

NOTES ON MATHEMATICS

MANVENDRA SOMVANSHI

manusomvanshi@hotmail.com

Updated on: December 11, 2022

CONTENTS

I BASICS

1	ELEMENTARY SET THEORY	4
2	RELATIONS	5
3	FUNCTIONS	6
4	CATEGORIES	12

II ABSTRACT ALGEBRA

1	INTRODUCTION	15
2	MORE ON PERMUTATION GROUP	20
3	LAGRANGE'S THEOREM	22
4	HOMOMORPHISM AND NORMAL SUBGROUPS	24
5	AUTOMORPHISMS	28
6	FREE GROUPS	30
7	GROUP ACTIONS	33
8	SYLOW'S THEOREMS	36
9	CLASSIFICATION OF FINITE GROUPS	40

III ANALYSIS

1	CONSTRUCTION OF REAL NUMBERS	42
2	METRIC SPACES AND EUCLIDEAN SPACE	47
3	SEQUENCES	53
4	CONTINUITY	56
5	DIFFERENTIATION	60
6	UNIFORM CONVERGENCE	63
7	RIEMANN INTEGRATION	64

IV SET THEORY AND LOGIC

1	SYMBOLIC LOGIC	69
1.1	Semantics of Propositional logic	70
1.2	Syntactics of Propositional logic	70
1.3	Proof Methods	71
1.4	Consistency	72
1.5	Soundness	73
1.6	Complete	74
2	FIRST ORDER LOGIC	76
2.1	Syntactics	76
2.2	Axioms, Rules of Inference and Replacement rules	77
2.3	Proof Methods	78
3	SET THEORY AS FIRST ORDER LOGIC	80
3.1	Zermelo-Frankel Axioms	80
3.2	Relations	81
3.3	Functions	82

V TOPOLOGY

1	TOPOLOGICAL SPACES AND CONTINUOUS FUNCTIONS	85
---	---	----

VI MEASURE THEORY

1	INTUITION OF MEASURE	87
2	FORMAL NOTION OF MEASURE	89
3	CARATHÉODORY THEOREM	96
4	LEBESGUE MEASURE	103
5	COMPLETE MEASURES	105
6	INTEGRATION	107

PART I
BASICS

1 ELEMENTARY SET THEORY

A set S is a collection of objects. The objects of a set are called the elements. The union of two sets S, T contains elements of both sets, and is written as $S \cup T$. The intersection of two sets contains the common elements of the two sets, and is represented $S \cap T$. The complement of a set $A (\subset S)$ with respect to some set S is represented as A' . The difference of set S from T is defined as $T - S = T \cap S'$. The algebra of these operations is as follows

- *Commutativity,*

$$\begin{aligned}A \cup B &= B \cup A \\A \cap B &= B \cap A\end{aligned}$$

- *Associativity,*

$$\begin{aligned}A \cap (B \cap C) &= (A \cap B) \cap C \\A \cup (B \cup C) &= (A \cup B) \cup C\end{aligned}$$

- *Distribution,*

$$\begin{aligned}A \cup (B \cap C) &= (A \cup B) \cap (A \cup C) \\A \cap (B \cup C) &= (A \cap B) \cup (A \cap C)\end{aligned}$$

- *De Morgan's rules,*

$$\begin{aligned}(A \cap B)' &= A' \cup B' \\(A \cup B)' &= A' \cap B'\end{aligned}$$

DEFINITION 1.1 The cartesian product of two sets A and B , denoted $A \times B$, is the set of ordered pairs $\{(a, b) | \forall a \in A \text{ and } b \in B\}$.

Other than these operations one can define a set operation called disjoint union, denoted $S \sqcup T$, which is loosely constructed in the following way: we first make copies S' and T' such that $S' \cap T' = \emptyset$ and then take their ordinary union. A more rigorous definition would be constructed later.

2 RELATIONS

DEFINITION 2.1 A relation on a set A is a subset C of the cartesian product $A \times A$. If $(x, y) \in C$ then it is denoted as xCy .

DEFINITION 2.2 An equivalence relation on a set A is a relation C on A such that:

- It is reflexive, i.e. $xCx \forall x \in A$.
- It is symmetric, i.e. if xCy then yCx .
- It is transitive, i.e. if xCy and yCz then xCz .

Generally the symbol \sim is used to denote an equivalence relation. For a given element $x \in A$ we also define a set called the equivalence class as:

$$E = \{y \mid y \sim x\}$$

PROPOSITION 2.3 Two equivalence classes E and E' are either disjoint or equal.

Proof | Let E be the equivalence class of x and E' be the equivalence class of x' . Assuming that $E \cap E'$ is non-empty, for all $y \in E \cap E'$ it follows that $y \sim x'$ and $y \sim x$. From symmetry and transitivity it follows that $x' \sim x$. Hence every element similar to x' will be similar to x . Hence $E' = E$, whenever $E \cap E'$ is non-empty. ■

DEFINITION 2.4 A partition of a set A is a collection of disjoint nonempty subsets of A whose union is all of A .

PROPOSITION 2.5 Given any partition \mathcal{D} of A , there is a unique equivalence relation C on A such that each element of \mathcal{D} is an equivalence class of C .

Proof | Consider a relation C defined as: xCy if both x and y belong to the same element of \mathcal{D} . Since x is always in the same element as itself, xCx is true for all x . If xCy , which means x is in the same subset as y . Since the converse is also true, yCx . If x is in the same subset as y and y is in the subset as z , then x is in the same subset as z . Hence xCy and yCz imply xCz . This means that C is an equivalent relation. Each element of \mathcal{D} can be viewed as an equivalence class of C .

Assume that there exist two equivalence relations C_1 and C_2 such that the set of each their equivalence classes is \mathcal{D} . Let E_1 and E_2 be equivalence classes of x with respect to relations C_1 and C_2 . E_1 and E_2 must be the same since we are claiming that both relations generate the identical collection of sets. Hence if yC_1x then yC_2x which implies that $C_1 = C_2$. ■

DEFINITION 2.6 The quotient of the set S , denoted S/\sim with respect to the equivalence relation \sim is the set of equivalence classes of S with respect to \sim .

3 FUNCTIONS

DEFINITION 3.1 A rule of assignment is a subset r of the cartesian product $C \times D$ of two sets, having the property that each element of C appears as the first ordinate of at most one ordered pair in r .

From this definition one can easily conclude that, if $r \subset C \times D$ and $(c, d), (c, d') \in r$ then $d = d'$. One can think of r as assigning an element $c \in C$, the element $d \in D$. The set C is called the domain of r and D is called the image set.

DEFINITION 3.2 A function f is a rule of assignment r , along with a set B which contains the image set of r . The domain of r is also the domain of f . The set B is called the range.

A function having a domain A and range B is written as $f : A \rightarrow B$. Given an element $a \in A$, $f(a)$ denotes a unique element in B , hence $(a, f(a)) \in r$.

DEFINITION 3.3 Given a function $f : A \rightarrow B$ and a subset $A_0 \subset A$, then a restriction of f to A_0 is the mapping $f|_{A_0} : A_0 \rightarrow B$ with rule:

$$\{(a, f(a)) | a \in A_0\}$$

DEFINITION 3.4 Given functions $f : A \rightarrow B$ and $g : B \rightarrow C$, the composite function is defined as $g \circ f : A \rightarrow C$, such that $g \circ f(a) = g(f(a))$. More formally, the rule of the function $g \circ f : A \rightarrow C$ is:

$$\{(a, c) | \forall b \in B, f(a) = b \text{ and } g(b) = c\}$$

DEFINITION 3.5 A function $f : A \rightarrow B$ is said to be injective if,

$$f(a) = f(a') \implies a = a'.$$

The function is called surjective if for each $b \in B$ there exists an $a \in A$ such that $b = f(a)$. If f is both injective and surjective it is said to be bijective.

PROPOSITION 3.6 For each bijective function $f : A \rightarrow B$, there exists a unique function, called the inverse function, $f^{-1} : B \rightarrow A$ such that $f \circ f^{-1}$ and $f^{-1} \circ f$ are both identity functions.

Proof | Since f is bijective for every $a \in A$ there exists a unique $b \in B$ (from injection), and for every $b \in B$ also there exists an $a \in A$ (from surjectivity). This implies that every $b \in B$ has a unique pre-image in A . Denote this pre-image by $f^{-1}(b)$. The rule of the inverse function is given by:

$$\{(b, f^{-1}(b)) | \forall b \in B\}$$

This proves the existence of inverse. Using the definition of composite function, the rule of the composite function $f \circ f^{-1}$ will be:

$$\{(b, b) | \forall b \in B\}.$$

Hence the composite is the identity function. Similarly the composite function $f^{-1} \circ f$ is also identity.

For proving the uniqueness, consider there exist two inverse functions, f^{-1} and \tilde{f}^{-1} , of f . Hence,

$$\begin{aligned} f(f^{-1}(b)) &= b, \\ \implies \tilde{f}^{-1}(f(f^{-1}(b))) &= \tilde{f}^{-1}(b), \end{aligned}$$

But since $\tilde{f}^{-1}(f(a)) = a$,

$$f^{-1}(b) = \tilde{f}^{-1}(b) \quad \forall b \in B$$

Hence the inverse is unique. ■

PROPOSITION 3.7 The inverse of a bijective function $f : A \rightarrow B$ is also bijective.

Proof | Let the inverse be f^{-1} . Let $b, b' \in B$ such that

$$\begin{aligned} f^{-1}(b) &= f^{-1}(b') \\ \implies f(f^{-1}(b)) &= f(f^{-1}(b')) \\ \implies b &= b' \end{aligned}$$

This shows that f^{-1} is injective. For proof of surjectivity, we can show that for each $a \in A$ there exists a $b (= f(a)) \in B$ such that $a = f^{-1}(b)$. This shows that f^{-1} is also bijective. ■

PROPOSITION 3.8 Let $f : A \rightarrow B$. If there are functions $g : B \rightarrow A$ and $h : B \rightarrow A$ such that $g(f(a)) = a \quad \forall a \in A$ and $f(h(b)) = b \quad \forall b \in B$, then f is bijective and $g = h = f^{-1}$.

Proof | Let $a, a' \in A$, such that

$$f(a) = f(a')$$

Using the function g ,

$$\begin{aligned} g(f(a)) &= g(f(a')), \\ \implies a &= a' \end{aligned}$$

hence f is an injective function. Now coming to surjectivity. Using the existence of h , we can show that for each $b \in B$ there exists $a (= h(b)) \in A$ such that $b = f(a)$. Hence f is a bijective function. For the final part of the proposition, since,

$$\begin{aligned} f(h(b)) &= b, \\ \implies g(f(h(b))) &= g(b), \\ \implies h(b) &= g(b) \quad \forall b \in B. \end{aligned}$$

And since the inverse is unique, they must also be equal to f^{-1} . ■

When there exists a bijection $f : A \rightarrow B$ then A and B are called *isomorphic*. This is sometimes represented as $A \simeq B$. Using the concept of isomorphism the notion of disjoint union can be made more rigorous.

DEFINITION 3.9 The disjoint union of two sets A and B is determined by constructing sets $A' \simeq A$ and $B' \simeq B$ such that $A' \cup B' = \emptyset$, and then determining the union $A' \cup B'$. Such sets can always be constructed for every set since $\{0\} \times A \simeq A$ and $\{1\} \times B \simeq B$ and $(\{0\} \times A) \cap (\{1\} \times B) = \emptyset$.

A less restrictive notion of invertibility is defined in terms of *left-invertible* and *right-invertible* functions. If for a function $f : A \rightarrow B$ there exists a $g : B \rightarrow A$ such that $g \circ f : A \rightarrow A$ is id_A then f is said to be left invertible. Similarly if there exists $h : B \rightarrow A$ such that $f \circ h : B \rightarrow B$ is id_B then f is called right invertible. The following is more general statement to proposition 3.8.

PROPOSITION 3.10 Let $f : A \rightarrow B$ be a function then:

- 1) f is injective if and only if it is left invertible.
- 2) f is surjective if and only if it is right invertible.

Proof | For statement 1, the forward implication follows from the fact that if f is injective then one can construct a $g : B \rightarrow A$ as follows: let $p \in B$ be a fixed point and

$$g(b) = \begin{cases} a, & \text{where } f(a) = b \\ p, & \text{when } b \text{ not in image of } A. \end{cases} \quad (3.1)$$

Clearly the function $g \circ f(a) = id_A(a) = a$. The backward implication for statement 1, is true because if g exists such that:

$$g \circ f(a) = a \quad \forall a \in A$$

then,

$$g \circ f(a) = a \neq a' = g \circ f(a') \implies f(a) \neq f(a')$$

which is the contrapositive of the statement required to prove.

Statement 2 can be proven in a similar way. Assuming that f is surjective, it follows that for each b there is at least one $a \in A$ such that $f(a) = b$. Choosing any one of these a for each b we can construct the map $h : B \rightarrow A$, $h(b) = a$. Hence it follows that $f \circ h(b) = f(a) = b = id_B(b)$. For the backward implication, assuming that there is an $h : B \rightarrow A$ such that $\forall b \in B, f \circ h(b) = id_B(b) = b$. Since $h(b) = a$ for some $a \in A$, it follows $f(a) = b$ for some $a \in A$. Hence f is surjective. ■

There is another way to look at bijective functions using the concept of monomorphisms and epimorphisms. This is a more fundamental and equivalent approach to defining bijections.

DEFINITION 3.11 A function $f : A \rightarrow B$ is said to be a monomorphism if the following holds:

$$\forall Z, \forall \alpha', \alpha'' : Z \rightarrow A, f \circ \alpha' = f \circ \alpha'' \implies \alpha' = \alpha''$$

PROPOSITION 3.12 A function $f : A \rightarrow B$ is a monomorphism if and only if it is injective.

Proof | Consider first the forward implication. Assuming f is a monomorphism, we know that for all sets Z and $\alpha', \alpha'' : Z \rightarrow A$,

$$f \circ \alpha' = f \circ \alpha'' \implies \alpha' = \alpha''$$

if $\alpha'(z) = a$ and $\alpha''(z) = a'$ then the above condition reduces to,

$$f(a) = f(a') \implies a = a'$$

Hence f is injective.

For the backward implication we assume that f is an injective function. Then we know that f is left-invertible, with inverse g . If,

$$\begin{aligned} f \circ \alpha' &= f \circ \alpha'' \\ \implies g \circ f \circ \alpha' &= g \circ f \circ \alpha'' \\ \implies \alpha' &= \alpha'' \end{aligned}$$

DEFINITION 3.13 A function $f : A \rightarrow B$ is said to be an epimorphism if,

$$\forall Z, \forall \beta', \beta'' : B \rightarrow Z, \beta' \circ f = \beta'' \circ f \implies \beta' = \beta''$$

PROPOSITION 3.14 A function $f : A \rightarrow B$ is an epimorphism if and only if it is surjective.

Proof | Let's first consider the forward implication:

$$(\forall Z, \beta', \beta'' : B \rightarrow Z, \beta' \circ f = \beta'' \circ f \implies \beta' = \beta'') \implies (\forall b \in B \exists a \in A \text{ such that } b = f(a))$$

The contraposition of this statment is:

$$(\exists b \in B, \forall a \in A, b \neq f(a)) \implies (\exists Z, \beta', \beta'' : B \rightarrow Z \text{ such that } \beta' \neq \beta'' \text{ \& } \beta' \circ f = \beta'' \circ f)$$

Assuming there exists $b \in B$ such that b is not in the image of f , let $Z = \{0, 1\}$, $\beta'(b) = 0, \forall b \in B$, and

$$\beta''(b) = \begin{cases} 0, & \text{if } b \text{ is in image of } f \\ 1, & \text{if } b \text{ is not in image of } f \end{cases}$$

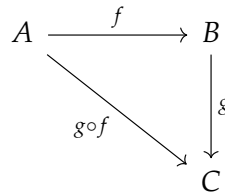
Clearly $\beta' \neq \beta''$ but we have $\beta' \circ f = \beta'' \circ f$. Hence the contrapositive is true, therefore proving the forward implication.

For the backward implication, assuming the function is surjective we also know that it would be right invertible. Let h be the right inverse then,

$$\begin{aligned} \beta' \circ f &= \beta'' \circ f \\ \implies \beta' \circ f \circ h &= \beta'' \circ f \circ h \\ \implies \beta' &= \beta'' \end{aligned}$$

Hence completing the proof.

Diagrams. Diagrams are graphical representations of a collection of sets and how they are operated on by functions. A diagram is said to be *commutative* if taking different paths between sets result in the same function. For example if $f : A \rightarrow B$ and $g : B \rightarrow C$, then here is a commutative diagram of A, B, C :



For injective functions \hookrightarrow is used, for surjective functions \twoheadrightarrow is used, and isomorphisms are represented by $\xrightarrow{\sim}$.

Canonical Decomposition. Let $f : A \rightarrow B$ be a function on A . Define the equivalence relation \sim on A as $a \sim a'$ iff $f(a) = f(a')$. A surjection $g : A \twoheadrightarrow A/\sim$ can be defined as $g(a) = [a]_{\sim}$. Also it is

possible to find an injection $h : \text{im} f \hookrightarrow B$, given by $h(b) = b$. Hence we have a diagram:

$$\begin{array}{ccc} A & \xrightarrow{f} & B \\ g \downarrow & & \uparrow h \\ A/\sim & & \text{im} f \end{array}$$

If we can find an isomorphism $i : A/\sim \xrightarrow{\sim} \text{im} f$ then the above diagram will commute. Consider the following proposition:

PROPOSITION 3.15 The function $i : A/\sim \rightarrow \text{im} f$ given by $i([a]_\sim) = f(a)$ is an isomorphism.

Proof | First we must check if i is a function. Let $[a]_\sim, [a']_\sim \in A/\sim$ then, $[a] = [a'] \implies f(a) = f(a') \implies i([a]_\sim) = i([a']_\sim)$. This means for each $[a]_\sim$ there is a unique image.

Injective. If $i([a]_\sim) = i([a']_\sim)$ then $f(a) = f(a')$ by definition of i . Further it implies that $a \sim a'$ by definition of the equivalence relation. Hence a and a' are in the same equivalence class, or $[a]_\sim = [a']_\sim$.

Surjective. Let $b \in \text{im} f$. Then there exists an $a \in A$ such that $f(a) = b$. Hence there exists $[a]_\sim$ such that $b = i([a]_\sim)$. ■

As a result of this proposition we have shown that any function f can be decomposed according to the commutative diagram:

$$\begin{array}{ccc} A & \xrightarrow{f} & B \\ g \downarrow & & \uparrow h \\ A/\sim & \xrightarrow{i} & \text{im} f \end{array}$$

This shows that any function can be written as a composition of injections, surjections, and isomorphisms. This decomposition is called the canonical decompositions.

DEFINITION 3.16 Let $f : A \rightarrow B$ be a function, and $A_0 \subset A$. Then define,

$$f(A_0) = \{b \mid b = f(a), a \in A_0\},$$

and

$$f^{-1}(B_0) = \{a \mid f(a) \in B_0\}.$$

Note that this definition is for all functions, not just bijective functions.

PROPOSITION 3.17 Let $f : A \rightarrow B$ be a function, and let $A_0 \subset A, B_0 \subset B$ then,

$$A_0 \subset f^{-1}(f(A_0)) \quad \text{and} \quad f(f^{-1}(B_0)) \subset B_0$$

Proof | For the first statement, let $a \in A_0$. Then $f(a) \in f(A_0)$. Which further implies, by definition, that $a \in f^{-1}(f(A_0))$. Hence,

$$\implies A_0 \subset f^{-1}(f(A_0))$$

For the second part of the proposition, let $b \in f(f^{-1}(B_0))$. This means that there exists $a \in f^{-1}(B_0)$ such that $b = f(a)$. Since $a \in f^{-1}(B_0)$, again by definition, $f(a) \in B_0$. Hence $b \in B_0$. Since this is true for any $b \in f(f^{-1}(B_0))$, we conclude that $f(f^{-1}(B_0)) \subset B_0$. ■

PROPOSITION 3.18 Let $f : A \rightarrow B$, $A_0, A_1 \subset A$, and $B_0, B_1 \subset B$. Then f^{-1} preserves:

- 1) inclusions
- 2) unions
- 3) intersections
- 4) differences

Proof | Preservation of inclusion: let $B_0 \subset B_1$. From the definition it follows that $f^{-1}(B_0) = \{a \mid f(a) \in B_0\}$. Since $B_0 \subset B_1$, if $f(a) \in B_0$ then $f(a) \in B_1$. Hence if $a \in f^{-1}(B_0)$ then $a \in f^{-1}(B_1)$. Hence $f^{-1}(B_0) \subset f^{-1}(B_1)$.

Proof of preservation of unions: the set $f^{-1}(B_0 \cup B_1) = \{a \mid f(a) \in B_0 \cup B_1\}$. While $f^{-1}(B_i) = \{a \mid f(a) \in B_i\}$. The union $f^{-1}(B_0) \cup f^{-1}(B_1) = \{a \mid f(a) \in B_0 \text{ or } f(a) \in B_1\}$, which is the same as $\{a \mid f(a) \in B_0 \cup B_1\}$. Hence the two sides are equivalent.

Proof for intersections and differences is very similar to the one for unions. ■

Unlike its inverse f only preserves inclusions and unions. Showing this is pretty easy. Also another property of functions is that $(g \circ f)^{-1}(C_0)$ is equivalent to $f^{-1}(g^{-1}(C_0))$ for functions $f : A \rightarrow B$, $g : B \rightarrow C$, and set $C_0 \subset C$.

4 CATEGORIES

A category is essentially a collection of 'objects' and of 'morphisms' between these objects, satisfying a list of natural conditions. These objects might be sets, groups, vector spaces, etc. Since there is simply no set of all sets (due to Russell's paradox), this collection of objects is just too 'big' to be called a set. The formal term used is a *class of objects*. The formal definition of categories is as follows.

DEFINITION 4.1 A category C consists of:

- 1) a class $\text{Obj}(C)$ of *objects* of the category.
- 2) for every two objects A, B of C , a set $\text{Hom}_C(A, B)$ of morphisms satisfying the following properties:
 - i) for every object A of C there exists (at least) one morphism $1_A \in \text{Hom}_C(A, A)$. This is the identity on A .
 - ii) one can compose morphisms: two morphisms $f \in \text{Hom}_C(A, B)$ and $g \in \text{Hom}_C(B, C)$ determine a morphism $gf \in \text{Hom}_C(A, C)$. For every triplet of objects A, B, C of C there is a function (of sets)

$$\text{Hom}_C(A, B) \times \text{Hom}_C(B, C) \rightarrow \text{Hom}_C(A, C)$$

- iii) this composition law is associative.
- iv) the identity morphisms are identities with respect to composition, i.e. if $f \in \text{Hom}_C(A, B)$ then

$$f1_A = f, 1_B f = f$$

- v) the sets $\text{Hom}_C(A, B)$ and $\text{Hom}_C(C, D)$ are disjoint unless $A = C$ and $B = D$.

One can make morphism diagrams similar to those of set functions. The set of morphisms from an object to itself are known as endomorphisms and are denoted $\text{End}(A)$. The subscript C will be dropped from now on, unless it is necessary to use it.

EXAMPLE 4.2 (Sets) As a first example consider the category Set defined as $\text{Obj}(\text{Set}) =$ the class of all sets, and $\text{Hom}(A, B) =$ the set of all set-functions from A to B . We must verify if this is a category. For every A there is an identity function $1_A : A \rightarrow A$, $1_A(a) = a$. Composition of set-functions is possible, the composition is known to be associative, and the identity function is identity with respect to the composition. The last property is also trivially true. Hence Set is indeed a category.

EXAMPLE 4.3 Consider this example.

Suppose S is a set and \sim is a reflexive and transitive relation. Then define a category C as:

- objects are elements of S ;
- $\text{Hom}(a, b)$, where a, b are objects, is the set consisting $(a, b) \in S \times S$ if $a \sim b$, and let $\text{Hom}(a, b) = \emptyset$ otherwise.

Verification that this is a category:

- 1) Since $a \sim a$ using reflexive property, $1_a = (a, a) \in \text{Hom}(a, a)$.
- 2) Given two morphisms $f = (a, b) \in \text{Hom}(a, b)$ and $g = (b, c) \in \text{Hom}(b, c)$ then using transitivity we know that $a \sim c$ and hence $gf = (a, c) \in \text{Hom}(a, c)$. Hence a composition exists.
- 3) The composition is clearly associative.

- 4) Let $f = (a, b) \in \text{Hom}(a, b)$, and we know that $1_a = (a, a)$ and $1_b = (b, b)$. Clearly $f1_a = (a, b)$ and $1_b f = (a, b)$.
- 5) Since each $\text{Hom}(a, b)$ has either one element (a, b) or is empty, any two set of morphisms will be disjoint.

