end{aligned}$$

□

8.4.3 Continuity in Metric Spaces

Proposition 8.4.3.1

Let $f : X \to Y$ be a funciton between metric spaces. The following are equivalent.

- f is continuous at $p \in X$
- For every sequence x_n such that $x_n \to p$, we have $f(x_n) \to f(p)$
- For every $\epsilon > 0$, there exists $\delta > 0$ such that

  $$x \in B_\delta(p) \implies f(x) \in B_\epsilon(f(p))$$

  Or equivalently, $f(B_\delta(p)) \subset B_\epsilon(f(p))$.

The notion of continuity follows from the general definition given for topological spaces. However we do have one special theorem concerning normed spaces.

Proposition 8.4.3.2

If $(Y, \|\cdot\|)$ is a normed vector space and $f, g : X \to Y$ are continuous then $f + g$ is continuous.

Definition 8.4.3.3: Lipschitz Continuity

A function $f : X \to Y$ is Lipschitz continuous if there exists $C \geq 0$ such that

$$d_Y(f(x), f(y)) \leq Cd_X(x, y)$$

for all $x, y \in X$. In this case, C is the Lipschitz constant.

Lemma 8.4.3.4

Lipschitz continuity implies continuity.

Proof. Set $\delta = \frac{\epsilon}{C}$ in the $\epsilon - \delta$ definition of continuity. □

8.4.4 Equivalent Metrics

We know investigate when will different metrics induce the same topology. The answer is reasonably straight forward considering we have the notion of homeomorphism at play.

Theorem 8.4.4.1

Let d_1, d_2 be two metrics on X. Then the following statements are equivalent.

- The open sets in (X, d_1) and (X, d_2) coincide
- For any metric space (Y, d_Y), a function $g : X \to Y$ is continuous from (X, d_1) to (Y, d_Y) if and only if g is continuous from (X, d_2) to (X, d_1)
- For any metric sapce (Y, d_Y), a function $f : Y \to X$ is continuous from (Y, d_Y) to (X, d_1) if and only if f is continuous from (Y, d_Y) to (X, d_2)

Definition 8.4.4.2: Topologically Equivalent Metrics

Two metrics d_1, d_2 on X are said to be topologically equivalent if the above statements are true.

Lemma 8.4.4.3

Topologically equivalent metrics induce the same metric topology and thus are the same topological space.

Definition 8.4.4.4: Lipschitz Equivalent Metrics

Two metrics d_1, d_2 on X are said to be Lipschitz equivalent if there exists $0 < c_1 \leq c_2 < \infty$ such that

$$c_1 d_1(x, y) \leq d_2(x, y) \leq c_2 d_1(x, y)$$

for all $x, y \in X$.

Lemma 8.4.4.5

Lipschitz equivalence implies topologically equivalence on metrics.

Recall the definition of norms in linear algebra.

Definition 8.4.4.6: Equivalent Norms

Two norms $\|\cdot\|_1$ and $\|\cdot\|_2$ for a vector space V over a field $F = \mathbb{R}$ or $\mathbb{C}$ are said to be equivalent if there exists $c_1, c_2 \in F$ such that for every $x \in V$,

$$c_1\|x\|_1 \leq \|x\|_2 \leq c_2\|x\|_1$$

Proposition 8.4.4.7

The equivalence on norms is an equivalent relation.

Proposition 8.4.4.8

Suppose that two norms are equivalent on a normed vector space, then they induce topologically equivalent metrics.

Proof. Suppose that $\|\cdot\|_1$ and $\|\cdot\|_2$ are equivalent. Then define their corresponding metrics by $d_1(x,y) = \|x-y\|_1$ and $d_2(x,y) = \|x-y\|_2$ for x, y in a normed vector space X. We show that the open sets coincide.

Suppose that $U \subseteq (X, d_1)$ is open. Then for every $x \in U$, there exists $r > 0$ such that $B_r(x) \subset U$. From the equivalent norms, we have that there exists c such that $\|x-y\|_2 \leq c\|x-y\|_1$ and thus

$$\left\{x \in X \middle| \|x-y\|_2 < \frac{r}{c}\right\} \subseteq \{x \in X | \|x-y\|_1 < r\}$$

Thus $B_{\frac{r}{c}}(x)$ in the d_2 metric is a subset of $B_r(x)$ in the d_1 metric. This means that we have constructed an open ball in (X, d_2) so that it is contained in U. Thus U is also open in (X, d_2).

Mirror this to show that the open sets of (X, d_2) must also be open sets of (X, d_1) using the fact that there exists c such that $\|x-y\|_1 \leq c\|x-y\|_2$ and we are done. □

Lemma 8.4.4.9

If X is a vector space and two norms induce topologically equivalent metrics, then the norms are equivalent.

8.4.5 Metrizability

A common question to ask is that whether every topology can be generated from a metric.

Definition 8.4.5.1: Metrizable Space

A topological space $(X, \mathcal{T})$ is said to be metrizable if there is a metric d on X that generates $\mathcal{T}$.

Proposition 8.4.5.2

The following are true about metrizability.

- Metrizability is a topological invariant.

- Metrizability is finitely productive.
- Metrizability is hereditary.

Proposition 8.4.5.3

The following are true about metric spaces.

- Every metric space is T_2, T_3 and T_4.
- Every metric space is first countable.

Theorem 8.4.5.4: Urysohn's Lemma

A topological space $(X, \mathcal{T})$ is normal if and only if for every pair of disjoint non-empty closed subsets $C, D \subseteq X$ there exists a continuous function $f : X \to [0,1]$ such that $f(x) = 0$ for all $x \in C$ and $f(x) = 1$ for all $x \in D$.

Theorem 8.4.5.5: Urysohn Metrization Theorem

Every second countable T_3 topological space is metrizable.

8.5 Compactness

8.5.1 Basic Definitions

Compactness is an important topological property that allows us to build contructive proofs. This stems from the fact that any size of open covers can be reduced to a finite number given the space is compact. One can think of compact spaces as being some sort of finite space, but instead of having a finite number of elements, it has a fintie number open sets that contain the entirety of the space. With finiteness in play we can do a lot more things such as extremums.

Definition 8.5.1.1: Open Covers

Let $(X, \mathcal{T})$ be a topological space, and let $\mathcal{U} \subseteq \mathcal{T}$ be a collection of open subsets of X. We say $\mathcal{U}$ is an open cover of X if $X = \bigcup_{U \in \mathcal{U}} U$

Definition 8.5.1.2: Subcovers

Let $(X, \mathcal{T})$ be a topological space. Let $\mathcal{U}$ be an open cover of X. If $\mathcal{V} \subseteq \mathcal{U}$ is an open cover of X, then $\mathcal{V}$ is a subcover of $\mathcal{U}$.

Definition 8.5.1.3: Compact Space

A topological space $(X, \mathcal{T})$ is said to be compact if every open cover of X has a finite subcover.

Below is a closed set characterization of compactness.

Theorem 8.5.1.4: Finite Intersection Property

Let $\mathcal{F}$ be a collection of non-empty closed subsets of a space X such that every finite subcollection of $\mathcal{F}$ has a non-empty intersection (finite intersection property). Then X is compact if and only if the intersection of all sets from $\mathcal{F}$ is non-empty.

Proof. Suppose that X is compact. Suppose for a contradiction that $\bigcap_{F \in \mathcal{F}} F = \emptyset$. Then

$$X \setminus \bigcap_{F \in \mathcal{F}} = \bigcup_{F \in \mathcal{F}} X \setminus F = X$$

which means that $\mathcal{F}$ is an open cover of X. By compactness, there exists a finite subcover of $\mathcal{F}$, namely $F_1, \dots, F_n$ such that $\bigcup_{k=1}^n F_k = X$. But then

$$\bigcup_{k=1}^n F_k = X$$
$$X \setminus \bigcup_{k=1}^n F_k = \emptyset$$
$$\bigcap_{k=1}^n X \setminus F_k = \emptyset$$

which is a contradiction of the finite intersection property.

Now suppose that every finite subcollection of $\mathcal{F}$ has a non empty intersection and that $\bigcap_{F \in \mathcal{F}} F \neq \emptyset$. Suppose for a contradiction that X is not compact. Let $\{U_\alpha | \alpha \in I\}$ be an open cover of X. Since X is not compact, every finite union of the form $\bigcup_{k=1}^n U_{\alpha_k} \neq X$ for $\alpha_1, \dots, \alpha_n \in I$ which implies that

$$\bigcap_{k=1}^n U_{\alpha_k} \neq \emptyset$$

This means that $\{X \setminus U_\alpha | \alpha \in I\}$ has the finite interseciton property and thus

$$\bigcap_{\alpha \in I} X \setminus U_\alpha \neq \emptyset$$

But then taking complements of X means that $\bigcup_{\alpha \in I} U_\alpha$ does not cover X which contradicts our assumption. □

8.5.2 Compactness and Closed and Continuous

Proposition 8.5.2.1

Let $f : X \to Y$ be continuous. If $K \subset X$ is compact, then $f(K)$ is compact.

Proof. Let K be compact. Let $\mathcal{U}$ be an open cover of $f(K)$. Since f is continuous, for all $U \in \mathcal{U}$, $f^{-1}(U)$ are open and forms a cover of K. Since K is compact, there exists a finite subcover of K, namely $f^{-1}(U_1), \dots, f^{-1}(U_n)$.

Let $y \in f(K)$. Then $y = f(x)$ for some $x \in K$. But $x \in f^{-1}(U_k)$ for some k. Thus $y \in U_k$. Thus $U_1, \dots, U_n$ are a finite subcover of $f(K)$. □

Lemma 8.5.2.2

Compactness is a topological invariant but is not hereditary.

Proof. From the above, if X, Y is homeomorphic then $Y = f(X)$ and we are done.

Considering $(0, 1) \in \mathbb{R}$ as a metric subspace of $[0, 1] \in \mathbb{R}$. □

Proposition 8.5.2.3

Let $(X, \mathcal{T})$ be a compact topological space. Let $C \subseteq X$ be a closed subset. Then C is compact.

Proof. Let U be a cover of C by open subsets of X. Then $U \cup X \setminus C$ is an open cover of X, thus has a finite subcover. This provides an open subcover of C since $X \setminus C$ is open and you can remove this element from the subcover. □

Proposition 8.5.2.4

Let $(X, \mathcal{T})$ be a compact Hausdorff space. Then $(X, \mathcal{T})$ is regular and normal.

Proposition 8.5.2.5

Let $(X, \mathcal{T})$ be a Hausdorff topological space. Let $K \subseteq X$ be compact. Then K is closed.

Proof. Let $a \in X \setminus K$. For each $x \in K$, there exists disjoint open sets U_x containing x and V_x containing a due to Hausdorff. Then $\{U_x | x \in K\}$ form an open cover of K, and thus has a finite subcover $U_{x_1}, \dots, U_{x_n}$ of K. Then

$$V = \bigcap_{k=1}^{n} V_{x_k}$$

is open and contains a and is disjoint from K. This means that $X \setminus K$ is open and thus K is closed. □

The point of the finite subcover in the above proof is to make sure that since the U_k are covering K, V_k being disjoint from U_k means that V_k will not overlap with K.

Proposition 8.5.2.6

Let $(X, \mathcal{T})$ be a compact topological space. Let $(Y, \mathcal{U})$ be a Hausdorff topological space. Then any continuous bijection $f : X \to Y$ is a homeomorphism.

Proof. We want to show that f^{-1} is continuous. Let $U \subset X$ be closed. Then U is compact by proposition 5.2.3. By 5.2.1, $f(U)$ is compact. Since Y is Hausdorff, $f(U)$ is closed by the above proposition. By the closed set characterization of continuous functions, f^{-1} is continuous. □

8.5.3 Productivity of Compactness

Proposition 8.5.3.1

Compactness is finitely productive.

Proof. We show that $(X, \mathcal{T})$ and $(Y, \mathcal{S})$ are compact then $X \times Y$ is compact. Consider the product topology on $X \times Y$ given by

$$\mathcal{B} = \{U \times V | U \in \mathcal{T}, V \in \mathcal{S}\}$$

Let $(x, y) \in W \subseteq T \times S$ be open, then by definition of a basis there exists $U \times V \in \mathcal{B}$ such that

$$(x, y) \in U \times V \subset W$$

Let $\mathcal{U}$ be an open cover of $X \times Y$.

Let $x \in X$. Then we can find $W_x \in \mathcal{U}$ such that $(x, y) \in W_x$. By the above, there exists $U_x \times V_x \in \mathcal{B}$ such that $(x, y) \in U_x \times V_x \subset W_x$. The sets U_x form an open cover of X, thus contains a finite subcover $U_{x_1}, \dots, U_{x_n}$. Let

$$N(y) = \bigcap_{k=1}^{n} V_{x_k}$$

This definition makes sense since $y \in N(y)$ is a neighbourhood of y that is open. We also have

$$X \times N(y) \subset \bigcup_{k=1}^{n} (U_{x_k} \times V_{x_k}) \subset \bigcup_{k=1}^{n} W_{x_k}$$

And thus $X \times N(y)$ has a finite subcover.

Since $\{N(y) | y \in Y\}$ forms an open cover, it has a finite subcover $N(y_1), \dots, N(y_m)$ that covers Y. Thus

$$X \times Y = \bigcup_{k=1}^{m} X \times N(y_k) \subset \bigcup_{k=1}^{m} \bigcup_{j=1}^{n} (U_{x_j} \times V_{x_k}) \subset \bigcup_{k=1}^{m} \bigcup_{j=1}^{n} W_{x_j k}$$

and $X \times Y$ has a finite subcover.

Repeated application of the proof proves that compactness is finitely productive. □

Theorem 8.5.3.2: Tychonov's Theorem

The product of any collection of compact spaces is compact.

8.5.4 Compactness in Metric Spaces

This section is dedicated to theorems related to compactness that is unique only to metric spaces. (Notice that metric spaces are Hausdorff thus some of the theorems above are already applicable to metric spaces)

Definition 8.5.4.1: Lebesgue Number

Let $\mathcal{U}$ be an open cover of a metric space X. A number $\delta > 0$ is called a Lebesgue number for $\mathcal{U}$ if for any $x \in X$ there exists $U \in \mathcal{U}$ such that $B_\delta(x) \subset U$.

Lemma 8.5.4.2

Every open cover $\mathcal{U}$ of a compact metric space X has a Lebesgue number.

Definition 8.5.4.3: Sequential Compactness

Let X be a metric space. Then X is said to be sequentially compact if any sequence of elements in X has a convergent subsequence.

Lemma 8.5.4.4

If X is sequentially compact that any open cover of X has a Lebesgue number.

Proposition 8.5.4.5

Let (X, d) be a metric space. Then the following are equivalent.

- X is compact
- X is sequentially compact
- X is closed and totally bounded

Since there is no notion of boundedness in a general topological space, we have the following theorem special to metric spaces.

Proposition 8.5.4.6

A compact subset of a metric space is bounded.

Proof. Let $a \in X$. Let $x \in K$. Then $x \in B_r(a)$ for all $r > d(a, x)$. Thus K is covered by the collection of open balls $B_r(a)$. Thus it has a finite subcover $B_{r_1}(a), \dots, B_{r_n}(a)$. Thus

$$K \subset \bigcup_{k=1}^n B_{r_k}(a) = B_{\max\{r_1,\dots,r_n\}}(a)$$

and we are done. □

Definition 8.5.4.7: Uniformly Continuous

A map $f : X \to Y$ between metric spaces is uniformaly continuous if for every $\epsilon > 0$, there exists $\delta > 0$ such that

$$d_X(x, y) < \delta \implies d_Y(f(x), f(y)) < \epsilon$$

for any $x, y \in X$.

Theorem 8.5.4.8

A continuous map from a compact metric into a metric space is uniformly continuous.

8.5.5 Compactness in $\mathbb{R}^n$

We arrive at an important characterization of compact sets in $\mathbb{R}^n$.

Theorem 8.5.5.1: Heine-Borel Theorem

A subset of $\mathbb{R}^n$ is compact if and only if it is closed and bounded.

Proof. Let K be a compact subset of $\mathbb{R}^n$. K is closed by proposition 5.2.5 and K is bounded by proposition 5.4.6.

Let K be a closed and bounded subset of $\mathbb{R}^n$. If K is bounded then $K \subset [-r, r]^n$ for some $r > 0$. I claim that $[-r, r]^n$ is compact. Once it is compact, applying 5.2.3 to the closed subset K of $[-r, r]^n$, we have that K is compact.

Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $[-r, r]$ by bolzano weierstrass it has a convergent subsequence. Thus $[-r, r]$ is sequentially compact and thus compact. Using the productivity of compact metric spaces, we have that $[-r, r]^n$ is compact thus we are done. □

This concludes the Bolzano weierstrass theorem for $\mathbb{R}^n$ by considering seuqential compactness. We also have another important theorem upcoming but we need another theorem prior to it.

Theorem 8.5.5.2

Let $f : X \to \mathbb{R}$ be a continiuous function from a non empty compact space X to $\mathbb{R}$. Then f is bounded and attains its bounds.

Proof. Since X is compact and f is continuous, $f(X) \subset \mathbb{R}$ is compact. By the Heine-Borel theorem, $f(X)$ is closed and bounded. But every closed and bounded subset of $\mathbb{R}$ contains its supremum and infinum. Thus f is bounded and attains its bounds. □

Theorem 8.5.5.3

All norms on $\mathbb{R}^n$ are equivalent. This measn that norms on $\mathbb{R}^n$ induce the same topology, the standard topology.

Proof. Let $\|\cdot\|$ be an arbitrary norm on $\mathbb{R}^n$. We show that it is equivalent to $\|\cdot\|_2$. Let $\{e_1, \dots, e_n\}$ be an orthonormal basis of $\mathbb{R}^n$. Then for any $x \in \mathbb{R}^n$, we have $x = \sum_{k=1}^n x_k e_k$. We have that

$$\begin{aligned}\|x\| &= \left\|\sum_{k=1}^{n} x_k e_k\right\| \\ &\leq \sum_{k=1}^{n} |x_k| \|e_k\| \\ &\leq \left(\sum_{k=1}^{n} |x_k|^2\right)^{\frac{1}{2}} \left(\sum_{k=1}^{n} \|e_k\|^2\right)^{\frac{1}{2}} \\ &= \left(\sum_{k=1}^{n} \|e_k\|^2\right)^{\frac{1}{2}} \|x\|_2\end{aligned}$$

Thus we have that $\|x\| \leq c_2\|x\|_2$.

Now we have that $\|x-y\| \leq c_2\|x-y\|_2$ for $x, y \in \mathbb{R}^n$. Define a map $f : (\mathbb{R}^n, \|\cdot\|_2) \to \mathbb{R})$ by $f(x) = \|x\|$. The above criteria means that f is continuous by choosing $\delta < \epsilon$. Since $\partial B_1(0)$ with the standard norm is a closed and bounded subset of $\mathbb{R}^n$, it is compact by the Heine-Borel theorem. From the above theorem, $f(\partial B_1(0))$ is bounded and attains its bounds. This means that $0 \leq c_1 < f(\partial B_1(0))$ for some c_1. But $c_1 \neq 0$ since if it is 0, then there exists $x \in \partial B_1(0)$ such that $\|x\|_2 = 0$.

This means that $\|x\|_2 = 1$ implies $\|x\| \geq c_1$. Let $y \in \mathbb{R}^n$ be arbitrary. Then since $\left\|\frac{y}{\|y\|_2}\right\| = 1$, we have that

$$\left\|\frac{y}{\|y\|_2}\right\| \geq c_1$$

and $\|y\| \geq c_1\|y\|_2$. Since this y is arbitrary, we are done. □

8.6 Connectedness

8.6.1 Basic Definitions

Definition 8.6.1.1: Connected Space

A topological space $(X,\mathcal{T})$ is said to be disconnected if there exists disjoint non-empty open subsets $A,B\subseteq X$ such that $X=A\cup B$, and $A\cap B=\emptyset$. If $(X,\mathcal{T})$ is not disconnected, then it is said to be connected.

Proposition 8.6.1.2

Let $(X,\mathcal{T})$ be a topological space. The following are equivalent.

- $(X,\mathcal{T})$ is disconnected
- There exists non-empty disjoint closed sets $A,B\subseteq X$ such that $X=A\cup B$
- There exists a non-trivial clopen subset of X.

Proof.

- (1) $\implies$ (2): Suppose that A,B are the sets satisfying the definition of a disconnceted space. Then since A is open, $B=X\setminus A$ is closed thus vice versa A is also closed.
- (2) $\implies$ (3): Again, since $B=X\setminus A$ is open, B is both open and closed.
- (3) $\implies$ (1): Let $A\subset X$ be both open and closed. Then $X\setminus A$ is also open and disjoint to A. Thus X is disconnected by A and $X\setminus A$.

□

Proposition 8.6.1.3

A topological space $(X,\mathcal{T})$ is disconnected if and only if there exists a continuous function $f:X\to\{0,1\}$ that is surjective.

Proof. Suppose that X is disconnected by A and B. Then define

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

Then f is continuous since every open set in $\{0,1\}$ maps to an open set in X. Clearly it is also surjective.

Now suppose that $f:X\to\{0,1\}$ is a surjective continuous function. Then define $A=f^{-1}(0)$ and $B=f^{-1}(1)$. A,B are open by continuity. Since f is surjective, we have $A\cup B=X$. Clearly they are disjoint else their common element will be mapped to both 0 and 1. Thus we are done. □

An equivalent formulation of the above proposition is that X is connected if and only if every continuous function from X to $\{0,1\}$ is constant.

8.6.2 Properties of Connectedness

Proposition 8.6.2.1

Let $(X,\mathcal{T})$ be connected. Let $(Y,\mathcal{U})$ be a topological space. Let $f : X \to Y$ be continuous. Then $f(X)$ is connected.

Proof. Suppose for a contradiction that $f(X)$ is disconnected. Suppose that it is disconnected by A, B. Then $f^{-1}(A)$ and $f^{-1}(B)$ is clearly disjoint since A and B are disjoint. Since $A \cup B = f(X)$, we must also have $f^{-1}(A) \cup f^{-1}(B) = X$. By continuity, $f^{-1}(A)$ and $f^{-1}(B)$ are open. Thus $f(X)$ is disconnected, a contradiction. □

Proposition 8.6.2.2

Let $(X,\mathcal{T})$ be a topological space. Let $\{C_i | i \in I\}$ is a non-empty connected subsets of X with the property that $\bigcap_{i\in I} C_i \neq \emptyset$. Then $\bigcup_{i\in I} C_i$ is connected.

Proof. Suppose that $x \in \bigcap_{i\in I} C_i$. Then $x \in C_i$ for all i. Let f be a continuous function from X to $\{0,1\}$. WLOG take $f(x) = 0$. Then since each C_i is connected, every $y \in C_i$ maps to 0. Then every $y \in \bigcup_{i\in I} C_i$ maps to 0 and thus f is constant and we are done. □

The following lemma is slightly different from the one above since we simply require the closure to have nonempty intersection. The two sets can be touching each other instead of intersecting each other to be connected.

Lemma 8.6.2.3

Suppose that C_1, C_2 are connected subsets of a topological space X and $\overline{C_1} \cap C_2 \neq \emptyset$. Then $C_1 \cup C_2$ is connected.

Proof. Suppose that $f : C_1 \cup C_2 \to \{0,1\}$ is continuous. WLOG take $f(C_1) = \{0\}$. Suppose for a contradiction that $f(C_2) = \{1\}$. Then $f^{-1}(1)$ is open thus there exists some $U \in \mathcal{T}$ such that $f^{-1}(1) = U \cap (C_1 \cup C_2)$. Now let $x \in \overline{C}_1 \cap C_2$. But $x \in C_2$ means that $f(x) = 1$ thus $x \in U \cap (C_1 \cup C_2)$ and $x \in U$. Then clearly

$$U \cap C_1 \neq \emptyset$$

since $x \in \overline{C}_1 \cup C_2$.

But $C_1 \subseteq C_1 \cup C_2$ thus we have

$$\begin{aligned} U \cap (C_1 \cup C_2) \cap C_1 &\neq \emptyset \\ f^{-1}(1) \cap C_1 &\neq \emptyset \end{aligned}$$

This is a contradiction since $f(C_1) = \{0\}$ thus we are done. □

Theorem 8.6.2.4

Let C and $\{C_i | i \in I\}$ be connected subsets of a topological space X and $C_i \cap \overline{C} \neq \emptyset$ for each i. Then

$$C \cup \bigcup_{i\in I} C_i$$

is connected.

Proof. Apply the above lemma to $C \cup C_i$. This is the possible since $C_i \cap \overline{C} \neq \emptyset$. Thus each $C \cup C_i$ is connected. Trivially each pair $(C \cup C_i) \cap (C \cup C_j)$ is nonempty since it contains C. Thus we can apply the proposition above the lemma and we are done. □

This theorem is slightly different then the one above since there is a central connected subset linking every other connected subset while in the first one, we simply require pairwise connectedness.

Corollary 8.6.2.5

If $C \subset X$ is a connected subset of a topological space X then so is any set K where $C \subseteq K \subseteq \overline{C}$.

Proof. Any K between C and $\overline{C}$ must contain some but not all of the boundary of C. Then we can apply the above theorem and we are done. □

8.6.3 Transferal of Connectedness

Proposition 8.6.3.1

Connectedness is not Hereditary.

Proposition 8.6.3.2

Connectedness is finitely productive.

Proof. Let X, Y be connected spaces. Let $y \in Y$. Define $C = X \times \{y\}$ and $C_t = \{t\} \times Y$. Then C is homeomorphic to T and C_t is homeomorphic to S. Thus they both are connected. Clearly we also have that $C_t \cap C \neq \emptyset$ since they both contain (x, y) and

$$X \times Y = C \cup \bigcup_{t \in X} C_t$$

thus $X \times Y$ is connected. □

Proposition 8.6.3.3

Connectedness is productive.

8.6.4 Path Connectedness

Definition 8.6.4.1: Paths

Let $(X, \mathcal{T})$ be a topological space. A path in X is a continuous function $p : [0, 1] \to X$. More specifically, given two points $a, b \in X$, a path p in X such that $p(0) = a$ and $p(1) = b$ is called a path from a to b.

Definition 8.6.4.2: Path Connectedness

A topological space $(X, \mathcal{T})$ is called path connected if for any distinct $a, b \in X$, there is a path from a to b.

Proposition 8.6.4.3

Every path connected space is connected.

Proof. Let $u \in X$. Consider any $v \in X$. Then there exists a path from u to v. Thus the image of the path is connected since it is the continuous image of $[0,1]$. Then $X = \{u\} \cup \bigcup_{v\in X} C_v$ and each C_v contains u thus X is connected. □

8.6.5 Connectedness on $\mathbb{R}^n$

Theorem 8.6.5.1

A subset of $\mathbb{R}$ is connected if and only if it is an interval.

Below is a partial converse of path connectedness implying connectedness over $\mathbb{R}^n$.

Theorem 8.6.5.2

Connected open subsets of $\mathbb{R}^n$ are path connected.

Theorem 8.6.5.3

Open subsets of $\mathbb{R}^n$ have open connected components.

Theorem 8.6.5.4

A subset U of $\mathbb{R}$ is open if and only if it is the disjoint union of countably many open intervals.

8.7 Completeness

8.7.1 Motivation and Definitions

Completeness is something that only appears in metric spaces.

Definition 8.7.1.1: Cauchy Sequence

We say that $\{x_n\} \subset (X, d)$ is a Cauchy sequence if for every $\epsilon > 0$, there exists some N such that $d(x_n, x_m) < \epsilon$ for all $n, m > \epsilon$.

Proposition 8.7.1.2

Every convergent sequence is Cauchy.

Proof. Let $(x_n)_{n\in\mathbb{N}}$ be a convergent sequence in a metric space X. Let $\epsilon > 0$, then from convergence we have that for $d(x_n, x) < \frac{\epsilon}{2}$ for all $n > N$ for some $N \in \mathbb{N}$. Then choosing the same N, we have that

$$d(x_n, x_m) \leq d(x_n, x) + d(x, x_m) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

thus we are done. □

We now give the definition of a complete space in terms of Cauchy sequences.

Definition 8.7.1.3: Complete Spaces

A metric space (X, d) is complete if any Cauchy sequence in X converges.

Proposition 8.7.1.4

Every compact metric space is complete.

Proof. Suppose that $(x_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in a compact metric space X. Then X being sequentially compact means that there exists a subsequence of $(x_n)_{n\in\mathbb{N}}$ such that it converges in X. But then clearly

$$d(x_n, x) \leq d(x_n, x_{n_k}) + d(x_{n_j}, x)$$

implies that $x_n \to x$ since in the inequality, the first part of the sum corresponds to the sequence being Cauchy and thus tends to 0, while the latter part correponds to the subsequence being convergent and thus tends to 0. □

8.7.2 Properties of Complete Spaces

Proposition 8.7.2.1

A subspace of a metric space is complete if and only if it is closed under a complete metric space.

Proof. Suppose that X is a metric space and $U \subset X$ is a complete metric space. Let $(x_n)_{n\in\mathbb{N}} \subset U$ and that $x_n \to x \in X$. Then $(x_n)_{n\in\mathbb{N}}$ is Cauchy thus it convergence to some $y \in U$. We will show that in fact $x = y$. This is true from the fact that

$$d|_U(x_n, y) = d(x_n, y)$$

Thus $(x_n)_{n\in\mathbb{N}}$ is in fact a sequence that converges in U. This shows that U is closed.

Now suppose that U is closed under a complete metric space X. Let $(x_n)_{n\in\mathbb{N}}$ be a Cauchy sequence in U. Then trivially it is also a Cauchy sequence in X and thus is convergent. Since U is closed, the limit is necessarily in U and thus U is complete. □

Theorem 8.7.2.2: Cantor's Intersection Theorem

Let X be a complete metric space. Let $S_1 \supseteq S_2 \supseteq \dots$ form a nested sequence of non-empty closed sets in X with the property that $\operatorname{diam}(S_n) \to 0$ as $n \to \infty$. Then

$$\bigcap_{n=1}^{\infty} S_n \neq \emptyset$$

Proof. For each $N \in \mathbb{N}$, choose $x_N \in S_N$. Then for all $n > N$, $x_n \in S_N$. Thus for $n, m > N$, we have that $d(x_n, x_m) \leq \operatorname{diam}(S_n)$. It follows that $(x_n)_{n\in\mathbb{N}}$ is Cauchy. Thus $x_n \to x$ for some $x \in X$. Since each S_n is closed and $x_n \in S_N$ for all $n > N$, we must have that $x \in S_n$ for each n. Thus $x \in \bigcap_{k=1}^{\infty} S_n$ is nonempty. □

Below are a few examples of complete spaces.

Proposition 8.7.2.3

$\mathbb{R}^n$ and $\mathbb{C}$ are both complete.

Proof. Let $(x_k)_{k\in\mathbb{N}}$ be a Cauchy sequence in $\mathbb{R}^n$. Denote the ith component of x_k by $x_{k,i}$. Then for every $\epsilon > 0$, there exists N such that

$$\|x_k - x_m\| = \left(\sum_{i=1}^{n} |x_{k,i} - x_{m,i}|^2\right)^{\frac{1}{2}} < \epsilon$$

for $k, m > N$. In particular, we have that each individual

$$|x_{k,i} - x_{m,i}| < \epsilon$$

for $m, n > N$. Thus $(x_{k,i})_{k\in\mathbb{N}}$ is a Cauchy sequence in $\mathbb{R}$. But we know that Cauchy sequences in $\mathbb{R}$ converges, thus $(x_{k,i})_{k\in\mathbb{N}}$ converges to $x_i \in \mathbb{R}$. Now define $x = (x_1, \dots, x_n)$, then

$$\|x_k - x\| = \left(|x_{k,i} - x_i|^2\right)^{\frac{1}{2}} < n\epsilon$$

by convergence of each individual component. Thus $(x_n)_{n\in\mathbb{N}}$ is a convergent sequence.

The proof for $\mathbb{C}$ is the same in considering $\mathbb{R}^2$. □

Proposition 8.7.2.4

Every normed vector space is complete.

8.7.3 Completion

The goal of this section is to attempt to complete a metric space by adding in the missing limits of a metric space.

Definition 8.7.3.1: Space of Bounded Real Functions

Denote $B(X)$ the space of all bounded real valued functions on a metric (topological) space X. This means that

$$B(X) = \{f : X \to \mathbb{R} | |f| \leq M \text{ for some } M \in \mathbb{R}\}$$

Proposition 8.7.3.2

The metric space with distance induced by the supremum norm

$$\|f\|_\infty = \sup_{x \in X} |f(X)|$$

for $f \in B(X)$ is complete.

Proof. Let $(f_n)_{n\in\mathbb{N}}$ be a Cauchy sequence in $B(X)$. Then for every $\epsilon > 0$, there exists N such that

$$\|f_n - f_m\|_\infty = \sup_{x \in X} |f_n(x) - f_m(x)| < \epsilon$$

for all $n, m > N$. In particular, for each $x \in X$, the property of supremum implies that $|f_n(x) - f_m(x)| < \epsilon$ for $n, m > N$. Thus $(f_n(x))_{n\in\mathbb{N}} \subset \mathbb{R}$ is Cauchy for each x. Since $\mathbb{R}$ is complete, $(f_n(x))_{n\in\mathbb{N}}$ converges for each $x \in X$.

Now define the function $f : X \to \mathbb{R}$ by

$$f(x) = \lim_{n\to\infty} f_n(x)$$

Then fix $\epsilon > 0$, we have that

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

for all $n > N$ by letting $m \to \infty$ from the fact that $|f_n(x) - f_m(x)| < \epsilon$. This N does not depend on x. Fix $\epsilon = 1$, then there exists $N_1 \in \mathbb{N}$ such that

$$\begin{aligned} |f(x) - f_n(x)| &\leq |f(x) - f_{N_1}(x)| \\ &\leq 1 + |f_{N_1}(x)| \end{aligned}$$

for all $x \in X$ and $n > N_1$ thus f is bounded. This means that $f \in B(X)$ and that $\|f_n - f\|_\infty < \epsilon$ for all $n > N$. □

Proposition 8.7.3.3

Any metric space X can be isometrically embedded into the complete metric space $B(X)$.

8.7.4 Compactness, Completeness and Totally Bounded

Definition 8.7.4.1: Totally Bounded

A metric space X is totally bounded if for any $\epsilon > 0$, there exists $B_\epsilon(p_k)$ for $k \in \{1, \dots, n\}$ such that

$$X \subseteq \bigcup_{k=1}^{n} B_\epsilon(p_k)$$

Theorem 8.7.4.2

A subspace Y of a metric space X that is complete is compact if and only if it is closed and totally bounded.

Theorem 8.7.4.3

A subspace Y of a complete metric space is totally bounded if and only if its closure is compact.

8.7.5 Contraction Mapping and Completion

Definition 8.7.5.1: Lipschitz Continuous

Let (X, d_X) and (Y, d_Y) be metric spaces and suppose that $f : X \to Y$. We say that f is a Lipschitz map if there is a constant $K \geq 0$ such that

$$d_Y(f(x), f(y)) \leq Kd(x, y)$$

for all x, y in X.

If $Y = X$ and $K \in [0, 1)$ then f is a contraction mapping.

Lemma 8.7.5.2

If $f : X \to Y$ is Lipschitz continuous then it is continuous.

Theorem 8.7.5.3: Contraction Mapping Theorem

Let X be a nonempty complete metric space and suppose that $f : X \to X$ is a contraction. Then f has a unique fixed point, meaning there is a unique $x \in X$ such that $f(x) = x$.

Proof. Let $x_0 \in X$ and define a sequence by $x_{n+1} = f(x_n)$ for $n \in \mathbb{N}$. Then we have that

$$d(x_{n+1}, x_n) \leq Kd(x_n, x_{n-1}) \leq \cdots \leq K^n d(x_1, x_0)$$

Then for any $k > n$, we have that

$$\begin{aligned} d(x_k, x_n) &\leq \sum_{i=n}^{k-1} d(x_{i+1}, x_i) \\ &\leq \sum_{i=n}^{k-1} K^i d(x_1, x_0) \\ &\leq \frac{K^i}{1-K} d(x_1, x_0) \end{aligned}$$

This is Cauchy since we can choose $\epsilon > 0$ such that $\frac{K^i}{1-K} < \epsilon$. Since X is complete, we have that $x_n \to x$ for some $x \in X$. Since f is continuous we have that $f(x_n) \to f(x)$. Now taking limits on

$$x_{n+1} = f(x_n)$$

we have that $x = f(x)$.

To prove uniqueness, note that if $f(x) = x$ and $f(y) = y$, then

$$d(x, y) = d(f(x), f(y)) \leq Kd(x, y)$$

which implies that $(1 - K)d(x, y) = 0$. Thus $x = y$. □

Another name for this theorem would be Banach's Fixed Point Theorem.