As an example of this kind of category consider the set $\{1, 2, 3\}$ along with the ordering \leq . The following is a commutative diagram of this category:

$$\begin{array}{ccc}
 1_1 \hookrightarrow 1 & \xrightarrow{f} & 2 \hookrightarrow 1_2 \\
 & \searrow gf & \downarrow g \\
 & & 3 \hookrightarrow 1_3
 \end{array}$$

PART II
ABSTRACT ALGEBRA

1 INTRODUCTION

DEFINITION 1.1 A group is a pair (G, \cdot) where G is a set and $\cdot : G \times G \rightarrow G$ is a binary operation such that:

- 1) G is closed under the operation \cdot .
- 2) \cdot is associative.
- 3) There exists $e \in G$ such that $a \cdot e = e \cdot a = a \forall a \in G$. This element is called the identity.
- 4) $\forall a \in G \exists b \in G$ such that $a \cdot b = b \cdot a = e$. b is called the inverse of a and is represented as a^{-1} .

PROPOSITION 1.2 The identity of a group (G, \cdot) is unique.

Proof | Let $e_1, e_2 \in G$ be two identities. Since e_1 is an identity:

$$e_1 \cdot e_2 = e_2$$

and since e_2 is an identity:

$$e_1 \cdot e_2 = e_1$$

Thus $e_1 = e_2$. ■

PROPOSITION 1.3 The inverse of every element of the group (G, \cdot) is unique.

Proof | Let $a_1, a_2 \in G$ both be inverse of a . Thus

$$\begin{aligned} a \cdot a_1 &= e \\ \implies a_2 \cdot a \cdot a_1 &= a_2 \\ \implies a_1 &= a_2. \end{aligned}$$

Hence the inverse is also unique. ■

PROPOSITION 1.4 Let (G, \cdot) be a group and $x, y \in G$, then there exists $w, z \in G$ such that $x = w \cdot y$ and $x = y \cdot z$.

Proof | Just choose $w = x \cdot y^{-1}$ and $z = y^{-1}x$. Then $w \cdot y = x \cdot y^{-1} \cdot y = x$, and $y \cdot z = y \cdot y^{-1} \cdot x = x$. ■

NOTATION 1.5 From now on the product between elements of any group will be written as xy instead of $x \cdot y$.

PROPOSITION 1.6 The inverse of $(xy)^{-1} = y^{-1}x^{-1}$ where $x, y \in G$.

Proof | Let $z \in G$ be the inverse of xy . Then:

$$\begin{aligned}xyz &= e \\ \implies yz &= x^{-1} \\ \implies z &= y^{-1}x^{-1}\end{aligned}$$

Also $zxy = y^{-1}x^{-1}xy = e$. ■

DEFINITION 1.7 $I_n = \{1, \dots, n\}$ where $n \in \mathbb{N}$ and $S_n = \{f : I_n \rightarrow I_n \mid \text{where } f \text{ is a bijection.}\}$.

NOTATION 1.8 Since the bijections on I_n can be viewed as permutations we use the following notation: if $1 \rightarrow k_1, 2 \rightarrow k_2, \dots, n \rightarrow k_n$ then,

$$f \doteq \begin{pmatrix} 1 & 2 & \dots & n \\ k_1 & k_2 & \dots & k_n \end{pmatrix}$$

Also another notation commonly used is as follows. Let $f \in S_3$ be a bijection given by:

$$f \doteq \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}$$

will be represented by $(12)(3)$ or just (12) .

PROPOSITION 1.9 (S_n, \circ) is a group (it's called the *Permutation group*).

Proof | Since the composition of two bijections is also a bijection S_n is closed under the composition, and since the composition is associative property 2 is also satisfied. Since the inverse function of a bijection always exists. and is itself a bijection property 4 is satisfied. The identity map is obviously a bijection, thus it is in S_n . ■

DEFINITION 1.10 The cardinality of a group is called the order.

PROPOSITION 1.11 Let (G, \cdot) be a finite group then $\forall a \in G \exists 0 \leq n \leq |G|$ such that $a^n = e$.

Proof | Let's assume that such an n does not exist. This means that each element a, a^2, a^3, \dots is distinct, because if $a^n = a^m \implies a^{n-m} = e$. This contradicts the fact that G is finite. If $n > |G|$ then it would contradict the fact that G has $|G|$ number of elements. Thus $\exists n \leq |G|$ such that $a^n = e$. ■

EXAMPLE 1.12 Let $n = 3$, then $I_3 = \{1, 2, 3\}$. Let the points represent the nodes of an equilateral triangle, as in fig. 1. Now consider the bijections in S_3 such that the triangle remains unchanged. These bijections are rotations about the center of the circle i.e. (123) , (132) , reflections about the medians i.e. (12) , (23) , (13) , and the identity map, id_3 . All the symmetries of the triangle can be generated by composing (123) and (23) in different ways.

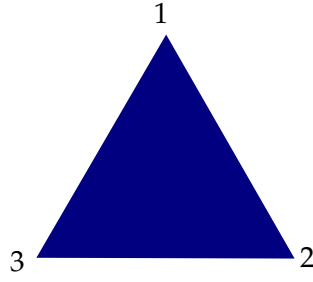


Figure 1: Geometric representation of I_3 .

EXAMPLE 1.13 Similar to the previous example consider the I_4 to represent a square. Then all the possible bijections in S_4 which take the square to itself are the rotation: (1234) , $(13)(24)$, (1432) , the reflections about diagonals: (13) , (24) , reflection along horizontal and vertical: $(14)(23)$, $(12)(34)$, and the identity map: id_4 .

EXAMPLE 1.14 Let \sim be an equivalence relation on \mathbb{Z} given by $a \sim b \iff a \bmod n = b \bmod n$. The group formed by the quotient set, \mathbb{Z}/\sim , under the operation \oplus_n defined as $a \oplus_n b = (a + b) \bmod n$ is called the *modulo- n group* and often represented as $(\mathbb{Z}/n\mathbb{Z}, +)$. For example the set $\mathbb{Z}/3\mathbb{Z} = \{0, 1, 2\}$, where technically each element represents an equivalence class of integers with remainder 0, 1, 2.

EXAMPLE 1.15 Let \otimes_n be an operation on $\mathbb{Z}/n\mathbb{Z}$ such that $a \otimes_n b = (ab) \bmod n$. We will abuse notation and just write ab instead of $a \otimes_n b$. Note that $\mathbb{Z}/n\mathbb{Z}$ is not a group under \otimes_n . First reason is that 0 does not have an inverse. Infact any number $a \in \mathbb{Z}/n\mathbb{Z}$ such that $\gcd(a, n) \neq 1$ will not have an inverse. This can be shown by contradiction. If $\exists b \in \mathbb{Z}/n\mathbb{Z}$ such that $ab = 1$ then $ab = kn + 1$. Now since $\gcd(a, n)$ divides ab and n , but does not divide 1, it divides LHS but not RHS leading to a contradiction. If we remove every element whose gcd with n is not 1 from $\mathbb{Z}/n\mathbb{Z}$ then we would get a group under \otimes_n . This group is represented by $((\mathbb{Z}/n\mathbb{Z})^*, \times)$. The cardinality of this group is given by the Euler totient function $\phi(n)$.

DEFINITION 1.16 A group is said to be *abelian* if the product commutes.

DEFINITION 1.17 A non-empty subset H of a group (G, \cdot) is said to be a subgroup if $(H, \cdot|_{H \times H})$ is a group.

PROPOSITION 1.18 A non-empty subset H of a group (G, \cdot) is a subgroup iff it is closed under $\cdot|_{H \times H}$ and if $a \in H$ then $a^{-1} \in H$.

Proof | (\implies) If H is assumed to be a subgroup then by definition it is a group and thus is closed, and an inverse exists for each element.


(\Leftarrow) If H is closed and for each $a \in H$ $a^{-1} \in H$ then definitely $e \in H$ since $aa^{-1} = e$. From the fact that G is a group it can be deduced that \cdot is associative and that $ae = ea = a$ and $aa^{-1} = a^{-1}a$. ■

DEFINITION 1.19 A group (G, \cdot) is said to be cyclic if $G = \{a^n \mid \forall n \in \mathbb{Z}\}$.

EXAMPLE 1.20 The group $(\mathbb{Z}, +)$ is a cyclic group since $\mathbb{Z} = \{1^n \mid \forall n \in \mathbb{Z}\}$. This is because any element $a \in \mathbb{Z}$ can be written either as the sum $1 + \dots + 1$ or $(-1) + \dots + (-1)$.

PROPOSITION 1.21 Every non-empty finite subset of a group (G, \cdot) that is closed under \cdot is a subgroup of G .

Proof | Let $H \subset G$ be non-empty and closed under \cdot . By non-emptiness there is some element $a \in H$. Since H is closed, all the powers of a must be in H as well. Since H is finite, using a similar argument as in proposition 1.11, there exists an $n \leq |H|$ such that $a^n = e$. This means that $a^{-1} = a^{n-1}$. Thus by proposition 1.18 H is a subgroup of G . ■

 As a direct result of the above proposition one can show that every closed subset of a finite group is a subgroup.

PROPOSITION 1.23 Every subgroup of $(\mathbb{Z}, +)$ is cyclic.

Proof | Let $H \subset \mathbb{Z}$ be a subgroup. In the case $H = \{0\}$, H is cyclic. Now consider H to be any non-trivial subgroup. Due to closure if $x \in H$ then $-x \in H$, thus there exists positive integers in H . Let d be the smallest positive integer in H and let $n \in H$. Using the division algorithm one can write $n = qd + m$ where $0 \leq m < d$. Again using closure since $d^q = qd \in H \implies d^{-q} = -qd \in H$. Thus $m = n - qd$. Since by definition d is the smallest positive number the only way to avoid a contradiction is $m = 0$. Thus $n = qd = d^q$ and H is cyclic. ■

PROPOSITION 1.24 Every subgroup of a cyclic group is cyclic.

Proof | Let (G, \cdot) be a cyclic group and $H \subset G$ be a non-trivial subgroup (claim is obviously true for trivial subgroup). Let x be the generator for G . Since H will only contain powers of x define a set $K = \{n \mid x^n \in H\}$. If $n, m \in K$ then $x^n, x^m \in H \implies x^{n+m} \in H \implies n + m \in K$. Also if $n \in K$ then $x^n \in H \implies x^{-n} \in H \implies -n \in K$. Thus $(K, +)$ is a subgroup of $(\mathbb{Z}, +)$. By proposition 1.23 K is cyclic. If d generates K then x^d generates H since $x^n \in H \implies n \in K \implies n = qd \implies x^n = (x^d)^q$. ■

DEFINITION 1.25 Let $X \subset G$ where (G, \cdot) is a group. Then X is said to generate G if $G = \{x_1^{n_1} \dots x_k^{n_k} \mid \forall x_i \in X, n_i \in \mathbb{Z}\}$. This is denoted by $G = \langle X \rangle$.

DEFINITION 1.26 The order of an element a of the group (G, \cdot) is defined to be the cardinality of the subgroup generated by a . The order is denoted by $\mathcal{O}(a)$.

PROPOSITION 1.27 Let $a \in G$ where (G, \cdot) is a group. Then the order of a is k iff k is the smallest positive integer such that $a^k = e$.

Proof | (\implies) Assuming that the subgroup generated by a has k elements, i.e. $\{e, a, a^2, \dots, a^{k-1}\}$. By closure a^k must be identified with one of the elements in $\langle a \rangle$. If $a^k = a^n$ where $1 \leq n \leq k$ then by cancelation $a^{k-1} = a^{n-1}$, implying that $\langle a \rangle$ has cardinality less than k , which contradicts our assumption. Thus the only remaining possibility is that $a^k = e$. Moreover since all a^i $0 \leq i \leq k-1$ are distinct it follows that k is the smallest positive integer such that $a^k = 1$.

(\impliedby) Assuming that k is the smallest positive integer such that $a^k = 1$, the subgroup $\langle a \rangle = \{e, a, a^2, \dots, a^{k-1}\}$. This follows from the argument that in case $a^n = a^m$ where $0 < m < n < k$ then $a^{n-m} = e$ contradicting the fact that k is the smallest such number. Thus all a^m $0 \leq m < k$ are distinct forming a subgroup of k . ■

PROPOSITION 1.28 If $a, b \in G$ such that $ab = ba$ then $\mathcal{O}(ab) = \mathcal{O}(a)\mathcal{O}(b)$.

Proof | Let $\mathcal{O}(a) = n$ and $\mathcal{O}(b) = m$ and without loss of generality assume that $n \leq m$. Consider the subgroup $\langle ab \rangle$. Since $ab \in \langle ab \rangle$ the power $(ab)^n = b^n a^n = b^n \in \langle ab \rangle$. Further it follows $b^n b^{m-n+1} = b^{m+1} = b \in \langle ab \rangle$. Similarly it can be shown that $a \in \langle ab \rangle$. Thus $\langle ab \rangle = \{a^i b^j \mid 0 \leq i < n, 0 \leq j < m\}$ where I have used the fact that $ab = ba$ (otherwise we would have additional terms like aba). Thus the number of elements in $\langle ab \rangle$ is nm . ■

2 MORE ON PERMUTATION GROUP

DEFINITION 2.1 A 2-cycle is an element of S_n which can be written as $(a_1 a_2)$. A 2-cycle is called a transposition.

DEFINITION 2.2 Two cycles $(a_1 \dots a_k), (b_1 \dots b_\ell) \in S_n$ are said to be disjoint if $a_i \neq b_j$ for all i, j .

PROPOSITION 2.3 Disjoint cycles commute under composition.

Proof | Let $f \doteq (a_1 \dots a_k), g \doteq (b_1 \dots b_\ell)$ be disjoint cycles then consider the composition $g \circ f$. Due to the disjointness any number m is either moved by f alone, by g alone, or not moved at all. If m is not moved by both then clearly $f \circ g(m) = g \circ f(m)$. If it is moved by f only, then $\exists j$ such that $f(m) = a_j$. It further follows that $f \circ g(m) = a_j = g(a_j) = g \circ f(m)$. Similarly this can be shown if m is only moved by g . Therefore $f \circ g = g \circ f$. ■

PROPOSITION 2.4 Set of all transpositions generates the group S_n .

Proof | Any element of S_n can be written as product of cycles. All that remains to be shown is that any cycle can be written as composition of transpositions. Since,

$$(a_1 \dots a_k) = (a_1 a_k) \dots (a_1 a_3)(a_1 a_2).$$

This completes the proof. ■

PROPOSITION 2.5 The set $\{(1k) \mid 2 \leq k \leq n\}$ generates S_n .

Proof | Since any element can be written as a composition of transpositions, all that we need to show is that any transposition can be written in terms of $(1k)$. Since,

$$(ab) = (1a)(1b)(1a).$$

This proof is complete. ■

PROPOSITION 2.6 The set $\{(kk+1) \mid 1 \leq k \leq n-1\}$ generates S_n .

Proof | Using the above proposition all we need to show is that $(1a)$ can be written in terms of $(kk+1)$. Since,

$$(1a) = (a-1a) \dots (23)(12)(23) \dots (a-1a)$$

. This proof is complete. ■

PROPOSITION 2.7 The set $\{(12), (12 \dots n)\}$ generates S_n .

Proof | This will be proven by showing that $(aa + 1)$ can be written in terms of (12) , $(1...n)$ and their powers. Since the map $(1...n)^{a-1}$ takes $1 \rightarrow a$ and the map $(1...n)^{1-a}$ takes a to 1. Thus,

$$(aa + 1) = (1...n)^{a-1}(12)(1...n)^{1-a}.$$

The compositions works as follows: $a \rightarrow 1 \rightarrow 2 \rightarrow a + 1$. This completes the proof. ■

DEFINITION 2.8 An element of S_n is said to be *even* if it can be written as a product of even number of transpositions. Similarly element is said to be *odd*.

PROPOSITION 2.9 Any element of S_n is either even or odd.

Proof | Define the polynomial P as:

$$P(x_1, \dots, x_n) = \prod_{i=1}^n \prod_{j>i}^n (x_i - x_j)$$

If $\alpha \in S_n$ then define αP as:

$$\alpha P(x_1, \dots, x_n) = \prod_{i=1}^n \prod_{j>i}^n (x_{\alpha(i)} - x_{\alpha(j)})$$

The terms in the polynomial αP are the same as P , the only difference would be the order of some may change, introducing a sign. Thus $\alpha P = P$ or $\alpha P = -P$. Clearly if $\alpha, \beta \in S_n$ then the sign change introduced would be the product of the sign change introduced by each. The sign introduced by the transposition (ab) is -1 (the only term that changes sign will be $x_a - x_b$), thus if $\alpha \in S_n$ is odd then α changes P by -1 , on the other hand if α is even then it does not change the sign of P . Since α is independent of the way we chose to represent it as products of transpositions, $\alpha P = \pm P$ will also be independent of the representation. Thus if α is even or odd in one representation it must be in all. ■

DEFINITION 2.10 The set of all even elements of S_n forms a subgroup of order $n!/2$ called the alternating group A_n .

PROPOSITION 2.11 A_n is generated by 3-cycles.

Proof | Every three cycle (abc) can be expressed as $(ac)(ab)$, and thus is even. Any element of A_n can be expressed in terms of products of even number of $(1a)$. Pair the adjacent transpositions in the following way: $(1a)(1b) = (1ba)$. Thus every element can be written as products of 3-cycles. ■

3 LAGRANGE'S THEOREM

DEFINITION 3.1 Let (G, \cdot) be a group and $H \subset G$ be a subgroup then define an equivalence relation \sim on G as follows: $a \sim b \iff ab^{-1} \in H$.

DEFINITION 3.2 Let (G, \cdot) be a group, and H be a subgroup of G then define $Ha = \{ga \mid g \in H\}$, where $a \in G$. H is said to be a right coset in G .

PROPOSITION 3.3 The equivalence class $[a]$ w.r.t. the equivalence relation \sim on (G, \cdot) is the as the set Ha .

Proof | If $b \in [a]$ then

$$\begin{aligned} ab^{-1} &= g \ (\in H), \\ \implies b &= g^{-1}a \text{ where } g^{-1} \in H \text{ since } H \text{ is a group,} \\ \implies b &\in Ha \implies [a] \subset Ha. \end{aligned}$$

On the other hand if $b \in Ha$ then,

$$\begin{aligned} b &= ga \\ \implies g^{-1} &= ab^{-1} \in H, \text{ again since } H \text{ is a subgroup} \\ \implies b &\in [a] \implies Ha \subset [a]. \end{aligned}$$


This completes the proof. ■

THEOREM 3.4 (Lagrange's Theorem) If G is a finite group and $H \subset G$ is a subgroup then the order of H divides order of G .

Proof | We consider the equivalence classes defined above. Let's say there are k distinct equivalence classes. Then $G = \cup_{j=1}^n Ha_j$, and $Ha_i \cap Ha_j = \emptyset$ if $i \neq j$ (since equivalence classes form a partition of the set). Let $f_a : H \rightarrow Ha$ be given by $f_a(g) = ga$. If $f_a(g) = f_a(h)$ then

$$\begin{aligned} ga &= ha \\ \implies g &= h \end{aligned}$$

Thus f_a is injective. Let $h \in Ha$, then $\exists g \in H$ such that $h = ga$. Thus $ha^{-1} = g \in H$. Hence for any $h \in Ha$ it is possible to find a $g \in H$ (given by ha^{-1}) such that $h = f_a(g)$. Therefore f_a is surjective as well, making f_a a bijection. This means that $|H| = |Ha|$. Since each Ha_j is disjoint the union has $k|H|$ elements. Thus $|G| = k|H|$. ■

 The number of right cosets H in G is called the index of H in G , denoted $i_G(H)$. As seen in the proof of Lagrange's theorem $i_G(H) = |G|/|H|$.

COROLLARY 3.6 Every group G of prime order, p , is cyclic.

Proof | From Lagrange's theorem any subgroup H of G can either be $\{e\}$ or G since only $1, p$ divide p . If $H = G$ and $a(\neq e) \in G$ then $\langle a \rangle$ forms a subgroup of G different from $\{e\}$. Thus $\langle a \rangle = G$, proving that G is cyclic. ■

COROLLARY 3.7 If $a \in G$, where G is a finite group then $\mathcal{O}(a)$ divides $|G|$.

Proof | Since $\mathcal{O}(a)$ is the cardinality of the subgroup generated by a , then by Lagrange's theorem it divides $|G|$. ■

COROLLARY 3.8 If G is a finite group then $a^{|G|} = e$ for all $a \in G$.

Proof | Let $a \in G$ with order k . From Lagrange's theorem $|G| = mk$. Then $a^{|G|} = (a^k)^m = e^m = e$. ■

THEOREM 3.9 (Euler) If a is relatively prime to n then $a^{\phi(n)} \bmod (n) = 1$, where $\phi(n)$ is the Euler totient function (defined as the number of coprimes of n).

Proof | Consider the group $((\mathbb{Z}/n)^*, \times)$. We have already seen that it has a cardinality $\phi(n)$. Then by the above corollary $(a \bmod (n))^{\phi(n)} = a^{\phi(n)} \bmod (n) = 1$. ■

COROLLARY 3.10 (Fermat's Little Theorem) If p is a prime and p does not divide a then $a^{p-1} \bmod (p) = 1$.

Proof | In Euler's theorem consider the case when $n = p$. Then the cardinality of the group $((\mathbb{Z}/p), \times)$ is $p - 1$. Thus $a^{p-1} \bmod (p) = 1$. ■

4 HOMOMORPHISM AND NORMAL SUBGROUPS

DEFINITION 4.1 Let G be a group and $A, B \subset G$, and $x \in G$. Then we define the following sets:

- 1) $Ax = \{ax \mid a \in A\}$,
- 2) $xA = \{xa \mid a \in A\}$,
- 3) $AB = \{ab \mid a \in A, b \in B\}$.

When A is a subgroup then Ax is called the right coset, and xA is called the left coset.

PROPOSITION 4.2 If $A, B \subset G$ and $x, y \in G$, where G is a group, then:

- 1) $(Ax)y = A(xy)$,
- 2) $(Ax)B = A(xB)$,
- 3) $(AB)x = A(Bx)$,
- 4) $(AB)C = A(BC)$.

Proof | It's pretty trivial, follows from the definitions. ■

PROPOSITION 4.3 If $A, B \subset G$ and $x \in G$, for some group G , then $A \subset B \implies$:

- 1) $Ax \subset Bx$,
- 2) $xA \subset xB$.

Proof | Define a function $f_x : G \rightarrow G$ defined as $f_x(g) = gx$. Clearly $f_x(A) = Ax$. Since inclusion is preserved under functions $A \subset B \implies f_x(A) \subset f_x(B) \implies Ax \subset Bx$. Similar proof for the second one. ■

DEFINITION 4.4 Let G and G' be two groups and $\phi : G \rightarrow G'$ then ϕ is called a homomorphism if $\phi(ab) = \phi(a)\phi(b)$, $\forall a, b \in G$. If ϕ is bijective and a homomorphism then it is called a group isomorphism. A group isomorphism from G to itself is called an automorphism.

LEMMA 4.5 Let ϕ be a homomorphism from $G \rightarrow G'$ then:

- 1) $\phi(e) = e'$
- 2) $\phi(a^{-1}) = \phi(a)^{-1}$, $\forall a \in G$.

Proof | Let $a \in G$, since $\phi(a) = \phi(a.e) = \phi(a)\phi(e)$ and $\phi(a) = \phi(e.a) = \phi(e)\phi(a)$ it follows that $\phi(e) = e'$. Similarly it can be shown that $\phi(a^{-1}) = \phi(a)^{-1}$. ■

LEMMA 4.6 If $\phi : G \rightarrow G'$ is a homomorphism then $\phi(G)$ is a subgroup of G' .

Proof | If $g', h' \in \phi(G)$ then $\exists g, h \in G$ s.t. $g' = \phi(g)$ & $h' = \phi(h)$. Thus $g'h' = \phi(g)\phi(h) = \phi(gh) \in \phi(G)$. Thus $\phi(G)$ is closed. Since $g \in G \implies g^{-1} \in G \implies \phi(g^{-1}) \in \phi(G) \implies \phi(g)^{-1} \in \phi(G)$. Thus $\phi(G)$ is a subgroup. ■

DEFINITION 4.7 Let $\phi : G \rightarrow G'$ be a homomorphism then the kernel of ϕ is defined as $\text{Ker}_\phi = \{a \in G \mid \phi(a) = e'\}$.

LEMMA 4.8 If $w \in \phi(G)$ such that $w = \phi(x)$ then $W := \{y \mid \phi(y) = w\} = \text{Ker}_\phi x$.

Proof | Since $\phi(yx^{-1}) = \phi(y)\phi(x^{-1}) = \phi(y)\phi(x)^{-1} = ww^{-1} = e'$. Thus $yx^{-1} \in \text{Ker}_\phi \implies y = kx$, $k \in \text{Ker}_\phi \implies W \subset \text{Ker}_\phi x$. If $y \in \text{Ker}_\phi x$ then $y = kx \implies \phi(y) = \phi(k)\phi(x) = w$ thus $\text{Ker}_\phi \subset W$. ■

THEOREM 4.9 If $\phi : G \rightarrow G'$ is a homomorphism then:

- 1) Ker_ϕ is a subgroup of G .
- 2) $g\text{Ker}_\phi g^{-1} \subset \text{Ker}_\phi$.

Proof | It's trivial:

- 1) Ker_ϕ is closed since $\phi(xy) = \phi(x)\phi(y) = e'$ if $x, y \in \text{Ker}_\phi$. The inverse exists since $\phi(x^{-1}) = \phi(x)^{-1} = e'$.
- 2) This is true since $\phi(gxg^{-1}) = \phi(g)\phi(x)\phi(g)^{-1} = e'$. ■

COROLLARY 4.10 A homomorphism $\phi : G \rightarrow G'$ is an injection iff $\text{Ker}_\phi = \{e\}$.


Proof | If the homomorphism is an injection then only one element will be mapped to e' , and since $\phi(e) = e'$ must be true the kernel just contains the identity. If the kernel is just identity then it means that only $\phi(e) = e'$. Since for any $w \in G$ we can define a W as in lemma 4.8, $W = \text{Ker}_\phi x \implies W = \{x\}$. Thus ϕ is an injection. ■

DEFINITION 4.11 (Normal subgroups) A subgroup H of group G is said to be normal if $Hx = xH$, $\forall x \in G$.

PROPOSITION 4.12 A subgroup H is a normal subgroup of group G iff either:

- 1) $xHx^{-1} \subset H$, $\forall x \in G$,
- 2) $HxHy = Hxy$, $\forall x, y \in G$.

Proof | 1) (\Leftarrow) If $xHx^{-1} \subset H \implies xH \subset Hx$. Also since $x^{-1}Hx \subset H \implies Hx \subset xH$. Thus $xH = Hx$. (\Rightarrow) If $xH = Hx \implies H = x^{-1}Hx$.
 2) (\Rightarrow) Since $HxHy = H(xH)y = H(Hx)y = (HH)(xy)$. Since $HH = \{hh' \mid h, h' \in H\}$ and H is a subgroup, it follows that $HH = H$. Thus $HxHy = Hxy$ if H is normal. (\Leftarrow) If $HxHy = Hxy \implies H(xHx^{-1}) = H \implies xHx^{-1} = H$ which proves that H is normal. ■

 Note that the kernel is always a normal subgroup.

PROPOSITION 4.14 Let G be a group and N be a normal subgroup of G . Define the relation \sim on G as before as $a \sim b \iff ab^{-1} \in N$. Then as seen before $Na = [a]$. Define a product between equivalence classes as $NaNb = Nab$. This product is well defined and the collection of all equivalence classes forms a group under this product. The collection of equivalence classes is denoted by G/N , and the group it forms under the product defined is called the Quotient group.

Proof | If $Na = Na'$ and $Nb = Nb'$ then $NaNb = Na'Nb'$, since $NaNb = Nab$ and $Na'Nb' = Na'b'$ it follows that $Nab = Na'b'$. Thus the product is well defined. Clearly set G/N is closed under this product due to proposition 4.12. The identity of the group is N , and the inverse of Na will be Na^{-1} . ■

PROPOSITION 4.15 There exists a homomorphism $\phi : G \rightarrow G/N$ such that $\text{Ker}_\phi = N$.

Proof | Consider the most natural map $\phi(g) = Ng$. Then the kernel is $\text{Ker}_\phi = \{g \mid \phi(g) = N\}$. Since $Ng = N \implies g \in N$ it follows that $\text{Ker}_\phi = N$. ■

💡 The order of the quotient group is the same as the index of N in G . From Lagrange's theorem it follows that for finite groups $|G/N| = |G|/|N|$.

THEOREM 4.17 (First Isomorphism Theorem) If $\phi : G \rightarrow G'$ is a surjective homomorphism with kernel Ker_ϕ then $G' \simeq G/\text{Ker}_\phi$.

Proof | Consider the map $\psi : G/\text{Ker}_\phi \rightarrow G'$ defined by $\psi(\text{Ker}_\phi a) = \phi(a)$. We prove that this is a group isomorphism.

- 1) (Well Defined). If $\text{Ker}_\phi a = \text{Ker}_\phi b$ then $ab^{-1} \in \text{Ker}_\phi$. It follows that $\psi(\text{Ker}_\phi a) = \phi(a) = \phi(ab^{-1})\phi(b) = \phi(b) = \psi(\text{Ker}_\phi b)$. Thus the map is well defined.
- 2) (Injective). If $\psi(\text{Ker}_\phi a) = \psi(\text{Ker}_\phi b)$. Then $\phi(a) = \phi(b) \implies \phi(ab^{-1}) = e'$. Thus $ab^{-1} \in \text{Ker}_\phi$ which further implies that $\text{Ker}_\phi a = \text{Ker}_\phi b$.
- 3) (Surjective). Surjectivity of the map can be seen by construction. If $\phi(a) \in G'$ then $\text{Ker}_\phi a \in G/\text{Ker}_\phi$ is mapped to $\phi(a)$.
- 4) (Homomorphism). Since $\psi(\text{Ker}_\phi a \text{Ker}_\phi b) = \psi(\text{Ker}_\phi ab) = \phi(ab) = \phi(a)\phi(b) = \psi(\text{Ker}_\phi a)\psi(\text{Ker}_\phi b)$, ψ is a homomorphism.

This completes the proof. ■

THEOREM 4.18 (Correspondence Theorem) Let $\phi : G \rightarrow G'$ be a surjective homomorphism, H' be a subgroup of G' and $H = \phi^{-1}(H')$ then:

- 1) H is a subgroup of G .
- 2) $\text{Ker}_\phi \subset H$
- 3) $H/\text{Ker}_\phi \simeq H'$.
- 4) H is normal if H' is normal.

Proof | Let $x, y \in H$ then by definition $\phi(x), \phi(y) \in H'$. This means that $\phi(x)\phi(y) \in H' \implies \phi(xy) \in H'$ since H' is a subgroup and ϕ is a homomorphism. Therefore $xy \in H$.

If $g \in \text{Ker}_\phi$ then $\phi(g) = e' \in H' \implies g \in H$. Thus $\text{Ker}_\phi \subset H$. Since $\phi|_H : H \rightarrow H'$ is a surjective homomorphism, by first isomorphism theorem $H/\text{Ker}_\phi \simeq H'$.

Let $g \in H$ and $x \in G$. Since H' is normal $\phi(x)\phi(g)\phi(x)^{-1} \in H' \implies \phi(xgx^{-1}) \in H' \implies xgx^{-1} \in H$. Therefore H is normal. ■

THEOREM 4.19 (Second Isomorphism Theorem) H, N be a subgroups of G and N be normal. Then:

- 1) HN is a subgroup of G .
- 2) $H \cap N$ is normal in H .
- 3) $H/H \cap N \simeq HN/N$.

Proof | First two are easy to show. Let $\phi : H \rightarrow HN/N$ defined by $\phi(h) = Nh$. This map is clearly a homomorphism. For any $g \in HN$ the coset $Ng = N(hn) = N(n'h') = Nh'$, thus for each $Ng \in HN/N \exists h' \in H$ s.t. $\phi(h') = Ng$, meaning that ϕ is onto. Also note that $\text{Ker}_\phi = H \cap N$. Thus by theorem 4.17 the last statement can be proven. ■

THEOREM 4.20 (Third Isomorphism Theorem) Let $\phi : G \rightarrow G'$ be a surjective homomorphism, N' be a normal subgroup of G' and $N = \phi^{-1}(N')$ then $G/N \simeq G'/N'$ and $G/N \simeq (G/\text{Ker}_\phi)/(N/\text{Ker}_\phi)$.

Proof | Firstly from theorem 4.18 it follows that N is normal and that $N/\text{Ker}_\phi \simeq N'$. Now let $\psi : G \rightarrow G'/N'$ defined by $\psi(g) = N'\phi(g)$. Clearly this is a surjective homomorphism. The kernel of this homomorphism is N . Thus by theorem 4.17 it follows that $G/N \simeq G'/N'$. Since $G/\text{Ker}_\phi \simeq G'$ and $N/\text{Ker}_\phi \simeq N'$, the last part of the theorem follows. ■

5 AUTOMORPHISMS

DEFINITION 5.1 An automorphism is an isomorphism from a group to itself. Let $\text{Aut}(G)$ represent the set of all automorphisms.

PROPOSITION 5.2 $\text{Aut}(G)$ is a group under composition.

Proof | It's trivial. ■

DEFINITION 5.3 The center of group is defined as $Z(G) = \{z \in G \mid \forall g \, zg = gz\}$.

DEFINITION 5.4 The set of all automorphisms of the form $\phi_g : x \mapsto gxg^{-1}$ are called inner automorphisms, and they form a subgroup of the group of automorphisms. It is represented as $\text{Inn}(G)$.

PROPOSITION 5.5 The group $\text{Inn}(G)$ is isomorphic to a quotient of G .

Proof | Construct a map from $G \rightarrow \text{Inn}(G)$ the following way: $\psi : g \mapsto \phi_g$.

1) Since

$$\psi(gh) = \phi_{gh} = \phi_g \phi_h = \psi(\phi_g) \psi(\phi_h),$$

ψ is a homomorphism.

2) For every $\phi_g \in \text{Inn}(G)$ there exists a g , so ψ is surjective.

3) The kernel of ψ is:

$$\text{Ker}_\psi = \{g \mid \psi(g) = \phi_e\} = \{g \mid gxg^{-1} = x, \forall x \in G\} = Z(G).$$


Thus using the first isomorphism theorem $\text{Inn}(G) \cong G/Z(G)$. ■

DEFINITION 5.6 Let H and N be groups, and let $\phi : H \rightarrow \text{Aut}(N)$ be a homomorphism. Then the semi-direct product $N \rtimes_\phi H$ is the set $N \times H$ equipped with the product:

$$(n_1, h_1) \star (n_2, h_2) = (n_1 \phi(h_1)(n_2), h_1 h_2).$$

PROPOSITION 5.7 The semi-direct $N \rtimes_\phi H$ is a group.

Proof | The closure and associativity part are trivial. The identity element is (e_N, e_H) . The inverse of (n, h) is $(\phi(h^{-1})(n^{-1}), h^{-1})$. ■

 In the case when $\phi : H \rightarrow \text{Aut}(N)$ is given by $\phi(h) = \text{id}$, the semidirect product is called the direct product. This is cause $(n_1, h_1)(n_2, h_2) = (n_1 n_2, h_1 h_2)$ in this case.

THEOREM 5.9 Let N, H be subgroups of G . Let N be normal in G , $NH = G$ and $N \cap H = \{e\}$, then $G \cong N \rtimes_\phi H$ where $\phi(h)(n) = hnh^{-1}$.

Proof | Construct the map $\psi : N \rtimes_{\phi} H \rightarrow NH$ given by $(n, h) \mapsto nh$.

1) Since

$$\psi((n_1, h_1)(n_2, h_2)) = \psi(n_1 h_1 n_2 h_1^{-1}, h_1 h_2) = n_1 h_1 n_2 h_2 = \psi(n_1 h_1) \psi(n_2 h_2),$$

ψ is a homomorphism.

2) For any $nh \in NH$, $(n, h) \in N \rtimes_{\phi} H$ is mapped to nh by ψ . So ψ is surjective.

3) The kernel of ψ is:

$$\text{Ker}_{\psi} = \{g \in N \rtimes_{\phi} H \mid \psi(g) = e\} = \{(n, h) \mid nh = e\} = N \cap H = \{e\}.$$

Thus $NH \simeq N \rtimes_{\phi} H$, thus $G \simeq N \rtimes_{\phi} H$. ■

6 FREE GROUPS

Let X be any set. We wish to construct a group using X . We can do this the following way:

- 1) If $X = \emptyset$ then the group $F = \{e\}$.
- 2) For non-empty sets, first choose a set, denoted X^{-1} , which is disjoint from X and $|X| = |X^{-1}|$ (for infinite sets the cardinality should be same).
- 3) Choose a bijection $f : X \rightarrow X^{-1}$. Denote the $f(x)$ by x^{-1} .
- 4) Find a set disjoint from $X \cup X^{-1}$ which has cardinality 1. Call the element of this set 1.

DEFINITION 6.1 A word on X is a sequence in $X \cup X^{-1} \cup 1$, (a_1, a_2, \dots) , such that $\exists N$ such that:

$$n > N \implies a_n = 1.$$

DEFINITION 6.2 The constant sequence $(1, 1, 1, \dots)$ is called the empty word. We denote it by 1 itself.

DEFINITION 6.3 A reduced word is a word such that:

- 1) If $a_n = x$ then $a_{n+1} \neq x^{-1}$ and vice versa.
- 2) If $a_k = 1$ then $a_i = 1, \forall i \geq k$.

NOTATION 6.4 Every reduced word is of the form $(x_1^{n_1}, \dots, x_k^{n_k}, 1, 1, 1, \dots)$ where $x_i \in X$ and $k \in \mathbb{N}$ and $n_i = \pm 1$. Thus we will formally write reduced words as $x_1^{n_1} \dots x_k^{n_k}$. Let $F(X)$ denote the set of all reduced words of X .

DEFINITION 6.5 Define the product of two reduced words as:

$$x_1^{n_1} \dots x_k^{n_k} y_1^{m_1} \dots y_j^{m_j}$$

where we remove every occurrence of terms of the form xx^{-1} or $x^{-1}x$. If all occurrences are such, then the product is the empty word 1. If x is a reduced word, we define $x1 = 1x = x$.

PROPOSITION 6.6 The set $F(X)$ is a group under the above product.

Proof | The identity of the group is clearly the empty word. The inverse of a reduced word $x_1^{n_1} \dots x_k^{n_k}$ is $x_k^{-n_k} \dots x_1^{-n_1}$. By definition closure is ensured. The only hard part is checking associativity.

To prove associativity, for each $x \in X$ and $n = \pm 1$ define $|x^n| : F \rightarrow F$ as:

$$1 \mapsto x^n$$

$$x_1^{n_1} \dots x_k^{n_k} \mapsto \begin{cases} x^n x_1^{n_1} \dots x_k^{n_k}, & \text{if } x^n \neq x_1^{-n_1} \\ x_2^{n_2} \dots x_k^{n_k}, & \text{otherwise.} \end{cases}$$

Clearly $|x^n|$ is a bijection with inverse $|x^{-n}|$. Let F_0 be the group generated by the set $\{|x| \mid x \in X\}$ under the composition of bijection. Consider the map $\phi : F(X) \rightarrow F_0$ given by $1 \rightarrow \text{id}_F$ and $x_1^{n_1} \dots x_k^{n_k} \mapsto |x_1^{n_1}| \dots |x_k^{n_k}|$. Clearly ϕ is surjective with the additional property that $\phi(xy) = \phi(x)\phi(y)$. Since the composition of bijections is associative the preimage of the products in $F[X]$ will also be associative. Moreover ϕ is a group isomorphism between F_0 and F . Also since the preimage of $\{|x| \mid x \in X\}$ is simply X , so we get that $F(x) = \langle X \rangle$. ■

THEOREM 6.7 Let $F(X)$ be the free group of X and let $i : X \rightarrow F(X)$ be the inclusion map. Let G be a group and $f : X \rightarrow G$ be some function. Then there exists a unique group homomorphism \tilde{f} such that the following diagram commutes:

$$\begin{array}{ccc} X & \xrightarrow{i} & F(X) \\ & \searrow f & \downarrow \tilde{f} \\ & & G \end{array}$$

Proof | Let $\tilde{f}(x_1^{n_1} \dots x_k^{n_k}) = f(x_1)^{n_1} \dots f(x_k)^{n_k}$ and $\tilde{f}(1) = e$, where $n_i = \pm 1$. Since $f(x_i)$ are elements of G , $f(x_i)^{n_i}$ are well defined.

1) Since

$$f(w_1 w_2) = f(w_1) f(w_2)$$

it follows that \tilde{f} is a homomorphism.

2) Let $x \in X$. Then $i(x)$ is the reduced word x in $F(X)$. Then $\tilde{f}(x) = f(x)$. Thus the diagram commutes.

3) Let $g : F(X) \rightarrow G$ be another homomorphism such that $g \circ i = f$. Since g is a homomorphism $g(1) = e$ and $g(x^{-1}) = g(x)^{-1}$ for $x \in X$. Thus

$$\begin{aligned} g(x_1^{n_1} \dots x_k^{n_k}) &= g(x_1^{n_1}) \dots g(x_k^{n_k}) = g(x_1)^{n_1} \dots g(x_k)^{n_k} = g \circ i(x_1)^{n_1} \dots g \circ i(x_k)^{n_k} = f(x_1)^{n_1} \dots f(x_k)^{n_k} \\ &= \tilde{f}(x_1^{n_1} \dots x_k^{n_k}). \end{aligned}$$

This means that \tilde{f} is unique. ■

COROLLARY 6.8 Every group G is a quotient of the free group.

Proof | Let G be a group and X be the set of generators of G . Let $j : X \rightarrow G$ be the restriction of identity automorphism of G . Then the following diagram commutes due to the above theorem:

$$\begin{array}{ccc} X & \xrightarrow{i} & F(X) \\ & \searrow j & \downarrow \tilde{f} \\ & & G \end{array}$$

where \tilde{f} is a unique homomorphism that takes $x \mapsto x \in G$. Since $G = \langle X \rangle$ we get that \tilde{f} must be a surjection. Then by first isomorphism theorem,

$$G \simeq F(X) / \text{Ker } \tilde{f}$$
■

DEFINITION 6.9 Let X be a set and $F(X)$ be the free group of X . A group G is said to be defined by the generators $x \in X$ and relations $y \in Y$ if $G \simeq F(X)/N$ where N is a normal subgroup of $F(X)$, and Y generates N . One says that $(X|Y)$ is the presentation of G .

PROPOSITION 6.10 Let $G = (X|Y)$ and $H = (X|Y')$ where $Y \subset Y'$. Then H is isomorphic to a quotient of G .

Proof | Since $Y \subset Y'$, it follows that $\langle Y \rangle \subset \langle Y' \rangle$. By definition the groups generated by Y, Y' are normal. We know that $G \simeq F(X)/\langle Y \rangle$ and $H \simeq F(X)/\langle Y' \rangle$. Thus by third isomorphism theorem:

$$G/\langle Y' \rangle / \langle Y \rangle \simeq F(X)/\langle Y \rangle / \langle Y' \rangle / \langle Y \rangle \simeq F(X)/\langle Y' \rangle \simeq H.$$

■

7 GROUP ACTIONS

DEFINITION 7.1 The action of a group G on a set S is a map $G \times S \rightarrow S$ such that

$$ex = x \text{ \& } (g_1g_2)x = g_1(g_2x), \forall x \in S$$

EXAMPLE 7.2 The action of S_n on $\{1, \dots, n\}$ is given by $(\sigma, x) \mapsto \sigma(x)$.

EXAMPLE 7.3 Let G be a group and H be a subgroup of G . Then the action of H on G given by $(h, g) \mapsto hg$ is called left translation.

EXAMPLE 7.4 If H is a subgroup of G , then the action $(h, g) \mapsto hgh^{-1}$ is called conjugation.

THEOREM 7.5 Let G act on a set S . The relation defined by:

$$x \sim x' \iff gx = x' \text{ for some } g \in G$$

is an equivalence relation.

Proof | Clearly $x \sim x$ since $ex = x$. Suppose $x \sim y$. Then there exists g such that $gx = y$. Multiplying by g^{-1} , $g^{-1}y = x$ and thus $y \sim x$. Suppose that $x \sim y$ and $y \sim z$. Then there exists g and h such that:

$$\begin{aligned} gx &= y \text{ \& } hy = z \\ \implies h(gx) &= z \\ \implies (hg)x &= z \\ \implies x &\sim z \end{aligned}$$

DEFINITION 7.6 The equivalence classes of \sim are called orbits of G on S . The orbit of G on $x \in S$ is denoted as \bar{x} .

THEOREM 7.7 Let G act on S . Then $G_x = \{g \in G \mid gx = x\}$ is a subgroup of G .

Proof | Suppose $g_1, g_2 \in G_x$, then

$$\begin{aligned} g_1x &= x \text{ \& } g_2x = x \\ \implies (g_1g_2)x &= g_1(g_2x) = g_1x = x. \end{aligned}$$

Thus G_x is closed. Suppose $gx = x$ then $g^{-1}x = x$. Therefore G_x is a subgroup.

DEFINITION 7.8 The subgroup G_x is called the stabilizer.

PROPOSITION 7.9 The cardinality of \bar{x} is the index of G_x in G , i.e. $[G : G_x]$.

Proof | Let G/G_x denote the set of cosets of G_x in G . Let

$$\begin{aligned}\phi : \bar{x} &\rightarrow G/G_x \\ gx &\mapsto gG_x\end{aligned}$$

be a map. This map is well defined since

$$\begin{aligned}gx &= hx \\ \implies x &= g^{-1}hx \\ \implies g^{-1}h &\in G_x \implies gG_x = hG_x.\end{aligned}$$

Clearly this map is also injective and surjective. Therefore the cardinality of \bar{x} is same as G/G_x . ■

DEFINITION 7.10 If the group acts on itself, with the group action being conjugation then the orbits of $x \in G$ are called conjugacy classes and the stabilizer of x is written as $C_G(x)$. Also the group $N_G(K) = \{g \in G \mid gKg^{-1} = K\}$ is called the normalizer of K subgroup K .

COROLLARY 7.11 Suppose G is a finite group and K is a subgroup of G :

- 1) The number of elements in the conjugacy class of x is $[G : C_G(x)]$ which divides $|G|$.
- 2) Suppose $G/\sim = \{\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n\}$ then

$$|G| = \sum_{i=1}^n [G : C_G(x_i)]$$

where \sim is defined in theorem 7.5.

- 3) the number of subgroups conjugate to K are $[G : N_G(K)]$.

Proof | Since the conjugacy classes are orbits and $C_G(x)$ is the stabilizer of x it follows from the above theorem that the number of elements in \bar{x} is $[G : C_G(x)]$. Since the group is finite by lagranges theorem we know that $|G|/[G : C_G(x)] = |C_G(x)|$. Similarly for 3 consider the set S of all subsets of G and let G act on S with conjugate group action. The number of subgroups conjugate to K is essentially the orbit of K , thus its cardinality would be $[G : N_G(K)]$, since $N_G(K)$ is the stabilizer of K . For the proof of 2, we use the fact G/\sim is a partition of G , therefore all \bar{x}_i are distinct and their union is G . Since the number of elements in each is $[G : C_G(x_i)]$ we get the result. ■

DEFINITION 7.12 Let $A(S)$ denote the permutation of S (i.e. the set of all bijections $S \rightarrow S$)

LEMMA 7.13 If a group G acts on a set S then this action induces a homomorphism $G \rightarrow A(S)$.

Proof | Let $\phi : G \rightarrow A(S)$ be defined as:

$$f_g(x) = \phi(g)(x) = gx$$

clearly $f_g \in A(S)$. Since

$$\phi(gh)(x) = f_{gh}(x) = g(hx) = g \circ h(x) = \phi(g) \circ \phi(h)(x)$$

Thus ϕ is a homomorphism. ■

THEOREM 7.14 (Cayley's Theorem) If G is a group then there is an injective homomorphism $G \rightarrow A(G)$.

Proof | Let G act on itself by the left translation. Then by the lemma above there is a homomorphism induced by this action. It is easy to show that the homomorphism is injective. ■

💡 For finite groups of order n it can be shown that $A(G)$ is isomorphic to a subgroup of S_n .

PROPOSITION 7.16 If G is a finite group and H is a subgroup such that $[G : H] = p$, where p is the smallest prime which divides $|G|$. Then H is normal in G .

Proof | Let S be the set of cosets of H in G . Let G act on S by the following action:

$$g(xH) = (gx)H.$$

There is a homomorphism induced by the action $G \rightarrow A(S)$. Since S has order p , $A(S) \simeq S_p$ (symmetric group). Let K be the kernel of the homomorphism induced by the action. Then

$$K = \{g \in G \mid \forall x \, gxH = xH\} \implies K \subset H.$$

By first isomorphism theorem G/K is isomorphic to some subgroup of S_p (subgroup because the homomorphism is not onto). Thus by Lagrange's theorem $|G/K|$ divides $p!$. Since $|G/K|$ divides G it follows that every factor of $|G/K|$ divides G . Since p is the smallest number that divides $|G|$, if $[G : K]$ is anything but $1, p$, we will get a contradiction. Since

$$[G : K] = [G : H][H : K] = p[H : K] \geq p \implies [G : K] = p$$

Therefore $[G : K] = p$ and $[H : K] = 1$, which means that $H = K$. Since K is normal in G (it's a kernel) it follows that H is normal. ■

8 SYLOW'S THEOREMS

LEMMA 8.1 If a group H of order p^n (p is prime) acts on a finite set S and if $S_0 = \{x \in S \mid \forall h \in H, hx = x\}$, then $|S| \equiv |S_0| \pmod{p}$.