Theorem 8.7.5.4: Picard-Lindelof Theorem

Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be Lipschitz continuous with

$$|f(x) - f(y)| \leq L|x - y|$$

where $x, y \in \mathbb{R}^n$. Then for any $x_0 \in \mathbb{R}^n$, the differential equation

$$\frac{dx}{dt} = f(x)$$

with initial condition $x(0) = x_0$ has a unique solution on $[-t, t]$ for any $Lt < 1$.

8.7.6 Cantor's Theorem

Theorem 8.7.6.1

If X is a complete metric space and $\{F_n | n \in \mathbb{N}\}$ is a collection of open dense subsets of X, then

$$F = \bigcap_{k=1}^{\infty} F_n t$$

is dense in X. Equivalently, if $\{G_n | n \in \mathbb{N}\}$ is a collection of nowhere dense subsets of a nonempty complete metric space X, then

$$\bigcup_{k=1}^{\infty} F_k \neq X$$

Lemma 8.7.6.2

The Cantor set is uncountable.

8.8 Notable Topologies

8.8.1 $\mathbb{R}$ and $\mathbb{C}$ with the Standard Topology

$\mathbb{R}$ and $\mathbb{C}$ being metric spaces allows a natural topology to be induced from the metric.

Definition 8.8.1.1: Standard Topology of $\mathbb{R}$ and $\mathbb{C}$

The standard topology on $\mathbb{R}$ induced by the Euclidean metric $d(x,y) = |x-y|$ consists of $B_r(a) = \{x \in \mathbb{R} | |x-a| < r\}$ for $a \in \mathbb{R}$ and $r > 0$ being open sets aside from $\emptyset$ and $\mathbb{R}$.

The standard topology on $\mathbb{C}$ induced by the Euclidean metric $d(x,y) = |x-y|$ consists of $B_r(a) = \{z \in \mathbb{C} | |z-a| < r\}$ for $a \in \mathbb{R}$ and $r > 0$ being open sets aside from $\emptyset$ and $\mathbb{R}$.

Proposition 8.8.1.2

The product topology of $\mathbb{R}^n$ and $\mathbb{C}^n$ is precisely $B_r(a) = \{x \in \mathbb{R}^n | |x-a| < r\}$ and $B_r(a) = \{z \in \mathbb{C} | |z-a| < r\}$.

Proposition 8.8.1.3

The subspace topology of $\mathbb{Z}$ induced by $\mathbb{R}$ is precisely $\mathcal{T}_Z = \{\{x_1, \dots, x_n\} | x_k \in \mathbb{Z}, n \in \mathbb{N}\}$.

Proposition 8.8.1.4

All the above topological spaces are Hausdorff.

8.8.2 Discrete and Indiscrete Topology

Definition 8.8.2.1: Discrete Topology

Let X be a set. The discrete topology on X is the topology where every subset of X is part of the topology:

$$\mathcal{T}_X = \{U | U \subseteq X\}$$

Proposition 8.8.2.2

Under the discrete topology, every set is both open and closed.

Definition 8.8.2.3: Indiscrete Topology

The indiscrete topology on a set X is the topology where the only open sets are $\emptyset$ and X.

8.8.3 Cofinite Topology

Definition 8.8.3.1: Cofinite Topology

The cofinite topology on a set X is a topology where every open set has its complement being finite:

$$\mathcal{T}_X = \{U \subseteq X | X \setminus U \text{ is finite }\} \cup \{\emptyset\}$$

Proposition 8.8.3.2

The cofinite topology on $\mathbb{R}$ is not Hausdorff.

8.8.4 The Space of Bounded Functions

Definition 8.8.4.1: Space of Bounded Real Functions

Denote $B(X)$ the space of all bounded real valued functions on a topological space X.

Proposition 8.8.4.2

The metric space with distance induced by the supremum norm

$$\|f\|_\infty = \sup_{x \in X} |f(X)|$$

for $f \in B(X)$ is complete.

Proposition 8.8.4.3

The space of all bounded continuious functions from a topological space T to $\mathbb{R}$, $C_B(T)$ is a closed subspace of $B(T)$ and thus is complete.

Corollary 8.8.4.4

Let $(X, \mathcal{T})$ be a nonempty compact topological space, then $C(T)$ is complete with maximum norm

$$\|f\|_\infty = \max_{x \in T} |f(x)|$$

8.9 Notable Metric Spaces

8.9.1 $\mathbb{R}^n$ on Different Metrics

Theorem 8.9.1.1

Let $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ and similarly for $y \in \mathbb{R}^n$. The following are all metrics of $\mathbb{R}^n$.

- l_p metric:
$$d_p(x, y) = \left(\sum_{k=1}^{n} (x_k - y_k)^p \right)^{1/p}$$
for $1 \le p < \infty$
- l_∞ metric:
$$d_\infty(x, y) = \max_{k \in \{1, \dots, n\}} \{|x_k - y_k|\}$$
- Jungle river metric on $\mathbb{R}^2$:
$$d_{\text{Jr}}(x, y) = \begin{cases} |x_2 - y_2| & \text{if } x_1 = y_1 \\ |y_2| + |x_2| + |x_1 - y_1| & \text{if } x_1 \neq y_1 \end{cases}$$
- French Railway Metric (Sunflower metric) on $\mathbb{R}^2$:
$$d_{\text{Fr}}(x, y) = \begin{cases} |x - y| & \text{if there exists } \lambda \in \mathbb{R} \text{ such that } y = \lambda x \\ |x| + |y| & \text{otherwise} \end{cases}$$
- Discrete Metric:
$$d_{\text{Discrete}}(x, y) = \begin{cases} 0 & \text{if } x = y \\ 1 & \text{if } x \neq y \end{cases}$$
- British Railway Metric on $\mathbb{R}^2$:
$$d(x, y) = \begin{cases} 0 & \text{if } x = y \\ |x| + |y| & \text{if } x \neq y \end{cases}$$

Do try and draw at least the unit ball for each of these metrics and see what happens (at least for $\mathbb{R}^2$).

Proposition 8.9.1.2

All l_p metrics are topologically equivalent.

Proof. The metric are all induced by the l_p norms and we know that they are equivalent. Equivalent norms induce topologically equivalent metrics and we are done. □

Proposition 8.9.1.3

Let (X, d) be a metric space. Then the function
$$d_{\text{B}}(x, y) = \min\{d(x, y), 1\}$$
for any $x, y \in X$ is a metric on X.

8.9.2 The Space of Continuous Functions

Definition 8.9.2.1

We denote $C([a,b])$ the space of real valued continuous functions whose domain is $[a,b]$.

Proposition 8.9.2.2

Let $f \in C([a,b])$. Define the supremum norm of f to be

$$\|f\|_\infty = \sup_{x\in[a,b]}]|f(x)|$$

Then the supremum norm is a norm on $C([a,b])$.

Proposition 8.9.2.3

Let $f \in C([a,b])$. Define the L^p norm of f to be

$$\|f\|_{L^p} = \left(\int_a^b |f(x)|^p \, dx \right)^{\frac{1}{p}}$$

for $p \in [1,\infty)$. Then the supremum norm is a norm on $C([a,b])$.

8.9.3 Sequence Space

Definition 8.9.3.1: Sequence Space

The sequence space l^p for $1 \leq p < \infty$ consists of all sequences $\{x_n\}$ such that

$$\sum_{k=1}^\infty |x_k|^p < \infty$$

If $p = \infty$ then l^∞ is the space of all bound sequences.

Proposition 8.9.3.2

The function

$$\|x\|_{l^p} = \left(\sum_{k=1}^\infty |x_k|^p \right)^{\frac{1}{p}}$$

on l^p space defines a norm on it.

If $p = \infty$ then $\|x\|_{l^\infty} = \sup_{k\in\mathbb{N}} |x_k|$ defines a norm on l^∞.

Proposition 8.9.3.3

l^p is a complete metric space with metric

$$d(\{x_n\},\{y_n\}) = \|x-y\|_{l^p}$$

Chapter 9

Geometry

9.1 Euclidean Geometry

9.1.1 Euclidean Space

The background of our study of geometry involves mainly three settings. In this chapter we will begin with the surface that mathematics students are most familiar with, namely the Euclidean Space. We will treat it as a metric space and investigate basic objects in traditional geometry such as lines, triangles and distance preserving transformations between them.

Definition 9.1.1.1: Euclidean Inner Product

Let $x, y \in \mathbb{R}^n$. The Euclidean inner product is defined to be

$$\langle x, y\rangle = \sum_{k=1}^{n} x_k y_k$$

The Euclidean inner product is just your standard product we investigated thoroughly in Linear Algebra. With this we have that $(\mathbb{R}^n, \langle\cdot,\cdot\rangle)$ becoming an Inner product space. The inner product naturally produces a norm, as well as a metric, as seem in Linear Algebra.

Definition 9.1.1.2: Euclidean Norm

Let $x \in \mathbb{R}^n$. The Euclidean Norm is defined to be

$$\|x\| = \sqrt{\langle x, x\rangle}$$

Definition 9.1.1.3: Euclidean Metric

Let $x, y \in \mathbb{R}^n$. The Euclidean Metric is defined to be

$$d(x, y) = \|x - y\|$$

These three notion would be useful in formulating properties of the Euclidean Space. Before we prove that the Euclidean Metric is a metric on $\mathbb{R}^n$, we need a lemma to make our lives easier.

Lemma 9.1.1.4: Translation Invariance of Euclidean Metric

For $x, y, z \in \mathbb{R}^m$, we have

$$d(x, y) = d(x - z, y - z)$$

Proof. Let $x, y, z \in \mathbb{R}^m$

$$\begin{aligned} d(x,y) &= \|x-y\| \\ &= \|(x-z)-(y-z)\| \\ &= d(x-z, y-z) \end{aligned}$$

Thus we are done. □

We now show that $\mathbb{R}^n$ is indeed a metric space with the distance function defined as above.

Theorem 9.1.1.5

Let $x, y \in \mathbb{R}^n$. $d(x,y) = \|x-y\|$ is a metric on $\mathbb{R}^n$.

Proof. We just have the triangle inequality to prove. By the translation invariance, this is equivalent to proving

$$d(0, y-x) \leq d(0, z-x) + d(z-x, y-x)$$

Rewriting this gives $d(0,u) \leq d(0,v) + d(v,u) \iff \|u\| \leq \|v\| + \|u-v\|$.

$$\begin{aligned} \|u\| &= \frac{\langle u, u\rangle}{\|u\|} \\ &= \frac{\langle v+u-v, u\rangle}{\|u\|} \\ &= \frac{1}{\|u\|}\left(\langle v, u\rangle + \langle u-v, u\rangle\right) \\ &= \left\langle v, \frac{u}{\|u\|}\right\rangle + \left\langle u-v, \frac{u}{\|u\|}\right\rangle \\ &\leq \left|\left\langle v, \frac{u}{\|u\|}\right\rangle\right| + \left|\left\langle u-v, \frac{u}{\|u\|}\right\rangle\right| \\ &\leq \|v\|\|\frac{u}{\|u\|}\| + \|u-v\| \cdot \|\frac{u}{\|u\|}\| \\ &= \|v\| + \|u-v\| \end{aligned}$$

□

As seen from the below proposition, we can see that $\mathbb{R}^n$ is a particularly well behaved space.

Proposition 9.1.1.6

$\mathbb{R}^n$ is an inner product space. Thus it is also a normed vector space and a metric space.

Proof. Trivial. □

Finally we will define the Euclidean Space. Beware that our definition is a bit vague and that it is not completely equal to $\mathbb{R}^n$

Definition 9.1.1.7: Euclidean Space

An Euclidean Space is a metric space which is isometric to $\mathbb{R}^n$, with the Euclidean metric for some $n \in \mathbb{N}$. We use the notation E^n to denote the Euclidean Space.

While we have a metric defined on E^n, it does not have addition and scalar multiplication, which makes it different from $\mathbb{R}^n$. However, it is by definition that an Euclidean Space can be associated with $\mathbb{R}^n$. This makes the discussion on Euclidean Space much more easier, simply by studying $\mathbb{R}^n$.

The reason we use the Euclidean Space instead of $\mathbb{R}^n$ is because we want our 0 to be avaliable anywhere, instead of it being fixated on point in the space.

9.1.2 Lines and Collinearity

Definition 9.1.2.1: Lines

Let $u, v \in \mathbb{R}^n, v \neq 0$. The line through u in the direction v is defined to be

$$L = \{u + \lambda v | \lambda \in \mathbb{R}\} \subset \mathbb{R}^n$$

Definition 9.1.2.2: Collinearity

We say that x, y, z are collinear if there is a line L with $x, y, z \in L$.

Lemma 9.1.2.3: Translation Preserves Collinearity

Let $x, y, z \in \mathbb{R}^n$ be distinct. They are collinear if and only if $x - z, y - z, 0$ are collinear.

Proof. x, y, z is collinear if and only if $z = (1 - \lambda)x + \lambda y \iff 0 = (1 - \lambda)(x - z) + \lambda(x - z)$. This means $x - z, y - z, 0$ are collinear □

Proposition 9.1.2.4

$x, y, z \in \mathbb{R}^n$ are collinear if and only if

$$d(x, y) = d(x, z) + d(z, y)$$

where $d(x, y)$ is the Euclidean metric.

Proof. Suppose that x, y, z are collinear with z in the middle. Then

$$z = (1 - \lambda)x + \lambda y$$

for some $\lambda \in (0, 1)$. Now we have

$$\begin{aligned} d(x, z) + d(z, y) &= \|x - z\| + \|z - y\| \\ &= \|x - (1 - \lambda)x - \lambda y\| + \|(1 - \lambda)x + \lambda y - y\| \\ &= \|\lambda(x - y)\| + \|(1 - \lambda)(x - y)\| \\ &= \lambda\|x - y\| + (1 - \lambda)\|x - y\| \\ &= \|x - y\| \\ &= d(x, y) \end{aligned}$$

Now assume that $d(x, y) = d(x, z) + d(z, y)$. We first translate it such that $d(x - z, y - z) = d(x - z, 0) + d(0, y - z)$ and set $u = x - z$ and $v = y - z$. Thus are new equality is $d(u, v) = d(u, 0) + d(0, v)$

$$\begin{aligned} \|u - v\| &= \|u\| + \|v\| \\ \|u - v\|^2 &= \|u\|^2 + 2\|u\|\|v\| + \|v\|^2 \\ \langle u - v, u - v\rangle &= \|u\|^2 + 2\|u\|\|v\| + \|v\|^2 \\ \|u\|^2 - 2\langle u, v\rangle + \|v\|^2 &= \|u\|^2 + 2\|u\|\|v\| + \|v\|^2 \\ -\langle u, v\rangle &= \|u\|\|v\| \\ \implies |\langle u, v\rangle| &= \|u\|\|v\| \end{aligned}$$

We know that this happens if and only if $v = \lambda u$ for some $\lambda \in \mathbb{R}$, which implies that $0, u, v$ are linear. □

Proposition 9.1.2.5

If T is an isometry, then for $x, y, z \in E^n$, x, y, z are collinear implies that $T(x), T(y), T(z)$ are also collinear.

9.1.3 Affine Maps

Definition 9.1.3.1: Affine Maps

A map $T : \mathbb{R}^n \to \mathbb{R}^m$ is affine if it is of the form

$$T(x) = Ax + b$$

for all $x \in \mathbb{R}^n$ for some linear map A and $b \in \mathbb{R}^k$.

Proposition 9.1.3.2

Given a map $T : \mathbb{R}^n \to \mathbb{R}^k$, the following are equivalent.

- T is affine
- For all $\lambda, \mu \in \mathbb{R}$ and $x, y \in \mathbb{R}^n$,
$$T(\lambda x + \mu y) - T(0) = \lambda(T(x) - T(0)) + \mu(T(y) - T(0))$$
- For all $\lambda \in \mathbb{R}$,
$$T(\lambda x + (1 - \lambda)y) = \lambda T(x) + (1 - \lambda)T(y)$$

Proof. Let $T : \mathbb{R}^n \to \mathbb{R}^k$ be a map.

- (1) $\iff$ (2): Define $L(x) = T(x) - T(0)$. Then T is affine if and only if L is linear, if and only if $L(\lambda x + \mu y) = \lambda L(x) + \mu L(y)$ if and only if $T(\lambda x + \mu y) - T(0) = \lambda(T(x) - T(0)) + \mu(T(y) - T(0))$.
- (2) $\implies$ (3): This is obtained by setting $\mu = 1 - \lambda$.
- (3) $\implies$ (2): We have
$$\begin{aligned} \lambda x + \mu y &= \frac{1}{2}(2\lambda x) + \frac{1}{2}(2\mu y) \\ T(\lambda x + \mu y) &= \frac{1}{2}T(2\lambda x) + \frac{1}{2}T(2\mu y) && \text{(By (3))} \end{aligned}$$
Also by (3), we have
$$\begin{aligned} 2\lambda x &= 2\lambda x + (1 - 2\lambda)0 \\ T(2\lambda x) &= 2\lambda T(x) + (1 - 2\lambda)T(0) \end{aligned}$$
And similarly,
$$T(2\mu y) = 2\mu T(y) + (1 - 2\mu)T(0)$$
Combining the three, we have
$$\begin{aligned} T(\lambda x + \mu y) &= \frac{1}{2}T(2\lambda x) + \frac{1}{2}T(2\mu y) \\ &= \frac{1}{2}(2\lambda T(x) + (1 - 2\lambda)T(0)) + \frac{1}{2}(2\mu T(y) + (1 - 2\mu)T(0)) \\ &= \lambda T(x) + \left(\frac{1}{2} - \lambda\right)T(0) + \mu T(y) + \left(\frac{1}{2} - \lambda\right)T(0) \\ T(\lambda x + \mu y) - T(0) &= \lambda(T(x) - T(0)) + \mu(T(y) - T(0)) \end{aligned}$$

□

Proposition 9.1.3.3

Let $L : \mathbb{R}^n \to \mathbb{R}^n$ be a linear map with matrix A with respect to the standard basis. Then the following are equivalent.

- L is an isometry
- $\|L(x)\| = \|x\|$ for all $x \in \mathbb{R}^n$
- $\langle L(x), L(y)\rangle = \langle x, y\rangle$ for all $x, y \in \mathbb{R}^n$
- A is orthogonal

Proof. Let $L : \mathbb{R}^n \to \mathbb{R}^k$ be linear.

- (1) $\implies$ (2): This is trivial since isometries are distance preserving
- (2) $\implies$ (3): This is given by the polarization identity
- (3) $\implies$ (4): We have that

  $$\begin{aligned}\left(A^TA\right)_{ij} &= \sum_{k=1}^n (A^T)_{ik}A_{kj} \\ &= \sum_{k=1}^n A_{ki}A_{kj} \\ &= \sum_{k=1}^n L(e_i)_k L(e_j)_k \\ &= \langle L(e_i), L(e_j)\rangle \\ &= \langle e_i, e_j\rangle && \text{(By (3))} \\ &= \delta_{ij}\end{aligned}$$

  Thus $A^TA = I$

- (4) $\implies$ (1): Suppose that $A^TA = I$. Then A is invertible and thus is a bijection. Thus we just need to show that $d(x, y) = d(L(x), L(y))$.

  $$\begin{aligned}\|L(x)\|^2 &= \langle L(x), L(x)\rangle \\ &= (L(x))^TL(x) \\ &= x^TA^TAx \\ &= x^Tx \\ &= \langle x, x\rangle \\ &= \|x\|^2\end{aligned}$$

  Thus $d(L(x), L(y)) = \|L(x) - L(y)\| = \|L(x - y)\| = \|x - y\| = d(x, y)$.

□

Theorem 9.1.3.4

Every Euclidean Isometry $T : \mathbb{R}^n \to \mathbb{R}^n$ is affine and has the form $T(x) = Ax + b$, where A is orthogonal.

Proof. We have shown that Euclidean Isometries are line preserving, thus by 1.3.2 they are affine and in the form $T(x) = Ax + b$. We now check that $G(x) = Ax = T(x) - b$ is an isometry.

G is distance preserving since

$$\begin{aligned} d(G(x),G(y)) &= \|G(x)-G(y)\| \\ &= \|T(x)-b-T(y)+b\| \\ &= \|T(x)-T(y)\| \\ &= \|x-y\| && (T \text{ is an isometry}) \\ &= d(x,y) \end{aligned}$$

We now show that G is bijective. Since T is an isometry, T is bijective and $T(x)-b$ is also bijective with $T^{-1}(x)=x+b$ and thus G is an isometry. By the above proposition, A is orthogonal. □

Proposition 9.1.3.5

Every orthogonal matrix $A \in \mathbb{R}^{n\times n}$ induces a set of Euclidean isometries $T(x)=Ax+b$ where $b \in \mathbb{R}^n$

Proof. If A is orthogonal, then $G(x)=Ax$ is an isometry by the above proposition. But so is $T(x)=Ax+b$ for any $b\in\mathbb{R}^n$ since G is bijective, T is bijective and

$$\begin{aligned} d(T(x),T(y)) &= \|T(x)-T(y)\| \\ &= \|Ax+b-Ay-b\| \\ &= \|G(x)-G(y)\| \\ &= \|x-y\| && (G \text{ is an isometry}) \\ &= d(x,y) \end{aligned}$$

□

9.1.4 Normal Form Theorem

This is to categorize that every isometry from $\mathbb{R}^n$ to $\mathbb{R}^n$ has the normal form under some orthonormal basis. By doing this, we can simply study the normal form of a linear map to derive results to every isometry of the Euclidean Space.

Lemma 9.1.4.1

Let L be an isometry of $\mathbb{R}^n$, and W a subspace of $\mathbb{R}^n$ such that $L(W) = W$. Then also $L(W^\perp)=W^\perp$

Proof. Let $v\in L(W^\perp)$.

$$\begin{aligned} v\in L(W^\perp) &\iff L^{-1}(v)\in W^\perp \\ &\iff \langle L^{-1}(v),w\rangle = 0\ \forall w\in W \\ &\iff \langle L(L^{-1}(v)),L(w)\rangle = 0\ \forall w\in W \\ &\iff \langle v,L(w)\rangle = 0\ \forall w\in W \\ &\iff \langle v,w\rangle = 0\ \forall w\in W \\ &\iff v\in W^\perp \end{aligned}$$

□

Geometrically this means that if a subspace is invariant, then its orthogonal complement is invariant as well under isometry.

Corollary 9.1.4.2

Let L be an isometry of $\mathbb{R}^n$. Let W be a subspace of $\mathbb{R}^n$ such that $L(W) = W$. Let $w_1, \dots, w_p$ be a basis for W and let $w_{p+1}, \dots, w_n$ be a basis for $W^\perp$. Then with respect to the basis, the matrix L has the form

$$\begin{pmatrix} L|_W & 0 \\ 0 & L|_{W^\perp} \end{pmatrix}$$

Proof. Note that A can be written in the form

$$\begin{pmatrix} L|_W & L|_{W^\perp} \end{pmatrix} = \begin{pmatrix} \text{pr}_W(L|_W) & \text{pr}_W(L|_{W^\perp}) \\ \text{pr}_{W^\perp}(L|_W) & \text{pr}_{W^\perp}(L|_{W^\perp}) \end{pmatrix}$$

where $\text{pr}_W(v)$ is the projection of $v \in V$ on to the subspace W and vice versa. Then by the above lemma $\text{pr}_W(L|_{W^\perp}) = 0$ and $\text{pr}_{W^\perp}(L|_W) = 0$. Thus

$$A = \begin{pmatrix} L|_W & 0 \\ 0 & L|_{W^\perp} \end{pmatrix}$$

□

Proposition 9.1.4.3

Let $\lambda \in \mathbb{C}$ be an eigenvalue of an isometry L of $\mathbb{R}^n$, with eigenvector $z \in \mathbb{C}^n$. Then the following are true.

- $|\lambda| = 1$
- $\overline{\lambda}$ is also an eigenvalue with eigenvector $\overline{z}$
- if μ is also eigenvalue not equal to λ with eigenvector w, then $\langle z, w\rangle = 0$

Proof. Let $z, w \in \mathbb{C}$.

- $\langle z, z\rangle = \langle Az, Az\rangle = \langle \lambda z, \lambda z\rangle = \lambda\overline{\lambda}\langle z, z\rangle$. Thus $|\lambda| = 1$
- $A\overline{z} = \overline{A}\overline{z} = \overline{Az} = \overline{\lambda}\overline{z}$ and $A = \overline{A}$ since A is real.
- Note that $\langle z, w\rangle = \langle Az, Aw\rangle = \langle \lambda z, \mu w\rangle = \lambda|\mu|\langle z, w\rangle$. Thus either $\langle z, w\rangle = 0$ or $\lambda|\mu| = 1$. If $\lambda|\mu| = 1$, then

$$\begin{aligned} \lambda|\mu| &= 1 \\ \lambda\mu|\mu| &= \mu \\ \lambda|\mu|^2 &= \mu \\ \lambda &= \mu \end{aligned}$$

  This is a contradiction since we assume $\lambda \neq \mu$. Thus we must have $\langle z, w\rangle = 0$

□

Note that the third point here implies that the eigenspaces has to be orthogonal to each other.

Lemma 9.1.4.4

If L is a linear isometry of $W \cong \mathbb{R}^2$ with eigenvalues $\lambda, \overline{\lambda} \in \mathbb{C} \setminus \mathbb{R}$, then L has an orthonormal basis in $\mathbb{R}^2$ with respect to which it has matrix

$$\begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$$

where $\lambda = \cos(\theta) + i\sin(\theta)$

Proof. Extend L to $\mathbb{C}^2$. Let $z = x + iy$ be an eigenvector with eigenvalue λ, with $x, y \in \mathbb{R}^2$. We know that $|\lambda| = 1$, so $\lambda = \cos(\theta) + i\sin(\theta)$ for some θ, and we have

$$L(x + iy) = (\cos(\theta) + i\sin(\theta))(x + iy)$$

Comparing real and imaginary parts give

$$L(x) = \cos(\theta)x - \sin(\theta)y$$

and

$$L(y) = \sin(\theta)x + \cos(\theta)y$$

Thus we have the given matrix.

Now we show that x, y can be an orthonormal basis. From the above, we have $\langle z, \overline{z}\rangle = 0$. Since $x = \frac{1}{2}(z + \overline{z})$ and $y = \frac{1}{2i}(z - \overline{z})$, we have

$$\begin{aligned} \langle x, y\rangle &= \frac{-1}{4i}\langle z + \overline{z}, z - \overline{z}\rangle \\ &= \|z\|^2 - \langle z, \overline{z}\rangle + \langle \overline{z}, z\rangle - \|\overline{z}\|^2 \\ &= 0 \end{aligned}$$

Thus x, y are orthogonal. We also have that
$0 = \langle z, \overline{z}\rangle = \langle x + iy, x - iy\rangle = \|x\|^2 - \|y\|^2 + 2i\langle x, y\rangle = \|x\|^2 - \|y\|^2$. Thus $\|x\| = \|y\|$. Thus our basis can be normalized by using $\frac{x}{\|x\|}$ and $\frac{y}{\|y\|}$ without changing the basis. □

Theorem 9.1.4.5: Normal Form Theorem

Let $L : \mathbb{R}^n \to \mathbb{R}^n$ be a linear isometry and A an orthogonal matrix such that $L(x) = Ax$. Then there exists an orthonormal basis of $\mathbb{R}^n$ with respect to which the matrix A is

$$A = \begin{pmatrix} I_k & & & & \\ & -I_m & & & \\ & & B_1 & & \\ & & & \ddots & \\ & & & & B_l \end{pmatrix}$$

where

$$B_i = \begin{pmatrix} \cos(\theta_i) & -\sin(\theta_i) \\ \sin(\theta_i) & \cos(\theta_i) \end{pmatrix}$$

Proof. Let $L : \mathbb{R}^n \to \mathbb{R}^n$ be an isometry. We prove the result by induction on n. Extend the domain of L such that it becomes a map from $\mathbb{C}^n \to \mathbb{C}^n$. By the fundamental theorem of algebra, $c_A(L)$ has at least one root, λ. We know that $|\lambda| = 1$. There are two cases.

- If $\lambda \in \mathbb{R}$, then $\lambda = \pm 1$. Choose an eigenvector $z = x + iy$, with $x, y \in \mathbb{R}^n$. We have that $\lambda(x + iy) = L(x + iy) = L(x) + iL(y)$, thus x, y respectively are both real eigenvectors of λ. At least one of these x, y is non zero, else $z = 0$. Now $\frac{x}{\|x\|}$ is a basis for $W = \mathbb{C} \cdot x$ and the matrix for L is

$$\begin{pmatrix} \pm 1 & 0 \\ 0 & L|_{W^\perp} \end{pmatrix}$$

- If $\lambda \in \mathbb{C} \setminus \mathbb{R}$, choose an eigenvector z. We must have $\overline{\lambda}$ is also an eigenvalue, with eigenvector $\overline{z}$. Let $W = E_\lambda \oplus E_{\overline{\lambda}}$. By the above lemma, there exists a real orthonormal basis for W. In terms of a basis for $W \oplus W^\perp$, the matrix for L is

$$\begin{pmatrix} \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} & 0 \\ 0 & L|_{W^\perp} \end{pmatrix}$$

By the above, in both cases, $L|_{W^\perp}$ is an isometry of $W^\perp$, so we can apply the induction step to $W^\perp$. Eventually, we will have decomposed V into mutually orthogonal subspaces $W_1, \dots, W_M$, all invariant under L. Either W_i has dimension 2 with a real orthonormal basis with respect to which the matrix of $L|_W$ has the form as in the above lemma, or W_i has dimenison 1, with normalized real basis, and L acts as 1 or -1. Since the W_i are mutally orthogonal, the basis consisting of the union of all these basis is orthonormal. By reordering the subspaces appropriately we obtain the above form. □

I would say that the main part of this theorem is that you essentially decompose $\mathbb{R}^n$ into $W_1 \oplus \cdots \oplus W_n$, where they are mutually orthogonal. It might be the case that an isometry in $\mathbb{R}^2$ or $\mathbb{R}^3$ allows some orthogonal to rotate or reflect, leaving the rest invariant. Thus performing each rotation or reflection in $\mathbb{R}^3$ is only transforming one of $W_1, \dots, W_n$.

9.1.5 Relation to Groups

Definition 9.1.5.1: The Orthogonal Group

The group of $n \times n$ real orthogonal matrices is defined to by

$$O(n, \mathbb{R}) = \{A \in \text{GL}(\mathbb{R}^n) | A^T A = I\}$$

Lemma 9.1.5.2

$O(n, \mathbb{R})$ is a group.

Proof. It is simple to check that $O(n, \mathbb{R})$ by closure, associativity, existence of inverse and identity. □

Proposition 9.1.5.3

$O(n, \mathbb{R})$ acts on $\mathbb{R}^n$ as a group action. In particular, for $M \in O(n, \mathbb{R})$, $A_M(x) = Mx$.

Definition 9.1.5.4: Isometry Group

Let (X, d) be a metric space. Then the group of isometries of (X, d) is the set

$$\text{Isom}(X, d) = \{f : X \to X | f \text{ is an isometry of } X\}$$

Lemma 9.1.5.5

For any metric space (X, d), $\text{Isom}(X, d)$ is a group.

Proof. Note that if $f, g, h \in \text{Isom}(X, d)$, then $g \circ f$ is an isometry, $h \circ g \circ f$ is associative, the identity mapping $f(x) = x$ for all $x \in X$ is the identity of the group and since f is bijective, there exists an inverse of f. □

Theorem 9.1.5.6

There is a group homomorphism between

$$\psi : \text{Isom}(\mathbb{R}^n, d) \to \mathbb{R}^n \rtimes_\phi O(n, \mathbb{R})$$

where ϕ is given by the natural action of $O(n)$ on $\mathbb{R}^n$ and

$$\psi(T) = (T(0), T - T(0))$$

and its inverse being

$$\psi^{-1}(b, A) = (T(x) = Ax + b)$$

9.1.6 Isometry Decomposition

Definition 9.1.6.1: Hyperplane

Let $V \subseteq \mathbb{R}^n$ be a vector subspace of $\mathbb{R}^{n-1}$. Let $b \in \mathbb{R}^n$. Let $v_n \in \mathbb{R}$ such that $v_n \perp V$. A hyperplane of $\mathbb{R}^n$ is an affine subspace of dimension $n-1$, which has the form

$$\Pi = V + b = \{b + v | v \in V\} = \{v \in V | \langle v, v_n \rangle = \langle v_n, b \rangle\}$$

where $V \subseteq \mathbb{R}^n$ and $b \in \mathbb{R}^n$

Definition 9.1.6.2: Fixed Points

Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be an isometry. Denote the set of fixed points of T to be

$$\text{Fix}(T) = \{x \in \mathbb{R}^n | T(x) = x\}$$

Definition 9.1.6.3: Reflection on Hyperplane

Let Π be a hyperplane. A reflection in Π is an Euclidean Isometry $\rho_\Pi : \mathbb{R}^n \to \mathbb{R}^n$ such that $\text{Fix}(\rho_\Pi) = \Pi$

Proposition 9.1.6.4

The reflection on hyperplane exists and is unique.

Proof. Let $\Pi = V + b$. Pick $v \in V^\perp$. Take a basis of $\mathbb{R}^n$ consisting of v together with a basis for V. With respect to the basis, define a linear map T with matrix

$$A = \begin{pmatrix} -1 & 0 \\ 0 & I_{n-1} \end{pmatrix}$$

Note that this fixes V and is not an identity. Let $S(x) = x - b$. I claim that $\rho = S^{-1} \circ T \circ S$ is a reflection. I prove that $\text{Fix}(\rho) = \Pi$. This means solving $x = Ax + (I - A)b$. We have that

$(A-I)(x-b)=0$. Solving this gives

$$x = \begin{pmatrix} 0+b_1 \\ t_2+b_2 \\ \ddots \\ t_n+b_n \end{pmatrix}$$

But then $x \in \Pi$ since x is of the form $b+v$ with $v \in V$. Thus ρ_Π exists.

Let ρ' that is not the identity mapping fixes Π. Then $R = T \circ \rho' \circ T^{-1}$ by 1.4.2 fixes V and fixes $V^\perp$. Thus $R(v) = \lambda v$ for some $\lambda \in \mathbb{R}$. Since R is an isometry, $|\lambda| = 1$. Since R is not the identity, $\lambda = -1$ and thus R has matrix A. Thus $\rho = \rho'$ and we have proved uniqueness. □

Lemma 9.1.6.5

Let $p \in \mathbb{R}^n$. Let $V \subseteq \mathbb{R}^n$ and $v \perp V$. Let ρ_Π be the reflection on a hyperplane $\Pi = V + b$. Then

$$\rho_\Pi(p) = p - 2\langle p-b, v\rangle v$$

Proof. Let d be the shortest Euclidean distance between P and the hyperplane. To obtain the reflection of P along the hyperplane, we calculate the vector that starts at P with magnitude d and direction towards the hyperplane, and perpendicular to it, then add it to P two times to obtain it.

Let the point where the tail of v is on Π be Q. Note that $QP = P - b$. Since $\|v\| = 1$, we have that the magnitude of the projection of QP on to v is given by $\langle P-b, v\rangle$. Thus using our method, we obtain that $\rho_\Pi(P) = P - 2\langle P-b, v\rangle v$. □

Lemma 9.1.6.6

Let $p, q \in \mathbb{R}^n$ be distinct. Then there exists a reflection ρ with $\rho(p) = q$.

Proof. Let $v = \frac{p-q}{\|p-q\|}$. Let $b = \frac{p+q}{2}$. I claim that $\Pi = \mathbb{R}v^\perp + b$ admits a reflection such that $\rho_\Pi(p) = q$.

$$\begin{aligned} \rho_\Pi(p) &= p - 2\left\langle p - \frac{p+q}{2}, \frac{p-q}{\|p-q\|}\right\rangle \frac{p-q}{\|p-q\|} \\ &= p - \langle p-q, p-q\rangle \frac{p-q}{\|p-q\|^2} \\ &= q \end{aligned}$$

Thus we are done. □

Theorem 9.1.6.7

The group $\text{Isom}(E^n)$ is generated by reflections. Moreover, any isometry of E^n is the product of at most $n+1$ reflections.