Proof | By definition $x \in S_0$ if and only if $|\bar{x}| = 1$. Suppose that $S = S_0 \cup \bar{x}_1 \cup \cdots \cup \bar{x}_n$, where \bar{x}_i are distinct and $|\bar{x}_i| > 1$ (this is true since \sim is an equivalence relation on S). Hence $|S| = |S_0| + |\bar{x}_1| + \cdots + |\bar{x}_n|$. Since $|\bar{x}_i| = [H : H_{x_i}]$, and since the index of a subgroup divides $|H| = p^n$, it follows that $|\bar{x}_i| = p^k$ for some non zero $k \leq n$. Therefore $p \mid |\bar{x}_i|$ for each i . This means that $S \equiv S_0 \pmod{p}$. ■

THEOREM 8.2 (Cauchy's Theorem) If G is a group whose order is divisible by a prime p then the group contains an element of order p .

Proof | Let $S = \{(a_1, \dots, a_p) \mid a_i \in G \text{ \& } a_1 \cdots a_p = e\}$. The order of S is $|G|^{p-1}$ (since the last element can be determined given the first $p-1$ elements). Since $p \mid |G|$ it follows that $p \mid |S|$. Let Z_p act on S by cyclic permutation,

$$k(a_1, \dots, a_p) = (a_{k+1}, a_{k+2}, \dots, a_p, a_1, a_2, \dots, a_k)$$

Since $aa^{-1} = a^{-1}a = e$, it follows that $a_{k+1}a_{k+2} \cdots a_k = e$ and therefore the cyclic permutations are in S . Clearly $0x = x$ and $k(k'x) = (k+k')x$ where $x \in S$. This shows that cyclic permutation is a valid action on S .

Consider the set S_0 as defined previously. Suppose $(a_1, \dots, a_p) \in S_0$ then it follows that every cyclic permutation is the same, which is only possible if $a_1 = a_2 = \cdots = a_p$. Since $(e, e, \dots, e) \in S_0$, it is not empty. We know that $0 \equiv S \equiv S_0 \pmod{p}$, therefore S_0 must have atleast p elements. Which means there exists $a \in G$ such that $(a, a, \dots, a) \in S_0$ such that that $a^p = e$. ■

DEFINITION 8.3 A group G in which every element has order p^k for some $k \geq 0$ and p prime is called a p -group. If G is a subgroup of some other group then G is called p -subgroup.

COROLLARY 8.4 G is a finite p -group if and only if $|G|$ is a power of p .

Proof | Suppose that there is some prime $q \neq p$ which divides the order of G . Then by Cauchy's theorem it follows that there exists $a \in G$ such that $a^q = e$. Since G is a p -group the order of a must be p^k for some $k > 0$. Since order is unique we have a contradiction here. Thus p is the only prime which divides $|G|$.

Suppose that $|G| = p^m$. Let $a \in G$. The order of the subgroup $\{e, a, a^2, \dots\}$ must divide p^m (Lagrange's theorem). This is only possible if the order of a is p^k where $k \leq m$. ■

PROPOSITION 8.5 The center $Z(G)$ of a non-trivial p -group G contains more than one element.

Proof | Using the corollary 7.11 it follows that:

$$|G| = |Z(G)| + \sum_{i=1}^m [G : C_G(x_i)]$$

where \bar{x}_i are distinct conjugacy classes and $[G : C_G(x_i)] = |\bar{x}_i| > 1$. Since each of $[G : C_G(x_i)]$ must divide $|G| = p^m$, it follows that $p \mid [G : C_G(x_i)]$. Therefore p must also divide $|Z(G)|$. Since $|Z(G)| \geq 1$, it must be atleast p . Which means it has more than one element. ■

LEMMA 8.6 If H is a p -subgroup of G then $[G : H] \equiv [N_G(H) : H] \pmod{p}$.

Proof | Let S be the set of all cosets of H in G . Then $|S| = [G : H]$. Let H act on S by left translation, i.e. $(g, xH) \mapsto (gx)H$. If $xH \in S_0$ then

$$\begin{aligned} gxH &= xH, \forall g \in H \\ x^{-1}gx &= H, \forall g \in H \\ \implies x^{-1}Hx &= H \\ \implies x &\in N_G(H). \end{aligned}$$

The cardinality of S_0 is thus the same as the number of cosets of H in $N_G(H)$. Therefore $|S_0| = [N_G(H) : H]$. By the lemma we have $|S| \equiv |S_0| \pmod{p}$, thus $[G : H] \equiv [N_G(H) : H] \pmod{p}$. ■

COROLLARY 8.7 If H is a p -subgroup of G such that p divides $[G : H]$ then $N_G(H) \neq H$.

Proof | Using the previous lemma we have

$$0 \equiv [G : H] \equiv [N_G(H) : H] \pmod{p}$$

Since $[N_G(H) : H]$ is atleast 1, we must have $[N_G(H) : H]$ to be atleast p . Therefore $N_G(H) \neq H$. ■

THEOREM 8.8 (Sylow's First Theorem) Let G be a group of order $p^n m$, $n \geq 1$, p prime and $\gcd(p, m) = 1$. Then G has a subgroup of order p^i for each $1 \leq i \leq n$ and moreover the subgroup of order p^i is normal in subgroup of order p^{i+1} , where $1 \leq i < n$.

Proof | By Cauchy's theorem we know that G has an element of order p , and thus a subgroup of order p . By induction suppose that G has a subgroup H of order p^i . Then by the previous lemma and corollary we know that $[N_G(H) : H] \neq H$ and $[G : H] \equiv [N_G(H) : H] \equiv 0 \pmod{p}$. This means $p \mid [N_G(H) : H]$. Since H is normal in $N_G(H)$ it follows that $N_G(H)/H$ is a group whose order is divisible by p . Again by Cauchy's theorem there is a subgroup of $N_G(H)/H$, of the form H_1/H , of order p . Since $|H_1| = |H||H_1/H| = p^i p = p^{i+1}$, the proof is complete. ■

DEFINITION 8.9 A subgroup of P of G is said to be a Sylow p -subgroup if it is the maximal p -subgroup; i.e. if $P \subset H \subset G$ and H is also a p -subgroup then $P = H$.

COROLLARY 8.10 Suppose G is a group of order $p^n m$ like before. Let H be a p -subgroup of G . Then:

- 1) H is a p -Sylow subgroup if and only if $|H| = p^n$.
- 2) Every conjugate of a Sylow p -subgroup is a Sylow p -subgroup.
- 3) If there is only one Sylow p -subgroup then it is normal in G .

Proof | Suppose that P is a Sylow p -subgroup of order p^i . Since by Sylow theorem we know that P must be normal in a subgroup of order p^{i+1} , which is a contradiction. Thus P cannot be a Sylow p -subgroup, unless $i = n$. Suppose now the converse. If $|H| = p^n$ then there cannot be any other subgroup of order p^k such that H is its subgroup. Thus H is a Sylow p -subgroup.


Supposing H is a Sylow p -subgroup, and let $K = gHg^{-1}$ for some $g \in G$. It is clear that K is a subgroup of p^n , since the order of H is also p^n . Therefore by 1 it follows that K is a Sylow p -subgroup.

In case there is only one Sylow p -subgroup, we have that $H = gHg^{-1}$ for every $g \in G$, since the conjugation is also a Sylow p -subgroup. Therefore H is normal. ■

THEOREM 8.11 (Sylow's Second Theorem) Suppose H is a p -subgroup of G and P is a Sylow p -subgroup of G . Then there exists $x \in G$ such that H is a subgroup of xPx^{-1} .

Proof | Let S be the set of all cosets of P . Let H act on S by left translation. Since P is a Sylow p -subgroup its index would not be divisible by p . Since $|S| = [G : P]$ it follows that $|S_0| \equiv [G : P] \pmod{p}$ and thus $|S_0| \neq 0$. Suppose that $xP \in S_0$, then:

$$\begin{aligned} hxP &= xP, \forall h \in H \\ xhx^{-1}P &= P, \forall h \in H \\ \implies xhx^{-1} &\in P, \forall h \in H \\ \implies xHx^{-1} &\subset P \implies H \subset x^{-1}Px. \end{aligned}$$

 In the case when H is itself a Sylow p -subgroup it can be shown that $H = xPx^{-1}$, which means that any two Sylow p -subgroups are conjugations of each other.

THEOREM 8.13 (Sylow's Third Theorem) If G is a finite group and p is a prime, then the number of Sylow p -subgroups of G divides $|G|$ and is of the form $kp + 1$.

Proof | Let S be the set of Sylow p -subgroups of G . Let G act on S by conjugation. Since the stabilizer of P is $N_G(P)$ and the orbit of P is all of S (by second Sylow's theorem), it follows that $|S| = [G : N_G(P)]$. Since $[G : N_G(P)]$ divides $|G|$ so does $|S|$.

Now let P act on S . Clearly $|S|$ at least contains P . Suppose $Q \in S_0$, then $xQx^{-1} = Q$ for all $x \in P$. This means that P is a subgroup of $N_G(Q)$. Since Q is normal in $N_G(Q)$ and since all Sylow p -subgroups are conjugate, it follows that $Q = xQx^{-1} = P$ for some $x \in N_G(Q)$. Thus S_0 only has one element P , implying that $|S| \equiv |S_0| \equiv 1 \pmod{p}$ or $|S| = kp + 1$. ■

PROPOSITION 8.14 If P is a Sylow p -subgroup of a finite group G , then $N_G(N_G(P)) = N_G(P)$.

Proof | Since P is normal in $N_G(P)$, P is the only Sylow p -subgroup in $N_G(P)$, and thus every conjugate of P in $N_G(p)$ is P itself. For all $x \in N_G(N_G(P))$ it follows

$$xN_G(P)x^{-1} = N_G(P) \implies xPx^{-1} \subset N_G(P) \implies xPx^{-1} = P \implies x \in N_G(P).$$

Thus $N_G(N_G(P)) \subset N_G(P)$. The other way is trivial. ■

9 CLASSIFICATION OF FINITE GROUPS

THEOREM 9.1 (Structure Theorem for finitely Generated Abelian Groups) If G is a finitely generated abelian group then there exists a unique tuple (m_1, \dots, m_k, s) such that

$$G \simeq \mathbb{Z}_{m_1} \times \mathbb{Z}_{m_2} \times \dots \times \mathbb{Z}_{m_k} \times \mathbb{Z}^s$$

where $m_i | m_{i+1}$.

Proof | Skipped for now... 

PROPOSITION 9.2 If G is a group of order pq where p, q are primes and $q \nmid p-1$ then $G \simeq \mathbb{Z}_{pq}$. If $q \mid p-1$ then either $G \simeq \mathbb{Z}_{pq}$ or $G \simeq T$ where T is generated by c, d where

$$o(c) = p, o(d) = q \text{ \& } dc = c^s d,$$

where $s \not\equiv 1 \pmod{p}$ and $s^q \equiv 1 \pmod{p}$.

PART III
ANALYSIS

1 CONSTRUCTION OF REAL NUMBERS

DEFINITION 1.1 (Ordered Set) An order on a set S is a relation $<$ with the following properties:

- 1) If $x, y \in S$ then either $x < y$, $y < x$, or $x = y$.
- 2) If $x, y, z \in S$, $x < y$, $y < z$ then $x < z$.

The set S is said to be ordered w.r.t. order $<$.

DEFINITION 1.2 (Bounds) Let S be an ordered set and $E \subset S$ if $\exists \alpha \in S$ such that $x \leq \alpha$, $\forall x \in E$ then α is called an upper bound of E , and E is said to be bounded from above.

Let S be an ordered set and $E \subset S$ if $\exists \alpha \in S$ such that $x \geq \alpha$, $\forall x \in E$ then α is called a lower bound of E , and E is said to be bounded from below.

A subset bounded from above and below is said to be bounded.

DEFINITION 1.3 (Supremum) Let S be an ordered set and $E \subset S$ be bounded from above. If $\alpha \in S$ such that α is an upper bound of E and $\gamma < \alpha$ implies that γ is not an upper bound. α is called the supremum and written as $\sup E$.

DEFINITION 1.4 (Infimum) Let S be an ordered set and $E \subset S$ be bounded from below. If $\alpha \in S$ such that α is a lower bound of E and $\gamma > \alpha$ implies that γ is not an upper bound. α is called the infimum and written as $\inf E$.

DEFINITION 1.5 An ordered set S is said to have the least-upper-bound property (l.u.b property) if every subset which is bounded from above has a supremum in S .

THEOREM 1.6 Let S be an ordered set with l.u.b property. Then every subset of S which is bounded from below has an infimum in S .

Proof | Let $E \subset S$ that is bounded from below and $L = \{\alpha \in S \mid \alpha \leq x, \forall x \in E\}$. Clearly the set L is bounded from above (since every element of E acts as an upper bound). Thus L has a supremum in S , $\beta = \sup L$. If $x < \beta$ then x is not an upper bound of L , thus $x \notin E$ (since every $x \in E$ is an upper bound of L by definition). Thus if $x \in E$ then $\beta \leq x$. This means that β is a lower bound for E . But since by definition it is the largest lower bound β is $\inf E$. ■

DEFINITION 1.7 (Fields) A field is a triplet, $(F, +, \times)$, where F is a set, and $+, \times : F \times F \rightarrow F$ such that:

- 1) $(F, +)$ is an abelian group. The identity of this group is denoted 0.
- 2) $(F - \{0\}, \times)$ is an abelian group. The identity of this group is denoted 1.
- 3) The "product" (i.e. \times operator) is distributive over the "addition" (i.e. $+$ operation).

The additive inverse of $a \in F$ is denoted $-a$ and the multiplicative is denoted a^{-1} or $1/a$.

PROPOSITION 1.8 (Field Properties) Let $x, y \in F$ then:

- 1) $0 \cdot y = 0$.
- 2) $(-x)y = -(xy)$.
- 3) $(-x)(-y) = xy$.

Proof | 1) Using the distribution property $y(0 + 0) = y \cdot 0 + y \cdot 0$. Adding $-y \cdot 0$ on both sides gives us $y \cdot 0 = 0$.

2) Again using distributive property:

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

3) In the above property just substituting $-y$ instead of y gives us $(-x)(-y) = xy$. ■

DEFINITION 1.9 An ordered field $(F, +, \times)$ is a field with an ordering $<$ on F such that

- 1) $y < z \implies x + y < x + z$.
- 2) If $x > 0, y > 0 \implies xy > 0$.

PROPOSITION 1.10 If $(F, +, \times)$ is an ordered field and then:

- 1) $x > 0 \implies -x < 0$ and vice versa.
- 2) If $x > 0$ and $y < z$ then $xy < xz$.
- 3) If $x < 0$ and $y < z$ then $xy > xz$.
- 4) If $x \neq 0$ then $x^2 > 0$.
- 5) $0 < x < y \implies 0 < 1/y < 1/x$.

Proof | 1) Since $x > 0$ and $x + (-x) = 0$, adding the inverse on boths sides gives $0 > -x$ (due to property 1 of ordered fields).

2) Since $z - y > 0$,

$$\begin{aligned} \implies z - y &> 0 \\ \implies x(z - y) &> 0, \text{ (using property 2 of ordered fields)} \\ \implies xz &> xy. \end{aligned}$$

3) If $x < 0$ then $-x > 0$. Applying the same method as above but multiplying $-x$ instead of x gives the result.

4) Since $x > 0$, by property 2 in definition of ordered field we can conclude that $x^2 > 0$.

5) Observe that if $xy > 0$ and $x > 0$ then either $y > 0$ or $y < 0$. If $y < 0$ then $-y > 0$ and $-xy > 0 \implies xy < 0$ leading to a contradiction. Thus if $xy > 0$ and $x > 0$ then $y > 0$. Since $x > 0$ and $x(1/x) = 1 > 0 \implies 1/x > 0$. Since $x < y$

$$\begin{aligned} \implies 1 &< y(1/x) \\ \implies 1/y &< 1/x \end{aligned}$$
■

THEOREM 1.11 There exists an ordered field \mathbb{R} with l.u.b. property. More over \mathbb{Q} is field isomorphic to some subset of \mathbb{R} .

To prove this theorem we will explicitly construct a field and show that it both contains \mathbb{Q} and has l.u.b property.

DEFINITION 1.12 A cut α is a subset of \mathbb{Q} such that:

- 1) α, α^c are not empty.
- 2) If $p \in \alpha, q \in \mathbb{Q}$, and $q < p$ then $q \in \alpha$.
- 3) For each $p \in \alpha$ there exists $r \in \alpha$ such that $p < r$.

PROPOSITION 1.13 If α is a cut then the following are true:

- 1) If $q \notin \alpha$ then $q > p$ for all $p \in \alpha$.
- 2) If $r \notin \alpha$ and $r < s$ then $s \notin \alpha$.

Proof | Both of these follow from 2 in definition 1.12:

- 1) The first statement is just the contrapositive of 2 in definition 1.12.
- 2) If $r \notin \alpha$ then $r > p$ for all $p \in \alpha$. Since $s > r \implies s > p, \forall p \in \alpha$. If we assume that $s \in \alpha$ then the third property is violated (i.e. a cut does not have a maximum element). Thus $s \notin \alpha$. ■

DEFINITION 1.14 Let \mathbb{R} be the collection of all cuts.

DEFINITION 1.15 Let $<$ be a relation on \mathbb{R} defined as $\alpha < \beta \iff \alpha \subsetneq \beta$.

PROPOSITION 1.16 The relation $<$ is an ordering on \mathbb{R} .

Proof | First we must prove that either $\alpha < \beta$, $\beta < \alpha$, or $\alpha = \beta$. Assuming that the later two are wrong, $\alpha \neq \beta$ and $\beta \not< \alpha$. The later can be rephrased as $\exists b \in \beta$ such that $b \notin \alpha$. But using 1 in proposition 1.13 then $a < b, \forall a \in \alpha$ & $b \in \beta$. Further using 2 in definition 1.12 we get that $a \in \alpha \implies a < b \implies a \in \beta$ where $b \in \beta$. Thus $\alpha \leq \beta$. Since we have assumed that $\alpha \neq \beta$, $\alpha < \beta$. Similarly it can be shown that if $\alpha \not< \beta$ and $\alpha \neq \beta$ then $\beta < \alpha$. If we assume that $\beta \not< \alpha$ and $\alpha \not< \beta$, then the former implies that $\alpha \subset \beta$ and the later implies that $\beta \subset \alpha$. Thus $\alpha = \beta$.

Finally if $\alpha < \beta$ and $\beta < \gamma$ then it is clear by definition that $\alpha < \gamma$. ■

PROPOSITION 1.17 Let A be some set, and if $\alpha_i, i \in A$ be cuts then

$$\bigcup_{i \in A} \alpha_i$$

is a cut.

Proof | Since α_i are cuts then clearly $\bigcup_{i \in A} \alpha_i$ is non-empty, and the compliment $\bigcap_{i \in A} \alpha_i^c$ is also non-empty (since cuts cannot be disjoint by the above proposition). If $p \in \bigcup_{i \in A} \alpha_i$ then $p \in \alpha_i$ for some i . It follows that if $q \in \mathbb{Q}$ and $q < p$ then $q \in \alpha_i$ implying that $q \in \bigcup_{i \in A} \alpha_i$. Similarly there exists $r \in \alpha_i$ such that $r > p$, implying that $\exists r \in \bigcup_{i \in A} \alpha_i$ such that $r > p$. Thus $\bigcup_{i \in A} \alpha_i$ is a cut. ■

PROPOSITION 1.18 The set \mathbb{R} has l.u.b. property.

Proof | Let $A \subset \mathbb{R}$ which is bounded from above. I claim that $\alpha_0 = \bigcup_{\alpha \in A} \alpha$ is the supremum of A . From the above proposition α_0 is a cut. Clearly α_0 is an upper bound for A since if $\alpha \in A$ then $\alpha < \bigcup_{\alpha \in A} \alpha \implies \alpha < \alpha_0$. Let $\gamma < \alpha_0$, then there exists $a \in \alpha_0$ s.t. $a \notin \gamma$. Thus there exists a cut $\alpha \in A$ such that $b \in \alpha$. Again since $<$ is an ordering on \mathbb{R} we must have $\gamma < \alpha$. This shows that α_0 is the least upper bound of A . ■

DEFINITION 1.19 Define "addition" on \mathbb{R} as $\alpha + \beta = \{a + b \mid a \in \alpha, b \in \beta\}$.

PROPOSITION 1.20 $(\mathbb{R}, +)$ is an abelian group.

Proof | We must prove the following: $\alpha + \beta$ is a cut, there exists an identity 0^* such that $\alpha + 0^* = \alpha + 0^* = \alpha$, and that for each α there exists an inverse such that $\alpha + (-\alpha) = 0^* = (-\alpha) + \alpha$. The associativity of the operator follows from the associativity of \mathbb{Q} . Also note that $+$ is commutative again due to commutativity of addition on \mathbb{Q} .

Let α, β be cuts. Then clearly $\alpha + \beta$ is non-empty. Since there exists $a' > a$ and $b' > b$ for all $a \in \alpha$ and $b \in \beta$ it follows that $a' + b' > a + b$ implying that $a' + b' \notin \alpha + \beta$. Thus $(\alpha + \beta)^c$ is also non-empty. If $p \in \alpha + \beta$ then $p = a + b$, $a \in \alpha$, $b \in \beta$. If $q < p = a + b \implies q - b < a \implies q - b \in \alpha$. Thus $q = (q - b) + b \in \alpha$. Also since there exists $a' \in \alpha$ and $b' \in \beta$ such that $a < a'$ & $b < b'$. Hence $p < a' + b'$ and $a' + b' \in \alpha + \beta$. Thus $\alpha + \beta$ is a cut.

Define $0^* = \{r \in \mathbb{Q} \mid r < 0\}$. If $p \in \alpha + 0^*$ then $p = a + r$ where $a \in \alpha$ and $r < 0$, this implies $p < a \implies p \in \alpha$. Thus $\alpha + 0^* \subset \alpha$. If $p \in \alpha$ then $\exists p' \in \alpha$ s.t. $p' > p$. Since $p - p' < 0 \implies p - p' \in 0^*$. Thus by definition $p = (p - p') + p' \in \alpha + 0^*$. Thus $\alpha \subset \alpha + 0^*$, and therefore $\alpha = \alpha + 0^*$. The commutativity proves that $\alpha + 0^* = 0^* + \alpha = \alpha$.

Define the inverse of α as $-\alpha = \{p \in \mathbb{Q} \mid \exists r > 0 \text{ s.t. } -r - p \notin \alpha\}$. If $p \in -\alpha$ then $\exists r > 0$ s.t. $-p - r \notin \alpha$. Since $-q - r > -p - r$, using 3 in proposition 1.13 we get that $-q - r \notin \alpha \implies q \in -\alpha$. If we set $t = p + (r/2)$ then $t > p$ and $-t - r/2 = -p - r \notin \alpha \implies t \in -\alpha$. Thus $-\alpha$ is a cut. If $p \in \alpha$ and $q \in -\alpha$ then $\exists r > 0$ s.t. $-q - r \notin \alpha$. Using the second property in proposition 1.13, $-q - r > p \implies p + q < -r < 0 \implies p + q \in 0^*$. Thus $\alpha + (-\alpha) \subset 0^*$. If $u \in 0^*$ then $u < 0$. Define $w = -u/2$. Clearly $w > 0$. Using Archimedean property in \mathbb{Q} we know that exists n such that $nw \in \alpha$ but $(n+1)w \notin \alpha$. Let $p = -(n+2)w$, then $-p - w \notin \alpha$ implying that $p \in -\alpha$. Thus $nw + p = nw - nw - 2w = -2w = v \in \alpha + (-\alpha)$. Hence $0^* \subset \alpha + (-\alpha)$, and therefore $\alpha + (-\alpha) = 0^*$. ■

One can also easily check that the field properties for $+$ are followed.

DEFINITION 1.21 Define a "product" on \mathbb{R}^+ (i.e. set of all cuts $\alpha > 0^*$) as:

$$\alpha\beta = \{p \in \mathbb{Q} \mid p \leq rs, \text{ for some } r \in \alpha, s \in \beta \text{ and } r, s > 0\}.$$

This definition is extended to all α, β in \mathbb{R} in the following way:

$$\alpha\beta = \begin{cases} \alpha\beta, & \alpha, \beta > 0^* \\ -(-\alpha)(\beta), & \alpha < 0^* \text{ \& \& } \beta > 0^* \\ -(\alpha)(-\beta), & \alpha > 0^* \text{ \& \& } \beta < 0^* \\ (-\alpha)(-\beta), & \alpha, \beta < 0^* \end{cases}$$

DEFINITION 1.22 Let $1^* = \{q \in \mathbb{Q} \mid q < 1\}$.

PROPOSITION 1.23 $(\mathbb{R} - \{0^*\}, \cdot)$ forms an abelian group.