Proof. Let T be an isometry. Let $P = T(0)$ and $Q = 0$. Then by the above, there exists a reflection R such that $R(T(0)) = 0$. Thus $L = R \circ T$ is a linear isometry. Choose an orthonormal basis $v_1, \dots, v_n$ so that by the normal form theorem, the matrix M has the following form

$$M = \begin{pmatrix} I_k & & & & \\ & -I_m & & & \\ & & B_1 & & \\ & & & \ddots & \\ & & & & B_l \end{pmatrix}$$

Denote J_i to be the $n \times n$ identity matrix, except -1 is at the (i,i)th spot. Note that this is the matrix of a reflection in $(\mathbb{R}v_i)^\perp$. Let B_θ be the matrix of rotation of degree θ. Then it can be decomposed into

$$B_i = A_i \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} \cos(\theta_i) & \sin(\theta_i) \\ \sin(\theta_i) & -\cos(\theta_i) \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

The matrix A_θ has eigenvalues 1 and -1, with corresponding eigenvectors $\left(\cos\left(\frac{\theta_i}{2}\right), \sin\left(\frac{\theta_i}{2}\right)\right)$ and $\left(\sin\left(\frac{\theta_i}{2}\right), -\cos\left(\frac{\theta_i}{2}\right)\right)$ respectively.

Define $C_i = J_{k+m+2i}$ and

$$D_i = \begin{pmatrix} I_{k+m+2i-2} & 0 & 0 \\ 0 & B_i & 0 \\ 0 & 0 & I_{n-k-m-2i} \end{pmatrix}$$

and

$$E_i = \begin{pmatrix} I_{k+m+2i-2} & 0 & 0 \\ 0 & A_i & 0 \\ 0 & 0 & I_{n-k-m-2i} \end{pmatrix}$$

Note that E_i is a reflection in the plane orthogonal to $\sin\left(\frac{\theta_i}{2}\right) v_{k+m+2i-1} - \cos\left(\frac{\theta_i}{2}\right) v_{k+m+2i}$. Then we have

$$M = J_{k+1} \circ \cdots \circ J_{k+m} \circ D_1 \circ \cdots \circ D_l$$

and thus

$$M = J_{k+1} \circ \cdots \circ J_{k+m} \circ E_1 \circ C_1 \circ \cdots \circ E_l \circ C_l$$

which is the product of at most n reflections. Since $T = R^{-1} \circ L$, T is the product of at most $n+1$ reflections. □

9.2 Spherical Geometry

9.2.1 The n-Sphere

Definition 9.2.1.1: n-Dimensional Sphere

Let $r \geq 0$. Define the n dimensional sphere of radius r to be

$$S^n(r) = \{(x_1, \dots, x_{n+1}) \in \mathbb{R}^{n+1} : \|x\| = r\}$$

Definition 9.2.1.2: Great Circle

A great circle is the intersection of $S^n(r)$ with a two dimensional vector subspace of $\mathbb{R}^{n+1}$

Definition 9.2.1.3: Antipodal Points

Let $P, Q \in S^n(r)$. We say that P and Q are antipodal if $Q = -P$.

Lemma 9.2.1.4

If $P, Q \in S^n(r)$ are not antipodal, then there exists a unique great circle containing P and Q.

Definition 9.2.1.5: Collinearity

Three distinct points on S^n are said to be collinear if there exists a great circle such that the three points lie on it.

Definition 9.2.1.6: Spherical Triangle

A triangle on $S^n(r)$ is three distinct points that are joined together by three great circles.

9.2.2 Spherical Metric

Definition 9.2.2.1: Spherical Metric

Define the Spherical Metric by

$$d_S(P, Q) = r \cos^{-1}\left(\frac{\langle P, Q \rangle}{r^2}\right)$$

where we take the domain of $\cos^{-1}$ to be $[0, \pi]$.

Proposition 9.2.2.2

Let $\triangle PQR$ be a spherical triangle of S^n. Then $\angle QOR = d(Q, R)$ and vice versa for the other angles.

Proof. By definition of a radian, the arc QR is equal to the angle $\angle QOR$. □

Definition 9.2.2.3: Spherical Angle

Suppose that two great circles on S^n intersect at one of the points P. Define the spherical angle P to be the smaller angle made on S^n between the two great circles. We write it as $\angle_S$.

Proposition 9.2.2.4

Let PQR be a spherical triangle of S^2. Suppose that $\alpha = \angle_S QPR$, $\beta = \angle_S PRQ$ and $\gamma = \angle_S PQR$. Let $a = \angle QOR = d_{S^2}(Q, R)$. Then

$$\cos(\alpha) = \cos(\beta)\cos(\gamma) + \sin(\beta)\sin(\gamma)\cos(a)$$

Proof. Note that isometries on $\mathbb{R}^3$ preserves the inner product, since spherical distance is defined by the inner product, the spherical distance is preserves. Thus we may perform an isometry such that $P = (0, 0, 1)$. We can then perform a rotation such that Q maps to a point on the xz plane. We can then write $Q = (\sin(\beta), 0, \cos(\beta))$ and $R = (\sin(\gamma)\cos(\alpha), \sin(\gamma)\sin(\alpha), \cos(\gamma))$. Calculating this gives our required formula. □

Proposition 9.2.2.5: Triangle Inequality

Let PQR be a spherical triangle of S^n whose sides are shorter arcs given by $\alpha, \beta, \gamma \leq \pi$. Then

$$\alpha \leq \beta + \gamma$$

with equality if and only if PQR is collinear.

Proof. Notice that by definition, $\alpha, \beta, \gamma \in [0, \pi]$ thus $\sin(\beta)$ and $\sin(\gamma)$ is greater than or equal to 0. Since $\cos(a) \geq -1$, we must have from the above proposition,

$$\begin{aligned} \cos(\alpha) &\geq \cos(\beta)\cos(\gamma) - \sin(\beta)\sin(\gamma) \\ &= \cos(\beta + \gamma) \end{aligned}$$

We split it into two cases. If $\beta + \gamma \leq \pi$, then $\cos(x)$ being decreasing function on $[0, \pi]$ implies $\alpha \leq \beta + \gamma$ and we are done. Equality is given as long as $\cos(a) = -1$, which means that $a = \pi$, which forces $\beta + \gamma = \pi$ from the inequality.

Now if $\beta + \gamma > \pi$, then the triangle inequality is trivial since $\alpha \leq \pi < \beta + \gamma$. Now equality occurs if and only if $a = \pi$. Then this is true if and only if $\cos(\alpha) = \cos(\beta + \gamma)$ which is true if and only if $\alpha + \beta + \gamma = 2\pi$ and we are done. □

Proposition 9.2.2.6

The spherical metric is indeed a metric on S^n. And thus S^n is a metric space.

Proof. □

9.2.3 Isometries in S^n

Proposition 9.2.3.1

An isometry on S^n preserves antipodal points and great circles.

Proposition 9.2.3.2

Every isometry in S^n can be extended to a Euclidean Isometry of $\mathbb{R}^{n+1}$. Every Euclidean isometry $T : \mathbb{R}^{n+1} \to \mathbb{R}^{n+1}$ can be restricted to an spherical isometry $T|_{S^n} : S^n \to S^n$.

Proposition 9.2.3.3

Every isometry $T : S^n \to S^n$ is of the form $T(x) = Ax$ for all $x \in S^n$, where $A \in O(n+1)$.

Corollary 9.2.3.4

There exists a group isomorphism

$$\text{Isom}(S^n, d_{S^n}) \cong O(n+1)$$

9.2.4 Geometry of the 2-Sphere

Proposition 9.2.4.1

Let C, D be two distinct great circles, then their intersection is a pair of antipodal points.

Theorem 9.2.4.2: Girad's Theorem

Let $\triangle PQR$ be a spherical triangle on S^2. With P, Q, R distinct points and such that the interior of $\triangle PQR$ does not intersect the great circles which contain the sides of $\triangle PQR$. Then the area of the triangle is equal to $\angle PQR + \angle PRQ + \angle QPR - \pi$.

The reason that we have this much constraints is for us to make sure that we are talking about the same triangle which should be the smallest triangle among the many triangles created from the great circles (or at least those without a segment of a great circle in it).

9.3 Hyperbolic Geometry

9.3.1 Lorentz Products

We first develop the necessary tools to develop the hyperbolic space.

Definition 9.3.1.1: Lorentz Inner Product

Let $x, y \in \mathbb{R}^n$. Define the Lorentz inner product to be

$$\langle x, y\rangle_L = -x_1y_1 + x_2y_2 + x_3y_3 + \cdots + x_ny_n$$

Definition 9.3.1.2: Lorentz Norm

Define the Lorentz norm of a vector $x \in \mathbb{R}^n$ to be

$$\|x\|_L = \sqrt{\langle x, x\rangle_L}$$

where ther square root to be positive or positive imaginary or 0.

Note that the Lorentz norm allows imnaginary numbers and it could also be negative and thus is not a norm in the usual sense.

Proposition 9.3.1.3: Hyperbolic Polarization Identity

Let $x, y \in \mathbb{R}^n$. Then

$$\langle x, y\rangle_L = \frac{1}{4}\|x+y\|_L^2 - \frac{1}{4}\|x-y\|_L^2$$

Definition 9.3.1.4: Classification of Vectors

Let $x \in \mathbb{R}^n$. We say that x is

- space-like if $\|x\|_L > 0$
- light-like if $\|x\|_L = 0$
- time-like if $\|x\|_L \in \mathbb{C} \setminus \mathbb{R}$

Definition 9.3.1.5: Lorentz Orthnormal

A set of vectors $v_1, \dots, v_n \in \mathbb{R}^n$ is Lorentz Orthgonal if $\langle v_i, v_j\rangle_L = 0$ for $i \neq j$. They are said to be Lorentz Orthonormal if

$$\langle v_i, v_j\rangle_L = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } 2 \leq i = j \leq n \\ -1 & \text{if } i = j = 1 \end{cases}$$

Lemma 9.3.1.6

If $v_1, \dots, v_n$ are Lorentz Orthonormal, then they form a basis of $\mathbb{R}^n$.

Proof. We just have to show that they are linearly independent. Suppose that $\sum_{k=1}^n a_kv_k = 0$ for some $a_1, \dots, a_n \in \mathbb{R}$. For each $i \in \{2, \dots, n\}$, we have that

$$0 = \left\langle \sum_{k=1}^n a_kv_k, v_i \right\rangle_L = a_i$$

and for $i = 1$ we have that the inner product is $-a_1$. Thus $a_1 = \cdots = a_n = 0$ and we are done. □

Definition 9.3.1.7: Lorentz Cross Product

Let $x, y \in \mathbb{R}^3$. Define the Lorentz cross product of x and y to be

$$x \times_L y = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} (x \times y) = \begin{vmatrix} -e_1 & e_2 & e_3 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{vmatrix}$$

An easy way to remember everything for Lorentz related operations is that the first element in a coordinate related operation is always negative.

Lemma 9.3.1.8: Hyperbolic Binet Cauchy Identity

For $x, y, z, w \in \mathbb{R}^3$, the following identity holds.

$$\langle x \times_L y, z \times_L w \rangle_L = -\langle x, z \rangle_L \cdot \langle y, w \rangle_L + \langle x, w \rangle_L \cdot \langle y, z \rangle_L$$

Now that we have mostly developed a sufficient amount linear algebra, we turn to study matrices that corresponds to Lorentz operations.

Definition 9.3.1.9: Lorentz Orthgonal Matrices

Let $J = \begin{pmatrix} -1 & 0 \\ 0 & I_{n-1} \end{pmatrix}$. A $n \times n$ real matrix is Loretnz orthgonal if

$$A^T J A = J$$

Proposition 9.3.1.10

The following are true for any $x, y \in \mathbb{R}^n$.

- $\langle x, y \rangle_L = x^T J y = \langle Jx, y \rangle$
- $x \times_L y = J(x \times y) = (Jy) \times (Jx)$

J will often replace the function of the identity in Hyperbolic space. From time to time we will see it reappear.

Proposition 9.3.1.11

Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear map. The following are equivalent.

- T is a Lorentz Transformation
- The matrix of T is Lorentz Orthogonal
- $\|T(x)\|_L = \|x\|_L$ for all $x \in \mathbb{R}^n$

Proof.

- (1) $\iff$ (2): Suppose that T is a Lorentz transformation. Let the matrix A represent T inthe standard basis so that it has columns $a_1, \dots, a_n$. This means that $T(e_k) = a_k$. Then we have

$$\begin{aligned}
A^TJA &= \begin{pmatrix} a_1^T \\ \vdots \\ a_n^T \end{pmatrix} J \begin{pmatrix} a_1 & \cdots & a_n \end{pmatrix} \\
&= \begin{pmatrix} a_1^T \\ \vdots \\ a_n^T \end{pmatrix} \begin{pmatrix} Ja_1 & \cdots & Ja_n \end{pmatrix} \\
&= \begin{pmatrix} a_1^TJa_1 & \cdots & a_1^TJa_n \\ \vdots & \ddots & \vdots \\ a_n^TJa_1 & \cdots & a_n^TJa_n \end{pmatrix} \\
&= \begin{pmatrix} \langle a_1, a_1\rangle_L & \cdots & \langle a_1, a_n\rangle_L \\ \vdots & \ddots & \vdots \\ \langle a_n, a_1\rangle_L & \cdots & \langle a_n, a_n\rangle_L \end{pmatrix}
\end{aligned}$$

Thus we can see that $A^TJA = J$ if and only if $T(e_1), \dots, T(e_n)$ are Lorentz orthonormal if and only if T is a Lorentz transformation.

- (1) $\implies$ (3): Suppose that T is a Lorentz transoformation. Then trivially

$$\|T(x)\|_L^2 = \langle T(x), T(x)\rangle_L = \langle x, x\rangle_L = \|x\|_L^2$$

Taking the positive square root gives our result.

- (3) $\implies$ (1): Using the polarization identity, we have that

□

9.3.2 Lorentz Transformations and $O(1, n)$

Definition 9.3.2.1: Lorentz Transformation

A map $\phi : \mathbb{R}^n \to \mathbb{R}^n$ is a Lorentz transformation if it preserves the Lorentz inner product:

$$\langle \phi(x), \phi(y)\rangle_L = \langle x, y\rangle_L$$

for all $x, y \in \mathbb{R}^n$. We say that ϕ is positive if both x and $\phi(x)$ has their first coordinate larger than 0.

Proposition 9.3.2.2

A map $T : \mathbb{R}^n \to \mathbb{R}^n$ is a Lorentz Transformation if and only if $\{T(e_1), \dots, T(e_n)\}$ is a Lorentz orthonormal basis and T is linear.

Proof. Suppose that T is a Lorentz transformation. Then since it preserves the Lorentz inner product and $e_1, \dots, e_n$ is trivially Lorentz orthonormal, we know that $T(e_1), \dots, T(e_n)$ are Lorentz orthonormal and we are done. Now let $x = \sum_{k=1}^n x_ke_k$ and $T(x) = \sum_{k=1}^n b_kT(e_k)$. Then

$$-b_1 = \langle T(x), T(e_1)\rangle_L = \langle x, e_1\rangle_L = -x_1$$

and

$$b_k = \langle T(x), T(e_k)\rangle_L = \langle x, e_k\rangle_L = -x_k$$

for $k \in \{2, \dots, n\}$ thus T is linear.

Suppose now that $T(e_1), \ldots, T(e_n)$ is Lorentz orthonormal and T is linear. Then

$$\begin{aligned}
\langle T(x), T(y)\rangle_L &= T(x)^T JT(y) \\
&= \left(\sum_{k=1}^n x_k T(e_k)\right)^T J \left(\sum_{k=1}^n y_k T(e_k)\right) \\
&= \sum_{k=1}^n \sum_{j=1}^n x_k y_j T(e_k)^T JT(e_j) \\
&= \sum_{k=1}^n \sum_{j=1}^n x_k y_j e_k J e_j \\
&= \left(\sum_{k=1}^n x_k e_k\right)^T J \left(\sum_{k=1}^n y_k e_k\right) \\
&= \langle x, y\rangle_L
\end{aligned}$$

□

Definition 9.3.2.3: Lorentz Group

The Lorentz group is the group of Lorentz orthogonal $n+1, n+1$ matrices denoted

$$O(1,n) = \{A \in M_{n+1\times n+1}(\mathbb{R})\}|A^T JA = J\}$$

Proposition 9.3.2.4

$O(1,n)$ is a group.

Proof. □

Proposition 9.3.2.5

Let $A \in M_{m\times n}(\mathbb{R})$. Then the following are equivalent.

- The map $T(x) = Ax$ is a Lorentz transformation
- $\|T(x)\|_L = \|x\|_L$ for all $x \in \mathbb{R}^n$
- A is Lorentz orthogonal.

Corollary 9.3.2.6

Every Lorentz Transformation corresponds to an element of $O(1,n)$ and every element of $O(1,n)$ gives rise to a Lorentz Transformation.

Proof. Let T be a lorentz transformation.

Now let T be a hyperbolic isometry. Define a Lorentz transformation S by $S|_{H^n} = T$ □

9.3.3 Positive Lorentz Transformations

Definition 9.3.3.1: Lorentz Transformation

A map $\phi : \mathbb{R}^n \to \mathbb{R}^n$ is a positive Lorentz transformation if the following are true

- ϕ is a Lorentz transformation
- x is time-like if and only if $T(x)$ is time-like for $x \in \mathbb{R}^n$.

Definition 9.3.3.2: Positive Lorentz Group

The positive Lorentz group is the set of all elements in $O(1, n)$ which maps positive time-like vectors bijectively to positive time-like vectors, denoted $O^+(1, n)$

Proposition 9.3.3.3

$O^+(1, n)$ is a group.

Proof. □

Lemma 9.3.3.4

Let $A \in O(1, n)$. Then $A \in O^+(1, n)$ if and only if $a_{1,1} > 0$.

This is extremely powerful since it gives a fairly easy way to check whether a Lorentz orthogonal matrix is an isometry.

Corollary 9.3.3.5

Every Positive Lorentz Transformation corresponds to an element of $O^+(1, n)$ and every element of $O^+(1, n)$ gives rise to a Positive Lorentz Transformation.

Proof. Let T be a lorentz transformation.

Now let T be a hyperbolic isometry. Define a Lorentz transformation S by $S|_{H^n} = T$ □

Lemma 9.3.3.6

For every set of linearly independent $a, b, c \in H^n$, there exists an element $A \in O^+(1, n)$ such that

$$\begin{aligned} Aa &= (1, 0, \dots, 0) \\ Ab &= (b_1, b_2, 0, \dots, 0) \text{ with } b_1 > 0 \\ Ac &= (c_1, c_2, c_3, 0 \dots, 0) \text{ with } c_1 > 0 \end{aligned}$$

Proof. Let $a, b, c, z_4, \dots, z_{n+1}$ be a basis for H^n. We construct a Lorentz orthonormal basis for $\mathbb{R}^{n+1}$. Since $a \in H^n$ implies $\|a\| = i$, we can choose $v_1 = a$. We find the basis by induction, where v_k is constructed by $a, b, c, z_4, \dots, z_k$. Now let $w_2 = b + \langle v_1, b\rangle_L v_1$. Observe that

$$\begin{aligned} \langle w_2, v_1\rangle_L &= \langle b, v_1\rangle_L + \langle v_1, b\rangle_L \langle v_1, v_1\rangle_L \\ &= \langle b, v_1\rangle_L - \langle v_1, b\rangle_L && (\langle v_1, v_1\rangle_L = -1) \\ &= 0 \end{aligned}$$

Thus w_2 is Lorentz orthogonal to v_1. Now just choose $v_2 = \frac{w_2}{\|w_2\|_L}$. Note that this is similar to the Gram-schimdt procedure in Euclidean Space. In general, suppose that $v_1, \dots, v_k$ are now made orthogonal. Define

$$w_{k+1} = a - \sum_{i=1}^{k} \langle z_{k+1}, v_i \rangle_L v_i$$

Then w_{k+1} will be orthogonal to all of $v_1, \dots, v_k$ thus we can choose $v_{k+1} = \frac{w_{k+1}}{\|w_{k+1}\|}$. Thus a Lorentz orthonormal basis $v_1, \dots, v_{n+1}$ is formed.

With respect to this basis, we have the desired results. □

9.3.4 Hyperbolic Space

Definition 9.3.4.1: Hyperbolic Space

The n dimensional hyperbolic space is the space

$$H^n = \{x \in \mathbb{R}^{n+1} | \|x\|_L = i \text{ and } x_1 > 0\}$$

along with the metric hyperbolic metric

$$d_{H^n}(x, y) = \cosh^{-1}(-\langle x, y \rangle_L)$$

In order to prove that H^n is a metric space, we need to first develop the notion of lines and triangles and angles in H^n.

Definition 9.3.4.2: Classification of Subspaces

We say that a vector subspace of $\mathbb{R}^n$ is

- time-like if V has at least one time-like vector
- space-like if every nonzero vector in V is space-like
- light-like otherwise

To check the type of subspace, we first find whether there is one time-like vector, we then check whether there is one light-like vector.

Proposition 9.3.4.3

There is a parametrization between $\mathbb{R}^2$ and H^2 given by

$$f(t, \theta) = (\cosh(t), \cos(\theta)\sinh(t), \sin(\theta)\sinh(t))$$

for $t \in [0, \infty)$ and $\theta \in [0, 2\pi]$ where $\mathbb{R}^2$ is in polar coordinates and H^2 in cartesian coordinates.

Proof. □

Definition 9.3.4.4: Hyperbolic Lines and Lorentz Planes

A Lorentz plane is a 2 dimensional vector subspace of $\mathbb{R}^{n+1}$ that contains a timelike vector. A hyperbolic line is the intersection of H^n with any Lorentz Plane.

Definition 9.3.4.5: Collinearity

Three points $x, y, z \in H^n$ are hyperbolically collinear if and only if there is a hyperbolic line L of H^n with $x, y, z \in L$.

We need the following theorem to show that d_{H^n} is a metric.

Lemma 9.3.4.6

For every $P \neq Q \in H^n$, there exists a unique hyperbolic line L containing P and Q.

Lemma 9.3.4.7

Every line through the origin in $\mathbb{R}^{n+1}$ intersects H^n in at most 1 point.

Proof. Let $y = \lambda x$ for some $x, y \in H^n$. Then by definition of H^n, we must have

$$\langle x, x\rangle_L = \langle y, y\rangle_L = -1$$

But the Lorentz inner product is linear. Thus $\langle y, y\rangle_L = \lambda^2\langle x, x\rangle_L$ and thus $\lambda = \pm 1$. Since both x_1 and y_1, the first component of x and y must be positive, we must have $\lambda = 1$ and we are done. □

Definition 9.3.4.8: Hyperbolic Triangles

A hyperbolic triangle, denoted $\triangle PQR$, consists of three distinct, non-collinear points $P, Q, R \in H^n$, and the finite hyperbolic line segments joining each pair of points, and the finite area enclosed by these lines.

Definition 9.3.4.9: Hyperbolic Angles

Let $\triangle PQR$ be a hyperbolic triangle in $H^2 \subset \mathbb{R}^3$. Define the hyperbolic angle a at P to be $a \in [0, \pi]$ such that

$$\cos(a) = \frac{\langle P \times_L R, P \times_L Q\rangle_L}{\|P \times_L R\|_L \|P \times_L Q\|_L}$$

To prove the triangle inqeuality for d_{H^n} we need the following theorem.

Theorem 9.3.4.10

Let $x, y, z \in H^n$ be distinct. Let $\alpha = d_{H^n}(z, y)$, $\beta = d_{H^n}(x, z)$, $\gamma = d_{H^n}(x, y)$ and a be the hyperbolic angle of $\triangle xyz$ at x. Then

$$\cosh(\alpha) = \cosh(\beta) \cdot \cosh(\gamma) - \sinh(\beta) \cdot \sinh(\gamma) \cdot \cos(a)$$

Proof. WLOG we can assume x, y, z takes the form $x = (1, 0, \dots, 0)$, $y = (\cosh(t_1), \sinh(t_1), 0, \dots, 0)$ and $z = (\cosh(t_2), \cos(\theta)\sinh(t_2), \sin(\theta)\sinh(t_2), 0, \dots, 0)$. This is because Loretnz transformation preserves the Loretnz inner product and thus the distance function remains unchanged. Since the trailing zeroes make no contributions to the calculations, we assume we are working on H^2. Now note that

$$x \times_L \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} 0 \\ -a_3 \\ a_2 \end{pmatrix}$$

Thus we have

$$\begin{aligned}\cos(a) &= \frac{\langle x \times_L y, x \times_L z\rangle_L}{\|x \times_L y\|_L \|x \times_L z\|_L} \\ &= \frac{1}{|\sinh(t_1)||\sinh(t_2)|}\left\langle \begin{pmatrix}0\\0\\\sinh(t_1)\end{pmatrix}, \begin{pmatrix}0\\-\sin(\theta)\sinh(t_2)\\\cos(\theta)\sinh(t_2)\end{pmatrix}\right\rangle_L \\ &= \cos(\theta)\end{aligned}$$

Now we have that

$$\begin{aligned}\cosh(\gamma) &= -\langle x, y\rangle_L = \cosh(t_1) \\ \cosh(\beta) &= -\langle x, z\rangle_L = \cosh(t_2) \\ \cosh(\alpha) &= -\langle z, y\rangle_L = \cosh(t_1)\cosh(t_2) - \cos(\theta)\sinh(t_1)\sinh(t_2)\end{aligned}$$

Combining these and using the fact that $\cos(a) = \cos(\theta)$ gives our result. □

Theorem 9.3.4.11

The hyperbolic metric is indeed a metric on H^n and thus the hyperbolic space is a metric space.

Proof.

- We first prove that the hyperbolic metric is well defined. This means that we need to prove that $-\langle x, y\rangle_L \geq 1$. Let $x, y \in H^n$. Denote $x' = (1, x_2, \dots, x_{n+1})$ and $y' = (1, y_2, \dots, y_{n+1})$. Then we have that

  $$x_1^2 = 1 + x_2^2 + \cdots + x_n^2 = \|x'\|^2$$

  and likewise for y_1^2. Note that the norm on the right is the usual Euclidean norm. By Cauchy Schwarz we have that $\langle x', y'\rangle^2 \leq \|x'\|^2\|y'\|^2 = (x_1y_1)^2$ which implies that

  $$|1 + x_2y_2 + \cdots + x_ny_n| \leq |x_1y_1| = x_1y_1$$

  Rearraging, we have that $\langle x, y\rangle_L \leq -1$ and thus

  $$-\langle x, y\rangle_L \geq 1$$

- We want to show that $d_{H^n}(x, y) \geq 0$ with equality if and only if $x = y$. Trivially the image of $\cosh^{-1}$ is greater than or equal to 0 and $\cosh^{-1}(u) = 0$ means that $u = 1$ and $-\langle x, y\rangle_L = 1$. From the above proof, in order for this to be true, we need $\langle x', y'\rangle^2 = \|x'\|^2\|y'\|^2$ which is true if and only if x is a multiple of y. But we know that every line passing through the origin only intersects H^n in at most one point. Thus $x = y$ and we are done.

- We want to showt that $d_{H^n}(x, y) = d_{H^n}(y, x)$ for all $x, y \in H^n$. But clearly the Lorentz inner product is symmetric, thus we are done.

- To show the triangle inequality, let $x, y, z \in H^n$. Let $\alpha = d_{H^n}(z, y)$, $\beta = d_{H^n}(x, z)$, $\gamma = d_{H^n}(x, y)$ and a be the hyperbolic angle of $\triangle xyz$ at x. Using the above theorem, we find that

  $$\cosh(\alpha) \leq \cosh(\gamma)\cosh(\beta) + \sinh(\beta)\sinh(\gamma) = \cosh(\beta + \gamma)$$

  Since $\cosh(x)$ is increasing on $[0, \infty)$, this implies that $\alpha \leq \beta + \gamma$. Thus we are done.

□

Proposition 9.3.4.12

Let $R \in \text{Isom}(H^n, d_H)$. Then there exists a unique $A \in O^+(1,n)$ with $R = T_A|_{H^n}$, where T_A is the linear map on $\mathbb{R}^{n+1}$ given by $T(x) = Ax$.

Corollary 9.3.4.13

There is a group isomorphism between $\text{Isom}(H^n)$ and $O^+(1,n)$.

Becareful that all isometries of H^n is not the Loretnz group but only of the positive Lorentz group.

9.3.5 Hyperbolic Lines

Lemma 9.3.5.1

If $x \in \mathbb{R}^n$ with $\langle x, x\rangle_L < 0$, and $w \in \mathbb{R}^n \setminus \{0\}$ with $\langle x, w\rangle_L = 0$, then $\langle w, w\rangle_L > 0$. In other words, any vector that is Lorentz orthogonal with a time-like vector must be a space-like vector.

Proof. Let $x' = x - x_1e_1$ and $w' = w - w_1e_1$. Then $\langle x', x'\rangle_L \geq 0$ and $\langle x, x\rangle_L = -x_1^2 + \langle x', x'\rangle_L$. This means that $x_1 \neq 0$. If $w_1 = 0$, then since $w \neq 0$, we must have $\langle w, w\rangle_L = \langle w', w'\rangle_L > 0$ and we are done.

Now suppose that $w_1 \neq 0$. Then $x_1w - w_1x$ has zero e_1 component. I claim that this vector is non-zero. Indeed if $x_1w = w_1x$, then

$$0 = \frac{x_1}{w_1}\langle x, w\rangle_L = \langle x, \frac{x_1}{w_1}w\rangle_L = \langle x, x\rangle_L < 0$$

Now we have that

$$\begin{aligned} 0 \leq \|x_1w - w_1x\|_L^2 &= \langle x_1w - w_1x, x_1w - w_1x\rangle_L \\ &= x_1^2\langle w, w\rangle_L + w_1^2\langle x, x\rangle_L \end{aligned}$$

Bu this means that

$$\langle w, w\rangle_L \geq -\frac{w_1^2}{x_1^2}\langle x, x\rangle_L > 0$$

and we are done. □

In layman terms, this means that the orthogonal subspace of a time-like vector is a subspace with all vectors being space-like.

Definition 9.3.5.2

Let L_1, L_2 be two distinct lines in H^2 with Lorentz plane Π_1 and Π_2 respectively. Let $v \in \Pi_1 \cap \Pi_2 \setminus \{0\}$. Let $V = \Pi_1 \cap \Pi_2$.

- If V is time like, L_1 and L_2 intersect at a point x and $V = \mathbb{R}v = \Pi_1 \cap \Pi_2$
- If V is space like, L_1 and L_2 are parallel and diverge
- If V is light like, L_1 and L_2 are ultraparallel

The following theorem shows that Euclid's Postulate does not hold in hyperbolic space.

Theorem 9.3.5.3

Let $x \in H^2$. Let L be a line in H^2. Then there are infinitely many lines in H^2 which pass through x and do not intersect L.

9.3.6 Hyperbolic Triangles and Angles

Proposition 9.3.6.1

The hyperbolic angles is invariant under hyperbolic isometries.

Proposition 9.3.6.2

Let $\triangle PQR$ be a hyperbolic triangle. Let a, b, c be the angles at P, Q, R respectively. Then

$$a + b + c = \pi - \text{Area}(\triangle PQR)$$

9.3.7 Normal Form Theorem

Lecture did not have time to thoroughly prove things and so do I. Every Lorentz Transformation must have one of the following four forms:

Lorentz Translation:

$$\begin{pmatrix} \cosh(\beta) & \sinh(\beta) & 0 \\ \sinh(\beta) & \cosh(\beta) & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

This transformation does not have a fixed point in H^2. Orientation is preserved.

Lorentz Glide:

$$\begin{pmatrix} \cosh(\beta) & \sinh(\beta) & 0 \\ \sinh(\beta) & \cosh(\beta) & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

This transformation does not have a fixed point in H^2. Orientation is also reversed

Rotation about the point $(1, 0, 0)$:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cosh(\beta) & \sinh(\beta) \\ 0 & \sinh(\beta) & \cosh(\beta) \end{pmatrix}$$

This transformation has one fixed point, namely $(1, 0, 0)$ and is orientation preserving.

Reflection along $x_2 = 0$:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

This transformation fixes the line $x_2 = 0$ and is orientation reversing.

9.4 Projective Geometry

9.4.1 Projective Space

Lemma 9.4.1.1

Let $\mathbb{F}$ be a field. The relation $\sim$ in $\mathbb{F}^{n+1}$ where $(x_0, \dots, x_n) \sim (y_0, \dots, y_n)$ if and only if $y_i = \lambda x_i$ for all $i \in \{1, \dots, n\}$ with $\lambda \in \mathbb{F}$ is an equivalence relation.

Definition 9.4.1.2: Projective Space

The equivalence relation $\sim$ on $\mathbb{F}^{n+1}$ induces the projective space with elements in it being 1 dimensional subspaces of $\mathbb{F}^n$, written as

$$\mathbb{P}(\mathbb{F}^{n+1}) = \frac{\mathbb{F}^{n+1} \setminus \{0\}}{\sim}$$

We use $\mathbb{P}^n$ to denote $\mathbb{P}(\mathbb{R}^{n+1})$. Also define the dimension of $\mathbb{P}^n$ to be n.

We use $\mathbb{R}^{n+1}$ as our vector space since every finite dimensional vector space is isomorphic to $\mathbb{R}^n$ for some n.

Proposition 9.4.1.3

There is a bijection between $\mathbb{P}^n$ and the set of lines through the origin in $\mathbb{R}^{n+1}$.

Proposition 9.4.1.4

We have that

$$\mathbb{P}^n \cong \mathbb{R}^n \cup \mathbb{R}^{n-1} \cup \cdots \cup \mathbb{R} \cup \{\infty\}$$

Proof. The proof is done by induction. We first consider $\mathbb{P}(\mathbb{R})$. Recall that $x \sim y$ if and only if $y = tx$ for some $t \neq 0$. This means that $[x]$ represents the same point as long as $x \in \mathbb{R} \setminus \{0\}$. We also simply identify $[0]$ with ∞.

We also consider $\mathbb{P}(\mathbb{R}^2)$ for better illustration. All elements of $\mathbb{P}(\mathbb{R}^2)$ are of the form $[x_0 : x_1]$. We consider the case that $x_0 \neq 0$. Since $x_0 \neq 0$, we can divide out x_0 to get the same coordinate $[1 : x_1] \in \mathbb{P}(\mathbb{R}^2)$. For each $x_1 \in \mathbb{R}$, $[1 : x_1]$ represents different coordinates in $\mathbb{P}(\mathbb{R}^2)$ thus we can easily identify it with $\mathbb{R}$. Now if $x_0 = 0$, $[0 : x_1]$ would represent the same coordinate for any $x_1 \in \mathbb{R}$. This is precisely the same situation as $\mathbb{P}(\mathbb{R})$. Thus we can virtually ignore the first coordinate and identify it with $\mathbb{P}(\mathbb{R})$. Thus we have shown that $\mathbb{P}(\mathbb{R}^2) = \mathbb{R}^2 \cup \mathbb{P}(\mathbb{R})$.

Now through the induction hypothesis, we just have to show that $\mathbb{P}(\mathbb{R}^n) = \mathbb{R}^n \cup \mathbb{P}(\mathbb{R}^{n-1})$. But this is done in a similar fashion. We can use $[1 : x_1 : \cdots : x_n]$ for $x_1, \dots, x_n \in \mathbb{R}$ to identify $\mathbb{R}^n$ and $\mathbb{P}(\mathbb{R}^{n-1})$ with $[0 : x_1 : \cdots : x_n]$. □

Theorem 9.4.1.5

We have that

$$\mathbb{P}(\mathbb{R}^n) = \frac{S^n}{\pm}$$

where ± is the equivalence relation of antipodal points.

9.4.2 Projective Linear Subspaces

Definition 9.4.2.1: Projective Linear Subspaces

Let $U \subseteq V$ be a vector subspace. Define the projective linear subspace to be

$$\mathbb{P}(U) = \frac{U \setminus \{0\}}{\sim} \subseteq \mathbb{P}(V)$$

with its dimension defined to be $\dim(\mathbb{P}(U)) = \dim(U) - 1$. We define $\dim(\emptyset) = -1$ for conventions.

Definition 9.4.2.2: Classification of Subspaces

Let $U \subseteq V$ be a vector subspace where $\dim(V) = n$. We say that $\mathbb{P}(U)$ is

- A single point if $\dim(U) = 1$, or $\dim(\mathbb{P}(U)) = 0$
- A line if $\dim(U) = 2$, or $\dim(\mathbb{P}(U)) = 1$
- A plane if $\dim(U) = 3$, or $\dim(\mathbb{P}(U)) = 2$
- A hyperplane if $\dim(U) = n - 1$, or $\dim(\mathbb{P}(U)) = n - 2$

In particular, note that lines in $\mathbb{R}^{n+1}$ become points. Let $x \in \mathbb{R}^{n+1}$ be fixed. Take the line $y = tx$ for $t \in \mathbb{R}$ as an example. Observe that in the projective space. $x \sim y$ if and only if $y = tx$ for some $t \in \mathbb{R}$. Naturally $y = tx$ becomes a point.