Proof | It's too tedious, but similar to that of addition (proof by "cause I said so"). ■

Similarly it can be shown that all ordered field properties are followed by this product. Also it can be shown that the product is distributive over addition. Thus $(\mathbb{R}, +, \cdot)$ is indeed an ordered field with least upper bound property. Thus the remaining part is that \mathbb{Q} is isomorphic to some subset of \mathbb{R} . This can be shown by mapping each rational r to the cut $r^* = \{p \in \mathbb{Q} \mid p < r\}$. It can be easily shown that products and additions are preserved under this map. This completes the proof for theorem 1.11.

COROLLARY 1.24 (Archimedean property) If $0 < x < y \in \mathbb{R}$ then $\exists n \in \mathbb{N}$ such that $nx > y$.

Proof | Let $A = \{nx \mid n \in \mathbb{N}\}$. If we assume that the corollary is false then y is an upper bound of A . Since \mathbb{R} has l.u.b. property A has a supremum, $a = \sup A$. Since $a - 1 < a$ it is not an upper bound of A . Hence $\exists m \in \mathbb{N}$ such that $a - 1 < m$. It follows further that $a < m + 1$ contradicting the fact that a is supremum and that the corollary is false. ■

COROLLARY 1.25 (Denseness of rationals in reals) Let $x < y \in \mathbb{R}$ then $\exists q \in \mathbb{Q}$ such that $x < q < y$.

Proof | Since $y - x > 0 \exists n \in \mathbb{N}$ such that $n(y - x) > 1$ (using archimedean property). Thus $ny - nx > 1$ meaning that there is an integer m such that $nx < m < ny$ (since there is an integer in every interval of length 1). Dividing by n we get $x < m/n < y$, proving the claim ■

2 METRIC SPACES AND EUCLIDEAN SPACE

DEFINITION 2.1 The pair (X, d) is said to be a *metric space* where X is some non-empty set and $d : X \times X \rightarrow \mathbb{R}_{\geq 0}$ is a function with the following properties:

- 1) $d(x, y) = 0 \iff x = y$.
- 2) $d(x, y) = d(y, x) \forall x, y \in X$.
- 3) $d(x, y) \leq d(x, z) + d(z, y)$.

DEFINITION 2.2 Let (X, d) be a metric space. Then:

- 1) A *neighborhood* of a point $x \in X$ is the set $N_r(x) \equiv \{p \mid d(x, p) < r\}$.
- 2) A point p is a *limit point* of set $E \subset X$ if *every* neighborhood of p contains a $q \neq p$ s.t. $q \in E$.
- 3) If $p \in E$ and p is not a limit point of E then p is called an *isolated point* of E .
- 4) E is *closed* if every limit point of E is in E .
- 5) p is said to be in the *interior* of E if $\exists \epsilon > 0$ s.t. $N_\epsilon(p) \subset E$.
- 6) E is *open* if every point of E is an interior point of E .
- 7) E is *perfect* if E is closed and every point of E is a limit point of E .
- 8) E is *bounded* if there exists an $M \in \mathbb{R}$ and $q \in X$ such that $d(p, q) < M, \forall p \in E$.
- 9) E is *dense* in X if every point of X is either a limit point of E , a point in E , or both.

PROPOSITION 2.3 A point x is a limit point of $E \subset X$ if and only if there exists a non-constant sequence $(x_n) \in E$ which converges to x .

Proof | Suppose x is a limit point. Then every neighborhood $N_{1/n}(x) \ni x_n (\neq x) \in E$ such that $x_n \in N_{1/n}(x)$. Since,

$$d(x_n, x) < 1/n \implies \lim_{n \rightarrow \infty} x_n = x.$$

Suppose that there exists a non-constant sequence $(x_n) \in E$ such that $x_n \rightarrow x$. Let $N_\epsilon(x)$ be some neighborhood of x . Since there exists N such that $n > N \implies d(x, x_n) < \epsilon$ it follows that for all $n > N, x_n \in N_\epsilon(x)$. ■

PROPOSITION 2.4 Every neighborhood is open.

Proof | Let $x \in X$ be some point and let $N_r(x)$ be some neighborhood of x . Let $p \in N_r(x)$. Choose $\epsilon < r - d(x, p)$. Then for any $y \in N_\epsilon(p)$,

$$d(x, y) \leq d(x, p) + d(p, y) < r.$$

Hence every point p is in the interior. ■

PROPOSITION 2.5 If p is a limit point of E then there are infinitely many points of E in any neighborhood of p .

Proof | Assume that some neighborhood has finitely many points of E , given by the set $S = \{y_1, \dots, y_n\}$. Then let $\delta < \min\{d(p, y_i) \mid y_i \in S\}$. Then there exists another point $y \in N_\delta(p)$ such that $y \in E$ (since p is a limit point). This is a contradiction, hence there are infinite points in every neighborhood. ■

COROLLARY 2.6 A finite set of points has no limit points.

Proof | If it had a limit point, then every neighborhood of that point must have infinite points. This isn't possible since it has only finitely many points. ■

PROPOSITION 2.7 Let $\{E_\alpha\}$ be any collection of points then:

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

Proof | It's simple. Just show if $x \in A$ then $x \in B$ and the converse, where A, B are the LHS and RHS of the above equation respectively. ■

PROPOSITION 2.8 A set is open iff its complement is closed.

Proof | Suppose E^c is closed. Let $x \in E$ then $x \notin E^c$ which means that x is not a limit point of E^c . Thus there exists a neighborhood N of x such that $N \cap E^c = \emptyset$. Thus $N \subset E$. Thus x has a neighborhood contained in E , making it an interior point. Thus E is open.

Suppose E is open. Let x be a limit point of E^c . Then for any neighborhood N of x there exists a point p s.t. $p \in N \cap E^c$. Thus no neighborhood of x is contained in E , thus x is not an interior point of E and since by assumption E is open it follows that $x \in E^c$. Thus E^c is closed. ■

COROLLARY 2.9 A set is closed iff its complement is open.

Proof | Directly follows from previous proposition. ■

PROPOSITION 2.10 Let X be a metric space and $\{G_\alpha\}$ be any collection of subsets, then:

- 1) If $\{G_\alpha\}$ is open then $\bigcup_{\alpha} G_\alpha$ is open.
- 2) If $\{G_\alpha\}$ is closed then $\bigcap_{\alpha} G_\alpha$ is closed.
- 3) If the collection $\{G_\alpha\}$ is finite and each set is open then $\bigcap_{\alpha=1}^n G_\alpha$ is open.
- 4) If the collection $\{G_\alpha\}$ is finite and each set is closed then $\bigcup_{\alpha=1}^n G_\alpha$ is closed.

Proof | 1) Let $x \in \bigcup_{\alpha} G_\alpha$ then there exists α s.t. $x \in G_\alpha$. Since G_α is open, every neighborhood N of x is contained in G_α and therefore in $\bigcup_{\alpha} G_\alpha$. Proving that $\bigcup_{\alpha} G_\alpha$ is open.

2) Let $x \in X$ be a limit point of $\bigcap_{\alpha} G_\alpha$. Then for every neighborhood N of x , $N \cap \bigcap_{\alpha} G_\alpha$ has infinitely many points. Thus $N \cap G_\alpha$ has infinitely many points (for all α). This means that x is a limit point of every G_α . Since G_α is closed $x \in G_\alpha$, for all α . Therefore $x \in \bigcap_{\alpha} G_\alpha$.

3) Let $x \in \bigcap_{\alpha=1}^n G_\alpha$ then for each α there exists a neighborhood N_α of x such that $N_\alpha \subset G_\alpha$. Let r be the minimum of the radii of the neighborhoods N_α , then $N_r(x) \subset \bigcap_{\alpha=1}^n G_\alpha$. Thus every x is in the interior of $\bigcap_{\alpha=1}^n G_\alpha$.

4) Just use the above proof for complements, and then use de-morgan law. ■

DEFINITION 2.11 Let (X, d) be a metric space, $E \subset X$, L be the set of limit points of E then $\bar{E} := E \cup L$ is called the closure of E .

PROPOSITION 2.12 The closure of any set is closed.

Proof | It is obvious since every limit point is in the set by definition. ■

PROPOSITION 2.13 $E = \bar{E}$ iff E is closed.

Proof | If E is closed then $L \subset E \implies \bar{E} = E \cup L = E$. If $\bar{E} = E$ then $\bar{E} \cap L = E \cap L$. Since $L \subset \bar{E}$ it follows that $L = E \cap L$. Thus $L \subset E$. ■

PROPOSITION 2.14 If $E \subset F$ and F is closed then $\bar{E} \subset F$.

Proof | Let x be a limit point of E . Since $E \subset F$, it follows that x is also a limit point of F (since it every neighborhood of x would contain a point E which is also in F). Since F is closed every limit point of F is in F . Thus if L is the set of limit points of E then $L \subset F$. Hence $\bar{E} = E \cup L \subset F$. ■

PROPOSITION 2.15 If $E \subset Y \subset X$ then E is open relative to Y iff $E = G \cap Y$ for some open set in G in X .

Proof | Suppose that $E = G \cap X$. Let $x \in E$. Then there exists a neighborhood $N_r(x)$ such that $N_r(x) \subset G$. Consider $N = N_r(x) \cap G$. Clearly this is a neighborhood of x relative to Y . Since $N \subset E$, x is an interior point of E .

Suppose E is open relative to Y . Then $\exists r_x > 0$ s.t. $V_{r_x}(x) := \{p \mid d(x, p) < r_x, p \in Y\} \subset E$. Clearly $V_{r_x}(x) = N_{r_x}(x) \cap Y \subset E$. Let $G = \bigcup_{x \in E} N_{r_x}(x)$. Clearly G is open. Thus $G \cap Y = E$ since $\bigcup_{x \in E} V_{r_x}(x) = E$. ■

DEFINITION 2.16 A collection $\{G_\alpha\}$ is said to be an open coering of $E \subset X$ if G_α are covered and $E \subset \bigcup_\alpha G_\alpha$. A subcovering is a subset of $\{G_\alpha\}$ which also covers E .

DEFINITION 2.17 A subset Y of metric space X is said to be compact if every open covering of Y contains a finite subcover.

PROPOSITION 2.18 If $K \subset Y \subset X$ then K is compact in X iff K is compact in Y .

Proof | Suppose K is compact in X . Let $\{H_\alpha\}$ be any open covering of K in Y . Then by the previous theorem each set $H_\alpha = G_\alpha \cap Y$, where G_α is open in X . Since K is compact in X there exists a finite subcover $\{G_1, \dots, G_n\}$. The collection $\{H_1 = G_1 \cap Y, \dots, H_n = G_n \cap Y\}$ will be a finite subcover of K in Y . Thus there is a finite subcover of every open cover in Y .

Conversly, suppose that K is compact relative to Y , then similarly using the same theorem it is possible to construct a finite subcovering of any opren cover of K in X . ■

PROPOSITION 2.19 Compact subsets of X are closed.

Proof | We will prove that the compliment of a compact subset K is open. Let $p \in K^c$ and $q \in K$. Let $\epsilon_q < d(p, q)/2$. Then the union $\bigcup_{q \in K} N_{\epsilon_q}(q)$ is an open covering of K . Since K is compact there some finite q_1, \dots, q_n s.t. $\{N_{\epsilon_i}(q_i)\}$ is also a covering of K (I have defined $\epsilon_i = d(p, q_i)$). Let $G = \bigcap_{i=1}^n N_{\epsilon_i}(p)$. G is an open neighborhood of p . Let $x \in G$ then $d(x, p) < d(p, q_i)$, $\forall 1 \leq i \leq n$. Since:

$$\begin{aligned} d(p, q_i) &\leq d(p, x) + d(q_i, x) \\ d(p, q_i) &< \frac{1}{2}d(p, q_i) + d(q_i, x) \\ \implies \frac{1}{2}d(p, q_i) &< d(q_i, x). \end{aligned}$$

Thus $x \notin N_{\epsilon_i}(q_i)$, and thus $x \notin K$. Meaning that $G \cap K = \emptyset$. Since G is an open neighborhood of p , and $G \subset K^c$ it follows that K^c is open. ■

PROPOSITION 2.20 Closed subsets of compact sets are compact.

Proof | Let $H \subset K$, where K is compact and H is closed. Let Ω be an open cover of H . Then $\Omega \cup \{H^c\}$ is also an open covering of K (this only works since H^c is open). Since K is compact there exists a finite subcovering Φ . If $H^c \in \Phi$ then $\Phi - \{H^c\}$ is a finite subcovering of H . ■

COROLLARY 2.21 If F is closed and K is compact then $F \cap K$ is compact.

Proof | Since K, H are closed $K \cap H$ is closed, and a subset of K . Thus it must be compact. ■

PROPOSITION 2.22 Let $\mathcal{K} = \{K_\alpha\}$ be a collection of compact sets such that every subcollection has non-empty intersection. Then $\bigcap_\alpha K_\alpha \neq \emptyset$.

Proof | Suppose that the intersection is empty. Then there exists $K_1 \in \mathcal{K}$ s.t. $K \cap K_\alpha = \emptyset$ whenever $K_\alpha \neq K_1$. This means that $K_1 \subset K_\alpha^c$. Thus $\{K_\alpha^c\}$ form an open cover of K_1 . This implies that there exists a finite subcover $\{K_{\alpha_i}^c \mid 0 \leq i \leq n\}$ which covers K_1 . Thus $K_1 \cap K_{\alpha_1} \cap \dots \cap K_{\alpha_n} = \emptyset$, which is a contradiction to the hypothesis. ■

COROLLARY 2.23 If $\{K_n\}$ are compact sets and $K_n \supset K_{n+1}$ then $\bigcap_{n \geq 1} K_n$ is non-empty.

PROPOSITION 2.24 Let $E \subset K$ where K is compact and E is infinte, then E has atleast one limit point in K .

Proof | Suppose there exists no limit point of E in K . This means that $\forall q \in K \exists N(q)$ s.t. $N(q)$ has at most one point of E (i.e. q). It is clear that $\{N(q)\}$ forms an open covering of K . Since E is infinite and $N(q)$ only has upto one point of E it is not possible to find a finite subcovering of E . Thus $\{N(q)\}$ has no finite subcovering of K (since $E \subset K$). This is contrary to the fact that K is compact. ■

PROPOSITION 2.25 If I_n are non-empty intervals in \mathbb{R} and $I_n \supset I_{n+1}$ then $\bigcap_{n \geq 1} I_n$ is non-empty.

Proof | Let $I_n = [a_n, b_n]$. Consider the set $\{a_n\}$. Clearly this is non-empty and bounded above by b_1 . Let the supremum of the set be x . Since

$$a_n < b_m$$

for any m , it follows that $x \leq b_m$. We also know that $a_m \leq x$. Thus $x \in I_m$ for all m . ■

COROLLARY 2.26 Let I_n be a sequence of non-empty k -cells such that $I_n \supset I_{n+1}$.

Proof | Follows directly from the proposition above. ■

THEOREM 2.27 Every k -cell is compact.

Proof | Let I be a k -cell with points (x_1, \dots, x_k) such that $a_j \leq x_j \leq b_j$. Define ϵ as:

$$\epsilon = \left(\sum_{j=1}^n (b_j - a_j)^2 \right)^{1/2}.$$

Clearly if $x, y \in I$ then $\|x - y\| < \epsilon$. Suppose that k -cells are not compact. Then there exists an open cover Ω which does not have a finite subcover of I . Let $c_j = (a_j + b_j)/2$, then cartesian products of $[a_j, c_j]$ and $[c_j, b_j]$ in different combinations produce 2^k k -cells. At least one of these k -cells cannot be covered using a finite subcovering of Ω , call this I_1 . Repeat the same process for I_1 to gain I_2 and so on. Thus we have a sequence of k -cells such that:

- 1) $I \supset I_1 \supset I_2 \cdots$;
- 2) Each I_n does not have a finite subcovering in Ω ;
- 3) If $x, y \in I_n$ then $\|x - y\| < 2^{-n}\epsilon$.

By the above proposition we know that there exists at least one x^* s.t. $x^* \in I_n$ for all n . Since Ω covers I there exists $G \in \Omega$ s.t. $x^* \in G$. Since G is open there exists $r > 0$ s.t. $N_r(x^*) \subset G$. Choose n large enough so that $2^{-n}\epsilon < r$. This means that $I_n \subset N_{2^{-n}\epsilon}(x^*) \subset G$. This is a contradiction since G alone covers I_n . ■

THEOREM 2.28 (Heine-Borel Theorem) If E is a subset of \mathbb{R}^k then the following are equivalent:

- 1) E is bounded and closed.
- 2) E is compact.
- 3) Every infinite subset of E has a limit point in E .

Proof | 2 follows from 1 since every bounded set can be contained inside a k -cell, and the fact that every closed subset of a compact set is compact. 3 follows from 2 due to proposition 2.23. All that is remaining is to show 1 from 3.

If E is not bounded then there is a sequence in $x_1, \dots, x_n \in E$ such that $|x_n| > n$ for each n . Clearly this does not have a limit point in \mathbb{R}^k , hence does not have a limit point in E . Thus 3 implies that E must be bounded. If E is not closed then there is a point x_0 which is a limit point of E but not in E . This means that there is a sequence in $(x_n) \in E$ which converges to x_0 . Since every infinite subset of E has a limit point in E , we have a contradiction. Thus E must be closed. ■

THEOREM 2.29 Every bounded infinite subset of \mathbb{R}^k has a limit point in \mathbb{R}^k .

Proof | If E is a bounded subset of \mathbb{R}^k then it is contained inside a k -cell I . Since I is compact every limit point of every subset is contained in I and thus in \mathbb{R}^k . ■

DEFINITION 2.30 (Seperable Sets) Two subsets $A, B \subset X$ are said to be seperated if $\bar{A} \cap B = A \cap \bar{B} = \emptyset$.

DEFINITION 2.31 (Connected) A subset $A \subset X$ is said to be connect if *cannot* be written as the disjoint union of non-empty seperated sets.

THEOREM 2.32 $E \subset \mathbb{R}$ is connected if and only if $x, y \in E$ & $x < z < y \implies z \in E$.

Proof | Suppose that E is connected and there exist $x, y \in E$ such that there is a $z \in (x, y)$ and $z \notin E$. Then let $A = (-\infty, z) \cap E$ and $B = (z, \infty) \cap E$. Since $A \subset (-\infty, z)$ and $B \subset (z, \infty)$ it follows that A and B are seperated. Since $A \cup B = E$, we have arrived at a contradiction.

To prove the converse suppose that E is not connected. Then $E = A \cup B$ where A and B are seperated. Let $x \in A$ and $y \in B$ and w.l.o.g. assume $x < y$. Let $z = \sup(A \cap [x, y])$. This means that z is a limit point of A and thus $z \in \bar{A}$. Since A and B are seperated $z \notin B$. Hence $x \leq z < y$.

- 1) If $z \notin A$ then $x < z < y$ and $z \notin E$.
- 2) If $z \in A$; then $z \notin \bar{B}$. Since z is not a limit point of B there is an open neighborhood $(z, z_1) \cap B = \emptyset$. Thus $z_1 \notin B$. Since z is a supremum of $A \cap [x, y]$ and $z_1 \in [x, y]$, it follows that $z_1 \notin A$. Therefore $z < z_1 < y$ but $z \notin E$. This completes the proof. ■

3 SEQUENCES

PROPOSITION 3.1 Let $x(n) = (x_1(n), \dots, x_k(n))$ be a sequence in \mathbb{R}^k then $x(n) \rightarrow x = (x_1, \dots, x_k)$ iff $x_j(n) \rightarrow x_j$.

Proof | Suppose $x_j(n) \rightarrow x_j$. Let $\epsilon > 0$. There exists N_j s.t.

$$n > N \implies \|x_j(n) - x_j\| < \frac{\epsilon}{k}.$$

Since

$$\|x(n) - x\| = \sqrt{\sum_{j=1}^k (x_j(n) - x_j)^2} \leq \sum_{j=1}^k |x_j(n) - x_j|,$$

it follows that for $N = \max\{N_j\}$:

$$n > N \implies \|x(n) - x\| < \epsilon.$$

Now suppose that $x(n) \rightarrow x$. Then there exists N such that:

$$n > N \implies \|x(n) - x\| < \epsilon.$$

Since,

$$|x_i(n) - x_i| \leq \sqrt{\sum_{j=1}^k (x_j(n) - x_j)^2} = \|x(n) - x\| < \epsilon$$

It follows that for each i ,

$$n > N \implies |x_i(n) - x_i| < \epsilon.$$

DEFINITION 3.2 If (p_n) is a sequence in X , and let (n_k) be a sequence in \mathbb{R} such that $n_1 < n_2 < \dots$ then (p_{n_k}) is a subsequence of (p_n) .

THEOREM 3.3 Let p_n be a sequence in a compact metric space X , then there is subsequence of p_n which converges in X .

Proof | Let $E = \{x \mid \exists n \text{ s.t. } x = p_n\}$. Suppose E is finite. Then by pigeon hole principle there must be a $p \in E$ such that for infinitely many n_i ,

$$p_{n_i} = \dots = p.$$

This is a subsequence which converges to a point in X . Now suppose that E is infinite. Since X is a compact space, every infinite subset of X must have a limit point in X . Thus there is some sequence $(x_n) \in E$ which converges to some $p \in X$. By definition $x_n = p_{n_k}$ for some k . ■

THEOREM 3.4 (Bolzano-Weistrass) Every bounded sequence in \mathbb{R}^k has a convergent subsequence.

Proof | Let x_k be a sequence in \mathbb{R}^k which is bounded. Since it is bounded it is contained inside a k -cell. Since k -cells are compact, and x_k is a sequence contained in x_k there exists a subsequence of x_k which converges to a point in the k -cell. ■

THEOREM 3.5 The set of all subsequential limits of a sequence in X is closed in X .

Proof | Let p_n be a sequence. Let $E = \{x \mid \exists (p_{n_k}) : p_{n_k} \rightarrow x\}$. Let z be a limit point of E . The set $N_{1/2n}(z)$ contains some $x \in E$. By definition there exists a subsequence p_{n_k} which converges to x . This means that there exists an N_n such that

$$k \geq N_n \implies d(x, p_{N_n}) < \frac{1}{2n}.$$

It follows that $d(z, p_{N_1}) \leq d(z, x) + d(x, p_{N_1}) < 1/n$. Repeating this for each n we get a subsequence p_{N_1}, p_{N_2}, \dots which converges to z by construction. Therefore $z \in E$. ■

PROPOSITION 3.6 Every convergent sequence is Cauchy.

Proof | Not hard to show. ■

PROPOSITION 3.7 Every compact space is complete.

Proof | Let x_n be a Cauchy sequence. Since x_n is a sequence in a compact space there exists a convergent subsequence x_{n_k} which converges to some $x \in X$. Let $\epsilon > 0$, then there exists N_1 and N_2 such that

$$\begin{aligned} k > N_1 &\implies d(x_{n_k}, x_k) < \epsilon/2 \\ k > N_2 &\implies d(x_{n_k}, x) < \epsilon/2 \end{aligned}$$

Since

$$d(x_k, x) \leq d(x_k, x_{n_k}) + d(x_{n_k}, x),$$

choosing $N > \max\{N_1, N_2\}$,

$$k > N \implies d(x_k, x) < \epsilon.$$

PROPOSITION 3.8 \mathbb{R}^k is complete.

Proof | If we can show that every cauchy sequence is bounded in \mathbb{R}^k we are done. There exists an N such that

$$n \geq N \implies x_n \in N_1(x_N).$$

Thus x_n is a bounded sequence. ■

4 CONTINUITY

DEFINITION 4.1 Let X, Y be metric spaces, $p \in X, E \subset X$, and let $f : E \rightarrow Y$. Then we write

$$\lim_{x \rightarrow p} f(x) = q$$

If there exists a $q \in Y$ which satisfies the following property:

$$\forall \epsilon > 0, \exists \delta > 0 : d_X(x, p) < \delta \implies d_Y(f(x), q) < \epsilon.$$

THEOREM 4.2 Let X, Y, E, f, p be the same as above. Then

$$\lim_{x \rightarrow p} f(x) = q \iff \lim_{n \rightarrow \infty} f(p_n) = q, \text{ for every sequence } p_n \rightarrow p$$

Proof | Suppose that as $x \rightarrow p, f(x) \rightarrow q$. Let $\epsilon > 0$, then

$$\exists \delta > 0 : d_X(x, p) < \delta \implies d_Y(f(x), q) < \epsilon.$$

Let p_n be any sequence such that $p_n \rightarrow p$. Thus:

$$\begin{aligned} \exists N : n > N &\implies d_X(p_n, p) < \delta \\ &\implies d_Y(f(p_n), q) < \epsilon. \end{aligned}$$

Therefore $f(p_n) \rightarrow q$.

Conversely suppose that

$$\exists \epsilon > 0 : \forall \delta > 0, d_X(x, p) < \delta \ \& \ d_Y(f(x), p) \geq \epsilon.$$

Let $x_n \in E$ be the point such that

$$d_X(x_n, p) < \frac{1}{n}$$

but since,

$$d_Y(f(x_n), p) \geq \epsilon$$

for all n , it follows that:

$$\lim_{n \rightarrow \infty} f(x_n) \neq q.$$

■



Since the limit of a sequence is unique, it follows that the limit of a function is also unique. This follows from the above theorem.

DEFINITION 4.4 Let X, Y, E, f, p be as before. Then f is said to be continuous at p if:

$$\forall \epsilon > 0, \exists \delta > 0 : d_X(x, p) < \delta \implies d_Y(f(x), f(p)) < \epsilon.$$

THEOREM 4.5 f is continuous at p if and only if

$$\lim_{x \rightarrow p} f(x) = f(p)$$

Proof | This follows from theorem 4.2. ■

PROPOSITION 4.6 Let X, Y, Z be metric spaces, $p \in X$, and let $f : E \rightarrow Y$, $g : f(E) \rightarrow Z$ be functions continuous at p and $f(p)$ respectively. Then $g \circ f$ is continuous at p .

Proof | Let $\epsilon > 0$. Since g is continuous at $f(p)$, there exists a δ_1 such that

$$d_Y(f(x), f(p)) < \delta_1 \implies d_Z(g(f(x)), g(f(p))) < \epsilon.$$

Now using the continuity of f at p , there exists $\delta > 0$ such that:

$$d_X(x, p) < \delta \implies d_Y(f(x), f(p)) < \delta_1.$$

Thus there exists $\delta > 0$ such that

$$d_X(x, p) < \delta \implies d_Z(g \circ f(x), g \circ f(p)) < \epsilon.$$

THEOREM 4.7 f is continuous on a metric space X if and only if $f^{-1}(F)$ is an open set if $F \subset Y$ is open.

Proof | Suppose that f is continuous every where and $F \subset Y$ is open. Let $p \in f^{-1}(F)$, then $f(p) \in F$ (by definition). Since F is open there exists an $\epsilon > 0$ such that $N_\epsilon(f(p)) \subset F$. Using continuity, there exists a δ such that

$$d_X(x, p) < \delta \implies d_Y(f(x), f(p)) < \epsilon.$$

Thus if $x \in N_\delta(p)$ then $f(x) \in N_\epsilon(f(p))$. Therefore $N_\delta(p) \subset f^{-1}(N_\epsilon(f(p)))$. But since $N_\epsilon(f(p)) \subset F$ it follows that $N_\delta(p) \subset f^{-1}(F)$.

Conversly suppose that $f^{-1}(F)$ is open whenever F is open. Let $\epsilon > 0$. Then we know that $f^{-1}(N_\epsilon(f(p)))$ is open. Clearly $p \in f^{-1}(N_\epsilon(f(p)))$, thus there exists a $\delta > 0$ such that $N_\delta(p) \subset f^{-1}(N_\epsilon(f(p)))$. Therefore:

$$d_X(x, p) < \delta \implies d_Y(f(x), f(p)) < \epsilon.$$

DEFINITION 4.8 A function $f : X \rightarrow \mathbb{R}^k$ is said to be bounded if $\|f(x)\| < M$ for some M .

PROPOSITION 4.9 If $f : X \rightarrow Y$ is a continuous function, and X is compact, then $f(X)$ is also compact.

Proof | Let V_α be an open covering on $f(X)$. Then we know that:

$$f(X) \subset \bigcup_{\alpha} V_{\alpha} \implies X = \bigcup_{\alpha} f^{-1}(V_{\alpha}).$$

Since f is continuous $f^{-1}(V_{\alpha})$ are open, and thus $f^{-1}(V_{\alpha})$ form an open cover of X . Since X is compact there exists a finite subcover $f^{-1}(V_{\alpha_i})$. Therefore:

$$X = \bigcup_{i=1}^n f^{-1}(V_{\alpha_i}) \implies f(X) \subset \bigcup_{i=1}^n V_{\alpha_i}.$$

Hence $f(X)$ is compact. ■

COROLLARY 4.10 If $f : X \rightarrow \mathbb{R}^k$, and X is compact then $f(X)$ is closed and bounded and f is a bounded function.

COROLLARY 4.11 If $f : X \rightarrow \mathbb{R}$, and X is compact and let

$$M = \sup_{x \in X} f(x) \ \& \ m = \inf_{x \in X} f(x).$$

Then there exists $p, q \in X$ such that $f(p) = M$ and $f(q) = m$.

Proof | Since X is compact, it follows that $f(X)$ is closed and bounded in \mathbb{R} . Since $f(X)$ is bounded the supremum and infimum would exist. Since there always exists a monotonous sequence $y_n \in f(X)$ such that $y_n \rightarrow M$, using the closedness we know that $M \in f(X)$. Therefore there exists p such that $f(p) = M$. Similarly for m . ■

PROPOSITION 4.12 If $f : X \rightarrow Y$ is bijective and X is compact, then f^{-1} is continuous.

DEFINITION 4.13 If $f : X \rightarrow Y$ where X, Y are metric spaces, we say that f is uniformly continuous on X if

$$\forall \epsilon > 0, \exists \delta > 0 : d_X(p, q) < \delta \implies d_Y(f(p), f(q)) < \epsilon.$$

Note that uniform continuity is defined for sets and not points. The δ in uniform continuity is only dependent on ϵ and not on p, q . In continuity δ may depend on the point.

THEOREM 4.15 If X is compact and $f : X \rightarrow Y$ is continuous function then f is uniformly continuous.

Proof | Let $\epsilon > 0$. Then for each $p \in X$ we can write that:

$$\exists \delta(p) : d_X(p, q) < \delta(p) \implies d_Y(f(p), f(q)) < \epsilon.$$

Let $V(p) = \{q \mid d_X(p, q) < \frac{1}{2}\delta(p)\}$. $\{V(p)\}$ is an open cover of X . Using compactness we can say that there exists a finite subcover $V(p_1), \dots, V(p_n)$ which covers X . Let $\delta = \min(\delta(p_1), \dots, \delta(p_n))/2$.

Then:

$$d_X(p, q) < \delta \implies d_Y(f(p), f(q)) < \epsilon.$$

■

PROPOSITION 4.16 If $f : X \rightarrow Y$ is continuous, and $E \subset X$ is connected then $f(E)$ is also connected.

Proof | Suppose that $f(E)$ is not connected, then $f(E) = A \cup B$ where A and B are separated. Let $G = E \cap f^{-1}(A)$ and $H = E \cap f^{-1}(B)$. Since $G \subset f^{-1}(A)$ it follows that $G \subset f^{-1}(\bar{A})$. Since f is continuous it follows that the latter is closed. Therefore

$$\begin{aligned} \bar{G} &\subset f^{-1}(\bar{A}) \\ f(\bar{G}) &\subset \bar{A} \end{aligned}$$

Since $f(H) = B$ and $\bar{A} \cap B = \emptyset$ it follows that $\bar{G} \cap H = \emptyset$. Similarly it can be shown that $G \cap \bar{H} = \emptyset$. This means E is not connected which is a contradiction. ■

THEOREM 4.17 Let $f : [a, b] \rightarrow \mathbb{R}$ be a continuous function. If $f(a) < f(b)$ and $f(a) < c < f(b)$ then there exists $x \in (a, b)$ such that $f(x) = c$.

Proof | Since $[a, b]$ is connected, it follows that $f([a, b])$ will also be connected, since f is continuous. Since every connected set in \mathbb{R} is an interval, it follows that $f([a, b])$ is an interval. If $f(a) < c < f(b)$ then $c \in f([a, b])$ therefore $f(x) = c$ for some x . ■

5 DIFFERENTIATION

In this section we will only discuss functions $f : [a, b] \rightarrow \mathbb{R}$.

DEFINITION 5.1 Let f be a function, and let $x \in [a, b]$ then define the quotient

$$\phi(t) = \frac{f(t) - f(x)}{t - x}, \quad t \neq x.$$

and let $f'(x)$ be

$$f'(x) = \lim_{t \rightarrow x} \phi(t)$$

Assuming that this limit exists. $f'(x)$ is called the derivative of f at x . So to f we are assigning a new function f' whose domain is all those points where the above limit exists.

PROPOSITION 5.2 If f is differentiable at x then f is continuous at x .

Proof | Since

$$\lim_{t \rightarrow x} f(t) - f(x) = \lim_{t \rightarrow x} \phi(t)(t - x) = f'(x) \cdot 0 = 0.$$

It follows that f is continuous. ■

PROPOSITION 5.3 Suppose f, g are functions which are differentiable at x . Then:

- 1) $(f + g)'(x) = f'(x) + g'(x)$
- 2) $(fg)'(x) = f(x)g'(x) + f'(x)g(x)$.

Proof | Since

$$\lim_{t \rightarrow x} \phi(t) = \lim_{t \rightarrow x} \frac{f(t) + g(t) - (f(x) + g(x))}{t - x} = \lim_{t \rightarrow x} \frac{f(t) - f(x)}{t - x} + \lim_{t \rightarrow x} \frac{g(t) - g(x)}{t - x}$$

it follows that $(f + g)'(x) = f'(x) + g'(x)$. For the second part let $h = fg$. Then

$$h(t) - h(x) = f(t)g(t) - f(x)g(x) = f(t)(g(t) - g(x)) + (f(t) - f(x))g(x)$$

Thus

$$\lim_{t \rightarrow x} \phi(t) = \lim_{t \rightarrow x} f(t) \lim_{t \rightarrow x} \frac{g(t) - g(x)}{t - x} + \lim_{t \rightarrow x} g(x) \lim_{t \rightarrow x} \frac{f(t) - f(x)}{t - x}.$$

PROPOSITION 5.4 Suppose $f : [a, b] \rightarrow \mathbb{R}$ and $g : f([a, b]) \rightarrow \mathbb{R}$, f is differentiable at x and g is differentiable at $f(x)$. Then $h = g \circ f$ is differentiable at x , and:

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

Proof | From the definition of derivative:

$$\begin{aligned}\phi(t) &= \frac{g(f(t)) - g(f(x))}{t - x} \\ &= \frac{g(f(t)) - g(f(x))}{f(t) - f(x)} \frac{f(t) - f(x)}{t - x}\end{aligned}$$

Therefore

$$\lim_{t \rightarrow x} \phi(t) = g'(f(x))f'(x).$$

DEFINITION 5.5 A function $f : X \rightarrow \mathbb{R}$ has a local minimum at p if there exists $\delta > 0$ such that $d(p, q) < \delta \implies f(q) < f(p)$. Similarly we define a local maximum.

PROPOSITION 5.6 If $f : [a, b] \rightarrow \mathbb{R}$ has a local maximum/minimum at x and f is differentiable at x then $f'(x) = 0$.

Proof | Suppose x is a point of local maximum then, there exists a δ in accordance to ???. Let $t \in (x - \delta, x)$, then:

$$\phi(t) = \frac{f(t) - f(x)}{t - x} \geq 0,$$

if $t \in (x, x + \delta)$ then

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

Since the limit exists, we must have $\lim_{t \rightarrow x} \phi(t) = 0$. Therefore $f'(x) = 0$.

PROPOSITION 5.7 Let h be a continuous function on $[a, b]$ and differentiable on (a, b) , and $h(a) = h(b)$ then there exists an $x \in (a, b)$ such that $h'(x) = 0$.

Proof | If h is a constant function then $h'(x) = 0$ everywhere on (a, b) . If $h(t) > h(a)$ for some t , then h attains its maximum at $x \in (a, b)$ (since $h(a) = h(b)$). Thus $h'(x) = 0$. Similarly if $h(t) < h(a)$ for some t then there exists a local minimum $x \in (a, b)$, where $h'(x) = 0$.

THEOREM 5.8 If f, g are continuous functions on $[a, b]$ and differentiable on (a, b) then there exists $x \in (a, b)$ such that

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

Proof | Let

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

then $h(a) = h(b)$ and h is continuous on $[a, b]$ and differentiable on (a, b) . Thus by the above theorem there exists an x such that $h'(x) = 0$. This proves our claim. ■

THEOREM 5.9 Suppose f is continuous on $[a, b]$ and differentiable on (a, b) then there exists an $x \in (a, b)$ such that

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

Proof | Use the previous theorem with $g(x) = x$. ■

6 UNIFORM CONVERGENCE

DEFINITION 6.1 A sequence of functions $f_n : X \rightarrow \mathbb{C}$ is said to be pointwise converging to f if

$$\forall x \ f(x) = \lim_{n \rightarrow \infty} f_n(x),$$

or, for all $\epsilon > 0$ there exists $N(x)$ such that

$$n > N(x) \implies |f_n(x) - f(x)| < \epsilon.$$

DEFINITION 6.2 A sequence of functions f_n are said to be uniformly converging to f if for all $\epsilon > 0$ there exists N such that

$$n > N \implies |f_n(x) - f(x)| < \epsilon, \forall x.$$



From here on if N depends on x I will write $N(x)$. Otherwise it means that N does not depend on x .

PROPOSITION 6.4 A sequence of functions f_n is uniformly convergent if and only if for each $\epsilon > 0$ there exists N such that $m, n > N \implies |f_n(x) - f_m(x)| < \epsilon$ for all $x \in E$.

Proof | Suppose that f_n is uniformly convergent to f . Then there exists N such that

$$n > N \implies |f_n(x) - f(x)| < \epsilon/2.$$

Using triangle inequality

$$\begin{aligned} |f_n(x) - f_m(x)| &< |f_m(x) - f(x)| + |f_n(x) - f(x)| \\ &< \epsilon. \end{aligned}$$

Conversely suppose that if f_n is cauchy convergent. Then there exists N such that

$$n, m > N \implies |f_n(x) - f_m(x)| < \epsilon.$$

Since $f_n(x)$ is a cauchy sequence in \mathbb{R} for each x , by completeness of \mathbb{R} we know that it converges to some number. Let $f(x)$ be this number. Then as $m \rightarrow \infty$ we get that

$$n > N \implies |f_n(x) - f(x)| < \epsilon.$$



7 RIEMANN INTEGRATION

DEFINITION 7.1 A partition P of an interval $[a, b]$ is a set of finite point $x_0 \leq \dots \leq x_n \in [a, b]$ such that $x_0 = a$ and $b = x_n$. We write $\Delta x_k = x_k - x_{k-1}$, where $1 \leq k \leq n$

DEFINITION 7.2 Let $f : [a, b] \rightarrow \mathbb{R}$ be bounded on $[a, b]$. Let P be a partition on $[a, b]$. Suppose

$$M_k = \sup_{x_{k-1} \leq x \leq x_k} f(x)$$

$$m_k = \inf_{x_{k-1} \leq x \leq x_k} f(x).$$

Then define the upper and lower integrals as

$$U(f, P) = \sum_{k=1}^n M_k \Delta x_k$$

$$L(f, P) = \sum_{k=1}^n m_k \Delta x_k$$

DEFINITION 7.3 Define the upper and lower Riemann integrals as:

$$\overline{\int_a^b} f(x) dx = \sup_P U(f, P)$$

$$\underline{\int_a^b} f(x) dx = \sup_P L(f, P)$$


The function f is said to be integrable if

$$\underline{\int_a^b} f(x) dx = \overline{\int_a^b} f(x) dx$$

and the Riemann integral of f on the interval $[a, b]$ is

$$\int_a^b f(x) dx = \underline{\int_a^b} f(x) dx = \overline{\int_a^b} f(x) dx$$

DEFINITION 7.4 If P is a partition then P^* is said to be a refinement of P if $P^* \supset P$. Given two partitions P_1 and P_2 then their common refinement P^* is defined as $P_1 \cup P_2$.

 It is obvious that $L(f, P) \leq U(f, P)$ in general.

PROPOSITION 7.6 If P^* is a refinement of P then

$$L(f, P^*) \geq L(f, P)$$

$$U(f, P^*) \leq U(f, P)$$

Proof | Let $P = \{x_1, \dots, x_n\}$ and suppose that $P^* = \{x_1, \dots, x_i, x^*, x_{i+1}, \dots, x_n\}$, then

$$L(f, P^*) = \sum_{k=1}^i m_k \Delta x_k + m'(x^* - x_i) + m''(x_{i+1} - x^*) + \sum_{k=i+2}^n m_k \Delta x_k.$$

Therefore

$$\begin{aligned} L(f, P^*) - L(f, P) &= m'(x^* - x_i) + m''(x_{i+1} - x^*) - m_{i+1}(x_{i+1} - x_i) \\ &= (m'' - m_{i+1})(x_{i+1} - x^*) + (m' - m_{i+1})(x^* - x_i) \geq 0. \end{aligned}$$

The last inequality is due to the fact that the infimum over a subset is greater than or equal to the infimum of the whole set. This shows that $L(f, P^*) \geq L(f, P)$. The $U(f, p)$ one follows in a similar manner. ■

PROPOSITION 7.7 The following inequality holds:

$$\int_a^b f dx \leq \overline{\int_a^b f dx}$$

Proof | Let P_1, P_2 be partitions and let P^* be their common refinement. It follows that

$$L(f, P_1) \leq L(f, P^*) \leq U(f, P^*) \leq U(f, P_2)$$

Keeping P_2 fixed if we take the supremum over P_1 ,

$$\int_a^b f dx \leq U(f, P_2),$$

Now taking the infimum over P_2 we get our result. ■

THEOREM 7.8 f is Riemann integrable if and only if for all $\epsilon > 0$ there exists a partition P such that

$$U(f, P) - L(f, P) < \epsilon.$$

Proof | Suppose that $\forall \epsilon > 0$ there exists a P such that

$$U(f, P) - L(f, P) < \epsilon.$$

Since

$$L(f, P) \leq \int_a^b f dx \leq \overline{\int_a^b f dx} \leq U(f, P)$$

it follows that

$$\overline{\int_a^b f dx} - \int_a^b f dx < \epsilon, \forall \epsilon > 0$$

Thus

$$\overline{\int_a^b f dx} = \underline{\int_a^b f dx}.$$

Now suppose that f is integrable. Using the fact that $\int_a^b f dx$ is supremum and infimum of $L(f, P)$ and $U(f, P)$ respectively, it follows that for each ϵ

$$\begin{aligned} U(f, P_1) - \int_a^b f dx &< \epsilon/2 \\ \int_a^b f dx - L(f, P_2) &< \epsilon/2 \end{aligned}$$

For some partitions P_1 and P_2 . Let P be the common refinement of P_1 and P_2 . Then

$$U(f, P) - L(f, P) < \epsilon.$$

■

PROPOSITION 7.9 If f is continuous on $[a, b]$ then it is integrable.

Proof | Since f is continuous on $[a, b]$ it follows that it is also uniformly continuous. Let $\epsilon > 0$. There exists a δ independent of x, y such that

$$|x - y| < \delta \implies |f(x) - f(y)| < \frac{\epsilon}{b - a}$$

Let P be a partition such that $\Delta x_i < \delta$. Therefore

$$U(f, P) - L(f, P) = \sum_{k=1}^n (M_k - m_k) \Delta x_k$$

But since $\Delta x_k < \delta$ it follows that $M_k - m_k < \epsilon/(b - a)$. Thus

$$U(f, P) - L(f, P) < \epsilon.$$

■

PROPOSITION 7.10 If f is a monotonic on $[a, b]$ then f is integrable.

Proof | Suppose that f is increasing. Choose a partition such that $\Delta x_i = (b - a)/n$ for some n . Since the function is increasing $M_i = f(x_i)$ and $m_i = f(x_{i-1})$. Therefore

$$U(f, P) - L(f, P) = \sum_{k=1}^n (M_i - m_i) \Delta x_i = \frac{b - a}{n} (f(b) - f(a))$$

Since n is arbitrary it follows that $U(f, P) - L(f, P)$ can be made arbitrarily small.

■

THEOREM 7.11 If f is bounded on $[a, b]$ and it has only finitely many discontinuities then f is integrable.

Proof | Suppose that f has only one discontinuity at x_0 . Let $\epsilon > 0$. Let P_1 be a partition of $[a, x_0 - 1/n]$ and let P_2 be a partition of $[x_0 + 1/n, b]$. Since f is continuous in both of these intervals it follows that

$$\begin{aligned} U(f, P_1) - L(f, P_1) &< \epsilon/2 \\ U(f, P_2) - L(f, P_2) &< \epsilon/2 \end{aligned}$$


Let M, m be the supremum and infimum of f in the interval $[x_0 - 1/n, x_0 + 1/n]$. Let $P = P_1 \cup P_2$, then

$$\begin{aligned} U(f, P) - L(f, P) &= (U(f, P_1) - L(f, P_1)) + (M - m) \frac{2}{n} + (U(f, P_2) - L(f, P_2)) \\ &< \epsilon + \frac{2(M - m)}{n} \end{aligned}$$

Since n is arbitrary, taking the limit $n \rightarrow \infty$ we get that f is integrable. This argument can be generalised to finitely many discontinuities. ■

THEOREM 7.12 Suppose that f is integrable and $m \leq f \leq M$, and ϕ is continuous on $[m, M]$, and $h = \phi \circ f$. Then h is integrable.

Proof | skipped for now... ■

 It is easy to show that the Riemann integral satisfies the following

- 1) It is linear
- 2)

$$\int_a^b f dx = \int_a^c f dx + \int_c^b f dx$$

- 3) If f is Riemann integrable then $|f|$ is Riemann integrable as well and

$$\left| \int_a^b f dx \right| \leq \int_a^b |f| dx.$$

THEOREM 7.14 If f_n is uniformly convergent f and f_n are Riemann integrable then, f is Riemann integrable and

$$\lim_{n \rightarrow \infty} \int_a^b f_n dx = \int_a^b f dx.$$

PART IV
SET THEORY AND LOGIC

DEFINITION 1.1 A logic system is defined as being composed of:

- 1) A language, which is:
 - i) A collection of symbols.
 - ii) A grammar, i.e. a set of rules determining valid statements.
- 2) A collection of axioms.
- 3) Rules for inference.
- 4) A model, i.e. an assignment of truth value to valid statements in the language. We also require the assignment to be such that all axioms are true.

EXAMPLE 1.2 Propositional logic is an example of a logic system, defined as follows:

- 1) Symbols:
 - i) Letters: $P, Q, R, \dots, P_1, P_2, \dots$.
 - ii) $\wedge, \vee, \neg, \rightarrow, \leftarrow, \leftrightarrow, (,)$.
- 2) The valid forms in the language are:
 - i) Atomic forms: P, Q, R, \dots
 - ii) If p is a valid form then $\neg p$ is also a valid form.
 - iii) If p, q are valid forms then $(p) \wedge (q), (p) \vee (q), (p) \rightarrow (q), (p) \leftrightarrow (q)$ are also valid forms.
- 3) The axioms of propositional logic are:
 - i) (FL1) $p \rightarrow (q \rightarrow p)$.
 - ii) (FL2) $p \rightarrow (q \rightarrow r) \rightarrow (p \rightarrow q \rightarrow (p \rightarrow r))$.
 - iii) (FL3) $\neg p \rightarrow \neg q \rightarrow (q \rightarrow p)$.
- 4) A valid form q is *inferred* from p_1, \dots, p_n if q can be written whenever p_1, \dots, p_n . Denote this by $p_1, \dots, p_n \implies q$. The rules of inference are as follows in propositional logic:
 - i) $p \rightarrow q, p \implies q$ (Modus Ponens).
 - ii) $p \rightarrow q, \neg q \implies \neg p$ (Modus Tolens).
 - iii) $(p \rightarrow q) \wedge (r \rightarrow s), (p \vee r) \implies (q \vee s)$ (Constructive Dilemma).
 - iv) $(p \rightarrow q) \wedge (r \rightarrow s), \neg q \vee \neg s \implies \neg p \vee \neg r$ (Destructive Dilemma).
 - v) $p \vee q, \neg p \implies q$ (Disjunctive Syllogism).
 - vi) $p \rightarrow q, (q \rightarrow r) \implies p \rightarrow r$ (Hypothetical Syllogism).
 - vii) $p, q \implies p \wedge q$ (Conjunction).
 - viii) $p \wedge q \implies p$ (Simplification).
 - ix) $p \implies p \vee q$ (Addition).

Along with these we also have the rules of replacement:

- i) $p \wedge q \wedge r \iff p \wedge (q \wedge r)$ and $p \vee q \vee r \iff p \vee (q \vee r)$.
- ii) $p \wedge q \iff q \wedge p$ and $p \vee q \iff q \vee p$.
- iii) $p \wedge (q \vee r) = p \wedge q \vee p \wedge r$ and $p \vee (q \wedge r) = p \vee q \wedge p \vee r$.
- iv) $p \rightarrow q \iff \neg q \rightarrow \neg p$.
- v) $p \iff \neg(\neg p)$.
- vi) $\neg(p \wedge q) \iff \neg p \vee \neg q$ and $\neg(p \vee q) \iff \neg p \wedge \neg q$.
- vii) $p \wedge p \iff p$ and $p \vee p \iff p$.
- viii) $p \rightarrow q \iff \neg p \vee q$.
- ix) $p \leftrightarrow q \iff (p \rightarrow q) \wedge (q \rightarrow p)$.

- 5) The model in propositional logic assigns each atomic form a value $\{T, F\}$. Given two valid forms p, q every model must satisfy the usual truth table which is assigned to the propositions $p \wedge q, p \vee q, p \rightarrow q, \neg p$.