Definition 9.4.2.3: Projective Cone and Span

Let V be a vector subspace and let $U \subset \mathbb{P}(V)$ be a subset. Define the projective cone of U to be

$$\tilde{U} = \bigcup_{v \in U} v$$

Define the span of U to be

$$\langle U \rangle = \mathbb{P}(\text{span}(\tilde{U}))$$

the smallest projective linear subspace of $\mathbb{P}(V)$ containing U.

Theorem 9.4.2.4: Dimemsion Theorem

Let $E, F \subset \mathbb{P}^n$ be projective linear subspaces. Then

$$\dim(E \cap F) = \dim(E) + \dim(F) - \dim\langle E, F \rangle$$

Proof. Let □

Theorem 9.4.2.5

Any two distinct lines L_1 and L_2 in $\mathbb{P}^2$ intersect in a point.

Proof. Note that L_1 and L_2 are projections of two planes A_1, A_2 through the origin. Since A_1, A_2 have dimension 2 and are distinct. Thus their intersection must have dimension 1 and their span has dimension 3. But $\langle L_1, L_2 \rangle = \mathbb{P}(\text{span}(A_1, A_2)) = \mathbb{P}(\mathbb{R}^3)$ and thus $\dim(\langle L_1, L_2 \rangle) = 2$. By the above theorem, we must have $\dim(L_1 \cap L_2) = 0$ and thus it is exactly a point. □

9.4.3 Projective Transformations

Definition 9.4.3.1: Projective Transformations

A projective transformation of $\mathbb{P}^n$ is a map $T|_A : \mathbb{P}^n \to \mathbb{P}^n$ defined by

$$T([x]) = [Ax]$$

where $A \in \text{GL}(n+1, \mathbb{R})$.

Definition 9.4.3.2: Projective General Linear Group

The projective general linear group of a vector space V over k with dimension n is the group of all invertible linear transformations unique up to scalar multiplication. That is, we say that $T_1 \sim T_2$ if $T_1 = \lambda T_2$ for some $\lambda \in k$ and $\lambda \neq 0$. It is denoted as $\text{PGL}(n+1, k)$.

Definition 9.4.3.3: Projective Frame of Reference

A projective frame of reference for $\mathbb{P}^n$ is an ordered set of $n+2$ points, $P_0, \dots, P_{n+1} \in \mathbb{P}^n$ such that any $n+1$ points are linearly independent. The standard frame of reference is just $[e_1], \dots, [e_{n+1}], [e_1 + \cdots + e_{n+1}]$.

Proposition 9.4.3.4

There is a bijection between projective transformations of $\mathbb{P}^n$ and projective frames of references of $\mathbb{P}^n$ as follows:

$$\phi(T) = \{T([e_1]), \dots, T([e_{n+1}]), T([e_1 + \cdots + e_{n+1}])\}$$

where T is the projective transformation.

This means that specifying any $n+2$ points in $\mathbb{P}^n$ such that they form a projective frame of reference induces a unique projective transformation.

9.4.4 Perspectivities

Definition 9.4.4.1: Perspectivities

Let Π_1, Π_2 be two hyperplanes in $\mathbb{P}^n$. A perspectivity $f : \Pi_1 \to \Pi_2$ from a point $O \in \mathbb{P}^n$ but not in Π_1 or Π_2 is a map given by

$$f(P) = \Pi_2 \cap \langle O, P \rangle$$

Lemma 9.4.4.2

Perspectivities are well defined.

Proof. We want to show that $f(P)$ is a point. We appeal to the dimension theorem. We have that

$$\dim(\Pi_2 \cap \langle O, P \rangle) = \dim(O) + \dim(\Pi_2) - \dim(\langle \Pi_2 \langle O, P \rangle \rangle)$$

Since O is necessarily not in Π_1, we have $\langle O, P \rangle$ is a projective line. Moreover, since O is not in Π_2, we must have that the span of the hyperplane Π_2 and a line $\langle O, P \rangle$ must be the projective space itself. Thus we have that

$$\dim(\Pi_2 \cap \langle O, P \rangle) = 0$$

This means that $f(P)$ must be a point in projective space. □

Definition 9.4.4.3: Cross Ratio

Let P, Q, R, S be distinct and ordered points on a projective line in $\mathbb{P}^n$. Define the cross ratio between the four projective points to be the ratio

$$(P, Q; R, S) = \left(\frac{p-r}{p-s}\right)\left(\frac{q-s}{q-r}\right)$$

where the ratio between vectors lying on the same projective line is defined to be the λ in $\frac{\lambda v}{v}$.

Proposition 9.4.4.4

Projective linear maps perserve perspectivity.

Proof. By linearity, we must have

$$\frac{T(\lambda v)}{T(v)} = \frac{\lambda v}{v}$$

□

Lemma 9.4.4.5

Perspectivities are projective linear maps.

9.4.5 Important Theorems

Definition 9.4.5.1: Triangles in Projective Space

Let $P, Q, R \in \mathbb{P}^n$ be distinct. The triangle $\triangle PQR$ in $\mathbb{P}^n$ is defined to be the three points and the three sides spanned by three pair of points.

Definition 9.4.5.2: In Perspective from a Point

Two triangles $\triangle PQR$ and $\triangle P'Q'R'$ in $\mathbb{P}^n$ are said to be in persepective from a point O if the lines passing through PP', QQ' and RR' intersect at a O.

Definition 9.4.5.3: In Perspective from a Line

Two triangles $\triangle PQR$ and $\triangle P'Q'R'$ in $\mathbb{P}^n$ are said to be in persepective from a line L if the points $PQ \cap P'Q'$, $QR \cap Q'R'$ and $PR \cap P'R'$ are collinear.

Note that this definition makes sense since any two lines in projective space must meet at one point.

Theorem 9.4.5.4: Desargues' Theorem

If $\triangle PQR$ and $\triangle P'Q'R'$ are two distinct triangles in $\mathbb{P}^n$ which are in perspective from a point, then they are also in perspective from a line.

Theorem 9.4.5.5: Pappus' Theorem

Let L and L' be distinct projective lines in $\mathbb{P}^2$. Let P, Q, R in L but not in L' and P', Q', R' in L' but not in L all be distinct points. Then the intersection points $PQ' \cap P'Q$, $PR' \cap P'R$ and $QR' \cap Q'R$ are collinear.

9.4.6 Axiomatic Projective Geometry

Just as Euclidean geometry has 5 axioms to follow, we can also establish axioms for projective geometry.

Definition 9.4.6.1: Axiomatic Projective Geometry

An axiomatic projective plane (P, L, I) consists of a set P called the set of points, a set L, the set of lines and a relation $I \subseteq P \times L$, the incidence relation. For $l_1, l_2 \in L$, define

$$l_1 \cap l_2 = \{p \in P | p \in l_1, p \in l_2\}$$

These sets must satisfy the following four axioms:

1. Every line contains at least three distinct points:

$$\forall l \in L \exists \text{ distinct } x, y, z \in P \text{ such that } (x, l) \in I, (y, l) \in I, (z, l) \in I$$

2. Every point is contained in at least three distinct lines:

$$\forall x \in P \exists \text{ distinct } l, m, n \in L \text{ such that } (x, l) \in I, (x, m) \in I, (x, n) \in I$$

3. Any two points span a unique line:

$$\forall x \neq y \in P \exists ! l \in L \text{ such that } (x, l) \in I, (y, l) \in I$$

4. Any two distinct lines intesect at a unique point:

$$\forall l \neq m \in I \exists ! x \in P \text{ such that } (x, l) \in I, (x, m) \in I$$

Part IV

The Fundamentals of Algebra

Chapter 10

Groups and Rings

10.1 Introduction to Groups

10.1.1 Basic Concepts of Groups

We begin our extensive study with the definition of binary operations.

Definition 10.1.1.1: Binary Operation

A binary operation $*$ on a set G is a function

$$*: G \times G \to G$$

While it is not our main object of study, it is important to lay out its definition to prevent confusion and as a good practice. Then comes the main object of our next 5 chapters.

Definition 10.1.1.2: Groups

A group is an ordered pair $(G, *)$ where G is a set and $*$ is a binary operation on G satisfying the following axioms.

- $a * (b * c) = (a * b) * c$ for all $a, b, c \in G$
- There exists an element $e \in G$, called an identity of G, such that for all $a \in G$ we have $a * e = e * a = a$
- For each $a \in G$ there is an element a^{-1} of G, called an inverse of a, such that $a * a^{-1} = a^{-1} * a = e$.

The 3 axioms of a group feels weird and out of place. There is no explicit implications or motivations what-so-ever in its formulation. We simply want the operation to be associative, as well as having an identity and inverse. It seems unclear as to what kinds of objects are in fact groups. Now before we begin with the theorems, we give name to groups that emmit extra structure.

Definition 10.1.1.3: Abelian Groups

A group $(G, *)$ is abelian if $a * b = b * a$ for all $a, b \in G$.

Simply put, it is a group where its elements are commutative with each other. There are also a dozen of immediate results just from the definition of a group.

Proposition 10.1.1.4

If G is a group under the operation $*$, then

- The identity of G is unique

- For each $a \in G$, a^{-1} is unique
- $(a^{-1})^{-1} = a$ for all $a \in G$
- $(a * b)^{-1} = (b^{-1}) * (a^{-1})$

Proof. Let G be a group.

- Let e, f be identities of G. Let $a \in G$. Since e is the identity, $ef = e$. Since f is the identity, $ef = f$. Thus $e = ef = f$.
- Suppose that $b, c \in G$ are inverses of a. Since groups are associative, $(ba)c = b(ac)$. From the left, $(ba)c = ec = c$. From the right, $b(ac) = be = b$. Thus $b = (ba)c = b(ac) = c$.
- Suppose that $a^{-1} = b$. Since the inverse of a is b, we have $ab = e$. But since $ab = e$, the inverse of b is a.
- Suppose that the inverse of a is c and the inverse of b is d. Then $(ab)(dc) = a(bd)c = ac = e$. Thus the inverse of ab is $dc = b^{-1}a^{-1}$.

□

Do be aware that since in general a group is not abelian, the inverse law above is rather important as taking the inverse also inverse the order of multiplication. The first three facts allows us to establish uniqueness of the identity and the inverse, as well as having the fact that the inverse $f(a) = a^{-1}$ is bijective. The final two items are of the nature of the elements's behaviour in its group.

Definition 10.1.1.5: Order of an Element

For a group G and $x \in G$ define the order of x to be the smallest integer n such that $x^n = 1$, and denote this integer by $|x|$. If no positive power of x is the identity, the order of x is defined to be infinity and x is said to be of infinite order.

Aside from the order of an element, we also have the order of the entire group, which is an easier definition.

Definition 10.1.1.6: Order of a Group

Let G be a group. The order of G is the number of elements that G has, denoted by $|G|$.

Lemma 10.1.1.7

Let G be a group. Let $k \in \mathbb{N}$. Then $a^k = 1$ if and only if $|a||k$

Proposition 10.1.1.8

Let G be a group. Then

$$|g^k| = \frac{n}{\gcd(k, n)} = \frac{\operatorname{lcm}(k, n)}{k}$$

for any $k \in \mathbb{N}$

Proof. The equality between the two expressions is trivial since $\gcd(m, n) \times \operatorname{lcm}(m, n) = m \times n$.

Now let $b = \frac{n}{\gcd(k,n)}$ and $c = \frac{k}{\gcd(k,n)}$. Then we must have $\gcd(b,c) = 1$. Now

$$\begin{aligned}(g^k)^b &= g^{kb} \\ &= g^{\gcd(k,n)cb} \\ &= g^{nc} \\ &= (g^n)^c \\ &= 1\end{aligned}$$

Thus we must have $\left|g^k\right|$ divides $\frac{n}{\gcd(k,n)}$. Now we know that $g^{k\left|g^k\right|} = (g^k)^{\left|g^k\right|} = 1$. Thus we must have n divides $k\left|g^k\right|$, meaning $\gcd(k,n)b$ divides $\gcd(k,n)c\left|g^k\right|$ and b divides $c\left|g^k\right|$. We know that $\gcd(b,c) = 1$ thus b divides $\left|g^k\right|$. From the fact that $b = \frac{n}{\gcd(k,n)}$ we have $\left|g^k\right|$ divides b as well and thus $b = \left|g^k\right|$ and

$$\left|g^k\right| = \frac{n}{\gcd(k,n)}$$

□

In the case that two elements commute, the order of their product is somehow determined by their individual orders.

Proposition 10.1.1.9

Let G be a group. Let $a, b \in G$ be elements that commute with each other. Then $|ab|\,\Big|\,\text{lcm}(|a|, |b|)$.

10.1.2 Subgroups

The idea of a subgroup is that some (not all) elements in a group can also form a group in and of itself. Most of the structure of a group carries on to its subgroup, while the converse may not necessarily be true.

Definition 10.1.2.1: Subgroups

If $(G, *)$ is a group and H is a subset of G such that $(H, *)$ is a group $(H, *)$. Then H is called a subgroup of G and is denoted $H \leq G$.

Lemma 10.1.2.2

Let G be a group and $H \leq G$. Then $1_H = 1_G$.

Proof. If $1_H \neq 1_G$ then there will be two identities in G. □

To check that a subset of a group is a subgroup, it is not necessary to go through the four axioms of a group as it is rather tedious. Instead, we may turn to an easier criterion which involves less work.

Theorem 10.1.2.3: Criterion for a Subgroup

Let $(G, *)$ be a group. A subset H of G is a subgroup if and only if

- If $a, b \in H$ then $a * b \in H$
- If $a \in H$ then $a^{-1} \in H$

Proof. We first prove the forward implication. Suppose that H is a subgroup. Then by definition of a group the above are all true.

Suppose that H is satisfies the above two conditions. We show that it implies the four definitions of a group. Since $a, b \in H$ implies $ab \in H$, closure is satisfied. Since $H \subseteq G$, and G is associative, H is also associative. If $a \in H$ then $a^{-1} \in H$ thus inverse exists. Finally since H is closed, $aa^{-1} = 1 \in H$ thus identity is in H. Thus H is a group. □

We thern have subgroups that exists within every group. They are called trivial since we usually do not want to deal with them.

Lemma 10.1.2.4: Trivial Subgroups

Let G be a group. Then G and $\{1\}$ are subgroups of G.

Proof. Both can easily be seen to satisfy the subgroup criterion. □

The final proposition gives a new way of generating a subgroup given two subgroups. In particular, their overlapping part is also a subgroup in itself.

Proposition 10.1.2.5

Let G be a group and H, K subgroups of G. Then $H \cap K$ is a subgroup of G.

Proof. Suppose that H, K are subgroups of G. We show that $H \cap K$ with the subgroup criterion. We have that

- $1 \in H$ and $1 \in K$ thus $1 \in H \cap K$
- Suppose that $a, b \in H \cap K$. Then since $ab \in H$ and $ab \in K$ we have $ab \in H \cap K$
- Suppose that $a \in H \cap K$ since H and K are subgroups respectively, $a^{-1} \in H$ and $a^{-1} \in K$ thus $a^{-1} \in H \cap K$.

The subgroup criterion is satisfied thus $H \cap K$ is a subgroup of G. □

10.1.3 Cyclic groups

Cyclic groups appear in a lot of diverse areas of mathematics such as topology and number theory. It structure is simple enough to be employed in different areas while the notions it encapsulates is broad.

Definition 10.1.3.1: Cyclic Subgroups

Let G be a group. Define

$$\langle g \rangle = \{g^k | k \in \mathbb{Z}\}$$

to be the cyclic subgroup generated by g. In this case, g is the generator of $\langle g \rangle$.

A group G is called cyclic if there exists some $g \in G$ such that $G = \langle g \rangle$.

The following few propositions give some basic results on cyclic subgroups.

Proposition 10.1.3.2

Let $g \in G$. Then $\langle g \rangle \leq G$.

Proof. We prove that it satifies the subgroup criterion.

- $g^0 = 1$ thus $1 \in \langle g \rangle$
- g^{-1} is in $\langle G \rangle$ is trivial
- Let $a, b \in \langle g \rangle$. Then $a = g^n$ and $b = g^m$ for some $n, m \in \mathbb{Z}$ thus $ab = g^{n+m} \in \langle g \rangle$.

The subgroup criterion is satisfied thus we are done. □

In particular, any nontrivial group has a cyclic subgroup simply by taking

Lemma 10.1.3.3

Cyclic groups are abelian.

Proof. Quite obvious since $g^n \cdot g^m = g^{n+m} = g^{m+n} = g^m \cdot g^n$. □

The following result would seem quite trivial. It makes sense to think that a subgroup of a cyclic group would be cyclic. It is however a non-trivial result and one of the first proofs that uses some notion in number theory, showing some linkage between the two subjects.

Proposition 10.1.3.4

Let $G = \langle g \rangle$ be cyclic. If $H \leq G$ then H is cyclic.

Proof. Let $H \leq G = \langle g \rangle$. Every element in H is of the form g^n for some $n \in \mathbb{Z}$. Take the smallest positive k such that $g^k \in H$. Consider another element $g^n \in H$. By the division algorithm, $n = qk + r$ for some $r \in \{0, \dots, k-1\}$. Thus we have

$$g^n = g^{kq+r} \iff g^r = (g^k)^{-q}(g^n)$$

Now since $g^k \in H$, $(g^k)^{-1} \in H$ and $(g^k)^q \in H$ thus $(g^k)^q \in H$ and $(g^k)^q(g^n) \in H$ by inverse and closure of a group. Thus $g^r \in H$. However this is a contradiction since k is the least power of g such that $g^k \in H$ but g^r with $r < k$ is now in H. Thus we must have $r = 0$ and $g^n = (g^k)^q$. This applies for every $g^n \in H$ thus $H = \langle g^k \rangle$. □

Proposition 10.1.3.5

Suppose that $G = \langle g \rangle$. Then $|G| = |g|$.

Proof. Suppose that $|g| = n$ Note that $g^{n+k} = g^k$ for any $k \in \{0, \dots, n-1\}$. Thus $\{g^n | n \in \mathbb{Z}\}$ has n distinct elements. Thus $|G| = n$.

If $|g| = \infty$ then $g^n \neq 1$ for all n not equal to 0. Thus $\{g^n | n \in \mathbb{Z}\}$ are all distinct and $|G| = \infty$. □

Recall that $|ab| \, | \, \text{lcm}(|a|, |b|)$ given that a and b commute. We have a stronger implication if we impose an extra condition.

Proposition 10.1.3.6

Let G be a group and $a, b \in G$ such that a, b commutes. If furthermore $\langle a \rangle \cap \langle b \rangle = \{1\}$, then

$$|ab| = \text{lcm}(|a|, |b|)$$

10.1.4 Homomorphisms and Isomorphisms

Homomorphisms and isomorphisms are important concepts not only in terms of the ability to classify groups with similar structure, but is also useful in the creation of new groups, such as subgroups and quotient groups. It is one of the central concepts in abstract algebra.

Definition 10.1.4.1: Homomorphism

Let $(G, *)$ and $(H, \circ)$ be groups. A map $\phi : G \to H$ such that

$$\phi(x * y) = \phi(x) \circ \phi(y)$$

for all $x, y \in G$ is called a homomorphism.

Proposition 10.1.4.2

Let G_1, G_2 be groups that are homomorphic. Then the following are true

- $\phi(1) = 1$
- $\phi(g^{-1}) = \phi(g)^{-1}$ for all $g \in G_1$
- $\phi(g^n) = \phi(g)^n$ for all $n \in \mathbb{N}$ and all $g \in G_1$
- Let H be a subgroup of G_1. Then $\phi(H)$ is a subgroup of G_2

Proof.

- Let $\phi : G \to H$ be a homomorphism.

$$\begin{aligned} \phi(1) &= \phi(1 \cdot 1) \\ \phi(1) &= \phi(1) \cdot \phi(1) \\ \phi(1)\phi(1)^{-1} &= \phi(1)\phi(1)\phi(1)^{-1} \\ 1 &= \phi(1) \end{aligned}$$

- We have that
$$\phi(g)\phi(g^{-1}) = \phi(gg^{-1}) = \phi(1) = 1$$
and
$$\phi(g)\phi(g)^{-1} = 1$$
Also since the inverse is unique, we have that $\phi(g^{-1}) = \phi(g)^{-1}$
- We have that $\phi(g) \cdots \phi(g) = \phi(g \cdots g)$ by applying the definition of homomorphism $n - 1$ times
- We already know that $1 \in \phi(H)$. Let $h_1, h_2 \in \phi(H)$, then there exists $g_1, g_2 \in H$ such that $\phi(g_1) = h_1$ and $\phi(g_2) = h_2$. Since $g_1 g_2 \in H$, we have that
$$h_1 h_2 = \phi(g_1)\phi(g_2) = \phi(g_1 g_2) \in \phi(H)$$
Also if $h \in \phi(H)$ and $\phi(g) = h$ and also $g \in H$ implies $g^{-1} \in H$ and $\phi(g^{-1}) \in \phi(H)$, we have
$$\phi(g)\phi(g^{-1}) = \phi(gg^{-1}) = \phi(1) = 1$$
thus by the subgroup criterion $\phi(H)$ is a subgroup.

□

Structures that can be seen carrying over would be power, inverses, subgroups and identities. But these are only some of the capabilities of a homomorphism.

Definition 10.1.4.3: Kernel

Let $\phi : G \to H$ be a homomorphism. Define the kernel of ϕ to be

$$\ker(\phi) = \{g \in G | \ker(g) = 1_H\}$$

This notion is exactly parallel to that of linear algebra. They are both the kernel in the sense that they are the element that maps to the identity. In fact, as readers will see in module theory, the entire theory of linear algebra is simply a special case of module theory.

Proposition 10.1.4.4

Let $\phi : G \to H$ be a homomorphism. Then $\ker(\phi)$ is a subgroup of G.

Proof. We prove closure, identity and inverse as stated by the subgroup criterion.

- $\phi(1_G) = 1_H$, thus we have $1_G \in \ker(\phi)$
- If $a, b \in \ker(\phi)$, then $\phi(a) = 1$ and $\phi(b) = 1$ and
$$\phi(ab) = \phi(a)\phi(b) = 1$$
thus $ab \in \ker(\phi)$.
- If $a \in \ker(\phi)$ then
$$1 = \phi(1) = \phi(aa^{-1}) = \phi(a)\phi(a^{-1}) = \phi(a^{-1})$$
By the subgroup criterion $\ker(\phi) \leq G$.

□

Now we impose additional constraints on homomorphisms.

Definition 10.1.4.5: Isomorphism

Let G, H be groups. A map $\phi : G \to H$ is said to be an isomorphism if ϕ is a bijective homomorphism.

Isomorphisms are stricter in the sense that the structure between the two groups it reflects has a higher compatibility than that of homomorphism.

Proposition 10.1.4.6: Properties of Isomorphism

Let G, H be groups. If a map $\phi : G \to H$ is isomorphic, then

- ϕ^{-1} is an isomorphism
- $|G| = |H|$
- $\ker(\phi) = \{1\}$
- G is abelian if and only if H is abelian
- G is cyclic if and only if H is cyclic
- $|x| = |\phi(x)|$
- G has a subgroup of order k if and only if H has a subgroup of order k.

Proof. Suppose that ϕ is a bijection of G and H.

- Since ϕ is bijective, ϕ^{-1} is also bijective thus it is well defined. But we also need to prove that ϕ^{-1} is a homomorphism. Let $h_1, h_2 \in H$. Then there exists unique $g_1, g_2 \in G$ such that $\phi(g_1) = h_1$ and $\phi(g_2) = h_2$. Then
$$\phi^{-1}(h_1h_2) = \phi^{-1}(\phi(g_1)\phi(g_2)) = \phi^{-1}(\phi(g_1g_2)) = g_1g_2 = \phi^{-1}(h_1)\phi^{-1}(h_2)$$
thus we are done.
- Since ϕ is bijective, $|G| = |H|$
- Since ϕ is bijective and $1 \in \ker(\phi)$, we must have $\ker(\phi) = \{1\}$
- Suppose that G is abelian. $\phi(a)\phi(b) = \phi(ab) = \phi(ba) = \phi(b)\phi(a)$. Thus H is abelian. Since ϕ is bijective it has a bijective inverse. Thus the backwards implication is proved.
- Suppose that $x \in G$ has order n. $1_H = \phi(1_G) = \phi(x^n) = \phi(x)^n$
- If $|x| = n$ then $\phi(x)^n = \phi(x^n) = \phi(1) = 1$ thus we know that $|\phi(x)|\Bigg|n$. However for any $k < n$, $\phi(x)^k = \phi(x^k) \neq 1$ since ϕ is an isomorphism and $\ker(\phi) = \{1\}$. Thus we must have $|\phi(x)^n| = 1$.
- Suppose that G has a subgroup of order k, then $\phi(G)$ is a subgroup as well and every element preserves its order, we have that $\phi(G) \leq H$ is order k. For the reverse statement, just take ϕ^{-1} to be the isomorphism.

□

Theorem 10.1.4.7: The First Isomorphism Theorem

Let $\phi : G \to H$ be a homomorphism.

- $\ker(\phi) \trianglelefteq G$
- $G/\ker(\phi) \cong \phi(G)$

Proof. We already know that $\ker(\phi) \leq G$. We show that for $n \in \ker(\phi)$, $gng^{-1} \in \ker(\phi)$. We have that
$$\phi(gng^{-1}) = \phi(g)\phi(n)\phi(g^{-1}) = \phi(g)\phi(g)^{-1} = 1 \in \ker(\phi)$$
thus we are done.

We construct an isomorphism f between cosets of G of $\ker(phi)$ and elements of $\phi(G)$. Write N for $\ker(\phi)$ Let $gN \in G/\ker(\phi)$. Define f by $f(gN) = \phi(g)$. I claim that this is the isomorphism we need.

We first show that this map is well defined. The point here is to show that any representative is fine. In particular, if $gH = hN$ are two representations of $thesamecoset$ and g, h is distinct, then we want $f(gN) = f(hN)$. Well $gN = hN$ if and only if $gh^{-1} \in N$. This means that $gh^{-1} \in \ker(\phi)$ and $\phi(gh)^{-1} = 1$. This means that $\phi(g)\phi(h)^{-1} = 1$ and $\phi(g) = \phi(h)$. This means that the map is well defined.

We now show that it is a homomorphism. Let gN and hN be cosets such that $gN \neq hN$. Then
$$f((gN)(hN)) = f((gh)N) = \phi(gh) = \phi(g)\phi(h) = f(gN)f(hN)$$
thus we are done.

Finally we show that it is an isomorphism. Firstly let $f(gN) = f(hN)$. Then

$$\begin{aligned} f(gN) &= f(hN) \\ \phi(g) &= \phi(h) \\ \phi(g)\phi(h)^{-1} &= 1 \\ \phi(gh^{-1}) &= 1 \end{aligned}$$

This means that $gh^{-1} \in \ker(\phi)$ thus $gN = hN$. Now let $\phi(g) \in \phi(G)$. Then quite obviously $f(gN) = \phi(g)$ thus it is surjective and bijective and we are done. □

Isomorphisms capture more sturcture than homomorphism such as orders, commutativity, subgroups of particular orders and whether the group is cyclic.

To provide more examples of group, we can construct the automorphism group from a group.

Definition 10.1.4.8: Automorphisms

Define the automorphisms group of a group G to be

$$\text{Aut}(G) = \{\phi : G \to G | \phi \text{ is an isomorphism }\}$$

Lemma 10.1.4.9

Let G be a group. Then the automorphism group $\text{Aut}(G)$ is indeed a group.

Proof. Composing bijective maps also gives bijective maps. This operation is naturally associative. Since it is a bijection it must also has an inverse and the identity map is also bijective. □

We will see it reappear later in when inner automorphisms is defined. For now, treat it as another example for you to work on as a group.

10.2 Quotient Groups

10.2.1 Cosets

Definition 10.2.1.1: Cosets

Let H be a subgroup of G. Let $g \in G$. Define

- $gH = \{gh|h \in H\}$ the left cosets of H generated by H
- $Hg = \{hg|h \in H\}$ the right cosets of H generated by$H$

Given a subgroup H of G and an element g of G, think of cosets as off-setting the subgroup in G by a factor of g. As we will soon see, they have nice properties such as the cosets in fact partition the group into equal parts. A good example would be the rotation subgroup of D_{2n}. Try and off set it by another rotation and you will realize it falls back to the same group. But if you off-set it with a reflection s then coset is not the subgroup.

Proposition 10.2.1.2

Let H be a subgroup of G and let $g_1, g_2 \in G$. Then the following are equivalent.

- $g_1H = g_2H$
- $g_2 \in g_1H$
- $g_1^{-1}g_2 \in H$

Proof. Let $H \leq G$.

- (1) $\implies$ (2) Suppose that $g_1H = g_2H$. Let $g_1h \in g_1H$. Then there exists k such that $g_1h = g_2k$. Then $g_2 = g_1hk^{-1}$ and since H is a subgroup, $hk^{-1} \in H$ and thus $g_2 \in g_1H$
- (1) $\implies$ (3) Similar to the above argument but we rewrite the expression as $g_1^{-1}g_2 = hk^{-1}$ thus $g_1^{-1}g_2 \in H$
- (3) $\implies$ (1) Let $g_2h \in g_2H$. By (3) there exists $k \in H$ such that $g_1^{-1}g_2 = k$. We then have

$$\begin{aligned} g_1^{-1}g_2 &= k \\ g_1^{-1}g_2h &= kh \\ g_2h &= g_1kh \\ g_2h &\in g_1H \end{aligned}$$

Since $kh \in H$. Thus $g_2H \subseteq g_1H$. Mirror the argument for $g_2^{-1}g_1 \in H$ and we have that $g_1H \subseteq g_2H$ and we are done

□

Proposition 10.2.1.3

Let H be a subgroup of G. Let $g_1, g_2 \in G$. Then $g_1H = g_2H$ if and only if $Hg_1 = Hg_2$.

Proof. Let $g_1H = g_2H$. Let $hg_1 \in Hg_1$. Since we know that $g_1g_2^{-1} \in H$, then $hg_1g_2^{-1} \in H$. Let $hg_1g_2^{-1} = k \in H$. Then $hg_1 = kg_2 \in Hg_2$ thus we have that $Hg_1 \subseteq Hg_2$. The argument is similar for $Hg_2 \subseteq Hg_1$ thus we have that $Hg_1 = Hg_2$. For the reverse, the argument is also similar. □

Proposition 10.2.1.4

Let H be a subgroup of G. Then the left (right) cosets of G partition G.

Proof. We first show that if $g_1H \neq g_2H$, then $g_1H \cap g_2H = \emptyset$. Suppose that $k \in g_1H \cap g_2H$, then by the above we have $g_1H = kH = g_2H$ thus a contradiction. Thus we must have $g_1H \cap g_2H = \emptyset$. Now we must have that

$$\bigcup_{g \in G} gH = G$$

since every gH must at least have $g \in gH$. Thus we are done. □

Proposition 10.2.1.5

Let H be a subgroup of G. Then the number of distinct left cosets is equal to the number of distinct right cosets.

Proof. Since all the distinct left cosets partition G and $g_1H = g_2H$ if and only if $Hg_1 = Hg_2$, we must have the same number of distinct right cosets as well in order to partition G. □

Proposition 10.2.1.6

Let H be a subgroup of a finite group G. Then for any $g \in G$, $|gH| = |H|$.

Proof. Let $g \in G$. We prove that the map $\phi : H \to gH$ defined by $\phi(h) = gh$ is a bijection. For injectivity, $gh_1 = gh_2$ implies $h_1 = h_2$ by applying g^{-1} on both sides. Let $k \in gH$, then there exists $h \in H$ such that $k = gh \in gH$. This proves surjectivity and we are done. □

Definition 10.2.1.7: Number of Distinct Cosets

Let H be a subgroup of G. Define $[G : H]$ the number of distinct left (right) cosets of H in G.

10.2.2 Lagrange's Theorem

Lagrange's theorem is a powerful theorem that has many applications. Some indirect results include the class equation and orbit stabilizer theorem as we will later see.

Theorem 10.2.2.1: Lagrange's Theorem

Let G be a finite group and H a subgroup of G. Then

$$|G| = [G : H]|H|$$

Proof. We know that distinct left cosets partition G and every left coset has $|H|$ elements thus $|G| = [G : H]|H|$. □

An immediate result is that orders of subgroups must divide the order of the group. But notice that this does not imply the converse: given any number n that divides the order of G, it does not necessarily mean that there exists a subgroup of G with order n. As always, it is good to think of counterexamples. A prototypical one would be A_4 (see chapter 5 for definition) that does not have subgroups of order 6. This is the smallest example.

Another immediate result of Lagrange's theorem is the following.

Corollary 10.2.2.2

Let G be a finite group and $g \in G$. Then the order of g divides the order of G.

Proof. For every $g \in G$, $\langle g \rangle \leq G$ thus by Lagrange's Theorem $|g| = |\langle g \rangle| \Big| |G|$ □

One of the primary results of this theorem appears in chapter 5. It states that any group of order prime p must be isomorphic to C_p, the cyclic group of order p. This is a very powerful result in the classification of fintie groups.

10.2.3 Normal Subgroups

Normal groups play a center role in homomorphisms and isomorphisms. We shall see that it has properties unique to itself, yet is a concept we have already defined before. In this section we will define the quotient group through normal subgroups.

Off setting a subgroup by an element from the left and from the right is usually not equivalent. There is however a case to study the subgroups where it IS actually equivalent.

Definition 10.2.3.1: Normal Subgroups

Let N be a subgroup of G. N is called a normal subgroup of G if $gN = Ng$ for all $g \in G$. We write $N \trianglelefteq G$ in this case.

Definition 10.2.3.2: Conjugates

Let H be a subgroup of G. The set

$$gHg^{-1} = \{ghg^{-1} | h \in H\}$$

is called the conjugate of H.

It does not need deep results to see that conjugations and normal subgroups are related notions.

Theorem 10.2.3.3: Normality Test

Let N be a subgroup of G. The following are equivalent.

- N is normal in G
- $gNg^{-1} \subseteq N$ for all $g \in G$
- $gNg^{-1} = N$ for all $g \in G$

Proof. Let $N \trianglelefteq G$.

- (1) $\implies$ (2) Let N be normal. Let $gng^{-1} \in gNg^{-1}$. Since N is normal there exists $k \in N$ such that $gn = kg$. Then $gng^{-1} = kgg^{-1} = k \in N$.
- (2) $\implies$ (1) Let $gNg^{-1} \subseteq N$. Then for all $n \in N$ there exists $k \in N$ such that $gng^{-1} = k$. Then $gn = kg$ for all n thus $gN = Ng$ and we are done.
- (2) $\implies$ (3) We only need to show that $N \subseteq gNg^{-1}$. Let $n \in N$. Then $ng \in Ng$. Since N is normal, there exists $k \in N$ such that $gk = ng$ and $k = g^{-1}ng$. But then $n = gkg^{-1}$ thus $n \in gNg$ and we are done.

- (3) $\implies$ (2) is trivial.

□

While the second item above is simply for ease of proving things, the important thing to take away here is that as long as conjugating results in the same group, that group would be normal.

We attempt to perform binary operations on cosets.

Theorem 10.2.3.4

Let N be a subgroup of G. Then the operation on cosets described by

$$(gN)(hN) = ghN$$

where $(gN)(hN) = \{ab|a \in gN, b \in hN\}$ is well defined if and only if N is a normal subgroup.

Proof. Suppose first that the operation is well defined. Let $g \in G$ and $n \in N$. By the normality test, showing that $gng^{-1} \in N$ is sufficient for N to be normal. But since $gn \in gN$ and $g^{-1} \cdot 1 \in g^{-1}N$, we have that $gng^{-1} \in gg^{-1}N = N$ thus we are done.
Now suppose that N is normal. Then we first show that $(gN)(hN) \subseteq ghN$. Let $gn \in gN$ and $hk \in hK$. Then since $hN = Nh$, there exists some $u \in N$ such that $nh = hu$. Then

$$(gn)(hk) = g(nh)k = ghuk \in ghN$$

and we are done. Now we show that $ghN \subseteq (gN)(hN)$. Let $ghn \in ghN$. Then $ghn = (g \cdot 1)(hn) \in (gN)(hN)$ thus we are done. □

Theorem 10.2.3.5: Quotient Group

Let N be a normal subgroup of G. Define the quotient group to be the group

$$G/N = \{gN|g \in G\}$$

with the operation $(gN)(hN) = (gh)N$. It is sometimes call the factor group.

Proof. Closure is proved in the above theorem. Associativity follows from the fact that G is associative. Identity is the normal subgroup N since $g \cdot 1 = g$. The inverse is $g^{-1}N$ since $gg^{-1} = 1$. □

Proposition 10.2.3.6

A subgroup N of the group G is normal if and only if it is the kernel of some homomorphism.