It is possible to replace the rules of inference only with MP. The resulting logical system is equivalent to propositional logic (shown in assignment 1).

The model part of a logic system is called *semantics*. Semantics is essentially assigning meaning to proposition. The inference rules, axioms, and replacement rules fall under *syntactics*.

1.1 Semantics of Propositional logic

DEFINITION 1.4 A tautology is a valid form which is true in all models. For example $q \wedge \neg q$. Similarly a contradiction is a statement which is false in all models of propositional logic. If a statement isn't a contradiction or tautology then it is called a contingency.

DEFINITION 1.5 Let $p_1, \dots, p_{\ell-1}, q$ be valid forms in propositional logic then:

- 1) If q is a tautology then write $\models q$.
- 2) We say that $p_1, \dots, p_{\ell-1}$ logically implies q if:

$$\models p_1, \dots, p_{\ell-1} \rightarrow q$$

- 3) When $p_1, \dots, p_{\ell-1}$ logically imply q we write:

$$p_1, \dots, p_{\ell-1} \models q$$

We call $p_1, \dots, p_{\ell-1}$ the premise and q the conclusion.

DEFINITION 1.6 Two statements are logically equivalent if $\models p \leftrightarrow q$.

PROPOSITION 1.7 All tautologies are logically equivalent, and all contradictions are logically equivalent.

1.2 Syntactics of Propositional logic

DEFINITION 1.8 A formal proof of q from the premise $p_1, \dots, p_{\ell-1}$ is a finite sequence of valid forms q_0, \dots, q_n such that:

- 1) q_i is either one of p_i ,
- 2) q_i is one of the axioms,
- 3) q_i follows from q_0, \dots, q_{i-1} using rules of inference/replacement rules.

If a formal proof of q exists from premise $p_1, \dots, p_{\ell-1}$ then we write $p_1, \dots, p_{\ell-1} \vdash q$.

DEFINITION 1.9 (Given in class) We say $p \implies q$ if $r \vdash p$ then $r \vdash q$.

THEOREM 1.10 $p \implies q$ if and only if $p \vdash q$.

Proof | Assuming $p \implies q$, it follows from definition that if $r \vdash p$ then $r \vdash q$. Thus the following proof sequence proves the forward implication: $p \vdash p, p \implies q, p \vdash q$. For the proof in other direction since $p \vdash q$ there is a sequence $p, q_0, \dots, q_{n-1}, q$. If $r \vdash p$ there is also a sequence

$r, p_0, \dots, p_{k-1}, p$. Thus the proof sequence $r, p_0, \dots, p_{k-1}, p, q_0, \dots, q_{n-1}, q$ is a proof from r to q . Thus $r \vdash q$, completing the proof. ■

DEFINITION 1.11 Converse of $p \rightarrow q$ is $q \rightarrow p$.

DEFINITION 1.12 Contrapositive of $p \rightarrow q$ is $\neg q \rightarrow \neg p$.

PROPOSITION 1.13 $p \rightarrow q \vdash \neg q \rightarrow \neg p$.

Proof |

$q_0 : p \rightarrow q$, given
 $q_1 : (p \rightarrow q) \rightarrow (\neg q \rightarrow \neg p)$, FL3
 $q_2 : \neg q \rightarrow \neg p$, MP.

Similarly it can be shown that the contrapositive implies the statement.

1.3 Proof Methods

LEMMA 1.14 If $\vdash q$ then $\vdash p \rightarrow q$.

Proof | If $\vdash q$ then there exists a sequence r_0, \dots, r_ℓ, q where r_i are either axioms or are inferred from r_0, \dots, r_{i-1} . Thus the sequence $r_0, \dots, r_\ell, q, q \vee \neg p, p \rightarrow q$ is a valid proof for $p \rightarrow q$. Thus $\vdash p \rightarrow q$. ■

THEOREM 1.15 (Deduction) $p \vdash q$ iff $\vdash p \rightarrow q$.

Proof | (Backward implication). If $\vdash p \rightarrow q$ then there exists a sequence $r_0, \dots, r_\ell, p \rightarrow q$ where r_i are either axioms or are inferred from rules of inference/replacement rules. Thus the sequence $r_0, \dots, r_\ell, p \rightarrow q, p, q$ is a valid proof of q given the premise p . Thus $p \vdash q$.

(Forward implication). Very long, refer to theorem 1.4.4 of O'Leary for proof. ■

COROLLARY 1.16 If $p_1, \dots, p_n, q \vdash r$ then $p_1, \dots, p_n \vdash q \rightarrow r$.

THEOREM 1.17 (Direct proof) If $p_1, \dots, p_n, q \vdash r$ then $p_1, \dots, p_n \implies q \rightarrow r$.

Proof | Using deduction we get $p_1, \dots, p_n \vdash q \rightarrow r$ and using the fact that \vdash and \implies are equivalent it follows that $p_1, \dots, p_n \implies q \rightarrow r$. ■

THEOREM 1.18 (Indirect Proof) $\neg q \rightarrow (p \wedge \neg p) \implies q$.

Proof |

$q_1 : \neg q \rightarrow (p \wedge \neg p)$, given
 $q_2 : \neg(p \wedge \neg p) \rightarrow \neg\neg q$, 1, contraposition
 $q_3 : \neg p \vee p \rightarrow q$, 2, DN
 $q_4 : p \rightarrow (p \rightarrow p)$, FL1
 $q_5 : \neg p \vee (\neg p \vee p)$, 4, Imp
 $q_6 : (\neg p \vee \neg p) \vee p$, 5, associativity
 $q_7 : \neg p \vee p$, 6, idem
 $q : q$, 3, 7 M.P.

Thus $\neg q \rightarrow (p \wedge \neg p) \vdash q$ which is equivalent to the claim. ■

1.4 Consistency

NOTATION 1.19 $p_0, \dots \vdash q$ if any subsequence $p_{i_1}, \dots, p_{i_k} \vdash q$. If no subsequence proves q then $p_0, \dots \not\vdash q$.

DEFINITION 1.20 A sequence of valid forms p_0, p_1, \dots is consistent if for every propositional form q , $p_0, p_1, \dots \not\vdash q \wedge \neg q$. We write this as $\text{Con}(p_0, \dots)$.

THEOREM 1.21 If p_0, \dots are valid forms then the following are equivalent:

- 1) $\text{Con}(p_0, \dots)$.
- 2) Any subsequence of p_0, \dots is consistent.
- 3) There exists a form p such that $p_0, \dots \not\vdash p$.

Proof | We prove that $1 \implies 2 \implies 3 \implies 1$.

- 1) Assume that there exists a subsequence p_{i_1}, \dots, p_{i_n} so that $p_{i_1}, \dots, p_{i_n} \vdash q \wedge \neg q$. Then it means that there is a subsequence such that a contradiction can be derived. Thus $\neg \text{Con}(p_0, \dots)$. Which is a contradiction. Thus every subsequence is consistent.
- 2) If every subsequence is consistent then there is no subsequence which derives the proposition of the form $q \wedge \neg q$. Thus $p_0, \dots \not\vdash q \wedge \neg q$.

Assume that p_0, \dots is not consistent. Then for all propositions q there exists a subsequence such that $p_{i_1}, \dots, p_{i_n} \vdash q \wedge \neg q$. Thus $p_{i_1}, \dots, p_{i_n}, r_0, \dots, r_\ell, q \wedge \neg q, q$ is a valid proof of q . Which shows that for all propositions q , $p_0, \dots \vdash q$. Which is a contradiction. ■

DEFINITION 1.22 A sequence p_0, \dots is maximally consistent if $\text{Con}(p_0, \dots)$ and for any $p \neq p_i$ we have $\neg \text{Con}(p, p_0, \dots)$.

Given any sequence which is not maximally consistent it is possible to construct a sequence which is maximally consistent by just adding all the implications of set in the set.

THEOREM 1.23 Every consistent sequence is a subsequence of maximally consistent sequence.

Proof | Let p_0, \dots be a consistent sequence and let q_0, q_1, \dots be the sequence of all propositional forms. Then define a new sequence in the following way:

$$r_{2k} = p_k, 0 \leq k$$

$$r_{2k+1} = \begin{cases} q_k, & \text{if } \text{Con}(q_k, r_0, \dots, r_{2k}, p_0, \dots) \\ p_k, & \text{otherwise} \end{cases}$$

Clearly p_0, p_1, \dots is a subsequence of r_0, r_1, \dots and that $\text{Con}(r_0, r_1, \dots)$ by definition. All that remains to show is that the sequence is maximal. Let q be some proposition, then $q = q_i$ for some $i \geq 0$. If $\text{Con}(q, r_0, \dots)$ then by construction $q = r_j$ for some $j \geq 0$ because it was added at step $2i + 1$. Thus if $q \neq r_i$ for some i then $\neg \text{Con}(q, r_0, \dots)$ (contrapositive of previous statement). ■

1.5 Soundness

DEFINITION 1.24 A logical system is sound if every theorem is a tautology.

LEMMA 1.25 All axioms of propositional logic are tautologies.

Proof | Easy to check using truth tables. ■

LEMMA 1.26 If $p \implies q$ then $p \rightarrow q$ is a tautology. Also if $p, q \implies r$ then $p \wedge q \rightarrow r$ is a tautology.

Proof | To show this we inotply have to check that all the rules of inference, and the rules of replcaement are tautologies. Since MP implies all the other rules of inference it is enough to check that MP, replacement rules are tautologies. This can be verified to be true by making the truth table for all of them. ■

LEMMA 1.27 If $p \rightarrow q$ and p are tautologies then q is a tautology.

Proof | If for any valuation function $v(p) = T$ and $v(p \rightarrow q) = T$ then from the truth table the only possibility is that $v(q) = T$, if v is a valid model. ■

THEOREM 1.28 (Soundness) If $\vdash p$ then $\models p$.

Proof | If $\vdash p$ then by definition there exists a proof sequence $q_0, q_1, \dots, q_n = p$ where each q_i is either an axiom, or $q_0, q_1, \dots, q_{i-1} \implies q_i$ using MP or replacement rules.

- 1) If q_i is an axiom then it is a tautology, by lemma 1.26.
- 2) If q_i is infered from q_0, \dots, q_{i-1} then also it is a tautology becauseof lemma 1.27.

Since $q_n = p$, it follows that p is a tautology. ■

COROLLARY 1.29 If $p_1, \dots, p_n \vdash q$ then $p_1, \dots, p_n \models q$.

Proof | This follows from soundness. The only additional part is that q_i 's in the proof sequence can now be one of p_i . But since we assume $v(p_i) = T$, it does not cause an issue in the proof. ■

COROLLARY 1.30 Propositional logic is consistent.

Proof | Since every theorem is a tautology, and $p \wedge \neg p$ is a contradiction it cannot be a theorem. Thus within propositional logic $\not\vdash p \wedge \neg p$. ■

1.6 Complete

DEFINITION 1.31 A logic system is complete if every tautology is a theorem.

LEMMA 1.32 If $\neg \text{Con}(\neg q, p_0, \dots)$ then $p_0, \dots \vdash q$.

Proof | If $\neg \text{Con}(p_0, \dots)$ then by theorem 1.21 it is possible to show that $p_0, \dots \vdash q$. Thus assume $\text{Con}(p_0, \dots)$. Since $\neg \text{Con}(\neg q, p_0, \dots)$ there exists some r such that:

$$\begin{aligned} \neg q, p_0, \dots &\vdash r \neg r, \text{ or,} \\ \neg q, p_{i_1}, \dots, p_{i_n} &\vdash r \neg r. \end{aligned}$$

$\neg q$ must show up in subsequence since the p_i 's are consistent. Thus there is a proof sequence: $p_{i_1}, \dots, p_{i_n}, \neg q, s_0, \dots, s_k, r \wedge \neg r$. By indirect proof one can show that the subproof $p_{i_1}, \dots, p_{i_n} \vdash q$ is valid. Thus $p_0, \dots \vdash q$. ■

LEMMA 1.33 If p_0, \dots is maximally consistent then for any q either $q = p_i$ or $\neg q = p_i$ for some $i \geq 0$.

Proof | Since p_0, \dots is consistent both q and $\neg q$ cannot be in the sequence. Thus assume that $\neg q$ is not. Thus by previous lemma we can show that since $\neg \text{Con}(\neg q, p_0, \dots)$ then $p_0, \dots \vdash q$. Since the sequence is maximal it must contain q . ■

HYPOTHESIS 1.34 (Induction hypothesis) Induction on propositional forms states that a property is true for all propositional forms if:

- 1) It is true for all atomic forms.
- 2) If it is true for p, q then it is true for $\neg p$ and $p \rightarrow q$. ($p \wedge q$, $p \vee q$ are not included here cause they can be expressed using \neg, \rightarrow).

In proving the later statement we assume that said property holds for p, q . This assumption is called the induction hypothesis.

LEMMA 1.35 If $\text{Con}(p_0, \dots)$ then there exists a valuation function v such that $v(p) = T$ if and only if $p = p_i$ for some i .

Proof | Since any consistent sequence can be extended to a maximally consistent one, let's assume that p_0, \dots is maximally consistent. Let X_0, \dots be sequence of all atomic forms. Then define the valuation function as follows:

$$v(X_i) = \begin{cases} T, & \text{if } X_i = p_j \text{ for some } j \geq 0 \\ F, & \text{otherwise} \end{cases}$$

Clearly, by construction, $v(X_j) = T$ iff $X_j = p_i$ for some i . Assume that $v(p) = T$ iff $p = p_i$ and $v(q) = T$ iff $q = p_i$ for i .

- 1) Assume $v(\neg q) = T$. Then $v(q) = F$. Thus by induction q is not in the list p_0, \dots . Thus $\neg q$ must be in the list (by previous lemma). Hence if $v(\neg q)$ then $\neg q = p_i$ for some i .
- 2) Conversely if $\neg q = p_i$ for some i then by consistency q is not in the sequence and hence by induction $v(q) = F$. Thus $v(\neg q) = T$. Hence $v(\neg q) = T$ iff $\neg q = p_i$ for some i .
- 3) Similarly case by case it can be shown that $v(p \rightarrow q) = T$ iff $p \rightarrow q = p_i$ for some i .

■

THEOREM 1.36 Propositional logic is complete.

Proof | Let's say that $\not\models p$. Then $\text{Con}(FL1, FL2, FL3, \neg p)$. It follows from the previous lemma that there exists a valuation function such that $v(p) = F$. Thus $\not\models p$ (we have proven the contrapositive). ■

2 FIRST ORDER LOGIC

2.1 Syntactics

DEFINITION 2.1 The symbols used are:

- 1) Variables: x, y, z, \dots
- 2) Constants: a, b, c, \dots
- 3) Quantifiers: \forall, \exists
- 4) Equals: $=$
- 5) Connectors: $\neg, \wedge, \vee, \rightarrow$.
- 6) Functions: $f(x_1, \dots, x_n)$.
- 7) Relations: $R(x_1, \dots, x_n)$.

DEFINITION 2.2 A term is:

- 1) either a variable
- 2) or a constant
- 3) or a function.

DEFINITION 2.3 A formula is:

- 1) $t_1 = t_2$ where t_i are terms
- 2) $R(x_1, \dots, x_n)$
- 3) If p is a formula then $\neg p$ is a formula
- 4) If p, q are formulas then any connector between them would be a valid formula.
- 5) $\forall x p$ and $\exists x p$ are formulas.

The rules for terms and formulas gives us the grammar of FOL.

DEFINITION 2.4 (Substitution for terms) If y is a variable then $y \frac{t}{x}$ is defined as:

$$y \frac{t}{x} \iff \begin{cases} t, & \text{if } y = x \\ y, & \text{otherwise.} \end{cases}$$

If c is a constant then

$$c \frac{t}{x} \iff c.$$

If f is a function then

$$f(x_1, \dots, x_n) \frac{t}{x} \iff f(x_1 \frac{t}{x}, \dots, x_n \frac{t}{x})$$

DEFINITION 2.5 Let t_i be terms and R be an relation. A variable is said to be free if:

- 1) A variable occurrence in $t_0 = t_1$ and $R(t_0, \dots, t_1)$ is free.
- 2) A variable occurrence of $\neg p$ is free if the occurrence is free in p .
- 3) A variable occurrence in $p \wedge q, p \vee q, p \rightarrow q$ is free if it is free in both p and q .
- 4) Any occurrence of x in $\forall x p$ and $\exists x p$ is bound.
- 5) Any occurrence of $x \neq y$ is free in $\forall y p$ and $\exists y p$ if the occurrence in p is free.

DEFINITION 2.6 A formula with no free variable is called a sentence.

DEFINITION 2.7 (Substitution in formulas) Let t_i be terms, R be a relation and p, q be valid formulas. Then:

- 1) $(t_0 = t_1) \frac{t}{x} \iff t_0 \frac{t}{x} = t_1 \frac{t}{x}$.

- 2) $R(t_0, \dots, t_n) \frac{t}{x} \iff R(t_0 \frac{t}{x}, \dots, t_n \frac{t}{x})$.
- 3) $(\neg p) \frac{t}{x} \iff \neg(p \frac{t}{x})$.
- 4) $(p \wedge q) \frac{t}{x} \iff p \frac{t}{x} \wedge q \frac{t}{x}$.
- 5) The rules for \forall, \rightarrow follow from the above two.
- 6) When the expression includes quantifier $Q \in \{\forall, \exists\}$:

$$(Qyp) \frac{t}{x} \iff \begin{cases} Qyp \frac{t}{x}, & \text{if } x \neq y \text{ and } y \text{ is not in } t \\ Qyp, & \text{otherwise.} \end{cases}$$

2.2 Axioms, Rules of Inference and Replacement rules

FOL is built on top of propositional logic in the sense that all sentences can be treated as propositional forms and therefore FL1, FL2, FL3, MP, and all the replacement rules are applicable in FOL. We only need a few more rules to deal with quantifiers, and equality.

AXIOM 2.8 The following are axiom schemas involving quantifiers:

- 1) $\forall x p \rightarrow p \frac{t}{x}$, where x can be substituted with t in p .
- 2) $\forall x (p \rightarrow q) \rightarrow \forall x p \rightarrow \forall x q$
- 3) $p \rightarrow \forall x p$ where x does not occur freely in p .
- 4) If ϕ is an axiom then $\forall x \phi$ is an axiom.

AXIOM 2.9 The following axioms are for $=$ symbol:

- 1) $x = x$.
- 2) $x = y \rightarrow (p \rightarrow p')$ where x occurs freely in p and p' is obtained by replacing any occurrence of x by y .

PROPOSITION 2.10 $x = y \rightarrow y = x$.

Proof | 1) $x = x$ (axiom)
 2) $x = y$ (given)
 3) $x = y \rightarrow (x = x \rightarrow y = x)$ (axiom)
 4) $x = x \rightarrow y = x$ (M.P on 2 and 3)
 5) $y = x$ (M.P. on 1 and 4)

PROPOSITION 2.11 $x = y, y = z \rightarrow x = z$.

Proof | 1) $y = z$ (axiom)
 2) $x = y$ (given)
 3) $x = y \rightarrow (y = z \rightarrow x = z)$ (axiom)
 4) $y = z \rightarrow x = z$ (M.P on 2 and 3)
 5) $x = z$ (M.P. on 1 and 4)

The only additional replacement rule is the Quantifier Negation:

PROPERTY 2.12 (Quantifier Negation) For any formula p ,

$$\neg \forall x p \iff \exists x \neg p$$

$$\neg \exists x p \iff \forall x \neg p$$

2.3 Proof Methods

Proof methods like Direct proof, Indirect proof (contradiction) are also valid in FOL. The additional proof methods are:

PROPOSITION 2.13 (Universal Generalization) If $\vdash p(a)$ then $\vdash \forall x p(x)$, where a is a some constant.

Proof | Suppose that $\vdash p(a)$. Then there is a sequence r_0, \dots, r_n such that each r_i :

- 1) Is either an axiom,
- 2) Or is inferred using M.P from some $r_j : p, r_k : t \rightarrow r_i$ where $j, k < i$,
- 3) Or it follows from a replacement rule on $r_j, j < i$.

Since proofs are of finite length, it is possible to find a new variable x which has not occurred in any of r_i . In the case where r_i is an axiom, $\forall x r_i$ is also an axiom. If r_i is derived from MP from t and $t \rightarrow r_i$ then using free generalization $t \rightarrow \forall x t$, and using universal MP $\forall x (t \rightarrow r_i) \rightarrow \forall x t \rightarrow \forall x r_i$. Thus using MP we get that $\forall x r_i$ is true whenever r_j, r_k is true. Therefore $\vdash \forall x p$. If r_i follows from a replacement rule, then the same replacement rule can be applied on $\forall x r_j$ to get $\forall x r_i$ since replacement rules only act on substrings. ■

PROPOSITION 2.14 (Existential Generalization) If $p_0, \dots \vdash q(c)$ then $p_0, \dots \vdash \exists x q(x)$, where p_i dont have any occurance of c .

Proof | Using direct proof we can write:

$$p_0, \dots \vdash q(c) \implies \vdash p_0, \dots \rightarrow q(c)$$

and

$$p_0, \dots \vdash \exists x q(x) \implies \vdash p_0, \dots \rightarrow \exists x q(x)$$

Using contraposition, it is enough to show that:

$$\text{if } \vdash (\neg q(c) \rightarrow \neg p_0 \wedge \dots) \text{ then } \vdash (\forall x \neg q(x) \rightarrow \neg p_0 \wedge \dots)$$

Thus it is enough to show that $\forall x \neg q(x) \vdash \neg q(c)$. This follows from universal instantiation. ■

PROPOSITION 2.15 (Existential Instantiation) If $q(c), p_0, \dots \vdash r$ then $\exists x q(x), p_0, \dots \vdash r$, where p_i do not have any occurance of c .

Proof | We need to show that:

$$q(c), p_0, \dots \vdash r \text{ then } \exists x q(x), p_0, \dots \vdash r$$

Using direct proof this can be written as:

$$q(c) \vdash p_0, \dots \rightarrow r \quad \text{then} \quad \exists x q(x) \vdash p_0, \dots \rightarrow r$$

It is enough to show that $\exists x q(x) \vdash q(c)$. Suppose that $\exists x q(x)$ does not infer $q(a)$ for any constant a . This means that for every a $\exists x q(x)$ infers $\neg q(a)$. Using universal generalization we get $\forall x \neg q(x)$, which is the same as $\neg \exists x q(x)$. This is a contradiction. This means that $\exists x q(x)$ infers $q(c)$ for some constant c . ■

PROPOSITION 2.16 (Proof by cases) Let $p \iff p_0 \vee \dots \vee p_n$. If $p_0 \rightarrow q, p_1 \rightarrow q, \dots, p_n \rightarrow q$ then $p \rightarrow q$.

Proof | This can be shown for two cases and then easily generalised.

$$\begin{aligned} (p_1 \rightarrow q) \wedge (p_2 \rightarrow q) &\iff (\neg p_1 \vee q) \wedge (\neg p_2 \vee q) \\ &\iff \neg p_1 \wedge \neg p_2 \vee q \\ &\iff \neg(p_1 \vee p_2) \vee q \\ &\iff p_1 \vee p_2 \rightarrow q. \end{aligned}$$
■

3 SET THEORY AS FIRST ORDER LOGIC

3.1 Zermelo Frankel Axioms

In first order logic of set theory every variable is considered to be a set. We introduce a new symbol \in , which is a connector between variables; $x \in y$ is interpreted semantically as " x is contained in y ".

AXIOM 3.1 (Extensionality) This axiom says that two sets are equal if all of their elements are equal.

$$\forall x \forall y (x = y \leftrightarrow \forall w (w \in x \leftrightarrow w \in y)).$$

AXIOM 3.2 (Empty Set) This axiom states that there exists a set which contains nothing else in it. Formally,

$$\exists y \forall x (\neg(x \in y)).$$

Let \emptyset be a witness to this axiom (meaning we instantiate it at \emptyset).

AXIOM 3.3 (Existence of pair) Given two sets x, y there exists a set containing exactly these two elements.

$$\forall x \forall y \exists z (\forall u (u \in z \leftrightarrow (u = x \vee u = y))).$$

Usually this set is denoted by $\{x, y\}$.

AXIOM 3.4 (Union of sets) Given a set x there is a set which is the union of all elements of x .

$$\forall x \exists y \forall z (z \in y \leftrightarrow \exists u (u \in x \wedge z \in u)).$$

This set is denoted as $\bigcup_{u \in x} u$.



Combining the axiom of unions with the axiom of pair: given x, y then pair $\{x, y\}$ exists and by axiom of union $x \cup y$ is also a set.

DEFINITION 3.6 We write $x \subset y$ if $\forall u (u \in x \rightarrow u \in y)$. If $x \subset y$ we say x is a subset of y .

AXIOM 3.7 (Existence of power set) Given any x there exists a set y which contains all subsets of x .

$$\forall x \exists y \forall z (z \in y \leftrightarrow z \subset x).$$

This set is generally denoted as $\mathfrak{P}(x)$.