Proof. Suppose that N is a normal subgroup of G. Then consider the homomorphism $\phi : G \to G/N$ defined by $\phi(g) = gN$. Then

$$\begin{aligned}\ker(\phi) &= \{g \in G|\phi(g) = N\}\\ &= \{g \in G|gN = N\}\\ &= \{g \in G|g \in N\}\\ &= N\end{aligned}$$

Thus we are done.
Let N be the kernel of the homomorphism $\phi : G \to H$. Then $\ker(\phi) = \{g \in G|\phi(g) = 1\} = N$.

By the normality test, showing that $gng^{-1} \in N$ is sufficient for $n \in N$ and $g \in G$. We have that

$$\phi(gng^{-1}) = \phi(g)\phi(n)\phi(g^{-1}) = \phi(g)\phi(g)^{-1} = 1$$

thus $gng^{-1} \in \ker(\phi) = N$ thus we are done. □

Undoubtedly aside from the construction of the quotient group, the above proposition is the main result of this section. It is equivalent to say that a normal subgroup is exactly the kernel of some homomorphism.

10.2.4 Normalizers

Sometimes, not every element in a group allows a subgroup H to normalize. Therefore can obtain a subset of G that contains all the elements that allows H to be normal.

Definition 10.2.4.1: Normalizers

Let G be a group. Let $S \subseteq G$. Then the normalizer is defined to be

$$N_G(S) = \{g \in G | gS = Sg\} = \{g \in G | gSg^{-1} = S\}$$

Proposition 10.2.4.2

Let G be a group and $S \subseteq G$. Then $N_G(S) \leq G$.

Proof. Let $g, h \in N_G(S)$. Then $h^{-1} \in N_G(S)$ since $h^{-1}S = Sh^{-1}$ if and only if $hS = Sh$ thus $h^{-1} \in N_G(S)$. Now we want $gh^{-1} \in N_G(S)$. We have

$$gh^{-1}S = g(Sh^{-1}) = (gS)h^{-1} = Sgh^{-1}$$

thus we are done. □

Proposition 10.2.4.3

Let G be a group and $N \subseteq G$. Then $N \trianglelefteq G$ if and only if $N_G(N) = G$.

Proof. Let $N \trianglelefteq G$. Trivially we already have $N_G(N) \leq G$. Let $g \in G$. Then since N is normal, we have $gN = Ng$ thus $g \in N_G(N)$ and we are done.

Let $N_G(N) = G$. Then for all $g \in G$, $gN = Ng$ thus N is normal. □

As motivated from above, $N_G(S)$ is just all the elements in G so that S is allowed to be normal. Then obviously if $N_G(S)$ is the entirety of G then S would be allowed to be normal.

10.2.5 More Isomorphism Theorems

Theorem 10.2.5.1: The Second Isomorphism Theorem

Let G be a group, let $A \leq G$ and $B \trianglelefteq G$. Then AB is a subgroup of G, $B \trianglelefteq AB$, $A \cap B \trianglelefteq A$ and $AB/B \cong A/A \cap B$.

Proof. We first prove that AB is a subgroup of G. By theorem 3.2.1, AB is a subgroup if and only if $AB = BA$. We have that

$$AB = \bigcup_{a \in A} aB = \bigcup_{a \in A} Ba = BA$$

and thus $AB \leq G$.

To show that $B \trianglelefteq AB$, we show that $abb'(ab)^{-1} \in B$ for any $b' \in B$ and $ab \in AB$. But this is trivial since $ab \in G$ and we know that since $B \trianglelefteq G$, $abb'(ab)^{-1} \in B$ thus we are done.

For $A \cap B \trianglelefteq A$, we want $kak^{-1} \in A$ for $k \in A \cap B$ and $a \in A$. But since $k \in A \cap B$ implies $k \in A$ and A is a subgroup, $kak^{-1} \in A$ thus we are done.

Now we construct an isomorphism between AB/B and $A/A \cap B$. Note that cosets of AB/B are of the form $(ab)B$ where $ab \in AB$. But $b \in B$ thus $bB = B$ and $abB = aB$. Define $\phi(aB) = a(A \cap B)$. We first show that this is well defined. Suppose that aB and $a'B$ are two different representations of the coset. We want $a(A \cap B) = a'(A \cap B)$ This happens if and only if $a'a^{-1} \in A \cap B$. $a'a^{-1} \in A$ is trivial since A is a subgroup. $a'a^{-1} \in B$ is true since $aB = a'B$ if and only if $a'a^{-1} \in B$. Thus ϕ is well defined.

Now suppose that $\phi(aB) = \phi(a'B)$. Then

$$\begin{aligned} \phi(aB) &= \phi(a'B) \\ a(A \cap B) &= a'(A \cap B) \end{aligned}$$

This is true if and only if $a'a^{-1} \in A \cap B$. In particular, $a'a^{-1} \in B$ thus $aB = a'B$ and we are done for bijectivity.

Now suppose that $a(A \cap B)$. Then in particular $\phi(aB) = a(A \cap B)$ thus surjectivity is proven. Thus ϕ is an isomorphism and we are done. □

Theorem 10.2.5.2: The Third Isomorphism Theorem

Let G be a group and H, K normal subgroups of G with H a subgroup of K. Then $K/H \trianglelefteq G/H$ and

$$(G/H)/(K/H) \cong G/K$$

Proof. Let $kH \in K/H$. Let $gH \in G/H$ I want to show that $(gH)(kH)(gH)^{-1} \in K/H$. But

$$(gH)(kH)(gH)^{-1} = gkg^{-1}H = k'H \in K/H$$

for some $k' = gkg^{-1}$. This is true since K/H is normal. Define $\phi : (G/H)/(K/H) \to G/K$ by $\phi((gH)(K/H)) = gK$. We first prove that it is well-defined. Let $(g_1H)(K/H)$ and $(g_2H)(K/H)$ be two representations of the same coset. Then

$$\begin{aligned} \phi((g_1H)(K/H)) &= \phi((g_2H)(K/H)) \\ g_1K &= g_2K \\ g_2^{-1}g_1K &= K \end{aligned}$$

And this is true if and only if $g_2^{-1}g_1 \in K$. But K is a subgroup thus we are done. Now let $\phi((g_1H)(K/H)) = \phi((g_2H)(K/H))$. Then

$$\begin{aligned} \phi((g_1H)(K/H)) &= \phi((g_2H)(K/H)) \\ g_1K &= g_2K \\ g_2^{-1}g_1K &= K \end{aligned}$$

But $g_2^{-1}g_1 \in K$ since K is a subgroup thus ϕ is injective. Surjectivity is trivial since for any $gK \in G/K$, $\phi((gH)(K/H)) = gK$ thus ϕ is an isomorphism. □

10.3 Group Actions

10.3.1 Group Actions

Definition 10.3.1.1: Groups Actions

Let G be a group and X a set. An action of G on X is a map $\cdot : G \times X \to X$ such that the following properties are satisfied.

- $(g_1 g_2) \cdot x = g_1 \cdot (g_2 \cdot x)$
- $1 \cdot x = x$

Theorem 10.3.1.2

Let G be a group acting on a set X. Let $\phi : G \to \text{Sym}(X)$ be defined by $\phi(g)(-) = g \cdot -$. We have that

- For each $g \in G$, $\phi(g) : X \to X$ is a permutation of X
- The mapping $\phi(g)(-) = g \cdot -$ is a homomorphism

Proof.

- To show that $\phi(g)$ is a permutation, we simply need to show that it is a bijection on itself. $\phi(g)$ is a bijection if and only if it has an inverse. For all $x \in X$, we have that

$$\begin{aligned}(\phi(g^{-1}) \circ \phi(g))(x) &= \phi(g^{-1})(\phi(g)(x)) \\ &= \phi(g^{-1}g)(x) \\ &= \phi(1)(x) \\ &= x\end{aligned}$$

Thus we have that $\phi(g^{-1}) = \phi(g)^{-1}$.

- We have that

$$\begin{aligned}\phi(g_1 g_2)(x) &= (g_1 g_2) \cdot x \\ &= g_1 \cdot (g_2 \cdot x) \\ &= \phi(g_1)(g_2 \cdot x) \\ &= \phi(g_1)(\phi(g_2)(x)) \\ &= (\phi(g_1) \circ \phi(g_2))(x)\end{aligned}$$

Thus we are done.

□

Definition 10.3.1.3: Kernel

Let G be a group acting on X. Define the kernel of the action to be the kernel of $\phi : G \to \text{Sym}(X)$, which is

$$\ker(G, X, \cdot) = \{g \in G | A_g(x) = x \text{ for all } x \in X\}$$

An action is faithful if $\ker(G, X, \cdot) = \{1\}$

Theorem 10.3.1.4: Cayley's Theorem

Every group G is isomorphic to a subgroup of S_n for some n.

Proof. Let X be any set. Let G act on X by $A_g(x) = g \cdot x$ for $x \in X$ and $g \in G$ and has kernel K. This action is faithful since $gx = x$ implies $g = 1$. By the first isomorphism theorem, $G \cong G/K \cong \text{im}(\phi) \leq\text{Sym}(X)$ thus we are done. □

10.3.2 Orbits and Stabilizers

Theorem 10.3.2.1

Let G be a group acting on a set X. The relation on X defined by

$$a \sim b \iff g \cdot b = a \text{ for some } g \in G$$

for $a, b \in X$ is an equivalence relation.

Proof. We check for the three criteria.

- (reflexivity) $a \sim a$ is obvious since $1 \cdot a = a$
- (symmetry) $a \sim b$ if and only if $b \sim a$ is true since $g \cdot b = a$ if and only if $g^{-1} \cdot a = b$
- (transitivity) $a \sim b$ and $b \sim c$ implies $(gh) \cdot c = a$ and $gh \in G$ thus $a \sim c$

□

Now fix an element $x \in X$. We consider all the possible elements that can be obtained by taking the action A_g on it for any g.

Definition 10.3.2.2: Orbits

Let G be a group acting on the nonempty set X. The equivalence class

$$\text{Orb}_x = \{g \cdot x | g \in G\} \in G/\sim$$

is called the orbit of G containing x.

Essentially a bunch of orbits are formed in a group. They would be closed on itself since it is a partition of the relation $\sim$. Conversely if we say that a and b are in the same orbit, then there must exists $g \in G$ for which perform the action of g on a gets to b.

Definition 10.3.2.3: Stabilizer

Let G be a group that acts on X. Let $x \in X$. Define the stabilizer of x to be

$$\text{Stab}_G(x) = \{g \in G | g \cdot x = x\}$$

Proposition 10.3.2.4

Let G be a group acting on X. Then the following are true.

- $\text{Stab}_G(x) \leq G$
- $\bigcap_{x \in X} \text{Stab}_G(x) = \ker(G, X)$

Proof.

- Let $g, h \in \text{Stab}_x$. Then $h^{-1} \in \text{Stab}_G(x)$ since $h^{-1} \cdot x = h^{-1} \cdot (h \cdot x) = (h^{-1}h) \cdot x = x$. Then

  $$\begin{aligned}(gh^{-1}) \cdot x &= g \cdot (h^{-1} \cdot x) \\ &= g \cdot x \\ &= x\end{aligned}$$

  Thus $gh^{-1} \in \text{Stab}_x$ and the subgroup criterion is satisfied.

- We show inclusion of the sets. Let $g \in \bigcap_{x \in X} \text{Stab}_G(x)$. Then for any $x \in X$, $g \cdot x = x$ and thus $g \in \ker(G, X)$. Let $g \in \ker(G, X)$, then for any $x \in X$, $g \cdot x = x$ which means that $g \in \text{Stab}_G(x)$ for all $x \in X$. Thus $g \in \bigcap_{x \in X} \text{Stab}_G(x)$.

□

Theorem 10.3.2.5: Orbit Stabilizer Theorem

Let G be a finite group acting on X. If $x \in X$, then

$$|\text{Orb}_x| = [G : \text{Stab}_G(x)]$$

In other words,

$$|G| = |\text{Orb}_x||\text{Stab}_G(x)|$$

Proof. The goal is to define a bijection between cosets of $\text{Stab}_G(x)$ and elements in Orb_x. Define $f : \text{Orb}_x \to G/G_x$ by $f(g) = g\text{Stab}_G(x)$. This map is clearly well defined.

Injectivity:
Let $g\text{Stab}_G(x) = h\text{Stab}_G(x)$. Then $g^{-1}h \in \text{Stab}_G(x)$ which means that $(g^{-1}h) \cdot x = x$. Then applying the action of g on both sides, we have

$$h \cdot x = g \cdot x$$

which proves that $g = h$.

Surjectivity:
For any $g\text{Stab}_G(x)$, just choose $g \in G$. Then clearly $f(g) = g\text{Stab}_G(x)$.

Thus $|\text{Orb}_x| = [G : \text{Stab}_G(x)]$. □

10.3.3 The Action of Conjugation

Definition 10.3.3.1: Conjugation

Let G be a group. Define conjugation to be an action of G on G where $\cdot : G \times G \to G$ is defined by

$$g \cdot x = gxg^{-1}$$

Define the conjugacy classes of a group to be the orbits of conjugation:

$$\text{Cl}(x) = \{gxg^{-1} | g \in G\} = \text{Orb}_x$$

Two elements are said to be conjugate to each other if they lie in the same conjugacy classes.

Define the centralizer of a group to be the stabilizer of conjugation:

$$C_G(x) = \{g \in G | gx = xg\} = \text{Stab}_G(x)$$

Define the center of a group to be the kernel of conjugation:

$$Z(G) = \{g \in G | gxg^{-1} = x \text{ for all } x \in X\} = \ker(G, G, \cdot)$$

Lemma 10.3.3.2

Conjugation is indeed an action of G on G.

Lemma 10.3.3.3

Let G be a group. Then $Z(G) \leq G$ and $C_G(x) \leq G$.

Proof. These two are a special case of proposition 4.2.4 □

Definition 10.3.3.4: Inner Automorophisms

Define the Inner automorphisms of a group to be the group of all bijective functions of G that are conjugation actions. Meaning

$$\text{Inn}(G) = \{g \cdot x = gxg^{-1} | g \in G\}$$

Proposition 10.3.3.5

Let G be a group. Then the following are true.

- $\text{Inn}(G) \leq \text{Aut}(G)$
- $G/Z(G) \cong \text{Inn}(G)$

Proof. Let G be a group.

- We show that every element in Inn(G) are automorphisms and that they satisfy the subgroup criterion. Let $A_g \in$Inn(G). Then $A_g(xy) = gxyg^{-1} = gxg^{-1}gyg^{-1} = A_g(x) \cdot A_g(y)$. For injectivity, let $A_g(x) = A_g(y)$. Then
$$\begin{aligned} A_g(x) &= A_g(y) \\ gxg^{-1} &= gyg^{-1} \\ x &= y \end{aligned}$$
To show injectivity, note that for any $u \in G$, $A_g(g^{-1}ug) = gg^{-1}ugg^{-1} = u$ thus A_g is bijective. Note that for any $g \in G$, $A_{g^{-1}}(x)$ is the inverse of $A_g(x)$. Now let $A_g(x), A_{h^{-1}}(x) \in$Inn($G$). Then
$$(A_g \circ A_{h^{-1}})(x) = A_{gh^{-1}}(x) = gh^{-1}xg^{-1}h \in \text{Inn}(G)$$
thus we are done.
- Define a homomorphism $f : G \to$Aut(G) by $f(g) = [A_g(h) = ghg^{-1}]$. This is indeed a homomorphism since $f(g_1g_2) = A_{g_1g_2} = A_{g_1}(A_{g_2}(h))$ and $f(g_1)f(g_2) = A_{g_1} \circ A_{g_2}$. Trivially im($f$) =Inn($G$). Also
$$\begin{aligned} \ker(f) &= \{g \in G | A_g(x) = x \text{ for all } x \in X\} \\ &= \{g \in G | gxg^{-1} = x \text{ for all } x \in X\} \\ &= \{g \in G | gx = xg \text{ for all } x \in X\} \\ &= Z(G) \end{aligned}$$
Thus by the first isomorphism theorem, $G/Z(G) \cong \text{Inn}(G)$.

□

Theorem 10.3.3.6: The Class Equation

Let G be a finite group and let $g_1, g_2, \dots, g_r$ be representatives of the distinct conjugacy classes of G not contained in $Z(G)$. Then

$$|G| = |Z(G)| + \sum_{i=1}^{r} |G : C_G(g_i)| = |Z(G)| + \sum_{i=1}^{r} |O_{x_i}|$$

Proof. Suppose that G acts on itself by conjugation. We know that the orbits partition the group G since it is a quotient of an equivalence relation. However, for any $g \in Z(G)$, $gxg^{-1} = x$ for any $x \in X$. Thus every element in $Z(G)$ is itself an orbit. Combining these facts, we have that

$$|G| = |Z(G)| + \sum_{i=1}^{r} |G : C_G(g_i)| = |Z(G)| + \sum_{i=1}^{r} |O_{x_i}|$$

□

10.3.4 Fixed Points

Definition 10.3.4.1: Fixed Points

Let G be a group acting on X. Let $g \in G$. We say that $x \in X$ is a fixed point of g if $g \cdot x = x$. The set of all fixed points of $g \in G$ is

$$\text{Fix}_X(g) = \{x \in X | g \cdot x = x\}$$

We say that $g \in G$ is fixed point free if $\text{Fix}_X(g) = \emptyset$.

Lemma 10.3.4.2: The Not Burnside Lemma

Let G be a finite group acting on a finite set X. Then

$$|\{O_x | x \in X\}| = \frac{1}{|G|} \sum_{g \in G} \text{Fix}_X(g)$$

Corollary 10.3.4.3

Let G be a finite group acting on a finite set X with $|X| > 1$. Suppose that G has precisely one orbit. Then G contains fixed point free element.

10.4 Products of Groups

We have quotient of groups which has the similar meaning to dividing out groups, algebraists have also formally defined the product of a group. However, there are three different notions of multplication, two of which we will see in this chapter.

10.4.1 External Direct Products

We first have the external product. It is the most direct and natural way to construct multiplication between two groups.

Proposition 10.4.1.1: External Product

Let $(G, *)$ and $(H, \circ)$ be groups. The cartesian product of G and H forms a group called the external (direct) product $G \times H$ with the operation

$$(g_1, h_1) \times (g_2, h_2) = (g_1 * g_2, h_1 \circ h_2)$$

where $g_1, g_2 \in G$ and $h_1, h_2 \in H$.

Proposition 10.4.1.2

If $G_1, \dots, G_n$ are groups, then

$$|G_1 \times \cdots \times G_n| = |G_1| \cdots |G_n|$$

Proof. Direct from the cardinality of the cartesian product. □

Theorem 10.4.1.3

Let $(g, h) \in G \times H$. If g and h has finite order r and s respectivelty, then

$$|(g, h)| = \operatorname{lcm}(r, s)$$

Proof. Trivially we have that $|(g,h)| \,|\, \operatorname{lcm}(r,s)$. Suppose that $|(g,h)| < \operatorname{lcm}(r,s)$. Then by definition of the lcm, either $r||(g,h)|$ or $s||(g,h)|$ but not both. But then maximally we only have one of the elements in (g,h) be the identity. Thus we must have $|(g,h)| = \operatorname{lcm}(r,s)$. □

Proposition 10.4.1.4

Let $G_1, \dots, G_n$ be groups and G their exrternal product. Then

$$G/G_i \cong G_1 \times \cdots \times G_{i-1} \times G_{i+1} \times \cdots \times G_n$$

Proof. Let $g = (g_1, \dots, g_n) \in G$. Take the function $f(gG_i) = (g_1, \dots, g_{i-1}, g_{i+1}, \dots, g_n)$ to be the isomorphism. Indeed the function is well defined. For any $g \in G$, gG_i is unique up to the ith element. It is a homomorphism since $f((gG_i)(hG_i)) = f(ghG_i)$ is easy to check.

To show injectivity, let $f(gG_i) = f(hG_i)$. Then

$$f(gG_i) = f(hG_i)$$
$$(g_1, \dots, g_{i-1}, g_{i+1}, \dots, g_n) = (h_1, \dots, h_{i-1}, h_{i+1}, \dots, h_n)$$

Note that the ith element does not matter since two $g_i \in G_i$ gives the same coset. Thus $gG_i = hG_i$ and we are done. Surjectivity is trivial since for any given $(g_1, \dots, g_{i-1}, g_{i+1}, \dots, g_n)$, take $g = (g_1, \dots, g_{i-1}, 1, g_{i+1}, \dots, g_n)$ and $f(gG_i) = (g_1, \dots, g_{i-1}, g_{i+1}, \dots, g_n)$ and we are done.

Thus the two groups are isomorphic. □

10.4.2 Internal Direct Products

While the external product is concerned of the product between two groups, the internal direct product is interested in how two subgroups of a group could be ”multiplied” to form the group itself. Essentially, we are trying to decompose a group into its internal direct product.

Definition 10.4.2.1: Internal Direct Product

Let G be a group and $H, K \leq G$. Define

$$HK = \{hk | h \in H, k \in K\}$$

If $G = HK \cong H \times K$ we say that HK is the internel direct product of G.

We will then discuss the conditions that makes HK the internal direct product.

Lemma 10.4.2.2

Let G be a group. Let $H, K \leq G$. Then

$$HK = \bigcup_{h \in H} hK$$

Proof. Let $hk \in HK$. Then $hk \in hK$ and $hk \in \bigcup_{h \in H} hK$ thus $HK \subseteq \bigcup_{h \in H} hK$. Now let $x \in \bigcup_{h \in H} hK$. Then there exists $h \in H$ such that $x \in hK$. This means that $x = hk \in hK$ for some $k \in K$. Thus $x \in HK$ and $\bigcup_{h \in H} hK \subseteq HK$. □

Theorem 10.4.2.3

Let G be a group. Let $H, K \leq G$. HK is a subgroup if and only if $HK = KH$.

Proof. First suppose that HK is a subgroup. Since $K \subseteq HK$ and $H \subseteq HK$, we have that $KH \subseteq HK$. by closure. To show the reverse. Let $hk \in HK$. Then there exists some $a \in HK$ such that $a^{-1} = hk$ since HK is a group. Then $a \in HK$ implies that $a = h'k'$ for some $h'k' \in HK$. Then

$$hk = a^{-1} = k'^{-1}h'^{-1} \in KH$$

thus we are done.

Now suppose that $HK = KH$. Let $h_1k_1 \in HK$ and $h_2k_2 \in HK$. We show that it satisfies the subgroup criterion.

- $1 \in HK$ is trivial
- For closure, Note that $k_1h_2 = h'k'$ since $KH = HK$. Then

  $$h_1k_1h_2k_2 = h_1h'k'k_2 = HK$$

  The last part is true since H, K are subgroups.

- Let $hk \in HK$. Then $(hk)^{-1} = k^{-1}h^{-1} \in KH = HK$

Thus we are done. □

Proposition 10.4.2.4

Suppose that G is a group with normal subgroups H, K such that $H \cap K = \{1\}$ and $G = HK$, then $G \cong H \times K$.

The final proposition here shows that really external direct products and internal direct products are more or less the same thing. If two groups can be decomposed into its internal direct products, we may as well write thew two subgroups into an external direct product, which makes it alot easier to perform operations on compared to the inner version.

10.4.3 Semidirect Product

Theorem 10.4.3.1: Semidirect Product

Let H, K be groups and let ϕ be a homomorphism from K into $\text{Aut}(H)$. Let G be the set of all ordered pairs (h, k) where $h \in H$ and $k \in K$. Then the operation

$$(h_1, k_1)(h_2, k_2) = (h_1 \cdot \phi_{k_1}(h_2), k_1k_2)$$

allows G to be a group. This operation is called the semidirect product and $G = H \rtimes_\phi K$.

Proof. Clearly closure is satisfied. We check associativity since it is asymetrical. We have that

$$\begin{aligned} ((h_1, k_1)(h_2, k_2))(h_3, k_3) &= (h_1 \cdot \phi_{k_1}(h_2), k_1k_2)\,(h_3, k_3) \\ &= (h_1 \cdot \phi_{k_1}(h_2) \cdot (\phi_{k_1k_2}(h_3)), k_1k_2k_3) \\ &= (h_1 \cdot \phi_{k_1}(h_2) \cdot (\phi_{k_1} \circ \phi_{k_2})(h_3), k_1k_2k_3) \\ &= (h_1 \cdot \phi_{k_1}(h_2 \cdot \phi_{k_2}(h_3)), k_1k_2k_3) \\ &= (h_1, k_1)(h_2 \cdot \phi_{k_2}(h_3), k_2k_3) \\ &= (h_1, k_1)((h_2, k_2)(h_3, k_3)) \end{aligned}$$

The identity in this group will be $(1_H, 1_K)$. Indeed we have that
$(h, k)(1_H, 1_K) = (h \cdot \phi_k(1_H), k) = (h, k)$ and $(1_H, 1_K)(h, k) = (\phi_{1_K}(h), k) = (h, k)$. The inverse of (h, k) is $(\phi_{k^{-1}}(h^{-1}), k^{-1})$ □

Theorem 10.4.3.2

Let $G = H \rtimes_\phi K$ be a semidirect product. Then $|G| = |H||K|$ and

$$H \cong \{(h, 1) | h \in H\} \leq G$$

and

$$K \cong \{(1, k) | k \in K\} \leq G$$

Proof. Clearly $H \cong \{(h, 1)|h \in H\}$. Notice that $(h_1, 1)(h_2, 1) = (h_1 \cdot (\phi(1))(h_2), 1) = (h_1h_2, 1)$ since $\phi(1)$ is the identity automorphism. This means that the multiplication structure of H is preserved. Since H itself is a group, this makes the set a subgroup of G.

Similarly, notice that $(1, k_1)(1, k_2) = (\phi(k_1)(1), k_1k_2) = (1, k_1k_2)$ □

Lemma 10.4.3.3

Let $G = H \rtimes_\phi K$ be a semidirect product. Then the following are true.

- $H \trianglelefteq G$
- $H \cap K = 1$

Proof.

- Let $(h, k) \in G$. Let $(h_0, 1) \in H$. Then

$$(h, k)(h_0, 1)(\phi_{k^{-1}}(h^{-1}), k^{-1}) = (h, k)(h_0 \cdot (\phi_{k^{-1}}(h^{-1})), k^{-1})$$

  Notice that the second term multiplies to 1 which means that the entire product lies in H. Thus we have proven $(h, k)H(h, k)^{-1} \subseteq H$

- From the above identification we clearly see that they are disjoint except from the identity.

□

These are the fundamental results that we want the semidirect product to have when we construct it.

Proposition 10.4.3.4

Let H, K be groups and $\phi : K \to \text{Aut}(H)$ be a group homomorphism. Then the following are equivalent characterizations of when the semidirect product is equivalent to the direct product:

- The identity map between $H \rtimes_\phi K \cong H \times K$ is a group isomorphism
- ϕ is the trivial homomorphism from K to $\text{Aut}(H)$
- $K \trianglelefteq H \rtimes K$

Theorem 10.4.3.5

Let G be a group with subgroups H, K such that $H \trianglelefteq G$ and $H \cap K = 1$. Let $\phi : K \to \text{Aut}(H)$ be the homomorphism defined by $\phi(k)(h) = khk^{-1}$. Then $HK \cong H \rtimes K$.

10.5 Important Groups to Note

In this section I will present a dozen of useful groups that appear frequently in different areas of mathematics. Although I am quite tempted to include the general linear group and its subrgoups in this section, I will leave them until the notion of a field is properly defined in subsequent chapters.

10.5.1 The Cyclic Group C_n

The cyclic group is one of the easiest groups to understand. Although it has been used in some of the proofs in above chapter, for tidyness and neatness I will state the official definition here.

Definition 10.5.1.1: The Cyclic Group C_n

Let $n \in \mathbb{N}$. The cyclic group of order n is defined to be the group

$$C_n = \{1, g, g^2, \dots, g^{n-1}\} = \langle g \rangle$$

where multiplication is defined by $g^m \cdot g^k = g^{m+k}$. It may also be denoted as Z_n

For completion and for the reader's own good, the following lemma is included.

Lemma 10.5.1.2

The cyclic group is indeed an abelian group satisfying the four axioms and the commutative rule.

Proof. Trivial. □

The following lemma reveals one of the many faces of the cyclic group.

Lemma 10.5.1.3

Let $n \in \mathbb{N}$. Then

$$C_n \cong (\mathbb{Z}/n\mathbb{Z}, +)$$

Proof. Define a homorphism by $\phi(g^k) = k + n\mathbb{Z}$ for $0 \leq k \leq n-1$. It is easy to check that this is an isomorphism. □

There are other forms of cycle groups in disguise such as the roots of unity $z^n = 1$. Its elements, the roots of this equation forms the same group.

Theorem 10.5.1.4

Let $n \in \mathbb{N}$ and $0 \leq k \leq n-1$ be an integer. Then g^k is a generator of C_n if and only if $\gcd(k, n) = 1$.

Proof. Suppose that g^k is a generator of C_n. Then $g = g^{ak}$ for some $a \in \mathbb{Z}$. Since $g^n = 1$, there exists $b \in \mathbb{Z}$ such that $g^{ak-bn} = g$ and $0 \leq ak - bn \leq n-1$. Then $ak - bn = 1$ has a solution if and only if $\gcd(k, n) = 1$ thus we are done.

Now suppose that $\gcd(k, n) = 1$. Then $\left|g^k\right| = \frac{n}{\gcd(k,n)=n}$ thus we are done. □

Lemma 10.5.1.5

Let $m, n \in \mathbb{N}$. Then

$$C_m \cong C_n$$

if and only if $m = n$.

Proof. Suppose that $C_m \cong C_n$. Then $m = |C_m| = |C_n| = n$. Now let $m = n$. Then $C_n \cong C_n$. □

Proposition 10.5.1.6

Let $m, n \in \mathbb{N}$. Then

$$C_m \times C_n \cong C_{mn}$$

if and only if $\gcd(m, n) = 1$.

Proof. Let $C_m \times C_n \cong C_{mn}$. Let g, h be the generator of C_m, C_n respectively. Then (g, h) is the generator of $C_{m,n}$ since isomorphisms map generators to generators. We have proved that $|\gcd(g, h)| = \text{lcm}(m, n) = \frac{mn}{\gcd(m,n)}$. But $|C_n \times C_m| = mn$ thus $mn = \frac{mn}{\gcd(m,n)}$ and $\gcd(m, n) = 1$.

Now let $\gcd(m, n) = 1$. Define a function $\phi : C_m \times C_n \to C_{m\times n}$ where $\phi(g^a, h^b) = g^a h^b$. I will show that it is an isomorphism. Firstly ϕ is a homomorphism since

$$\phi((g^{a_1}, h^{b_1})()) =$$

□

10.5.2 The Dihedral Group D_{2n}

Definition 10.5.2.1: The Dihedral Group D_{2n}

Let $n \in \mathbb{N}$. The dihedral group D_{2n} of order $2n$ is defined to be

$$D_{2n} = \langle a, b | a^n = 1, b^2 = 1, ab = ba^{-1} \rangle$$

Lemma 10.5.2.2

The dihedral group is indeed a group satisfying the four axioms.

Proof. Trivial. □

10.5.3 The Klein Four Group K_4 and Quaternion Group Q_8

The Klein Four Group is the only other group of order 4 that is not isomorphic to the cyclic group.

Definition 10.5.3.1: The Klein Four Group K_4

The Klein four group of order 4 is defined to be the group

$$K_4 = \{1, a, b, c\}$$

where multiplication is defined by

×	1	a	b	c
1	1	a	b	c
a	a	1	c	b
b	b	c	1	a
c	c	b	a	1

Treating c in the definition as ab makes it easy to see that K_4 is in fact D_4 in disguise.

Lemma 10.5.3.2

The Klein four group K_4 is isomorphic to D_4 but not isomorphic to C_4.

Definition 10.5.3.3: The Quaternion Group Q_8

The quaternion group of order 8 is defined to be the group

$$Q_8 = \langle -1, i, j, k | (-1)^2 = 1, i^2 = j^2 = k^2 = ijk = -1 \rangle$$

Lemma 10.5.3.4

Q_8 is not isomorphic to D_4 and C_8.

10.5.4 The Permutation Group S_n

Definition 10.5.4.1: The Permutation Group S_n

Let $n \in \mathbb{N}$. Let X be a set of n elements usually denoted $X = \{1, \dots, n\}$. The permutation group S_n is defined to be

$$S_n = \{\phi : X \to X | \phi \text{ is bijective}\}$$

$\phi \in S_n$ are represented with a matrix:

$$\begin{pmatrix} 1 & 2 & \cdots & n \\ \phi(1) & \phi(2) & \cdots & \phi(n) \end{pmatrix}$$

We have an alternate notation for elements in the permutation group.

Definition 10.5.4.2: Cycle Notation

For an element $\phi \in S_n$, we can represent the permutation by (multiple) cycles such as

$$\begin{pmatrix} 1 & 3 & 4 \end{pmatrix} \begin{pmatrix} 2 & 5 \end{pmatrix} \in S_5$$

In here, this element is equivalent to

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 5 & 4 & 1 & 2 \end{pmatrix}$$

In general, for any element $\begin{pmatrix} 1 & 2 & \cdots & n \\ \phi(1) & \phi(2) & \cdots & \phi(n) \end{pmatrix}$ in S_n, we can write it in cycle notation $\begin{pmatrix} 1 & \phi(1) & \phi(1) & \dots \end{pmatrix}$. Once the composition returns to 1, we start another cycle by considering any element not in the current cycle and perform compositions of ϕ on it.

Cycle notations make multplication of elements in S_n easier. Suppose that we want to find the following product

$$\tau \cdot \psi = \begin{pmatrix} 2 & 4 & 7 & 8 \end{pmatrix} \begin{pmatrix} 1 & 3 \end{pmatrix} \begin{pmatrix} 5 & 6 \end{pmatrix} \cdot \begin{pmatrix} 1 & 3 & 7 \end{pmatrix} \begin{pmatrix} 2 & 5 & 6 & 9 \end{pmatrix} \begin{pmatrix} 4 & 10 \end{pmatrix}$$

We start with any element, say 1. Then observe that ψ maps 1 to 3. We then see that 3 maps to 1 in τ so 1 maps to itself and we can omit the trivial cycles. Now for 2, we see that 2 maps to 5 in ψ and 5 maps to 6 in τ. So we can write down $\begin{pmatrix} 2 & 6 & \cdots \end{pmatrix}$. Now instead of choosing another element to begin. We follow up on 6 to see that 6 maps to 9. This means we can continue to chain and write $\begin{pmatrix} 2 & 6 & 9 & \cdots \end{pmatrix}$. Once you reach the end, you begin a new cycle with another element not in previous cycles.

Lemma 10.5.4.3

The permutation group is indeed a group satisfying the four axioms.

Proof. Trivial. □

Lemma 10.5.4.4

Let $n \in \mathbb{N}$. Then the order of S_n is $n!$.

Proof. Using the matrix form, we see that $\phi(1), \dots \phi(n)$ is a permutation of $1, \dots, n$. In particular, if $\phi(1)$ has n possibilities, then $\phi(2)$ has $n-1$ options etc. Then there will be $n!$ different permutations. □

Proposition 10.5.4.5

Let $n \in \mathbb{N}$. Then S_n is non-abelian if and only if $n \geq 3$.

Proof. We first show if $n \geq 3$ then S_n is non-abelian. By fixing $n-3$ elements to map to themselves in S_n, we see that $S_3 \leq S_n$. Now S_3 is not abelian since $\begin{pmatrix}1 & 2\end{pmatrix} \cdot \begin{pmatrix}1 & 3\end{pmatrix} = \begin{pmatrix}1 & 3 & 2\end{pmatrix}$ but $\begin{pmatrix}1 & 3\end{pmatrix} \cdot \begin{pmatrix}1 & 2\end{pmatrix} = \begin{pmatrix}1 & 2 & 3\end{pmatrix}$.

Now for $n = 1$ and $n = 2$, the fact they are abelian is trivial. □

Definition 10.5.4.6: Cycle Types

Let $\phi \in S_n$. We say that ϕ has cycle type $2^{r_2}3^{r_3}\cdots$ if it has exactly r_k k cycles for $k \geq 2$.

Lemma 10.5.4.7

A cycle of k elements has order k.

Proof. Let τ be a cycle of order k. Notice that multiplying n times of the same element means that we are taking the ith element in τ to the $i+n$ modulo k element in the cycle. If $n = k$ then every element goes to itself, which means that τ^k is just the permutation that maps every element to iself, which is the identity of the permutation group. □

Definition 10.5.4.8: Transpositions

A transposition is a cycle of order 2.

Notice that the following lemma means that we will have a third notation for elements of S_n.

Lemma 10.5.4.9

A cycle with k elements can be decomposed into a product of $k-1$ transpositions.

Proof. For a cycle $\begin{pmatrix}x_1, \dots, x_k\end{pmatrix}$. Notice that the product

$$\begin{pmatrix}x_{k-1} & x_k\end{pmatrix} \cdots \begin{pmatrix}x_2 & x_3\end{pmatrix} \begin{pmatrix}x_1 & x_2\end{pmatrix}$$

returns the original cycle. □

Based on the number of transpositions an element can be decomposed to, we classify them into odd and even permutations.

Definition 10.5.4.10: Odd/Even Permutations

A permutation $\phi \in S_n$ is said to be even if it has an even number of transpositions, and odd if it has an odd number of transpositions.

Proposition 10.5.4.11

Suppose that $\phi \in S_n$ consists of k cycles with order $n_1, \dots, n_k$. Then

$$|\phi| = \text{lcm}(n_1, \dots, n_k)$$

Definition 10.5.4.12: The Alternating Group A_n

Let $n \in \mathbb{N}$. The alternating group A_n is defined to be

$$A_n = \{\phi \in S_n | \phi \text{ is even}\}$$

Theorem 10.5.4.13

The alternating group is the kernel of the homomorphism $f : S_n \to \{-1, 1\}$. $\{-1, 1\}$ is defined by ordinary multiplication and f is given by

$$f(\phi) = \begin{cases} 1 & \text{if } \phi \text{ is even} \\ -1 & \text{if } \phi \text{ is odd} \end{cases}$$

Proposition 10.5.4.14

Let $n \in \mathbb{N}$. Then

$$S_n/\{-1, 1\} \cong A_n$$

and thus $|A_n| = \frac{n!}{2}$.

Proposition 10.5.4.15

Let $\tau, \psi \in S_n$ be a permutation. Then the conjugate $\psi\tau\psi^{-1}$ is obtained by replacing the elements x of the cycle τ with $\psi(x)$.

Proposition 10.5.4.16

Two permutations in S_n are conjugate if and only if they have the same cycle types.

10.6 Introduction to Rings

10.6.1 Basic Concept of Rings

Definition 10.6.1.1: Rings

A ring R is a set together with two binary operations $+$ and $\times$ such that

- $(R, +)$ is an abelian group.
- $a, b \in R$ implies $ab \in R$
- $a \times (b \times c) = (a \times b) \times c$ for all $a, b, c \in R$
- There exists an element $1 \in G$, called a multiplicative identity of R, such that for all $a \in R$ we have $a * 1 = 1 * a = a$
- $(a + b) \times c = (a \times c) + (b \times c)$ for all $a, b, c \in R$
- $a \times (b + c) = (a \times b) + (a \times c)$ for all $a, b, c \in R$

Proposition 10.6.1.2

Let R be a ring.

- $0a = a0 = 0$ for all $a \in R$
- $(-a)b = a(-b) = -(ab)$ for all $a, b \in R$
- $(-a)(-b) = ab$ for all $a, b \in R$
- $-a = (-1)a$

Proof. Trivial. □

As with subgroups and groups, we have subrings of a ring.

Definition 10.6.1.3: Subring

A subring of the ring R is a subgroup of R that is closed under multiplication.

Theorem 10.6.1.4: Subring Criterion

Let R be a ring and S a subset of R. S is a subring of R if

- $a, b \in S \implies ab \in S$
- (subgroup criterion) $a, b \in S \implies a - b \in S$
- $1 \in S$

Proof. Suppose that R is a ring and S is a subring. Then the subgroup criterion is already satisfied. If $a, b \in S$ then $ab \in S$ since S is a ring. Identity also exists since it is a ring.

Now suppose that S is a subset that satisfies the three criterion. By the subgroup criterion S is an abelian subgroup. Associativity and distributivity is automatically satisfied. The closure property is also satisfied by the first item. The identity is also satisfied by the third item thus we are done. □

Proposition 10.6.1.5

Let R be a ring and I, J subrings of R. Then $I \cap J$ is also a subring of R.

Proof. We show it using the subring criterion. The subgroup criterion is met since I, J are subgroups. If $a, b \in I \cap J$ then $ab \in I$ and $ab \in J$ by closure of rings thus $ab \in I \cap J$. Since I, J are subrings, $1 \in I$ and $1 \in J$ thus we are done. □

Similar to the order of a group, we have the characteristic of a ring.

Definition 10.6.1.6: Characteristic

Let R be a ring. If there exists a positive integer n such that $nx = 0$ for $x \in R$, then the smallest of such integer is called the characteristic of R. If there is no such integer then it has characteristic 0.

10.6.2 Division Rings

Definition 10.6.2.1: Unit

Let R be a ring. An element $a \in R$ is a unit in R if there is some $b \in R$ such that $ab = ba = 1$. The set of units in R is denoted $R^\times$.

Definition 10.6.2.2: Division Rings

A ring R is a division ring if every non-zero element of R is a unit in R.

In particular, division ring as the name suggests, allows division in the ring. This is because since every element in R is a unit, for any $a, b \in R$, we have that $a = a(b^{-1}b) = (ab^{-1})(b)$. It looks ugly but if we let $k = ab^{-1}$, then we are in fact expressing $a = kb$ for any $a, b \in R$. Unfortunately we do not have the notion of size in a general division ring so it is not particularly useful that we can do this.

An important type division rings will be the following.

Definition 10.6.2.3: Fields

A commutative division ring is called a field.

Fields are studied extensively in field theory, where its unique properties as a field is discussed. In groups and rings, we perform treatment on fields by considering its properties as a ring, rather than a field.

Proposition 10.6.2.4

Let $(R, +, \cdot)$ be a ring. Then R is a field if and only if $(R, +)$ and $(R \setminus \{0\}, \cdot)$ are abelian groups and the distributive law $a(b + c) = ab + ac$ holds.

Proof. Let R be a field. Then $(R, +)$ is already an abelian group. Since every field is a division ring, every element except 0 is a unit and thus $(R, \cdot)$ is a group. Since a field is also commutative, $(R, \cdot)$ is abelian. Since every field is a ring, the distributive law already holds and we are done.

Suppose that the three criteria is satisfied. Then R is a division ring since every nonzero element has an inverse in the group $(R, \cdot)$. It is also commutative since $(R, \cdot)$ is abelian thus we are done. □

I give the distributive law here although it seems redundant. Sometimes we want to check that a set with two binary operations is a field. In this case, we will need the distributive law to link the two binary operations in to action.

10.6.3 Integral Domains

We first proof some equivalent characterizations of the cancellation law.

Proposition 10.6.3.1: Cancellation Law

The following are equivalent in a ring R.

- If $ab = 0$ in R then $a = 0$ or $b = 0$
- If $ab = ac$ and $a \neq 0$ then $b = c$
- If $ba = ca$ and $a \neq 0$ then $b = c$

Proof. Let R be a ring.

- (1) $\implies$ (2) We have that $ab = ac$ implies $ab - ac = 0$ by additive inverse and $a(b - c) = 0$ by distributivity and we are done. (1) $\implies$ (3) is similar.
- (2) $\implies$ (1) Rewriting $ab = ac$ as $a(b - c) = 0$ gives the result. (3) $\implies$ (1) is similar.

□

If a domain satisfies the above cancellation law, we give it a fancy name.

Definition 10.6.3.2: Domain

A ring that is not equal to the 0 ring is a domain if the above condition is satisfied.

Note that we necessarily need $0 \neq 1$ for this to work or else the cancellation law will fail automatically. Now we give another fancy name when we flavour our domain with commutativity.

Definition 10.6.3.3: Integral Domain

A commutative domain is called an integral domain.

Integral domains appear commonly and we will work with them extensively throughout the next chapter. For now, see the following propsition.

Proposition 10.6.3.4

Every finite integral domain is a field.

Proof. Let $\{0, a_1, \dots, a_n\}$ be elements of a finite integral domain. We simply show that every element has a mupltiplicative inverse. Fix $1 \leq i \leq n$. Notice that the n products $a_i a_j$ for $1 \leq j \leq n$ are all distinct and nonzero. This means that $a_i a_1, \dots, a_i a_n$ is a permutation of $a_1, \dots, a_n$. Since one of $a_1, \dots, a_n$ is the multiplicative identity, we must have that for some j, $a_i a_j = 1$ thus we are done. □

Definition 10.6.3.5: Zero Divisors

A nonzero element $a \in R$ is called a zero divisor if there exists $b \in R$, where $b \neq 0$ such that $ab = 0$ or $ba = 0$.

Proposition 10.6.3.6

A commutative ring is an integral domain if it has no zero divisors.

Proof. Suppose that there are no zero divisors in a commutative ring R. Then for every $a \in R$, if $ab = 0$ for some $b \in R$, then $b = 0$. This amounts to saying the cancellation law. □

Below lists two of the many relationships between the fancy rings.

Proposition 10.6.3.7

Every field is an integral domain.

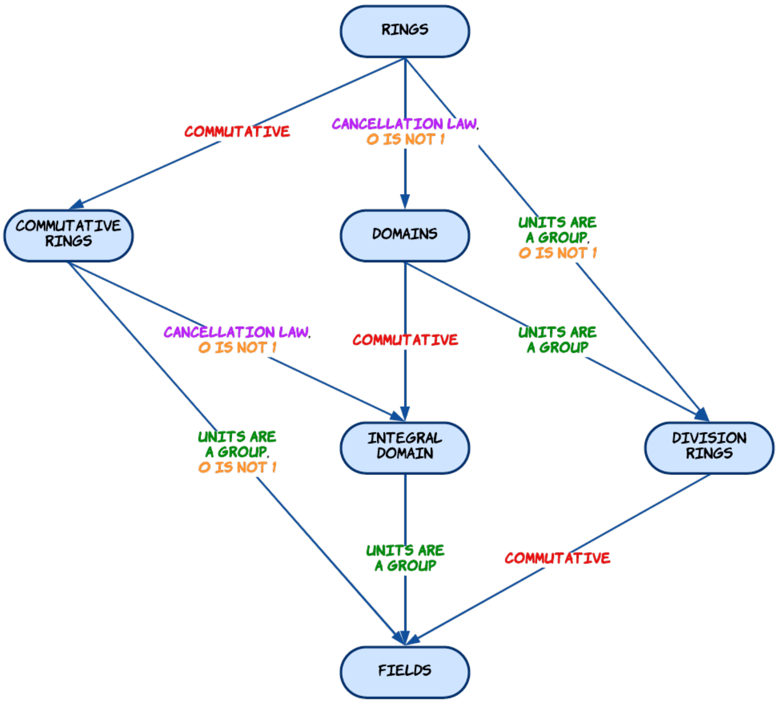
Proof. Every field is necessarily commutative. It remains to show that the cancellation law applies. However, since every field is a division ring, this means that every nonzero element is a unit. Let R be a field and $a, b \in R$ such that $ab = 0$. Then $a^{-1}ab = a^{-1}0$ thus $b = 0$ and we are done. □

Proposition 10.6.3.8

Every division ring is a domain.

Proof. Similar to the above proof, since every element is a unit, the cancallation law must apply to every element. □

The following diagram shows roughly the relationship between different types of rings.



I like that there is a subtle sense of symmetry here. The conditions on the arrow is the condition for the ring in the tail to be a ring in the arrow head. For instance, if a domain R satisfies the property that elements in R are commutative, then R is an integral domain. On the other hand, every integral domain will necessarily be a domain.

10.6.4 Ring Homomorphisms and Ideals

Definition 10.6.4.1: Ring Homomorphism

Let R, S be rings. A ring homomorphism is a map $\phi : R \to S$ such that

$$\phi(a+b) = \phi(a) + \phi(b)$$

and

$$\phi(ab) = \phi(a)\phi(b)$$

for all $a, b \in R$. A bijective ring homomorphism is called a ring isomomorphism.

Definition 10.6.4.2: Kernel

The kernel of the ring homomorphism ϕ, denoted $\ker(\phi)$ is defined as

$$\ker(\phi) = \{r \in R | \phi(r) = 0\}$$

Definition 10.6.4.3: Ideals

Let R be a ring. Let $I \subset R$ and let $r \in R$. Define $rI = \{ra | a \in I\}$ and $Ir = \{ar | a \in I\}$

- I is a left ideal of R if $(I, +)$ is a subgroup of $(R, +)$ and $rI \subseteq I$ for all $r \in R$
- I is a right ideal of R if $(I, +)$ is a subgroup of $(R, +)$ and $Ir \subseteq I$ for all $r \in R$
- A subset I that is both a left ideal and right ideal is called an ideal of R.

Lemma 10.6.4.4

Let I be an ideal of a ring R. Then $I = R$ if and only if there exists a unit $u \in R$ such that $u \in I$.

Proof. Let $I = R$. Then trivially $u \in I$. Now let $u \in I$ be a unit. Let $x \in R$, I want to show that $x \in I$ thus completing the proof that $R \subseteq I$. Now $x = x(u^{-1}u) = (xu^{-1})u$. Since $xu^{-1} \in R$ and $u \in I$, their muplication will also be in I, thus $x \in I$. □

This lemma is also very useful if you consider the fact that the mupltiplicative identity is also a unit.

Proposition 10.6.4.5

If I, J are ideals of a ring R then

$$I + J = \{i + j | i \in I, j \in J\}$$

and

$$I \cap J = \{r \in R | r \in I, r \in J\}$$

are both ideals of R.

Proof. Let R be a ring and I, J its ideals.

- We first show that $I + J$ is an ideal. Let $i_1 + j_1, i_2 + j_2 \in I + J$. Then since R is an abelian group an I, J are subgroups of R,

  $$i_1 + j_1 + i_2 + j_2 = (i_1 + i_2) + (j_1 + j_2) \in I + J$$

  thus closure is satisfied. Associativity is satisfied since it inherits from R. We also have $0 \in I, J$ thus $0 \in I + J$. The inverse of $i + j \in I + J$ is $-i - j \in I + J$ since $-i \in I$ and $-j \in J$. Thus $I + J$ is a subgroup of R.

Now let $r \in R$ and $i + j \in I + J$. Then $r(i + j) = ri + rj$ by distributivity. Since I, J are ideals, there exists $s \in I$ such that $s = ri$ and $t \in J$ such that $t = rj$. Then $r(i + j) = ri + rj = s + t \in I + J$ thus we are done.

- $I \cap J$ are already subgroups of R. Let $k \in I \cap J$. Let $r \in R$. Then $rk \in r(I \cap J)$. Then there exists $i \in I$ and $j \in J$ such that $i = rk = j$. Then $i = j \in I \cap J$ thus we are done.

□

Proposition 10.6.4.6

The kernel of any ring homomorphism is an ideal.

Proof. Kernels are trivially subgroups of $(R, +)$. Let $r \in R$. Let $x \in \ker(\phi)$. Then $\phi(rx) = \phi(r)\phi(x) = \phi(r) \cdot 0 \in \ker(\phi)$ thus we are done. □

Proposition 10.6.4.7: Quotient Ring

Let I be an ideal of a ring R. Then the cosets of I in R,

$$R/I = \{r + I | r \in R\}$$

is a ring with addition defined the same as quotient groups

$$(r + I) + (s + I) = (r + s) + I$$

and multiplication defined as

$$(r + I)(s + I) = \{ab \in R | a \in r + I, b \in s + I\} = rs + I$$

called the quotient ring of R by I.

Proof. Since $(I, +)$ is a subgroup of $(R, +)$, we know that R/I is already an abelian group. We simply show that multplication is well defined. Suppose that $x_1 + I = x_2 + I$ and $y_1 + I = y_2 + I$. Then

$$\begin{aligned}(x_1 + I)(y_1 + I) &= x_1y_1 + I \\ &= (x_1y_1 - x_1y_2 + x_1y_2 - x_2y_2 + x_2y_2) + I \\ &= (x_1(y_1 - y_2) + (x_1 - x_2)y_2 + x_2y_2) + I \\ &= x_2y_2 + I\end{aligned}$$

since $y_1 - y_2 \in I$ and $x_1 - x_2 \in I$. This shows that taking different represetatives of the cosets of an ideal does not matter for multiplication.

Now we show that $1 + I$ is the muplicative identity of R/I. Let $x + I \in R/I$. We have that $(1 + I)(x + I) = x + I$ thus we are done. Associativity and distributivity follows from the fact that R has these properties. □

Theorem 10.6.4.8

If I is an ideal of R, the map $\phi : R \to R/I$ defined by $\phi(r) = r + I$ is a surjective ring homomorphism with kernel I.

Proof. Let $a, b \in R$. Then

$$\begin{aligned}\phi(a+b) &= (a+b)+I\\ &= (a+I)+(b+I)\\ &= \phi(a)+\phi(b)\end{aligned}$$

Now let $r \in R$. Then

$$\begin{aligned}\phi(ra) &= ra+I\\ &= (r+I)(a+I)\\ &= \phi(r)\phi(a)\end{aligned}$$

Now we show that ϕ is surjective. Let $r + I \in R/I$. Then trivially $\phi(r) = r + I$ thus we are done.

Finally we show that $\ker(\phi) = I$. Let $a \in \ker(\phi)$. Then $\phi(a) = a + I = I$. This means that $a \in I$ and we have shown that $\ker(\phi) \subseteq I$. Now suppose that $a \in I$. Then $\phi(a) = a + I = I$. Thus $a \in \ker(\phi)$ and we are done. □

10.6.5 Types of Ideals

Lemma 10.6.5.1

Let R be a commutative ring and $a \in R$. Then $(a) = \{ar : r \in R\}$ is an ideal.

Proof. Let $s, t \in (a)$. Then $s = ar_1$ and $t = ar_2$ for some $r_1, r_2 \in R$. We show that (a) is a subgroup of R. We have

$$s + t = a(r_1 + r_2) \in (a)$$

Identity is also in (a) since $0 \in R$ and

$$a \cdot 0 = 0 \in (a)$$

To show inverse we have that $u = -ar_1$ and

$$s + u = ar_1 - ar_1 = 0$$

By the subgroup criterion, (a) is a group. We now show that $r(a) \subseteq (a)$. Let $r_1ar_2 \in r(a)$. Then

$$r_1ar_2 = ar_1r_2 \in (a)$$

since R is commutative. Thus (a) is an ideal. □

Definition 10.6.5.2: Principal Ideals

Let R be a commutative ring with identity. Let I be an ideal of R. Then an ideal I of the form

$$I = (a)$$

for some $a \in I$ is called a principal ideal.

Definition 10.6.5.3: Maximal Ideals

A proper ideal M of a ring R is a maximal ideal of R is the ideal M is not a proper subset of any ideal of R except R itself.

Becareful that maximal ideals are not necessarily unique. A typical example would be the fact that (2) and (3) are principle ideals that are both maximal in $\mathbb{Z}$.

Definition 10.6.5.4: Prime Ideals

A proper ideal P in a commutative ring R is called a prime ideal if $ab \in P$ implies $a \in P$ or $b \in P$.

Proposition 10.6.5.5

Let R be a commutative ring with identity and M an ideal of R. Then M is maximal if and only if R/M is a field.

Proof. Suppose that M is a maximal ideal. Let $x \notin R$. We show that $x+M$ has an inverse. We know that $M+(x)$ is an ideal containing M. Since M is maximal, we must have $M+(x)=R$. This means that $1 \in M+(x)$ which means there exists $m \in M$ and $r \in R$ such that $1=m+rx$. Now consider $r+M$. We have that

$$\begin{aligned}(x+M)(r+M) &= xr+M \\ &= (1-m)+M \\ &= 1+M\end{aligned}$$

Now suppose that R/M is a field. Let J be an ideal such that $I \subseteq J \subseteq R$ and $I \neq J$. Now let $x \in J \setminus I$. Then $I+x$ has a multiplicative inverse in R/I since $I+x \neq I$, the additive identity. Let the inverse be $I+y$. Then

$$I+1=(I+x)(I+y)=I+xy$$

Trivially $xy \in (x)$, thus $1 \in I+(x) \subseteq J$. But if the identity is in the ideal, the ideal is equal to the ring thus we are done. □

Proposition 10.6.5.6

Let R be a commutative ring with identity not equal to 0. Then P is a prime ideal in R if and only if R/P is an integral domain.

Proof. Suppose that P is a prime ideal. Then let $(a+P)(b+P)=P$. Since we also have that $(a+P)(b+P)=ab+P$. This means that $ab \in P$ thus either $a \in P$ or $b \in P$ which in turns leads to either $a+P=P$ or $b+P=P$ and thus the cancellation law applies.

Now suppose that R/P is an integral domain. Let $ab \in P$. Since R/P is an integral domain, we have that

$$(a+P)(b+P)=ab+P=P$$

means that either $a+P=P$ or $b+P=P$ which means that either $a \in P$ or $b \in P$ and we are done. □

Corollary 10.6.5.7

Every maximal ideal in a commutative ring with identity is also a prime ideal.

Proof. If M is maximal then R/M being a field implies that R/M is an integral domain thus M is prime. □

The proof is rather inconstructive in the sense that it does not provide a good insight to the structure, the reasoning behind why every maximal ideal is prime. For a more structure-revealing proof, consider the following alternative approach:

Let M be a maximal ideal of a commutative ring R. Let $ab \in M$ but $a \notin M$. Then $M + (a) = R$ since M is maximal. Then $1 \in R$ means that there exists $k \in M$ and $r \in R$ such that $k + ra = 1$. Mutplying by b gives $kb + rab = b$. Since $kb \in M$ and $rab \in M$, $b \in M$ and we are done.

10.7 Integral Domains

10.7.1 Field of Fractions

Definition 10.7.1.1: Fractional Equivalence

Let R be an integral domain. Let

$$S = \{(a, b) | a, b \in R \text{ and } b \neq 0\}$$

Define a relation on S by $(a, b) \sim (c, d)$ if $ad = bc$.

Lemma 10.7.1.2

The relation $\sim$ between elements of S is an equivalence relation.

Proof. Since R is an integral domain, symmetry is satisfied. Clearly it is reflexive since $ab = ba$. For transitivity, suppose that $ad = bc$ and $cf = de$. Then $adcf = bcde$ and by cancellation law, $af = be$. □

Proposition 10.7.1.3

The set of equivalence classes of S of an integral domain R, under the equivalence relation $\sim$, together with the operations of addition and multiplication defined by

$$(a, b) + (c, d) = (ad + bc, bd)$$

and

$$(a, b) \cdot (c, d) = (ac, bd)$$

is a field, called the field of fractions, denoted $Q(R)$.

Proof. Note that R is a field if and only if $(R, +)$ and $(R \setminus \{0\}, \cdot)$ are commutative groups and the distributive law holds.

- Let $(a, b), (c, d) \in S$. Then $ad + bc, bd \in R$ since R is closed thus $(ad + bc, bd) \in S$. Let (e, f) also be in S. Then

  $$\begin{aligned}((a, b) + (c, d)) + (e, f) &= (ad + bc, bd) + (e, f)\\ &= ((ad + bc)f + (bd)e, bdf)\\ &= (adf + bcf + bde, bdf)\end{aligned}$$

  and

  $$\begin{aligned}(a, b) + ((c, d) + (e, f)) &= (a, b) + (cf + de, df)\\ &= (a(df) + b(cf + de), b(df))\\ &= (adf + bcf + bde, bdf)\end{aligned}$$

  Thus associativity is satisfied. I claim that $(0, 1)$ is an identity. We have $(a, b) + (0, 1)(a \cdot 1 + b \cdot 0, b \cdot 1) = (a, b)$. If $(a, b) \in S$ then $(-a, b) \sim (a, -b)$ is an inverse. We have

  $$(a, b) + (-a, b) = (ab - ab, b^2) = (0, b^2) \sim (0, 1)$$

  Finally we have

  $$\begin{aligned}(a, b) + (c, d) &= (ad + bc, bd)\\ &= (da + cb, db) && (R \text{ is an Integral Domain})\\ &= (cb + da, db) \qquad = (c, d) + (a, b) && (R \text{ is an abelian group})\end{aligned}$$

  Thus we have shown that $(S, +)$ is an abelian group.

- We now show that $(S, \cdot)$ is an abelian group. Let $(a, b), (c, d), (e, f) \in S$. Then $(a, b) \cdot (c, d) = (ac, bd) \in S$ sincve $ac, bd \in R$ by closure of rings. Thus the closure property is satisfied. Associativity is inherited from R since elements in S are pairs of R. I claim that the identity is $(1, 1) \sim (k, k)$ for any $k \in R$. We have

$$(a, b) \cdot (1, 1) = (a \cdot 1, b \cdot 1) = (a, b)$$

If $(a, b) \in S$ then its inverse is (b, a). We have

$$(a, b) \cdot (b, a) = (ab, ba) = (ab, ab) = (1, 1)$$

thus $(S, \cdot)$ is a group. Now to show abelian, we have

$$(a, b) \cdot (c, d) = (ac, bd) = (ca, db) = (c, d) \cdot (a, b)$$

since R is an integral domain. Thus we have shown that $(S, \cdot)$ is an abelian group.

- Finally we show distributivity. Let $(a, b), (c, d), (e, f) \in S$. Then

$$\begin{aligned}(a, b) \cdot ((c, d) + (e, f)) &= (a, b) \cdot (cf + de, df) \\ &= (acf + ade, bdf)\end{aligned}$$

and

$$\begin{aligned}(a, b) \cdot (c, d) + (a, b) \cdot (e, f) &= (ac, bd) + (ae, bf) \\ &= (acbf + bdae, bdbf) \\ &= (acf + ade, bdf) && \text{(equivalence relation)}\end{aligned}$$

Thus we are done.

□

In particular, the field of fractions of $\mathbb{Z}$ is precisely $\mathbb{Q}$.

Lemma 10.7.1.4

For any integral domain R, R is a subring of $Q(R)$.

Proof. Define a function $\phi : R \to Q(R)$ by $\phi(r) = \frac{r}{1}$. Then ϕ_R is the identity homorphism thus $\phi(R) = R$. Then $\phi(R)$ is trivially a subring of $Q(R)$ and we are done. □

10.7.2 Divisibility

Definition 10.7.2.1: Division

Let R be a commutative ring and let $a, b \in R$ with $b \neq 0$. a is said to be a multiple of b if there exists an element $x \in R$ with $a = bx$. In this case b is said to divide a or be a divisor of a, written $b|a$.

Proposition 10.7.2.2

Let R be a commutative ring. Let $x, y \in R$ and $y \neq 0$. Then the following are equivalent.

- $x|y$
- $y \in (x)$
- $(y) \subseteq (x)$

Proof. Let R be a commutative ring. Let $x, y \in R$ and $y \neq 0$.

- (1) $\implies$ (2) Suppose that $y = kx = xk$ for some $k \in R$. Then $y \in (x) = \{ax|a \in R\}$ by definition.
- (2) $\implies$ (3) Suppose that $y \in (x)$. Then there exists $k \in R$ such that $y = kx = xk$. To show that $(y) \subseteq (x)$, let $ry \in (y)$. Then $ry = rkx$ and $rk \in R$ thus $ry \in (x)$ thus we are done.
- (3) $\implies$ (1) Suppose that $(y) \subseteq (x)$. Then there exists $k \in R$ such that $y = kx$. Then we are done.

□

Definition 10.7.2.3: Greatest Common Divisor

A greatest common divisor of a and b is a non-zero element d such that

- $d|a$ and $d|b$
- If $c|a$ and $c|b$ then $c|d$

It is denoted $\gcd(a, b)$.

Definition 10.7.2.4: Least Common Multiple

A least common multiple of a and b is a non-zero element l such that

- $a|l$ and $b|l$
- If $a|m$ and $b|m$ then $l|m$

It is denoted $\text{lcm}(a, b)$.

Unfortunately, these numbers do not always exists for any a, b in a general ring. We will prove their existence in principal ideal domains later.

Proposition 10.7.2.5

Let R be a commutative ring. Let $x, y \in R$ such that they are nonzero. If the ideal generated by a and b, namely (a, b) is a principal ideal (d), then d is the gcd of a and b.

Proof. Suppose that $(a, b) = (d)$ for some $d \in R$. Then $a \in (d)$ and $b \in (d)$ already implies that $d|a$ and $d|b$. Suppose that $c|a$ and $c|b$. This means that $a \in (c)$ and $b \in (c)$. Since $d \in (a, b)$ there exists $r, s \in R$ such that $ra + sb = d$. This means that $d \in (c)$ thus $c|d$. □

This explains the notation that (a, b) is often used to denote the gcd.

Definition 10.7.2.6: Associates

Let R be a commutative ring. Let $x, y \in R$. We say that x and y are associates if $x|y$ and $y|x$. We denote it as $x \sim y$.

Proposition 10.7.2.7

Let R be a integral domain. Let $x, y \in R$. Then the following are equivalent.

- $x \sim y$

- $(x) = (y)$
- There exists a unit $u \in R$ such that $x = qy$.

Proof.
- (1) $\implies$ (2): Suppose that $x \sim y$ then $x|y$ and $y|x$ which means that $(x) \subseteq (y)$ and $(y) \subseteq (x)$.
- (2) $\implies$ (3): Suppose that $(x) = (y)$. Then since $x \in (y)$, there exists $s \in R$ such that $x = sy$ and likewise $y = tx$. Then $x = stx$ and since R is an integral domain, $st = 1$ which means that s, t are units.
- (3) $\implies$ (1): Suppose that $x = qy$ for some unit q. Then clearly $y|x$. Since q is a unit, $q^{-1}x = y$ and thus $x|y$ which means that x and y are associates.

□

10.7.3 Primes and Irreducibles

Primes and irreducibles are two similar concepts, their difference is only made clear in Euclidean domains that are not principles ideal domains which we will see both notions later.

Definition 10.7.3.1: Irreducibles

Let D be an integral domain. A nonzero element $p \in D$ that is not a unit is said to be irreducible if $p = ab$ implies a or b is a unit.

Definition 10.7.3.2: Primes

Let D be an integral domain. A nonzero element $p \in D$ that is not a unit is said to be a prime if $p|ab$ implies $p|a$ or $p|b$.

Lemma 10.7.3.3

Let D be an integral domain. Let p be a non-unit. Then p is prime if and only if (p) is a prime ideal.

Proof. Suppose that p is prime. Suppose that $rp \in (p)$. Then $p \in P$ and we are done. Now suppose that (p) is a prime ideal. Suppose that $p|ab$. Then $pd = ab$ for some $d \in D$ thus $ab \in (p)$. WLOG take $a \in (p)$ by definition of prime ideal. Then we are done since $a \in (p)$ implies $p|a$. □

Proposition 10.7.3.4

Let D be an integral domain and $p \in D$. If p is a prime then p is irreducible.

Proof. Let p be a prime in D. Suppose that $p = ab$. Then trivially $p|ab$ thus $p|a$ or $p|b$. WLOG take $p|a$. Trivially $a|p$ since $a|ab$. This means that a and p are associates. Thus $p = aq$ for some unit q. Then since $aq = ab$ and integral domains have cancellation law, we must have $q = b$ which means that b is a unit. Thus p is irreducible. □

10.7.4 Unique Factorization Domains

Definition 10.7.4.1: Unique Factorization Domains

An integral domain D is a unique factorization domain if the following are true

- Let $a \in D$ such that $a \neq 0$ and a is not a unit. Then a can be written as the product of irreducible elements in D.
- Let $a = p_1 \cdots p_r = q_1 \cdots q_s$, where p_i and q_j are irreducible. Then $r = s$ and there is a permutation such that $p_i = q_{\pi(i)}$ for $i \in \{1, \dots, r\}$.

Notice that in general integral domains, primes are not the same as irreducibles. But they have the nice property that they coincide in UFDs, which is why they are put in the chapter on UFDs here. Below gives a full converse to the relation between prime and irreducibles we gave above under the umbrealla that is UFDs.

Proposition 10.7.4.2

Let D be a UFD and $p \in D$. Then p is a prime if and only if p is irreducible.

Proof. We have already shown the forward implication. Now let p be irreducible. We show that (p) is prime. Let $ab \in (p)$. Then there exists $d \in D$ such that $pd = ab$. We can factorize ab and pd respectively into a product of irreducibles elements in D. But since they are equal, by uniqness of fatorization, p is exactly an associate of one of the irreducibles in the factorization of ab. If p is in the factorization of a then $p|a$ and we are done. Otherwise it is in b and we are also done. □

Notice that in the above proof, the fact that every prime is irreducible does not use the properties of UFD. This means that this is true in general integral domains.

10.7.5 Principal Ideal Domains

Definition 10.7.5.1: Principal Ideal Domains

A principal ideal domain is an integral domain in which every ideal is principal, meaning every ideal is of the form

$$(a) = \{ra : r \in R\}$$

Proposition 10.7.5.2

Let R be a PID and $x, y \in R$. Then $\gcd(x, y)$ and $\text{lcm}(x, y)$ exists and there exists $r, s \in R$ such that

$$\gcd(x, y) = rx + sy$$

Proof. Let $x, y \in R$. Then $(x) + (y)$ is an ideal of R, thus it must be prciniple, say $(d) = (x) + (y)$. Similarly, $(x) \cap (y)$ is also an ideal, say $(l) = (x) \cap (y)$.

We prove that d and l are the gcd and lcm respectively. Trivially, $(x) \subseteq (d)$ and $(y) \subseteq (d)$ implies $d|x$ and $d|y$. Also for any z such that z divides x and y, $(x) \subseteq (z)$ and $(y) \subseteq (z)$ thus $(d) \subseteq (z)$. The proof is similar for (l).

Since $(d) = (x) + (y)$, there exists $r, s \in (x), (y)$ respectively such that $d = rx + sy$ and we are done. □

Proposition 10.7.5.3

Let R be a PID and (p) a nonzero ideal in R. Then the following are equivalent.

- (p) is maximal
- p is irreducible
- p is prime ((p) is prime)

We have seen that every maximal ideal is a prime ideal. We have seen that if (p) is a prime ideal then p is prime. We have also seen that every prime is irreducible. We show separately that if p is irreducible then p is a prime, and also if p is prime then (p) is maximal.

For the first part, suppose that p is irreducible. Suppose that $p|ab$ for $a, b \in R$. By the above proposition, $d = \gcd(p, a)$ exists. Then for some $t \in R$, $p = dt$. Since p is irreducible, either d or t is a unit. We consider both cases. If t is a unit, then p and d are associates and thus $p|d$. Since $d|a$, we have that $p|a$ and we are done. Now if d is a unit, then $d = ra + sp$ for some $r, s \in R$. Multplying both sides with b gives $db = rab + spb$. Since $p|ab$ and $p|spb$, we have that $p|db$. Then $pu = db$ for some unit $u \in R$. Since d is a unit, we have that $d^{-1}pu = b$, which means that $p|b$ and we are done.

Now we show that if p is a prime then (p) is maximal. We know that (p) is a prime ideal. Suppose that $(p) \subseteq (q) \subseteq R$ for some ideal q of R. Since $p \in (q)$, $p = tq$ for some $t \in R$. Since $p \in (p)$, we have that $tq \in (p)$. Now (p) is prime implies that either $t \in (p)$ or $q \in (p)$. If $q \in (p)$ then $(q) \subseteq (p)$ and thus $(q) = (p)$ and we are done. If $t \in (p)$, then $t = rp$ for some $r \in R$, which means that $p = rpq$. Then $rq = 1$ by cancallation law in integral domains. This means that $1 \in (q)$ thus $(q) = R$ and we are done.

Proposition 10.7.5.4

Every PID is a UFD.

Proof. Suppose that D is a principal ideal domain. Suppose for a contradiction that x, a non-unit cannot be factorized into a product of irreducibles. Clearly x is not irreducible else a contradiction. Then there exists x_1, y_1 non unit such that $x = x_1y_1$. Since x is assumed to be not a product of irreducibles, WLOG take x_1 to be not irreducible. Then we can repeat the process to get non units x_2y_2 such that $x_1 = x_2y_2$. Notice that $(x) \subset (x_1)$ is a proper containment of ideals if $x_1|x$. Then we have a chain of ideals

$$(x) \subset (x_1) \subset (x_2) \subset \dots$$

I claim that

$$I = \bigcup_{k=0}^{\infty} (x_k)$$

is an ideal. Indeed if $r, s \in I$, then $r \in (x_m)$ and $s \in (x_n)$ for some $m, n \in \mathbb{N}$. WLOG rtake $m \leq n$. Then $(x_m) \subseteq (x_n)$ implies $r, s \in (x_n)$ thus $r + s \in (x_n) \subseteq I$. Also if $t \in R$, then $tr \in (x_m) \subseteq I$ thus I is indeed an ideal.

Since R is a PID, there exists some $d \in I$ such that $I = (d)$. This also means that $d \in (x_m)$ for some $m \in \mathbb{N}$. Then this means that $I = (d) \subseteq (x_m)$. This proves that the chain eventually stops and this is a contradiction since we assumed that the chain of ideals are properly contained.

This menas that x can indeed be factorized into a product of irreducibles. □

Notice that the key in the proof is that the union of the countably finite principal ideals is again a principal ideals which allows the infinte chain of ideals to stop.

10.7.6 Euclidean Domains

Technically, division algorithms can exist in general domains so long that it has the notion of division. But without a measurement of size to guarantee division is taking larger numbers into smaller numbers, we cannot promise that division algorithms will halt eventually. Therefore we restrict the notion of division algorithm only to integral domains that has a notion of size.

Definition 10.7.6.1: Euclidean Valuation

Let R be an integral domain. A function $f : R \setminus \{0\} \to \mathbb{N} \cup \{0\}$ is said to be a Euclidean Valuation of R if

- $f(ab) \geq f(b)$ for all $a, b \in R \setminus \{0\}$
- For all $a, b \in R$ with $b \neq 0$, there exists q, r such that

  $$a = qb + r$$

  with $r = 0$ or $f(r) < f(b)$

In the above definition the second item is simply the division algorithm, with size of a number decided by the function f. Thus the Euclidean domain is simply an integral domain that possess a division algorithm.

Definition 10.7.6.2: Euclidean Domain

An integral domain R is said to be a Euclidean Domain that admits a Euclidean Valuation

Theorem 10.7.6.3

Let R be a Euclidean Domain. Let a, b be nonzero elements of R. Let $d = r_n$ be the last nonzero remainder in the Euclidean Algorithm for a and b. Then d is the greatest common divisor of a and b and (d) is generated by a and b. In particular, there exists $x, y \in R$ such that $d = ax + by$.

Proposition 10.7.6.4

Every Euclidean Domain is a PID.

In general, we have that

$$\text{Fields} \subset \text{Euclidean Domains} \subset \text{PID} \subset \text{UFD} \subset \text{Integral Domains}$$

in which all containments are strict.

10.8 Polynomials

10.8.1 Polynomials over General Rings

In this section we formulate the basic theory of generating polynomials from a ring.

Definition 10.8.1.1: Indeterminates

Let R be a ring. A symbol x is called an indeterminate over R if

$$a_0 + a_1x + a_2x^2 + \cdots + a_nx^n = 0$$

where $a_i \in R$, implies that $a_i = 0$ for each i.

Definition 10.8.1.2: Polynomial over a Ring

Any expression of the form

$$f(x) = \sum_{k=0}^{n} a_kx^k$$

where x is an indeterminate and $a_0, \dots, a_n \in R$ and $a_n \neq 0$ is called a polynomial over R. We define the the degree of f in this case to be n.

Definition 10.8.1.3: Ring of Polynomials

Let R a ring. Define $R[x]$ to be the set of all polynomials over R.

Proposition 10.8.1.4

Let R be a commutative ring with identity. Then $R[x]$ is a commutative ring with identity.

Proof. Since $R \leq R[x]$ by considering all the constant polynomials, the identity of R is also the identity of $R[x]$. Let $f, g \in R[x]$. Then the coefficient of x^n in $f(x)g(x)$ is

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

Since $a_kb_{n-k} = b_{n-k}a_k$, we have that $f(x)g(x) = g(x)f(x)$. □

Lemma 10.8.1.5

Let R be a ring. Then

$$\frac{R[x]}{(x)} \cong R$$

Proof. Taking the natural homomorphism $\phi : R[x] \to \frac{R[x]}{(x)}$ where (x) is the ideal, we have that $\ker(\phi) = (x)$. By the first ring isomorphism theorem,

$$\frac{R[x]}{(x)} \cong \phi(R[x])$$

But $\phi(R[x])$ is isomorphic to R since every polynomial in $R[x]$ is a sum of a constant function and an element in (x). □

Whenever something similar to the above notion is seen, say $\frac{\mathbb{Q}[x]}{x-3}$, one can think of it as, whenever I see a factor of $x - 3$ in $\mathbb{Q}[x]$, treat it as 0. So if you want to find the image of $x^2 - 4x + 5$ in the ring, we simply see that $x^2 - 4x + 5 = (x-3)(x-1) + 2 = 2$ since $x - 3$ is treated as 0. Recalling that this is the quotient ring, we essentially quotient out all polynomials that has this factor which leads to $x^2 - 4x + 5$ being treated as the same as the element 2 in the ring.

Theorem 10.8.1.6: Evaluation Theorem

Let R be a ring and let a be an element in the center $Z(R)$ of R. Define a mapping $\phi_a : R[x] \to R$ by

$$\phi_a\left(\sum_{k=0}^{n} c_k x^k\right) = \sum_{k=0}^{n} c_k a^k$$

Then ϕ_a is an onto ring homomorphism.

The evaluation maps gives a useful ring homomorphism to construct quotient rings.

10.8.2 Polynomials over Integral Domains

Proposition 10.8.2.1

If R is an integral domain then $R[x]$ is an integral domain. In particular the units in R are also units in $R[x]$.

Proof. Commutativity is clear since for $f, g \in R[x]$, coefficients of the product fg inherits commutativity from R thus $fg = gf$. Now we show that cancellation law exists in $R[X]$. Let $f, g \in R[x]$ such that $fg = 0$. Suppose for a contradiction that $f \neq 0$ and $g \neq 0$. This means that $\deg(f) = n$ and $\deg(g) = m$ for some $n, m \neq 0$. Consider the coefficient of x^{n+m} in fg, which is $a_n b_m$. Since $fg = 0$, $a_n b_m = 0$ and by cancellation law in R either $a_n = 0$ or $b_m = 0$. This contradicts the fact that $\deg(f) = n$ and $\deg(g) = m$ thus we are done.

The second part is trivial since $R \leq R[x]$ and they have the same identity. □

Proposition 10.8.2.2

Let R be an integral domain. Then if $f \neq 0$ and $g \neq 0$ in $R[x]$, then

$$\deg(fg) = \deg(f) + \deg(g)$$

Definition 10.8.2.3: Irreducible Polynomial

Let F be a field. A non constant polynomial $f \in F[x]$ is irreducible over F if whenever $f(x) = g(x)h(x)$ with $g, h \in F[x]$, then one of g or h is a unit.

This defintion corresponds to the same definition of irreduciblesgiven in the UFD section.

10.8.3 Polynomials over UFDs

The primary goal of this chapter is to compute results relating to finding out ireducible polynomials in UFD, rather than investigating the structure of polynomial rings.

Definition 10.8.3.1: Primitive

An element $0 \neq f \in R[x]$ where R is a unique factorization domain is called primitive if $\gcd(a_0, a_1, \dots, a_n) = 1$.

The reason that we have this notion is to prevent polynomials such as $5x - 5$ to be discussed. This clearly will not be irreducible since there is a factor of 5. Likewise if the gcd of all its coefficients are not 1, then you can factor out the gcd from the entire polynomial.

Proposition 10.8.3.2

The productive of two primitive polynomials is also primitive.

Proposition 10.8.3.3: Eisentein's Criterion

Let R be a UFD. Let $f \in R[x]$ be primitive. Suppose there is a prime $p \in R$ such that p does not divide a_n but $p|a_i$ for $0 \leq i \leq n-1$ and p^2 does not divide a_0. Then f is irreducible in $R[x]$.

Theorem 10.8.3.4

Let R be a UFD with field of fractions $Q = Q(R)$. Then a primitive polynomial in $R[x]$ is irreducible if and only if it is irreducible in $Q[x]$.

An immediate corollary for when $R = \mathbb{Z}$ is as follows.

Theorem 10.8.3.5: Gauss' Lemma

A primitive irreducible polynomial in $\mathbb{Z}[x]$ remains irreducible in $\mathbb{Q}[x]$

Proposition 10.8.3.6

R is an UFD if and only if $R[x]$ is an UFD.

10.8.4 Polynomials over a Field

This section will mainly be revisiting old notations with $F[x]$.

Theorem 10.8.4.1: Division Algorithm

Let F be any field and let f and g be polynomials in $F[x]$. Assume that $f \neq 0$ and that the leading coefficient of f is a unit in R. Then uniquely determined polynomials q and r exist in $F[x]$ such that

- $g = qf + r$
- Either $r = 0$ or $\deg(r) < \deg(f)$

In particular, deg is an Euclidean Valuation of $F[x]$ and $F[x]$ is a Euclidean domain.

The above theorem is equivalent to saying that $F[x]$ is a Euclidean domain as long as F is a field. Trivially, this also means that $F[x]$ is both a principal ideal domain and a unique factorization domain.

We give an alternate proof showing that $F[x]$ is a principal ideal domain.

Let I be a nontrivial ideal of $F[x]$. Let $f \in I$ be nonzero such that the degree of f is as small as possible. I claim that $(f) = I$. We already have that $(f) \subseteq I$ since $f \in I$. Now we show that $I \subseteq (f)$. So let $g \in I$. By the division algorithm, write $g = fq + r$ for some q, r such that $\deg(r) < \deg(f)$ or $\deg(r) = 0$. If $r \neq 0$, then $r = g - fq \in I$ since $f, g \in I$. But then $\deg(r) < \deg(f)$ means that f is not of smallest degree in I, a contradiction. Thus $g = fq$, which means that $g \in (f)$.

This is a constructive proof in the sense that if we would like to know the sole generator of an ideal in a polynomial ring $F[x]$ of a field, we simply take the polynomial of lowest degree.

Since we have shown that polynomials over a field are euclidean domains, the following theorems and definitions are also trivial.

Lemma 10.8.4.2

Let F be a field and suppose that $p \in F[x]$. Then the ideal generated by p is maximal if and only if p is irreducible.

Proof. Proved when we introduced irreducibility. □

Proposition 10.8.4.3: Greatest Common Divisor

Let f and g be nonzero polynomials in $F[x]$, where F is a field. Then a uniquely determined polynomial d exists in $F[x]$ satisfying the following conditions.

- d is monic
- d divides both f and g
- If h divides both f and g, then h divides d
- $d = uf + vg$ for some polynomials $u, v \in F[x]$

Proposition 10.8.4.4

Let $p \in F[x]$ be irreducible, F a field. If p divides a product $f_1 f_2 \cdots f_n$ of nonzero polynomials in $F[x]$, then p divides one of f_i.

Theorem 10.8.4.5: Unique Factorization Theorem

If F is a field, let f be a nonconstant polynomial in $F[x]$. Then

- $f = ap_1 p_2 \cdots p_n$, where $a \in F$ and p_i is monic and irreducible for all i
- The factorization is unique up to the order of the factors

We now begin dicussion of new notions that only polynomial rings based on a field will have.

Theorem 10.8.4.6: Factor Theorem

Let F be a field. Let $a \in R$. Let $f \in F[x]$. Then $f(a) = 0$ if and only if $f = (x - a)q$ for some $q \in F[x]$.

Proof. Clearly if f is of the form $f = (x - a)q$ then $f(a) = 0$.

Now suppose that $f(a) = 0$. We apply the division algorithm on f with $x - a$ to get

$$f(x) = (x - a)q(x) + r(x)$$

where either $r(x) = 0$ or $\deg(r) < 1$. This means that $r(x) = k$ for some constant $k \in F$. Since $f(a) = 0$, we have that $(a - a)q(a) + k = 0$ which means that $k = 0$ and we are done. □

Proposition 10.8.4.7

Let F be a field and let f be a nonzero polynomial in $F[x]$ of degree n. Then f has at most n roots in R.

Theorem 10.8.4.8: Remainder Theorem

Let F be a field. Let $a \in F$. Let $f \in F[x]$. If f is divided by $(x-a)$, the remainder is $f(a)$.

Proof. By division algorithm, there exists $q, r \in F[x]$ such that $f(x) = (x-a)q(x) + r(x)$. Evaluating x at a gives $f(a) = r(a)$. □

10.8.5 Number Fields

This section to be removed to Field Theory.

Definition 10.8.5.1: Algebraic Elements

Let $\alpha \in \mathbb{C}$. We say that α is algebraic over $\mathbb{Q}$ if there exists $f \in \mathbb{Q}[x]$ such that $f(\alpha) = 0$ where f is not the identity. Otherwise α is said to be transcendental.

Proposition 10.8.5.2

If α is an algebraic element of $\mathbb{C}$, then there exists a unique nonzero irreducible polynomial $f \in \mathbb{Q}[x]$ with leading coefficient 1 such that $f(\alpha) = 0$.

Using the first isomorphism theorem for rings, we se that the image of the evaluation map

$$\text{im}(\phi_\alpha) \cong \mathbb{Q}[x]/(f)$$

Since f is irreducible, $\mathbb{Q}[x]/(f)$ is a field, called a number field.

Definition 10.8.5.3: Number Fields

For any $\alpha \in \mathbb{C}$ an algebraic element, we define

$$\mathbb{Q}[\alpha] = \mathbb{Q}[x]/(f)$$

to be the number field containing α.

Let us look at an example.

Suppose we want to quotient out the polynomial $x^2 - 4x + 3$ in $\mathbb{Q}[x]$. Then notice that 3 and 1 are roots of the quadratic. So we have that

$$\mathbb{Q}[x]/(x-3) \cong \phi_3(\mathbb{Q}[x])$$

where ϕ_3 is the evaluation map at 3. This is proven as follows: We show that $\ker(\phi_3) = (x-3)$ the ideal and thus we can apply the first ring isomorphism theorem.

Suppose that $f \in \ker(\phi_3)$. Then $f(3) = 0$ and since $f \in \mathbb{Q}[x]$, f has a linear factor $x - 3$ by by division algorithm which we will see later. Then $f \in (x-3)$. Now if $g \in (x-3)$, then $g(x) = (x-3)f(x)$ for some $f \in \mathbb{Q}[x]$ and clearly $g(3) = 0$. Thus $\ker(\phi_3) = (x-3)$.

Now we reapply the first ring isomorphism theorem which the evaluation map

$$\phi_1 : \frac{\mathbb{Q}[x]}{(x-3)} \to \mathbb{Q}$$

with the same method and we get

$$\frac{\mathbb{Q}[x]}{(x-3)(x-1)} \cong \phi_1(\phi_3(\mathbb{Q}[x]))$$

10.9 Important Rings to Note

10.9.1 The Number Fields $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ and its Polynomial Rings

Theorem 10.9.1.1

The number fields $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ are all commutative rings with identity. In particular,

- $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ are Fields
- $\mathbb{Z}$ is an Euclidean Domain

Lemma 10.9.1.2

The field of fractions of $\mathbb{Z}$ is equivalent to $\mathbb{Q}$. Meaning $Q(\mathbb{Z}) = \mathbb{Q}$.

10.9.2 The Modulo Rings $\mathbb{Z}/n\mathbb{Z}$ and the Finite Fields $\mathbb{F}_p$

Theorem 10.9.2.1

The modulo rings $\mathbb{Z}/n\mathbb{Z}$ is a quotient ring that is commutative with kernel $n\mathbb{Z}$.

Theorem 10.9.2.2

$\mathbb{Z}/p\mathbb{Z}$ is a field with characteristic p if and only if p is a prime.

10.9.3 The Matrix Groups $GL(n, F)$, $SL(n, F)$ and $O(n)$

Theorem 10.9.3.1: The Matrix Groups

Let F be a field. Denote $GL(n, F)$ the set of all $n \times n$ matrices over F with nonzero determinant, called the general linear group, is a group. $SL(n, F)$, the special linear group, defined to be

$$SL(n, F) = \{M \in GL(n, F) | \det(M) = 1\}$$

is a subgroup of $GL(n, F)$. The orthogonal group, defined to be

$$O(n) = \{M \in GL(n, F) | M^T M = MM^T = I\}$$

is a subgroup of $SL(n, F)$.

Lemma 10.9.3.2

Let F be a field. Check that $O(n) \leq SL(n, F) \leq GL(n, F)$.

Chapter 11

Introduction to Number Theory

11.1 Properties of the Integers

11.1.1 Divisibility

We begin our study of number theory with divisibility.

Definition 11.1.1.1: Divisibility

Let $a, b \in \mathbb{Z}$. We define the relation

$$a|b$$

if and only if there exists some $k \in \mathbb{Z}$ such that $b = ak$. We say that a divides b in this case.

The definition is vey simple. The intuition is straight forward as well. Savour this moment as the subject increases its difficulty exponentially.

Proposition 11.1.1.2

Let $d, m, n \in \mathbb{Z}$. The relation $|$ has the following properties and thus is a partial order in $\mathbb{N}$.

- (Reflexivity) $n|n$
- (Antisymmetry) $m|n$ and $n|m \implies m = n$
- (Transitivity) $d|n$ and $n|m \implies d|m$
- (Linearity) $d|n$ and $d|m \implies d|(an + bm)$ for any $a, b \in \mathbb{Z}$
- $1|n$
- $n|0$

Proof. We prove antisymmetry and transitivity and leave the others for the reader. Let $m, n, d \in \mathbb{Z}$.

- (Antisymmetry) If $m|n$ and $n|m$ then there exists some $k_1, k_2 \in \mathbb{N}$ such that $n = k_1m$ and $m = k_2n$ thus $n = k_1k_2n$. Then $k_1k_2 = 1 \implies k_1 = k_2 = 1$ and $m = n$
- (Linearity) If $d|n$ and $n|m$ then there exists $k_1k_2 \in \mathbb{N}$ such that $n = k_1d$ and $m = k_2n$. Then $m = k_2k_1d$ thus $d|m$

□

These properties will come up again and again and will be the foundation of number theory. It is safe to say that number theory is built upon the notion of divisibility.

11.1.2 The Division Algorithm

This section is dedicated to develop the Euclidean algorithm, a means to find the greatest common divisor. The gcd is a central notion in number theory as well.

Definition 11.1.2.1: Greatest Common Divisor

Suppose that $m, n \in \mathbb{Z}$. A number $d \in \mathbb{N}$ such that

- $d \geq 0$
- $d|m$ and $d|n$
- $e|a$ and $e|b \implies e|d$

is called the greatest common divisor of m and n, denoted $\gcd(m, n)$.

In contrast to the greatest common divisor, we also have the lowest common multiple. Although they work as a pair, we often see the notion of gcd come up more than lcm.

Definition 11.1.2.2: Lowest Common Multiple

Suppose that $m, n \in \mathbb{Z}$. A number $l \in \mathbb{N}$ such that

- $l \geq 0$
- $m|l$ and $n|l$
- $m|e$ and $n|e \implies l|e$

is called the lowest common multiple of m and n, denoted $\text{lcm}(m, n)$.

Beware that both of these definitions does not imply the uniqueness of such a number. However, with a little work, we will see that both of them are indeed unique. Readers should think about whether the existence of these numbers is guaranteed as well.

Proposition 11.1.2.3

Let $m, n \in \mathbb{Z}$. $\gcd(m, n)$ and $\text{lcm}(m, n)$ are unique.

Proof. By the third property of both numbers, we must have if c, d are $\gcd(m, n)/\text{lcm}(m, n)$, then $c|d$ and $d|c$ thus $c = d$ and $\gcd(m, n)/\text{lcm}(m, n)$ is unique. □

We will see more on gcd and lcm when we deal with factorization. For now, we turn our heads to the division algorithm. This algorithm proves to us that upon dividing two integers, as long as they are not divisible by one or the other, you can always guarantee a remainder smaller than the divident.

Theorem 11.1.2.4: The Division Algorithm

Let $a \in \mathbb{N}$ and $b \in \mathbb{Z}$ with $b \neq 0$. Then there exists unique $q, r \in \mathbb{Z}$ such that

$$b = aq + r$$

with $0 \leq r < a$.

Proof. We prove existence first by considering three cases.
Cases 1: b is divisible by a. If b is divisible by a then there exists $k \in \mathbb{Z}$ such that $b = ka$ thus $k = q$ and $r = 0$.

Case 2: b is positive and a does not divide b. Let

$$S = \{b - ka \in \mathbb{N} | k \in \mathbb{N}\}$$

Then $S \subseteq \mathbb{N}$ thus we can apply the well-ordering principle to S. Let r be the least natural number in S. Then $r \in S$ implies $r = b - ka$ for some $k \in \mathbb{N}$. Thus $b = ka + r$ for some k and r. We show that $r < a$. Suppose for a contradiction that $r \geq a$. Then $u = r - a \in \mathbb{N}$ and

$$b = ka + r \implies b = ka + (u - a) \implies b = (k - 1)a + u$$

thus $u \in S$ and $u < r$, contradicting the fact that r is the least element in S. Thus $r \leq a$. If $r = a$, then

$$b = ka + a \implies b = (k + 1)a$$

which means that $a|b$ which is false in our case. Thus we must have $r < a$.

Case 3: b is negative and a does not divide b. Then apply the exact same argument to the number $-b$ to get $(-b) = ka + r$ and $b = -ka - r$. Let $k' = -k - 1$ and $r' = -r + a$. Then

$$b = -ka - r = k'a + a + r' - a = k'a + r'$$

Since we have $0 \leq r < a$, we have $-a < -r \leq 0$ and $0 < r' \leq a$. Again $r' \neq a$ or else $a|b$ which contradicts our assumption.

We now prove uniqueness. Suppose that $b = aq_1 + r_1$ and $b = aq_2 + r_2$. Then $r_1 - r_2 = a(q_2 - q_1)$. We know that $-a < r_1 - r_2 < a$ thus $-a < a(q_2 - q_1) < a$ and $-1 < q_2 - q_1 < 1$ which is impossible for integers q_1, q_2 unless $q_1 = q_2$. If $q_1 = q_2$ then $r_1 = r_2$ and we are done. □

The division algorithm does not require b to be larger than a. In fact, if a is larger than b, then the division algorithm simply gives a itself as the remainder. Before we reach our conclusion, we need one more proposition.

Proposition 11.1.2.5

Suppose that $m \geq n > 0$ are natural numbers with $m = qn + r$ for some $q, r \in \mathbb{N}$. Then

$$\gcd(m, n) = \gcd(n, r)$$

Proof. Suppose that $d = \gcd(m, n)$. Then we know that $d < n$ from definition. We want to show that d satisfies the three results of a gcd but in terms of n and r. Since $d|n$ and $d|m$, by linearity we must have $d|r$.

Now suppose for a contradiction that there exists e such that e is a common divisor of n and r and $e > d$. Then $e|n$ and $e|r$ by definition thus $e|m$ by linearity. $e|m$ and $e|n$ implies that e is a larger common divisor of m and n than d. However this is not possible since d is assumed to be the largest among the common divisors. This is a contradiction thus $d = \gcd(n, r)$ and we are done. □

Theorem 11.1.2.6: Euclid's Algorithm

Suppose that $m \geq n > 0$ are natural numbers. We have the following inequalities.

$$m = nq_1 + r_1 \text{ with } 0 < r_1 < n$$

$$n = r_1 q_2 + r_2 \text{ with } 0 < r_2 < n$$

$$r_1 = r_2 q_3 + r_3 \text{ with } 0 < r_3 < n$$

$$\cdots\cdots\cdots\cdots\cdots$$

$$r_{k-2} = r_{k-1}q_k + r_k \text{ with } 0 < r_k < r_{k-1}$$

$$r_{k-1} = r_k q_{k-1}$$

From this, we have $r_k | r_{k-1}$, $r_k | r_{k-2} \dots r_k | n$ and $r_k | m$.

Proof. The first part of the results is due to the repeated use of the division algorithm. For the second part, we have

$$\gcd(m,n) = \gcd(n, r_1) = \gcd(r_1, r_2) = \cdots = \gcd(r_{k-1}, r_k) = r_k$$

and we are done. □

Lemma 11.1.2.7: Bezout's Lemma

Let $a, b \in \mathbb{Z}$ such that they are not both 0. Then there exists $x, y \in \mathbb{Z}$ such that

$$ax + by = \gcd(a, b)$$

Proof. Reconstruct x and y using the Euclidean Algorithm. This is possible since $\gcd(m,n) = r_k$ and every $r_1, \dots, r_{k-1}$ has a factor of r_k in it. □

Lemma 11.1.2.8

Let $a, b \in \mathbb{Z}$ such that they are not both 0. Then the equation

$$ax + by = \gcd(a, b)$$

has an infinite number of integer solutions.

Proof. Using Bezout's Lemma, we conclude that (x_0, y_0) is a solution to the equation. But then

$$(x_0 - bt, y + at)$$

are also solutions for $t \in \mathbb{Z}$ since

$$a(x_0 - bt) + b(y + at) = ax + by = \gcd(a, b)$$

□

11.1.3 Unique Factorization

Definition 11.1.3.1: Prime Numbers

We say that $n \in \mathbb{N}$ is a prime number if and only if it has exactly two factors, which is 1 and n. Else n is composite.

Lemma 11.1.3.2

Every integer is divisible by a prime.

Proof. If the integer is a prime then it divides itself. If the integer is not a prime then it has some other factor $k < n$ not equal to 1 or n. If k is prime then we are done. If k is not prime then there is another non trivial factor $k_1 < k$. Repeat this process until you reach a prime. This is always possible since the integers are well ordered integers between 1 and n are finite. □

Lemma 11.1.3.3

Every integer $n > 1$ can be written as a product of primes.

Proof. If n is a prime that we are already done. If n is not a prime then we know that it is divisible by a prime p. Then repeat this procedure on $\frac{n}{p}$ until the remaining integer is a prime. □

Theorem 11.1.3.4

There is an infinite number of primes.

Proof. Suppose for a contradiction that there is only a finite number of primes $p_1, \dots, p_n$. Then I claim that $p = p_1, \cdots p_n + 1$ is a prime. □

Proposition 11.1.3.5: Euclid's Lemma

Suppose that $p, m, n \in \mathbb{N}$, with p prime and $m, n > 1$. Suppose that $p|mn$. Then p divides at least one of m or n.

Proposition 11.1.3.6

Suppose that p is a prime such that $p|a_1a_2 \cdots a_k$. Then $p|a_i$ for some $i \in \{1, 2, \dots, k\}$.

Proof. Treat $a_2 \cdots a_k$ as one integer. By Euclid's lemma, p either divides a_1 or $a_2 \cdots a_k$. If p divides a_1 we are done. If it doesn't then $p|a_2 \cdots a_k$. Repeat this procedure until one of a_i is divisible by p or we reach $p|a_{k-1}a_k$. Then by Euclid's lemma $p|a_{k-1}$ or $p|a_k$ and we are done. □

Proposition 11.1.3.7

Let $d, m, n \in \mathbb{Z}$. If $\gcd(m, d) = 1$, then $d|mn$ implies $d|n$.

Theorem 11.1.3.8: Fundamental Theorem of Arithmetic

Suppose that $n \neq 0$ is a natural number. Then there exists exactly one prime factorization for every n, meaning that the decomposition

$$n = \prod_{k=1}^{n} p_k^{s_k}$$

where p_k is prime exists and is unique.

Theorem 11.1.3.9

Suppose that $m, n \in \mathbb{N}$. Suppose that

$$m = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_r^{\alpha_r}$$

$$n = p_1^{\beta_1} p_2^{\beta_2} \cdots p_q^{\beta_q}$$

with $p_1 = 2$, $p_2 = 3$, $p_3 = 5\ldots$. Without loss of generality $r \leq q$. Then

$$\gcd(m,n) = p_1^{\min(\alpha_1,\beta_1)} p_2^{\min(\alpha_2,\beta_2)} \cdots p_q^{\min(\alpha_q,\beta_q)}$$

$$\operatorname{lcm}(m,n) = p_1^{\max(\alpha_1,\beta_1)} p_2^{\max(\alpha_2,\beta_2)} \cdots p_q^{\max(\alpha_q,\beta_q)}$$

Proof. This is direct from the definition of $\gcd(m,n)$ and $\operatorname{lcm}(m,n)$ and the fact that $p_k^{\min(\alpha_k,\beta_k)}|m$ and n but $p_k^{\min(\alpha_k,\beta_k)+1}$ either does not divide m or n. The proof for $\operatorname{lcm}(m,n)$ is similar. □

Theorem 11.1.3.10

Suppose that m and n are natural numbers. Then

$$\gcd(m,n) \times \operatorname{lcm}(m,n) = m \times n$$

Proof. Since $\min\{a,b\} \cdot \max\{a,b\} = ab$, from the above theorem, we have that $\gcd(m,n) \times \operatorname{lcm}(m,n) = m \times n$ and we are done. □

11.2 Congruences

11.2.1 Modular Arithmetic

Definition 11.2.1.1: Modulo Notation

We say that $a \in \mathbb{Z}$ is congruent to $b \in \mathbb{Z}$ modulo $n \in \mathbb{N}$ if and only if $m|(a-b)$. We write it as $a \equiv b \pmod{n}$.

Proposition 11.2.1.2

The congruence relation is an equivalence relation. We denote the equivalence class as

$$\mathbb{Z}/n\mathbb{Z}$$

with elements in it as either $m \in \mathbb{Z}/n\mathbb{Z}$, $[m] \in \mathbb{Z}/n\mathbb{Z}$ or $m + n\mathbb{Z} \in \mathbb{Z}/n\mathbb{Z}$.

Proof. The three facts

- (Reflexivity) $a \equiv a \pmod{m}$
- (Symmetry) $a \equiv b \pmod{m}$ if and only if $b \equiv a \pmod{m}$
- (Transitivity) $a \equiv b \pmod{m}$ and $b \equiv c \pmod{m} \implies a \equiv c \pmod{m}$

are obvious to prove. □

Group and ring theory play a very important role in abstract algebra. Abstract algebra is practically invented to investigate properties of integers.

Proposition 11.2.1.3

Suppose that $a, b, c, d \in \mathbb{Z}$. Then

- (Addition) $a \equiv b \pmod{m}$ and $c \equiv d \pmod{m} \implies a + c \equiv b + d \pmod{m}$
- (Multiplication) $a \equiv b \pmod{m}$ and $c \equiv d \pmod{m} \implies ac \equiv bd \pmod{m}$

and thus $\mathbb{Z}/n\mathbb{Z}$ form a ring.

Proof. Easy expansion involving rewriting the modulo definition into its divisibility equivalence. □

Proposition 11.2.1.4

If $ac \equiv bc \pmod{m}$ and $\gcd(c, m) = 1$ then

$$a \equiv b \pmod{m}$$

Proof. If $ac \equiv bc \pmod{m}$, then $ac = bc + km$ for some $k \in \mathbb{Z}$. Then $c(a-b) = km$. But m does not divide c so m must divide $a - b$. Thus $a \equiv b \pmod{m}$. □

Proposition 11.2.1.5

If $ac \equiv bc \pmod{m}$ and $\gcd(c, m) = d$ then

$$a \equiv b \pmod{m/d}$$

Proof. If $ac \equiv bc \pmod{m}$, then $ac = bc + km$ for some $k \in \mathbb{Z}$. Then $c(a-b) = km$ and $\frac{c}{d}(a-b) = k\frac{m}{d}$. The same thing happens since $\frac{m}{d}$ does not divide $\frac{c}{d}$ so it must divides $a-b$. Thus $a \equiv b \pmod{m/d}$ and we are done. □

11.2.2 Linear Congruences

Lemma 11.2.2.1

If $\gcd(a, m)$ does not divide b, then

$$ax \equiv b \pmod{m}$$

has no solutions.

Proof. This lemma is equivalent to asking whether

$$ax - my = b$$

has integer solutions, which has no solution according to Bezout's lemma. □

Lemma 11.2.2.2

If $(a, m) = 1$, then

$$ax \equiv b \pmod{m}$$

has exactly one solution modulo m.

Proof. The question is equivalent to finding integers x, y such that $ax = by + m$ holds. Rewriting this gives $ax - by = m$ which is Bezout's lemma. Thus existence of solution is guaranteed.

We need to show that there are no other solutions modulo m. □

Corollary 11.2.2.3

Let $m \in \mathbb{N}$ and $a \in \mathbb{Z}$. Then a has an inverse modulo m if and only if $\gcd(a, m) = 1$.

Proof. From the above lemma, we know that if $\gcd(a, m) = 1$ then $ax \equiv 1 \pmod{m}$ has exactly one solution thus we are done.

If $ax \equiv 1 \pmod{m}$ has a unique solution, then $ax = 1 + km$ for some $k \in \mathbb{Z}$. This is just rewriting bezout's lemma with $ax - km = 1$ thus we know that this means that $\gcd(a, m) = 1$. □

Lemma 11.2.2.4

Let $d = (a, m)$. If $d|b$, then

$$ax \equiv b \pmod{m}$$

has exactly d solutions.

Theorem 11.2.2.5: Chinese Remainder Theorem

Let $m_1, \dots, m_k \in \mathbb{N}$ be pairwise coprime and let $a_1, \dots, a_k \in \mathbb{Z}$. Then there exists $x \in \mathbb{Z}$, unique to modulo $\prod_{i=1}^k m_k$ such that

$$x \equiv a_i \pmod{m_i}$$

for $1 \leq i \leq k$. This solution is given by

$$x = \sum_{t=1}^k a_t M_t y_t$$

where $M_t = \prod_{j \neq t} m_j$ and $M_t y_t \equiv 1 \pmod{m_t}$. Any other integer z is a solution to the answer as long as $x \equiv z \pmod{m_1 \times \cdots \times m_k}$

Proof. We show that x is indeed congruent to a_i modulo m_i for $1 \leq i \leq k$. Note that for $t \neq i$, m_i is a factor of M_t. Thus for $t \neq i$, $a_t M_t y_t \equiv 0 \pmod{m_i}$. Thus

$$\begin{aligned} x &\equiv \sum_{t=1}^k a_t M_t y_t \pmod{m_i} \\ &\equiv a_i M_i y_i \\ &\equiv a_i \end{aligned}$$

Since $M_i y_i \equiv 1 \pmod{m_i}$.

Now we show uniqueness. Suppose that x, y are two solutions, then $x - y$ is divisible by $m_1, \dots, m_k$. As $m - 1, \dots, m_k$ are coprime, we have that $m_1 \cdots m_k | x - y$ thus x is in fact congruent to y. □

In practice, you are suppose to find y_i by yourself using the fact that $M_i y_i \equiv 1 \pmod{m_i}$. An algorithm for solving for the system of linear congruences is given as follows:

Step 1: Convert the system of linear congruences $a_i x \equiv b_i \pmod{m_i}$ into the form $x_i \equiv c_i \pmod{m_i}$ by finding the inverse of a_i modulo m_i.
Step 2: Compute $M_t = \frac{1}{m_t} \Pi_{i=1}^k m_k$
Step 3: Find y_t from $M_t y_t \equiv 1 (\text{mod } m_i)$
Step 4: Find x from $x = \sum_{t=1}^k a_t M_t y_t$

11.2.3 Multiplicative Functions

Definition 11.2.3.1: Multiplicative Functions

We say that $f : \mathbb{Z} \to \mathbb{Z}$ is a multiplcative function if $(m, n) = 1$ implies $f(mn) = f(m)f(n)$.

Definition 11.2.3.2: Sum and Number of Divisors

Let $n \in \mathbb{N}$. Denote

$$d(n) = \sum_{d|n} 1$$

the number of positive divisots of n and

$$\sigma(n) = \sum_{d|n} d$$

the sum of the positive divisors of n.

Theorem 11.2.3.3

$d(n)$ and $\phi(n)$ are multiplicative. Meaning if $n = \prod_{i=1}^{k} p_i^{r_i}$ is the prime decomposition of $x \in \mathbb{N}$, then

$$d(n) = \prod_{i=1}^{k} d(p_i^{r_i})$$

and

$$\sigma(n) = \prod_{i=1}^{k} \sigma(p_i^{r_i})$$

Definition 11.2.3.4: Euler's Totient Function

Let $n \in \mathbb{N}$. Define the euler totient function to be

$$\phi(n) = \sum_{\substack{(d,m)=1\\ d\leq m}} 1 = |\{k \in \mathbb{N} | \gcd(k, n) = 1, 1 \leq k \leq n\}|$$

the number of positive integers less than or equal to itself that is relatively prime. In particular, $\phi(m)$ is the order of the group $(\mathbb{Z}/m\mathbb{Z})^\times$

It is clear that the order of the group $(\mathbb{Z}/m\mathbb{Z})^\times$ has to exclude all elements in $\mathbb{Z}/m\mathbb{Z}$ that has a multplicative inverse as a ring, which is exactly the elements $k + m\mathbb{Z} \in \mathbb{Z}/m\mathbb{Z}$ with $\gcd(k, m) = 1$

Theorem 11.2.3.5: Euler's Theorem

Suppose that $m \geq 1$ and $(a, m) = 1$. Then

$$a^{\phi(m)} \equiv 1 \pmod m$$

Proof. Easy proof by considering Lagrange's theorem. □

Lemma 11.2.3.6

Let p be a prime. Then

$$\phi(p^n) = p^{n-1}(p - 1)$$

Proof. $\phi(p) = p - 1$ is trivial since a prime is coprime to all numbers except 1. Let $n \geq 1$, Then the positive integers less than and not coprime with p^n are exactly $p, 2p, 3p, \dots, p^{n-1}p$. There are p^{n-1} of them. Thus we have that

$$\begin{aligned}\phi(p^n) &= \text{all numbers less than } p^n - \text{ less than and not coprime with } p^n\\ &= p^n - p^{n-1}\\ &= p^{n-1}(p - 1)\end{aligned}$$

□

Theorem 11.2.3.7

$\phi(n)$ is multiplicative. Meaning if $n = \prod_{i=1}^{k} p_i^{r_i}$ is the prime decomposition of $x \in \mathbb{N}$, then

$$\phi(n) = \prod_{i=1}^{k} \phi(p_i^{r_i})$$

Proof. We appeal to the Chinese Remainder Theorem for Rings. We have that $(p\mathbb{Z})$ and $(q\mathbb{Z})$ are coprime ideals in $\mathbb{Z}$ since $(p\mathbb{Z})+(q\mathbb{Z})=(1)=\mathbb{Z}$ from the fact that there exists $x,y\in\mathbb{Z}$ such that $px+qy=1$ from Bezout's lemma. Notice also that $(p\mathbb{Z})\cap(q\mathbb{Z})=(pq\mathbb{Z})$We can thus apply the Chinese Remainder Theorem for Rings and have that

$$\mathbb{Z}/p\mathbb{Z}\times\mathbb{Z}/q\mathbb{Z}\cong\mathbb{Z}/pq\mathbb{Z}$$

Now notice that in ring products, $(r,s)\in R\times S$ is a unit if and only if r is a unit in R and s is a unit in S. Thus we have that the number of units in $\mathbb{Z}/pq\mathbb{Z}$ is exactly the product of the number of units in $p\mathbb{Z}$ and the number of units in $q\mathbb{Z}$. Since the number of units in a $\mathbb{Z}/m\mathbb{Z}$ is exactly $\phi(m)$, we are done. □

Corollary 11.2.3.8

If $n=\prod_{i=1}^k p_i^{r_i}$ is the prime decomposition of $x\in\mathbb{N}$, then

$$\phi(n)=\prod_{i=1}^k p_i^{r_i-1}(p_i-1)=n\prod_{i=1}^k\left(1-\frac{1}{p_i}\right)$$

Proof. This is direct from the fact that ϕ is multiplicative and that $\phi(p^n)=p^{n-1}(p-1)$. □

Theorem 11.2.3.9

If $n\geq 1$, then

$$\sum_{d|n}\phi(d)=n$$

11.2.4 Special Congruences

Lemma 11.2.4.1

If $\gcd(a,m)=1$, then the least residues of $a,2a,\dots,(m-1)a$ are

$$1,2,\dots,m-1$$

in some order.

Proof. We show that no two residues in the set $\{a,2a,\dots,(m-1)a\}$ is congruent to complete the proof. Suppose for a contrary that there exists $1\leq r,s\leq m-1$ such that $ra\equiv sa\pmod{p}$. Then $(r-s)a\equiv\pmod{p}$ and $\gcd(a,m)=1$ implies $r\equiv s\pmod{p}$. Thus r and s in fact are the same element in the set $\{a,2a,\dots,(m-1)a\}$ and we are done. □

Theorem 11.2.4.2: Fermat's Theorem

If p is a prime and $\gcd(a,p)=1$. Then

$$a^{p-1}\equiv 1\pmod{p}$$

Proof. Using the above lemma, we find that

$$a \cdot \dots \cdot (p-1)a \equiv 1 \cdot \dots \cdot (p-1) \pmod{p}$$
$$(p-1)!a^{p-1} \equiv (p-1)! \pmod{p}$$

Since $(p-1)!$ and p are relatively prime, we can cancel it out to get

$$a^{p-1} \equiv 1 \pmod{p}$$

□

We will see a vast generalization of Fermat's theorem soon involving general modulo instead of primes. It involves the notion of groups.

Lemma 11.2.4.3

The congruence equation

$$x^2 \equiv 1 \pmod{p}$$

has exactly two solutions, 1 and $p-1$.

Proof. It is easy to check that 1 and $p-1$ are indeed solutions of the congruence equation. Now let r be a solution to the linear congruence. Then

$$(r-1)(r+1) \equiv 0 \pmod{p}$$

Hence either $p|r+1$ or $p|r-1$. This means that either $r \equiv -1 \pmod{p}$ or $r \equiv 1 \pmod{p}$ thus we are done. □

Lemma 11.2.4.4

Let p be an odd prime. For every $a \in \{1, \dots, p-1\}$, there exists a unique $b \in \{1, \dots, p-1\}$ such that $ab \equiv 1 \pmod{p}$ such that eventually we can pair up the numbers in $\{1, \dots, p-1\}$ so that they are inverses of each other.

Moreover, the only elements with inverse as itself is precisely 1 and $p-1$.

Proof. Notice that since p is a prime, $\gcd(a, p) = 1$ for any a in the set. We have proven that this guarantees an inverse for a that is unique up to modulo p. The above lemma has also shown that $x^2 \equiv 1 \pmod{p}$ precisely have two solutions. □

Theorem 11.2.4.5: Wilson's Theorem

p is a prime if and only if

$$(p-1)! \equiv -1 \pmod{p}$$

Proof. From the above lemma, we pair up elements in the set $\{1, \dots, p-1\}$ so that multiplication in the congruence relation gives 1. Then we have that

$$\begin{aligned}(p-1)! &\equiv 1 \cdot 1 \cdot (p-1) \pmod{p} \\ &\equiv -1\end{aligned}$$

Now suppose that $(p-1)! \equiv -1 \pmod{p}$. Suppose for a contradiction that p is not prime. Then $p = ab$ for some $a, b \in \mathbb{N}$ where $a, b < p$. Then since $(p-1)!$ necessarily contains one multiple of a and b, $(p-1)!$ will contain a copy of p in it and thus is divisible by p, a contradiction. □

We will see a proof of similar style when we encounter Euler's Criterion.

Lemma 11.2.4.6

Let p be prime. Let $1 \leq k < p$ be a positive integer. Then

$$\binom{p}{k} \equiv 0 \pmod{p}$$

Proof. Clear there is a factor of p in $\frac{p!}{k!(p-k)!}$ since p is a prime and p is not contained in $k!$ or $(p-k)!$. □

Notice that the proof goes wrong if we relax the conditions to general numbers instead of prime numbers because divisors of p in this case could lie in $k!$ or $(p-k)!$.

Lemma 11.2.4.7: Power-Up Lemma

Let p be a prime. Let $k \in \mathbb{N}$. Suppose that $a \equiv b \pmod{p^k}$. Then

$$a^p \equiv b^p \pmod{p^{k+1}}$$

Proof. Suppose that $a = b + cp^k$ for some $c \in \mathbb{Z}$. Then we have that

$$\begin{aligned} a^p &\equiv (b + cp^k)^p \pmod{p^{k+1}} \\ &\equiv \sum_{k=0}^{p} \binom{p}{k} b^k c^{p-k} p^{k(p-k)} \pmod{p^{k+1}} \\ &\equiv b^p \pmod{p^{k+1}} \end{aligned}$$

using the binomial theorem. □

Corollary 11.2.4.8

Let p an odd prime. Let $k \geq 2$ be an integer. Then

$$(1+ap)^{p^{k-2}} \equiv 1 + ap^{k-1} \pmod{p^k}$$

where $a \in \mathbb{Z}$.

11.2.5 Order and Primitive Roots

Definition 11.2.5.1: Order

Let $m \in \mathbb{Z}$ and $a \in \mathbb{Z}$ such that $\gcd(a, m) = 1$. Define the order of a modulo m to be the smallest natural number d such that

$$a^d \equiv 1 \pmod{m}$$

In particular, the order of a modulo m is equivalent to saying the order of $a \in (\mathbb{Z}/m\mathbb{Z})^\times$.

Theorem 11.2.5.2

Let $\gcd(a, m) = 1$ and a has order d modulo m. Then $a^n \equiv 1 \pmod{m}$ if and only if $d|n$.

Proof. Suppose that $a^n \equiv 1 \pmod{m}$. By the division algorithm, we have that $n = dq + r$ for some $q, r \in \mathbb{Z}$ and $0 \leq r < d$. Then

$$\begin{aligned} a^n &\equiv a^{dq+r} \pmod{m} \\ &\equiv (a^d)^q \cdot a^r \pmod{m} \\ &\equiv a^r \pmod{m} \end{aligned}$$

This mean that $a^r \equiv 1 \pmod{m}$ which means $r = 0$ since $r < d$ and d is the order of a.

Now suppose that $d|n$. Then

$$\begin{aligned} a^n \equiv (a^d)^{n/d} &\pmod{m} \\ 1 &\pmod{m} \end{aligned}$$

Thus we are done. □

Lemma 11.2.5.3

If $\gcd(a, m) = 1$ and a has order d modulo m, then $d|\phi(m)$.

Proof. Apply Euler's theorem and the above theorem. □

Theorem 11.2.5.4

If the order of a modulo m is t then $a^r \equiv a^s \pmod{m}$ if and only if $r \equiv s \pmod{t}$.

Proof. Suppose that $a^r \equiv a^s \pmod{m}$. WLOG let $r \geq s$. Then $a^{r-s} \equiv 1 \pmod{m}$ which is true if and only if $r - s$ is a multiple of t. The process can be reversed for the if part. □

Theorem 11.2.5.5

Let a have order d modulo m. Let $u \in \mathbb{N}$. Then

$$\text{ord}_m(a^k) = \frac{\text{ord}_m(a)}{\gcd(k, \text{ord}_m(a))}$$

Proof. This is proven in groups and rings. □

Definition 11.2.5.6: Primitive Root

We say that a is a primitive root of m if $(a, m) = 1$ and the order of a modulo m is $\phi(m)$. In particular, a being a primitive root of m is equivalent to saying that a generates $(\mathbb{Z}/m\mathbb{Z})^\times$

Corollary 11.2.5.7

If p is a prime then there are exactly $\phi(p-1)$ primitive roots modulo p.

Proof. □

Theorem 11.2.5.8

If g is a primitive root of m, then g^t is a primitive root modulo p if and only if $(t, \phi(m)) = 1$.

Notice that this is a rather strong statement. Once we are successful in finding one primitive root, we will be able to find all other primitive roots.

Proposition 11.2.5.9

Let $n \in \mathbb{N}$ such that there exists at least one primitive root modulo n. Then the number of primitive roots modulo n is exactly $\phi(\phi(n))$.

Proof. Notice that from the above theorem, g^t is a primitive root modulo p if and only if $\gcd(t, \phi(m)) = 1$. Thus we need to count how many numbers are coprime to $\phi(m)$, which is exactly $\phi(\phi(m))$. □

Theorem 11.2.5.10

Let p be an odd prime. Then there exists a primitive root $g \in \mathbb{Z}$ modulo p such that g^{p-1} does not equal to 1 congruent to p^2. Moreover, any such g is a primitive root modulo any power of p.

Theorem 11.2.5.11

Let $m \geq 2$ be an integer. If $m = 2$ or 4 or $m = p^k$ or $m = 2p^k$ for some $k \in \mathbb{N}$ and p and odd prime, then there exists a primitive root modulo p. Otherwise, there isn't.

To show that g is a primitive root of p, we usually use the definition, which is to show that the order of g is $\phi(p)$. And to do this, we consider all factors of $\phi(p)$ and simply show that powers of g of those factors are not the identity.

11.3 Quadratic Congruences

11.3.1 Quadratic Residues

Proposition 11.3.1.1

Let $A, B, C \in \mathbb{Z}$. Solving
$$Ax^2 + Bx + C \equiv 0 \pmod{p}$$
is equivalent to solving $y^2 \equiv a \pmod{m}$ where y is linear to x, given that p is a prime number.

Proof. Take A to be indivisible by p. Else the quadratic congruence deforms into a linear congruence. Then A must have a moduo inverse since p is a prime. Then we can write the congruence as $x^2 + A'Bx + A'C \equiv 0 \pmod{p}$. If $A'B$ is even, we can complete the square and we are done. If $A'B$ is odd, then replace $A'B$ with $p + A'B$ and it is even. □

Definition 11.3.1.2: Quadratic Residue

If there exists x_0 to be the solution to $x^2 \equiv a \pmod{m}$, then a is said to be a quadratic residue modulo m. If there are no solutions, then a is said to be a quadratic non-residue modulo m

Proposition 11.3.1.3

Suppose that p is an odd prime. If p does not divide a then $x^2 \equiv a \pmod{m}$ has either two or zero solutions modulo m.

Definition 11.3.1.4: Legendre Symbol

Let p be an odd prime and $a \in \mathbb{Z}$. The Legendre Symbol is defined to be
$$\left(\frac{a}{p}\right) = \begin{cases} 0 & \text{if } p|a \\ 1 & \text{if } p \not| a \text{ and } a \text{ is a quadratic residue modulo } p \\ -1 & \text{if } a \text{is a quadratic non-residue modulo } p \end{cases}$$

Theorem 11.3.1.5: Euler's Criterion

Let p be prime. Let $a \in \mathbb{Z}$ with $p \not| a$. Then
$$a^{(p-1)/2} \equiv \left(\frac{a}{p}\right) \pmod{p}$$

Proof. By Wilson's theorem, we have that $(p-1)! \equiv -1 \pmod{p}$. Thus we just have to prove that
$$(p-1)! \equiv -\left(\frac{a}{p}\right) a^{\frac{p-1}{2}} \pmod{p}$$
First suppose that $\left(\frac{a}{p}\right) = 1$. Let x be the intger that solves this quadratic residue. Notice that
$$\begin{aligned} x(p-x) &\equiv -x^2 \pmod{p} \\ &\equiv -a \pmod{p} \end{aligned}$$

Now also notice that since p does not divide a, for all numbers between 1 and $p-1$ except the x and $p-x$ that solves the quadratic congruence, we can pair them up so that multiplication between the two elements yield a. This means that we have

$$\begin{aligned}(p-1)! &\equiv -a \prod_{\substack{1\le k\le p-1\\ k\notin\{x,p-x\}}} k \pmod p\\ &\equiv -a\cdot a^{\frac{p-3}{2}} \pmod p\\ &\equiv -a^{\frac{p-1}{2}} \pmod p\end{aligned}$$

Thus we are done.

Now suppose that $\left(\frac{a}{p}\right) = -1$. Then similar to the above, we get $p-1$ pairs since none are quadratic residues thus

$$(p-1)! \equiv a^{\frac{p-1}{2}} \pmod p$$

Thus we are done. □

This criterion is a powerful statement in the sense that we can know whether a is a quadratic residue of p by direct computation.

Proposition 11.3.1.6

Let p be an odd prime. Let $a, b \in \mathbb{Z}$. Then

- If $a \equiv b \pmod p$, then $\left(\frac{a}{p}\right) = \left(\frac{b}{p}\right)$
- $\left(\frac{ab}{p}\right) = \left(\frac{a}{p}\right)\left(\frac{b}{p}\right)$

Corollary 11.3.1.7

Let p be an odd prime. Then

$$\left(\frac{-1}{p}\right) = \begin{cases}-1 & \text{if } p\equiv 3 \pmod 4\\ 1 & \text{if } p \equiv 1 \pmod 4\end{cases}$$

Proof. A simple application of Euler's criterion. □

Proposition 11.3.1.8: Gauss's Lemma

Let p be an odd prime. Let a be an integer such that $\gcd(a,p) = 1$. Consider the integers $a, 2a, 3a, \dots, \frac{p-1}{2}a$ and their least positive residues modulo p. Let n be the number of these residues that are greater than $\frac{p}{2}$. Then

$$\left(\frac{a}{p}\right) = (-1)^n$$

Proof. Define a set $P_1 = \{1, \dots, \frac{p-1}{2}\}$ and $P_2 = \{\frac{p+1}{2}, \dots, p-1\} = \{\frac{1-p}{2}, \dots, -1\}$. We only consider modulo relations thus they are equivalent sets. Let $k_1, k_2 \in P_1$ be distinct. Notice that $1 \le k_1 + k_2 \le p-1$. Without loss of generality take $k_1 > k_2$. Then since p does not divide $k_1 + k_2$ and $k_1 - k_2$, and that p does not divide a, then p does not divide $k_1a \pm k_2a$ by Euclid's lemma.

Since this is true for arbitrary k_1, k_2, taking modulo p for general ka for $k \in P_1$ will allow ka to lie within the set $P_1 \cup P_2$. Also notice that taking distinct elements for k will result in a distinct modulo. Suppose that among these k, m of them fall into P_1 and n of them fall into P_2. Say they fall into $A \subset P_1$ and $B \subset P_2$ respectively (Observe that $P_1 \cap P_2 = \emptyset$ implies $A \cap B = \emptyset$, also $A \cup B = P_1$). Recalling that P_2 has two equivalent definitions, we have

$$\begin{aligned}\prod_{k=1}^{\frac{p-1}{2}}(ka) &\equiv \left(\prod_{\substack{1 \leq k \leq \frac{p-1}{2}\\ k \in A}} k\right) \times \left((-1)^n \prod_{\substack{1 \leq k \leq \frac{p-1}{2}\\ k \in B}} k\right) \pmod{p}\\ &\equiv (-1)^n \prod_{k=1}^{\frac{p-1}{2}} \pmod{p}\end{aligned}$$

and also

$$\prod_{k=1}^{\frac{p-1}{2}}(ka) \equiv a^{\frac{p-1}{2}} \prod_{k=1}^{\frac{p-1}{2}} k \pmod{p}$$

Finally, notice that the product is invertible modulo p, thus we can cancel it on both sides to get

$$(-1)^n \equiv a^{\frac{p-1}{2}} \pmod{p}$$

and we are done. □

Lemma 11.3.1.9

Let p be an odd prime. Then

$$\left(\frac{2}{p}\right) = \begin{cases} -1 & \text{if } p \equiv \pm 3 \pmod{8}\\ 1 & \text{if } p \equiv \pm 1 \pmod{8}\end{cases}$$

Proof. We apply Gauss's lemma direct by counting the number of elements in $\{2, 4, \dots, p-1\}$ that upon taking modulo p, ends in the set $\{\frac{p+1}{2}, \dots, p-1\}$. But these are precisely the elements $\frac{p+1}{2}, \dots, p-1$, which has $\lfloor \frac{p+1}{4} \rfloor$ elements. The result follows immediately. □

Theorem 11.3.1.10: The Quadratic Reciprocity Theorem

Let $p \neq q$ be odd primes. Then

$$\left(\frac{q}{p}\right) = \begin{cases} -\left(\frac{p}{q}\right) & \text{if } p \equiv q \equiv 3 \pmod{4}\\ \left(\frac{p}{q}\right) & \text{otherwise}\end{cases}$$

11.3.2 Primitive Roots and Quadratic Residues

Proposition 11.3.2.1

Let p be a prime. If g is a primitive root modulo p then g is a quadratic non-resuidue of p.

Proof. Trivially if there exists x such that $x^2 \equiv g \pmod{p}$, then since the order of x divides $\phi(p)$, the order of g divides $\frac{\phi(p)}{2}$. But the order of a primitive root must be $\phi(p)$. □

Definition 11.3.2.2: Fermat Primes

We say that n is a Fermat prime if n is prime and $n = 2^{2^k} + 1$ for some $k \in \mathbb{N}$.

Theorem 11.3.2.3

If p is a Fermat prime then g is a primitive root modulo p if and only if g is a quadratic non-residue of p.

11.4 Diophantine Equations

11.4.1 Introduction to Diophantine Equations

Definition 11.4.1.1: Diophantine Equations

Diophantine Equations are polynomial equations in two or more unknowns with integer coefficients, where we only consider integer solutions.

We will consider diophantine equations of the form:

$z = x^2 + y^2$ and the cases with three squares and four squares.
$z^2 = x^2 + y^2$ the pythagoreas triples.
$ax^2 + by^2 = cz^2$ which is ternary quadratic equations.

11.4.2 Lattices

Definition 11.4.2.1: Lattice

Let $n \in \mathbb{N} \setminus \{0\}$. We say that Λ is a lattice in $\mathbb{R}^n$ if

$$\Lambda = \left\{ \sum_{k=1}^{n} a_k v_k \middle| v_k \in \mathbb{Z} \right\}$$

where $v_1, \dots, v_n$ are linearly independet vectors. In this case, we say that $v_1, \dots, v_n$ forms a basis for Λ.

Theorem 11.4.2.2: Minkowski's Theorem

Let Λ be a lattice in $\mathbb{R}^n$. Let $S \subseteq \mathbb{R}^n$ be a symmetric, convex set whose volumes exceeds $2^n \det(\Lambda)$. Then S contains a non-zero lattice point.

Corollary 11.4.2.3: Strong Minkowski's Theorem

Let Λ be a lattice in $\mathbb{R}^n$. Let $S \subseteq \mathbb{R}^n$ be a symmetric, convex, compact set whose volumes is greater than or equal to $2^n \det(\Lambda)$. Then S contains a non-zero lattice point.

11.4.3 Sum of Squares

We discuss numbers that can be written as a sum of two squares, in other words, solving the diophantine equation $z = x^2 + y^2$.

Theorem 11.4.3.1

Let $p \equiv 1 \pmod 4$ be prime. Then p is a sum of two squares.

Proof. We know that -1 is a quadratic residue of p. So let m be an integer such that $m^2 \equiv -1 \pmod p$. Define a lattice by $\Lambda = \text{span}\{(1, m), (0, p)\}$. Consider another set

$$S = \{(x, y) \in \mathbb{R}^2 | x^2 + y^2 < 2p\}$$

Clearly S is symmetric and convex. It also has volume $2\pi p$ thus is larger than 2^2 times the volume of Λ which is $\det(\Lambda) = p$. By Minkowski's theorem, there exists $(x, y) = a(1, m) + b(0, p) \in \Lambda$ for some a, b such that $(x, y) \in S$ which is nonzero.

Since

$$\begin{aligned} x^2 + y^2 &\equiv a^2 + a^2m^2 \pmod p \\ &\equiv a^2(1+m^2) \pmod p \\ &\equiv 0 \pmod p \end{aligned}$$

Since $x^2 + y^2 < 2p$, naturall we have $x^2 + y^2 = p$. □

Lemma 11.4.3.2

If $a, b \in \mathbb{N}$ are each sum of two squares, then so is ab.

Proof. A simple proof involving algebraic manipulations. □

Lemma 11.4.3.3

Let $x, y \in \mathbb{Z}$ and suppose a prime $p \equiv 3 \pmod 4$ divides $x^2 + y^2$. Then $p|x$ and $p|y$.

Corollary 11.4.3.4

If $n \in \mathbb{N}$ is a sum of two squares and $p \equiv 3 \pmod 4$ is a prime divisor of n, then $p^2|n$ and $\frac{n}{p^2}$ is a sum of two squares.

Theorem 11.4.3.5: Two Square Theorem

Let $n \in \mathbb{N}$. Then n can be expressed as a sum of two squares if and only if for every prime

$$p \equiv 3 \pmod 4$$

in the prime decomposition of n, p has an even power in the factorization.

We turn to discuss diophnatine equations of the form $w = x^2 + y^2 + z^2$

Theorem 11.4.3.6: Three Square Theorem

A positive integer is a sum of three squares if and only if its not of the form $4^a(8b + 7)$ where $a, b \in \mathbb{N} \cup \{0\}$.

Finally we also consider the case with four squares.

Lemma 11.4.3.7

If $a, b \in \mathbb{N}$ are each sum of four squares, then so is ab.

Proof. A matter of algebra manipulation. □

Theorem 11.4.3.8: Lagrange's Four Square Theorem

Any positive integer is a sum of four squares.

Proof. By the above lemma and the fact that $2 = 1^2 + 1^2 + 0 + 0$, we just have to show that any odd prime has a decomposition into a sum of four squares. Let p be an odd prime. We show that there exists $x^2 + y^2 + w^2 + z^2$ divisible by p and lies between $(0, 2p)$, which completes the proof.

Define two sets $A = \{a^2 | a \in \mathbb{Z}/p\mathbb{Z}\}$ and $B = \{-(1+b^2) | b \in \mathbb{Z}/p\mathbb{Z}\}$. They each have cardinality $\frac{p-1}{2}$ and thus must intersect at say $a_0 \in A$ and $-(1+b_0^2) \in B$. Then we have the modulo equation

$$a^2 + b^2 \equiv -1 \ (\text{mod } p)$$

Now define

$$\Lambda = \{(x, y, w, z) \in \mathbb{Z}^4 | z \equiv ax + by \ (\text{mod } p) \text{ and } w \equiv ay - bx \ (\text{mod } p)\}$$

Then

$$\begin{aligned} x^2 + y^2 + z^2 + w^2 &\equiv x^2 + y^2 + (ax+by)^2 + (bx-ay)^2 \ (\text{mod } p) \\ &\equiv x^2 + y^2 + (a^2+b^2)x^2 + (a^2+b^2)y^2 \ (\text{mod } p) \\ &\equiv (a^2+b^2+1)(x^2+y^2) \ (\text{mod } p) \\ &\equiv (a^2+b^2+1)(x^2+y^2) \ (\text{mod } p) \\ &\equiv 0 \ (\text{mod } p) \end{aligned}$$

We now show that Λ can be applied the Minkowski theorem (The symmetric convex set is defined below). Notice that

$$\left\{ \begin{pmatrix} 1 \\ 0 \\ a \\ -b \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ b \\ a \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ p \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \\ p \end{pmatrix} \right\}$$

forms a basis for the lattice Λ, and that $\det(\Lambda) = p^2$. Now

$$D = \{(x, y, z, w) \in \mathbb{R}^4 | x^2 + y^2 + z^2 + w^2 < 2p\}$$

is symmetric and convex. It also has volume

$$\frac{\pi^2}{2}(\sqrt{2p})^4 = 2\pi^2 p^2 > 2^4 \det(\Lambda) = 2^4 p^2$$

Applying Minkowski's theorem gives a nonzero element of Λ thus we are done. □

In general, to find out the sum of four squares of a natural number, we first try and find the biggest square contained in the number, then try and find a three or two square decomposition. If it does not work, try the second biggest square and vice versa.

11.4.4 Gaussian Integers

We study the ring of Gaussian integers here, which is actually a type of field extension. We attempt to classify all primes in the ring of Gaussian Integers.

Definition 11.4.4.1: Gaussian Integers

Define the ring of Gaussian integers to be

$$\mathbb{Z}[i] = \{a + bi | a, b \in \mathbb{Z}\}$$

Define the norm of $z = a + bi \in \mathbb{Z}[i]$ to be $N(z) = a^2 + b^2$.

Notice that $\mathbb{Z} \leq \mathbb{Z}[i]$. In fact, the Guassian integers are simply complex numbers with integer coefficients. We are trying to extend the notion of divisibility and prime numbers to an algebraically complete field. Moreover, by the definition of norm, we are in fact discussing equations of the form $x^2 + y^2 = z$.

Proposition 11.4.4.2

The ring of Gaussian integers is a Euclidean domain, and thus is also a PID and UFD.

Recall that we have the notion of division in a commutative ring. As ina analysis, we have defined the notion of distance in the complex numbers.

Definition 11.4.4.3: Gaussian Primes

We say that $z \in \mathbb{Z}[i]$ is a Gaussian prime if $z \neq 0$, z is not a unit and that for any $x, y \in \mathbb{Z}[i]$ such that $z|xy$, then either $z|x$ or $z|y$.

This is precisely the definition of a prime or irreducible element in integral domains.

Proposition 11.4.4.4

Let $z \in \mathbb{Z}[i]$ be a Gaussian prime. Then any unit multiple of z is also a Gaussian prime, as well as $\overline{z}$

Lemma 11.4.4.5

Let $z \in \mathbb{Z}[i]$ and $N(z)$ is a prime number in $\mathbb{Z}$. Then z is a Gaussian prime.

Theorem 11.4.4.6

Let $z \in \mathbb{Z}[i]$. Then z is a Gaussian prime if and only if z is in one of the following forms:

- $1 + i$
- p is a prime in $\mathbb{Z}$ such that $p \equiv 3 \pmod 4$
- $N(z)$ is a prime in $\mathbb{Z}$ such that $N(z) \equiv 1 \pmod 4$
- Any unit multple of the above

Below gives a list of ways to detect whether p is a prime for $z \in \mathbb{Z}[i]$.

Proposition 11.4.4.7

Let $z \in \mathbb{Z}[i]$. Let $p \in \mathbb{Z}$ be a prime divisor of $N(z)$.

- If $p = 2$ then $1 + i|z$
- If $p \equiv 3 \pmod 4$ then $p|z$
- If $p = x^2 + y^2$ with $x, y \in \mathbb{Z}$ then either $x + yi$ or $x - yi$ divides z

11.4.5 Pythagorean Triples

Definition 11.4.5.1: Pythagorean Triples

A pythagorean triple is a solution $(x, y, z) \in (\mathbb{N} \setminus \{0\})^3$ such that $x^2 + y^2 = z^2$. A pythagorean triple is said to be primitive if $\gcd(x, y, z) = 1$.

Lemma 11.4.5.2

Exactly one of x, y, z in a pythagorean triple is even.

Proof. Obviously all three cannot be odd numbers at the same time. If they both are even then by taking modulo 2 we see that $0 \equiv 1 \pmod 2$ which is a contradiction. □

Lemma 11.4.5.3

If x, y, z is a pythagorean triple then z is odd.

Proof. For odd or even x, y such that not both are even, $x^2 + y^2 \equiv 1 \pmod 2$. □

Lemma 11.4.5.4

Let $r, s, t \in \mathbb{N}$ such that $r^2 = st$. If $\gcd(s, t) = 1$, then s, t are both perfect squares.

Proof. Result is clear from the fact that s and t have no common prime numbers in their decomposition, and that every prime in the prime decomposition of r^2 must have an even number of power. □

Theorem 11.4.5.5

Let $(x, y, z) \in (\mathbb{N} \setminus \{0\})^3$. Then (x, y, z) is a pythagorean triple if and only if $x = u^2 - v^2$, $y = 2uv$ and $z = u^2 + v^2$ for some coprime $u, v \in \mathbb{N} \setminus \{0\}$ such that at least one of u, v is even and that $u > v$.

Proof. Suppose that (x, y, z) is a pythagorean triple with y even. Since y is even, $y = 2r$ for some $r \in \mathbb{N}$. Thus $y^2 = 4r^2$. From $y^2 = z^2 - x^2$, we have that $4r^2 = (z + x)(z - x)$. Since x, z are both odd, $z + x$ and $z - x$ are both even. Thus $z + x = 2s$ and $z - x = 2t$ for some $s, t \in \mathbb{N}$. Solving the the system gives $z = s + t$ and $x = s - t$. Now we have that $4r^2 = 4st$ thus $r^2 = st$.

I claim that s and t here are relatively prime. Suppose for a contradiction that there exists $d \in \mathbb{N}$ such that $d|s$ and $d|t$. From the system of equations, we see that $d|z$ and $d|x$. But we already assumed that x, y, z are relatively prime. Thus the contradiction arises. Now we can apply the above lemma to see that $s = u^2$ and $t = v^2$ for some $u, v \in \mathbb{N}$. Thus $y = \sqrt{4r^2} = \sqrt{4st} = 2uv$. From the system of equations we also have that $z = u^2 + v^2$ and $x = u^2 - v^2$ and we are done.

The only if part is simple in substituting the expressions and seeing that they are indeed equal. It remains to show that u and v are coprime. □

Theorem 11.4.5.6

The equation $x^4 + y^4 = z^2$ has no solutions over $\mathbb{N} \setminus \{0\}$.

Proof. Notice that this equation is just $(x^2)^2 + (y^2)^2 = z^2$ which is just a question of pythagorean triples. Suppose that x^2, y^2, z is a primitive pythagorean triple that satisfies the equation. By the solution of pythagorean triples, there exists coprime u, v with $u > v$ such that $x^2 = u^2 - v^2$, $y = 2uv$, $z^2 = u^2 + v^2$.

I claim that in fact u must be odd and v even. Because if this was the case, then $x^2 = u^2 - v^2 \equiv -1 \pmod 4$ which is impossible. Rearranging this equation gives $x^2 + v^2 = u^2$, which is another pythagorean triple. So using the solution gives a pair (m, n) such that $x = m^2 - n^2$, $v = 2mn$, $u = m^2 + n^2$.

Now since $y^2 = 2uv$ where u, v are coprime and u is odd, we can deduce that u is a square and

v is twice a square, say $u = r^2$ and $v = 2s^2$. Thus we can rewrite equations for u, v as $2s^2 = 2mn$ and $r^2 = m^2 + n^2$. Finally, since $2s^2 = 2mn$ and m, n coprime, they must both be squares as well, say $m = M^2$ and $n = N^2$. Thus M, N, r is yet another triple satisfying $M^4 + N^4 = r^2$. This complete the infinite descent and we are done. □

We give partial solutions to the general equation $x^n + y^n = z^n$.

Corollary 11.4.5.7

If n is a multiple of 4 then then the equation $x^n + y^n = z^n$ has no solution over $\mathbb{N} \setminus \{0\}$.

Proof. Take $n = 4k$, then $x^n + y^n = z^n$ is just $(x^k)^4 + (y^k)^4 = (z^{2k})^2$ which is the theorem above in disguise. □

Theorem 11.4.5.8

Let p be an odd prime such that $q = 2p + 1$ is prime. Then the equation

$$x^p + y^p = z^p$$

has no integer solutions for which p does not divide xyz.

Theorem 11.4.5.9

Let p be an odd prime. Assume that there exists $x, y, z \in \mathbb{Z}$ such that

$$x^p + y^p + z^p = 0$$

where x, y, z each is indivisble by p. Then

$$2^{p-1} \equiv 1 \pmod{p^2}$$

11.4.6 Ternary Quadratic Equation

Definition 11.4.6.1: Squarefree

We say that a number is squarefree if it is not a square of an integer.

Theorem 11.4.6.2

Let $a, b, c \in \mathbb{N} \setminus \{0\}$ be squarefree and pairwise coprime. Then the equation

$$ax^2 + by^2 = cz^2$$

has a non-trivial integer solution if and only if

- bc is a quadratic residue modulo a
- ac is a quadratic residue modulo b
- $-ab$ is a quadratic residue modulo c

Proof. Firstly suppose that (x, y, z) is a non-trivial solution to $ax^2 + by^2 = cz^2$ that is coprime. Multiplying by c gives

$$acx^2 + bcy^2 = (cz)^2$$

Taking modulo a gives

$$bcy^2 \equiv (cz)^2 \pmod 4$$

I claim that $\gcd(a, y) = 1$. Indeed suppose for a contradiction that $p|a$ and $p|y$. Then $p|cz^2$. Since $\gcd(a, c) = 1$ we must have $p|z$, which is a contradiction. Thus $\gcd(a, y) = 1$. This means that we can take the inverse of y modulo a and get

$$bc \equiv (y^{-1}cz)^2 \pmod a$$

Similarly, the other quadratic resdiues are proven in the same way.

Now for the other direction, we use the strong form of Minkwoski's theorem. Let $r, s, t \in \mathbb{Z}$ solve the quadratic congruences for bc, ac and $-ab$ respectively. Suppose that

$$\Lambda = \{(x, y, z) \in \mathbb{Z}^3 | by \equiv rz \pmod a, cz \equiv sx \pmod b, ax \equiv ty \pmod c\}$$

Then we must have

$$bry \equiv r^2 z \equiv bcz \pmod a$$

Thus as $\gcd(a, b) = 1$ we have $ry \equiv cz \pmod a$. Thus

$$\begin{aligned} ax^2 + by^2 - cz^2 &\equiv by^2 - cz^2 \equiv \pmod a \\ &\equiv rzy - cz^2 \pmod a \\ &\equiv z(ry - cz) \pmod a \\ &\equiv 0 \pmod a \end{aligned}$$

Similarly, we can show that $ax^2 + by^2 - cz^2$ is divisible by b and c. As a, b, c are pariwise coprime, we must have

$$ax^2 + by^2 - cz^2 \equiv 0 \pmod{abc}$$

Thus we can write the congruences in Λ into

$$\Lambda = \{(x, y, z) \in \mathbb{Z}^3 | z \equiv r_a^{-1} by \pmod a, z \equiv c_b^{-1} sx \pmod b, x \equiv a_c^{-1} ty \pmod c\}$$

where r_a^{-1} denotes the inverse of r modulo a and so on.

By the CRT, we can write the congruences into $y \equiv \nu x \pmod c$ and thus

$$z \equiv \tau x + \rho y \pmod{ab}$$

for some $\nu, \tau, \rho \in \mathbb{Z}$. Now, □

11.4.7 Waring's Problem

Definition 11.4.7.1: Warring's Problem

Let $k \in \mathbb{N}$. Consider the equation

$$n = \sum_{i=1}^{s} x_i^k$$

for any $n \in \mathbb{N}$. Warring's problem consists of finding $g(k)$, the least $s \in \mathbb{N}$ for which any $n \in \mathbb{N}$ can be expressed as a sum of s powers of k.