AXIOM 3.8 (Restricted Comprehension) Given a set x and a relation $\phi(z)$, then the elements of x satisfying $\phi(z)$ forms a set.


$$\forall x \exists y \forall z (z \in y \leftrightarrow (z \in x \wedge \phi(z))).$$

Note that this is an axiom schema. Also note that that $\phi(z)$ has no free occurrence of x , otherwise the definition will be recursive. We denote this set as $\{z \in x \mid \phi(z)\}$.

DEFINITION 3.9 Suppose x, y are sets. Let $\phi(z)$ be the property $z \in x$. Then we define $x \cap y$ as $\{z \in y \mid z \in x\}$.

AXIOM 3.10 (Axiom of Regularity) This axiom says that every non-empty set x has atleast one $y \in x$ such that y does not contain any element of x . This axiom prohibits paradoxes like Russell's paradox which occur due to infinite descent. Formally

$$\forall x (\neg(x = \emptyset) \rightarrow \exists y (y \in x \wedge y \cap x = \emptyset)).$$

 Consider the set $\{x\}$. Since it only contains one element, by regularity axiom we have that $x \cap \{x\} = \emptyset$. This means that $\neg \exists y (y \in x \wedge y = x)$ (if this were true then the intersection would be non-empty), which further implies $\neg x \in x$. Thus axiom of regularity prohibits sets which contain themselves.

DEFINITION 3.12 We write $\exists! x \phi(x)$ if $\exists x (\phi(x) \wedge \forall w (\phi(w) \leftrightarrow w = x))$. This is semantically read as "there exists a unique x ".

AXIOM 3.13 (Axiom of Replacement) This axiom is about functions. It says that suppose for all $w \in x$ there exists a unique u such that the property $\phi(w, u)$ holds, then the collections of all such u is contained in a set.

$$\forall x (\forall w (w \in x \rightarrow \exists! u (\phi(w, u))) \rightarrow \exists y \forall u (u \in y \leftrightarrow \exists w (w \in x \wedge \phi(w, u))))$$

DEFINITION 3.14 Given a set x the singleton set $\{x\}$ is defined as an instance of $\exists y \forall z (z \in y \leftrightarrow z = x)$. Similarly the set $x \cup \{x\}$ is an instance of $\exists y \forall z (z \in y \leftrightarrow \forall w (w = x \vee w \in z))$. The set $x \cup \{x\}$ is called the successor of x and is denoted $S(x)$.

AXIOM 3.15 (Axiom of Infinity) There exists a set which consists infinitely many elements. Formally we define this set inductively, claiming that if x is in the set then so is the successor of x .

$$\exists y (\emptyset \in y \wedge \forall x (x \in y \rightarrow S(x) \in y))$$

3.2 Relations


DEFINITION 3.16 We define an ordered pair (x, y) to be the set $\{\{x\}, \{x, y\}\}$.

PROPOSITION 3.17 Suppose $(x, y) = (u, v)$ then $x = u$ and $y = v$.

proof by Kuratowski's trick | Since $\bigcap_{w \in (x, y)} w = \{x\}$ and similarly $\bigcap_{w \in (u, v)} w = \{u\}$, it follows that $\{x\} = \{u\}$ and thus $x = u$. Define

$$\delta(p) = \{z \in \bigcup_{w \in p} w \mid \neg(\bigcup_{w \in p} w = \bigcap_{w \in p} w) \rightarrow \neg(z \in \bigcap_{w \in p} w)\}$$

Clearly $\delta((x, y)) = \{y\}$ and $\delta((u, v)) = \{y\}$. Since $(x, y) = (u, v)$ it follows that $v = y$. ■

 Suppose that $u \in x$ and $v \in y$. Then $\{u\}, \{u, v\} \subset x \cup y$. Thus $\{u\}, \{u, v\} \in \mathfrak{P}(x \cup y)$ which further means that $(u, v) \in \mathfrak{P}(\mathfrak{P}(x \cup y))$.

DEFINITION 3.19 Given two sets x, y define

$$x \times y = \{z \in \mathfrak{P}(\mathfrak{P}(x \cup y)) \mid \exists u \exists v (u \in x \wedge v \in y \wedge z = (u, v))\}$$

DEFINITION 3.20 A relation r is a subset of $x \times y$. If $(u, v) \in r$ then we write urv .

DEFINITION 3.21 (Reflexive, Symmetric and Transitive) A relation $r \subset x \times x$ is said to be reflexive if $\forall u(uru)$. It is said to be symmetric if $\forall u \forall v (urv \rightarrow vru)$. It is said to be transitive if $\forall u \forall v \forall w (urv \wedge vrw \rightarrow urw)$. If r is all three then it is called an equivalent relation.

DEFINITION 3.22 An equivalence class w.r.t an equivalence relation $r \subset x \times x$ and $u \in x$ is defined as

$$[u]_r = \{v \in x \mid urv\}$$

PROPOSITION 3.23 If r is an equivalent relation then $\forall u \forall v (\neg([u]_r \cap [v]_r = \emptyset) \rightarrow [u]_r = [v]_r)$.

Proof | If $[u]_r \cap [v]_r \neq \emptyset$ then $\exists w(urw \wedge vrw)$. But since $vrw \rightarrow wrv$ and $urw \wedge wrv \rightarrow urv$, it follows that urv . Therefore $[u]_r = [v]_r$. ■

DEFINITION 3.24 The quotient of a set x w.r.t. an equivalence relation r is given by $x/r = \{z \in x \mid \exists u(u \in x \wedge z = [u]_r)\}$.

DEFINITION 3.25 A partial order is a relation r on a set x which is reflexive, transitive and anti-symmetric, i.e. $\forall u \forall v (urv \wedge vru \rightarrow u = v)$.


DEFINITION 3.26 A strict partial order r on a set x is transitive and $\forall u(\neg uru)$.

DEFINITION 3.27 A partial order is said to be a total order if $\forall u \forall v (urv \vee vru \vee v = u)$.

3.3 Functions

DEFINITION 3.28 A function is a relation f on $x \times y$ which satisfies the property

$$\forall u(u \in x \rightarrow \exists! v((u, v) \in f)).$$

 Note that due to the axiom of replacement we know that the collection of all values v such that $(u, v) \in f$ is also a set.

DEFINITION 3.30 Define

$$y^x = \{f \in \mathfrak{P}(x \times y) \mid \forall u(u \in x \rightarrow \exists! v((u, v) \in f))\}$$

DEFINITION 3.31 Given a function $f \in y^x$ and subset $z \subset x$ then we define the image of z as

$$f(z) = \{v \in y \mid \exists u(u \in z \wedge (u, v) \in f)\}.$$

Similarly we define the inverse image of a subset $z \subset y$ is defined as

$$f^{-1}(z) = \{u \in x \mid \exists v(v \in z \wedge (u, v) \in f)\}.$$

PART V
TOPOLOGY

1 TOPOLOGICAL SPACES AND CONTINUOUS FUNCTIONS

DEFINITION 1.1 A topological space is a pair (X, \mathcal{T}) where X is a set and $\mathcal{T} \subset \mathfrak{P}(X)$ such that:

- 1) Both X and \emptyset are in \mathcal{T} .
- 2) The union of elements of any subset of \mathcal{T} is in \mathcal{T} .
- 3) The intersection of elements of any finite subset of \mathcal{T} is in \mathcal{T} .

\mathcal{T} is said to be a topology on the set X .



Abuse of Notation By definition a topological space is (X, \mathcal{T}) , but for convinience of typing I will generally just say X is a topological space given that no confusion arises.

DEFINITION 1.3 An open set of a topological space X is a set which belongs to the topology \mathcal{T} .

DEFINITION 1.4 If (X, \mathcal{T}) and (X, \mathcal{T}') are topological spaces such that $\mathcal{T} \subset \mathcal{T}'$ then \mathcal{T}' is said to be a finer than \mathcal{T} . In case $\mathcal{T} \subsetneq \mathcal{T}'$ then \mathcal{T}' is said to be strictly finer. Also \mathcal{T} may be said to coarser than or strictly coarser than \mathcal{T}' respectively in the above two cases. \mathcal{T} is comparable to \mathcal{T}' if $\mathcal{T} = \mathcal{T}'$.

Generally it is very hard to completely specify every element in the topology. Thus we use the concept of a basis and specify the topology in terms of these basis.

DEFINITION 1.5 A basis of a topology on X is a collection, \mathcal{B} , of subsets of X such that:

- 1) $\forall x \in X \exists B \in \mathcal{B} (x \in B)$.
- 2) $\forall B_1, B_2 \in \mathcal{B} \forall x \in B_1 \cap B_2 \exists B_3 \in \mathcal{B} (x \in B_3 \wedge B_3 \subset B_1 \cap B_2)$.

PART VI
MEASURE THEORY

1 INTUITION OF MEASURE

Consider any interval of \mathbb{R} , $(a, b]$. Intuitively we define the length of this interval as $\lambda((a, b]) = b - a$. Now is it possible to extend this concept of length to any subset of \mathbb{R} ? For that we would wish to find a function λ such that the following properties are true:

- 1) $\lambda : \mathfrak{P}(\mathbb{R}) \rightarrow \mathbb{R}^+$ is a set function.
- 2) If $I = (a, b]$ is any interval in \mathbb{R} then $\lambda(I) = b - a$.
- 3) Length of union of two disjoint subsets $A, B \in \mathfrak{P}(\mathbb{R})$ must be the sum of their individual lengths, i.e. $\lambda(A \cup B) = \lambda(A) + \lambda(B)$. This can be extended to any countable union of pairwise disjoint subsets.
- 4) The translation of any subset $A \in \mathfrak{P}(\mathbb{R})$ must have the same length, i.e. $\forall x \in \mathbb{R}, \lambda(A + x) = \lambda(A)$.



CONJECTURE 1.1 The function λ as described above does not exist.

To prove this we must first prove some other propositions. We use the following notations: let \sim be an equivalence relation on \mathbb{R} defined as:

$$x \sim y \text{ if } x - y \in \mathbb{Q}$$

Let $[x]$ denote the equivalence classes of x , and let $\Lambda = \mathbb{R} / \sim$.

PROPOSITION 1.2 The set Λ is uncountable.

Proof | Let $\alpha \in \Lambda$ be an equivalence class. Let $x \in \alpha$ be a fixed point. Then for each $y \in \alpha$ it is possible to find a unique rational number given by $x - y$. Hence α is a countable set. Since the countable union of countable sets is countable, but \mathbb{R} is uncountable, it follows that Λ must be uncountable. ■

Let $\Omega \subset \mathbb{R}$ be set constructed in the following way: for each $\alpha \in \Lambda$ we know that a point can be found between $(0, 1)$; so take one such point from each α and put it in the set Ω . From this construction it is easy to see that $\Omega \subset (0, 1)$.

PROPOSITION 1.3 Let $p, q \in \mathbb{Q}$, then either $\Omega + q = \Omega + p$ or $\Omega + q \cap \Omega + p = \emptyset$.

Proof | Let's say $\Omega + q \cap \Omega + p \neq \emptyset$. Then for $a, b \in \Omega$ we can find an $x \in \Omega + q \cap \Omega + p$ such that $x = a + p = b + q$. Hence $a - b = q - p$ for all $a, b \in \Omega$. Since $q - p$ is rational, $a - b$ will also be rational. Hence a and b belong to the same equivalence class. But since we only chose one element from each equivalence class in the construction of Ω , we must have $a = b$. Hence $q = p$, making the two sets in question equal. ■

Hence from this proposition we can say that if $q \neq p$ then $\Omega + q \cap \Omega + p = \emptyset$.

PROPOSITION 1.4 Let λ be a length function as defined above. If $A \subset B \subset \mathbb{R}$ then $\lambda(A) \leq \lambda(B)$.

Proof | Since $B = A \cup (B - A)$, and A and $B - A$ are disjoint,

$$\begin{aligned} \lambda(B) &= \lambda(A \cup (B - A)) \\ &= \lambda(A) + \lambda(B - A) \end{aligned}$$

$$\geq \lambda(A)$$

Hence proving our claim. ■

Now we are ready to prove conjecture 1.1.

proof of conjecture 1.1 | Consider the union of sets $\Omega + q$:

$$\bigcup_{\substack{q \in \mathbb{Q} \\ -1 < q < 1}} \Omega + q$$

From proposition 1.3 we know that this is a union of disjoint sets. Hence,

$$\lambda \left(\bigcup_{\substack{q \in \mathbb{Q} \\ -1 < q < 1}} \Omega + q \right) = \sum_{\substack{q \in \mathbb{Q} \\ -1 < q < 1}} \lambda(\Omega + q)$$

Using property 4 of λ ,

$$\lambda \left(\bigcup_{\substack{q \in \mathbb{Q} \\ -1 < q < 1}} \Omega + q \right) = \sum_{\substack{q \in \mathbb{Q} \\ -1 < q < 1}} \lambda(\Omega) = 0$$

Let $x \in (0, 1)$. Let $a \in \Omega \cap [x]$. Then we know that $x - a = q$ for some rational q . Since $a \in \Omega$ implies $a \in (0, 1)$, the range of q must be $-1 < q < 1$. Hence $x = a + q$ for some $-1 < q < 1$ implying that

$$x \in \bigcup_{\substack{q \in \mathbb{Q} \\ -1 < q < 1}} \Omega + q$$

further implying that,

$$(0, 1) \subset \bigcup_{\substack{q \in \mathbb{Q} \\ -1 < q < 1}} \Omega + q$$

Using proposition 1.4,

$$1 \leq \lambda \left(\bigcup_{\substack{q \in \mathbb{Q} \\ -1 < q < 1}} \Omega + q \right)$$

Hence we have arrived at a contradiction. This shows that a function λ with the properties 1,2,3,4 as given above does not exist. ■

Hence this shows that to construct a general notion of length (called the *measure*) we must let go of one of the four properties: 1,2,3, or 4. Since 2,3,4 are essential for a notion of length, we change 1 to be the following:

1) $\lambda : \mathcal{B}(\subset \mathfrak{P}(\mathbb{R})) \rightarrow \mathbb{R}^+$ is a set function.

This means that we are discarding the notion that all subsets of \mathbb{R} can be assigned a length.

2 FORMAL NOTION OF MEASURE

DEFINITION 2.1 A class of subsets, \mathcal{A} , of a set Ω is said to be a semi-algebra if:

- 1) $\Omega \in \mathcal{A}$,
- 2) closed under finite intersections,
- 3) The compliment of any set in \mathcal{A} can be expressed as unions of finite pairwise disjoint sets in \mathcal{A} .

DEFINITION 2.2 A class of subsets, \mathcal{A} , of a set Ω is said to be an algebra if:

- 1) $\Omega \in \mathcal{A}$,
- 2) closed under finite intersections,
- 3) Closed under compliment.

DEFINITION 2.3 A class of subsets, \mathcal{A} , of a set Ω is said to be a σ -algebra if:

- 1) $\Omega \in \mathcal{A}$,
- 2) closed under countable intersections,
- 3) Closed under compliment.

PROPOSITION 2.4 Let Ω be a set and $\mathcal{A}_\alpha \subset \mathfrak{P}(\Omega)$ be algebras, where $\alpha \in I$ (no assumptions have been made on I). Then

$$\mathcal{A} = \bigcap_{\alpha \in I} \mathcal{A}_\alpha$$

is also an algebra.

Proof | Since $\Omega \in \mathcal{A}_\alpha, \forall \alpha \in I$, implies that $\Omega \in \mathcal{A}$. If $A_1, \dots, A_n \in \mathcal{A}$ then $A_1, \dots, A_n \in \mathcal{A}_\alpha$ for any $\alpha \in I$. Since \mathcal{A}_α is an algebra, it follows that $\bigcap_{j=1}^n A_j$ is in \mathcal{A}_α for any $\alpha \in I$; hence it is also in \mathcal{A} . If $A \in \mathcal{A}$ then it is in every \mathcal{A}_α and hence its compliment is in every \mathcal{A}_α . ■



The above proposition also applies to σ -algebras as well. Essentially the same argument applies, just that instead of finite sets we have countable intersection, i.e. $n \rightarrow \infty$. To denote that something applies to both algebras and σ -algebras we use the notation $(\sigma-)$ algebra.

DEFINITION 2.6 A class \mathcal{C} of subsets of set Ω is said to generate an $(\sigma-)$ algebra \mathcal{A} if $\mathcal{C} \subset \mathcal{A}$ and if for any $(\sigma-)$ algebra $\mathcal{A}' \supset \mathcal{C}$ implies that $\mathcal{A} \subset \mathcal{A}'$.

PROPOSITION 2.7 Every class $\mathcal{C} \subset \mathfrak{P}(\Omega)$ generates an $(\sigma-)$ algebra.

Proof | Let $\mathcal{A}_\alpha, \alpha \in I$ be all the $(\sigma-)$ algebras which contain the class \mathcal{C} . Then we know that,

$$\mathcal{A} = \bigcap_{\alpha \in I} \mathcal{A}_\alpha$$

is also an $(\sigma-)$ algebra, and it will contain \mathcal{C} . From the definition of intersection it follows that $\mathcal{A} \subset \mathcal{A}_\alpha$. Hence \mathcal{A} is the $(\sigma-)$ algebra generated by \mathcal{C} . ■

LEMMA 2.8 If \mathcal{S} is a semi-algebra and \mathcal{A} is the algebra generated by \mathcal{S} then

$$A \in \mathcal{A} \iff \exists \text{ pairwise disjoint } E_1, \dots, E_n \in \mathcal{S} \text{ such that } A = \bigcup_{j=1}^n E_j$$

Proof | (\Leftarrow) Assuming that A is finite union of disjoint sets $E_1, \dots, E_n \in \mathcal{S}$ we need to show that $A \in \mathcal{A}$. Since E_1, \dots, E_n are in \mathcal{S} it follows that they are also in \mathcal{A} . It further follows that the complement of each $E_j \in \mathcal{A}$. Since

$$\left(\bigcap_{j=1}^n E_j^c \right)^c = \bigcup_{j=1}^n E_j,$$

and algebras are closed under finite intersections, $A \in \mathcal{A}$.

(\Rightarrow) Let \mathcal{B} be the class defined as:

$$\mathcal{B} = \{B \mid \text{where } B = \bigcup_{j=1}^n F_j, F_j \in \mathcal{S} \text{ are pairwise disjoint.}\}$$

If we can show that \mathcal{B} is an algebra containing \mathcal{S} then by definition of generated algebras $\mathcal{A} \subset \mathcal{B}$. This shows that any element of \mathcal{A} can be expressed as a finite union of disjoint sets. Hence all that remains is to show that \mathcal{B} is an algebra containing \mathcal{S} .

- 1) Clearly by the definition, any element of \mathcal{S} is also in \mathcal{B} . Hence $\mathcal{S} \subset \mathcal{B}$ and hence $\Omega \in \mathcal{B}$.
- 2) Let $B_1, \dots, B_n \in \mathcal{B}$ then

$$\begin{aligned} \bigcap_{j=1}^n B_j &= \bigcap_{j=1}^n \bigcup_{i=1}^m F_{ji} \\ &= \bigcup_{i=1}^m \bigcap_{j=1}^n F_{ji}, \text{ using definition of } \mathcal{B} \\ &= \bigcup_{i=1}^m E_i, \text{ where, } E_i = \bigcap_{j=1}^n F_{ji} \end{aligned}$$

Since \mathcal{S} is closed under finite intersections, this shows that \mathcal{B} is closed under finite intersections.

- 3) Let $B \in \mathcal{B}$. Then,

$$\begin{aligned} B^c &= \left(\bigcup_{i=1}^m F_i \right)^c \\ &= \bigcap_{i=1}^m F_i^c \\ &= \bigcap_{i=1}^m \bigcup_{j=1}^n E_{ij} \text{ (using property 3 of semi-algebras)} \\ &= \bigcup_{j=1}^n E_j, \text{ where } E_j = \bigcap_{i=1}^m E_{ij} \end{aligned}$$


Since \mathcal{S} is closed under finite intersections, this shows that \mathcal{B} is closed under complement.

\mathcal{B} is indeed an algebra containing \mathcal{S} , hence completing our proof. ■

DEFINITION 2.9 Let \mathcal{C} be a class of subsets of Ω such that $\emptyset \in \mathcal{C}$, and let $\mu : \mathcal{C} \rightarrow \mathbb{R}^+$ be a function such that:

- 1) $\mu(\emptyset) = 0$,
- 2) If $E_1, \dots, E_n \in \mathcal{C}$ are pairwise disjoint and if $\bigcup_{j=1}^n E_j \in \mathcal{C}$ then $\mu(\bigcup_{j=1}^n E_j) = \sum_{j=1}^n \mu(E_j)$.

then μ is said to be an additive measure.


 Observe that if we have a $A \in \mathcal{C}$ such that $\mu(A) < \infty$ then:

$$\begin{aligned}\mu(A \cup \emptyset) &= \mu(A) + \mu(\emptyset) \\ \mu(A) &= \mu(A) + \mu(\emptyset) \\ \implies \mu(\emptyset) &= 0\end{aligned}$$

Hence the first condition is just a consequence of the second if a subset with finite measure exists. Secondly observe that if $E \subset F \in \mathcal{C}$ and $F - E \in \mathcal{C}$ then:

$$\mu(E \cup F - E) = \mu(F) = \mu(E) + \mu(F - E)$$

this means that $\mu(E) \leq \mu(F)$, the equality being true when $\mu(E) = \infty$. In the case where $\mu(E) < \infty$ we have the identity $\mu(F - E) = \mu(F) - \mu(E)$. This property is called monotonicity.

 Observe that if A, B are any sets in \mathcal{C} and $A \cup B \in \mathcal{C}$ then additivity implies that $\mu(A \cup B) \leq \mu(A) + \mu(B)$, since

$$\mu(A \cup B) = \mu(A \cup (B - A)) = \mu(A) + \mu(B - A) \leq \mu(A) + \mu(B). \text{ (using monotonicity)}$$

EXAMPLE 2.12 Let Ω be any non-empty set and let $X_1, X_2, \dots \in \Omega$. Also let $a_1, a_2, \dots \geq 0$ be some constants. Then define a measure $\mu : \mathcal{C} \subset (\mathfrak{P}(\Omega)) \rightarrow \mathbb{R}^+$ as:

$$\mu(A) = \sum_{j \geq 1} a_j 1\{X_j \in A\}$$

where,

$$1\{X_j \in A\} = \begin{cases} 1, & \text{if } X_j \in A \\ 0, & \text{if } X_j \notin A \end{cases}$$

It is easy to see that this measure is indeed additive.

DEFINITION 2.13 Let \mathcal{C} be a class of subsets of Ω such that $\emptyset \in \mathcal{C}$, and let $\mu : \mathcal{C} \rightarrow \mathbb{R}^+$ be a function such that:

- 1) $\mu(\emptyset) = 0$,
- 2) If $E_1, E_2, \dots \in \mathcal{C}$ are pairwise disjoint and if $\bigcup_{j \geq 1} E_j \in \mathcal{C}$ then $\mu(\bigcup_{j \geq 1} E_j) = \sum_{j \geq 1} \mu(E_j)$.

then μ is said to be a σ -additive measure.

EXAMPLE 2.14 Let $\Omega = (0, 1)$ and $\mathcal{C} = \{(a, b] \mid 0 \leq a < b < 1\} \cup \{\emptyset\}$. Define a function $\mu : \mathcal{C} \rightarrow \mathbb{R}^+$ as:

$$\mu(a, b] = \begin{cases} \infty, & \text{if } a = 0 \\ b - a, & \text{if } a \neq 0 \end{cases}$$

Clearly since a subset with finite measure exists $\mu(\emptyset) = 0$. Also since,

$$(a, b] = \bigcup_{j=1}^n (a_j, a_{j+1}], \text{ where } a_1 = a \text{ \& } a_n = b$$

when $a = 0$, $a_1 = 0$ and hence applying the measure on both sides we get ∞ . When $a \neq 0$, so are none of the a_j and hence:

$$\mu(a, b] = b - a = (a_2 - a_1) + \dots + (a_n - a_{n-1}) = \sum_{j=1}^n \mu(a_j, a_{j+1}]$$


Hence μ is additive. But it is possible to show that μ is not σ -additive. Consider for example the interval $(0, 1/2]$, and let $x_1 = 1/2, x_2, \dots$ be a monotonic decreasing sequence in $(0, 1)$ which converges to 0. Then

$$(0, 1/2] = \bigcup_{j \geq 1} (x_{j+1}, x_j]$$

Clearly $\mu(0, 1/2] = \infty$, but $\mu(x_{j+1}, x_j] = x_{j+1} - x_j$ which is finite.

DEFINITION 2.15 Let \mathcal{C} be a class of subsets of Ω and $\mu : \mathcal{C} \rightarrow \mathbb{R}^+$ be any set function. Then,

- 1) μ is said to be *continuous from below* at $E \in \mathcal{C}$ if $\forall (E_n)_{n \geq 1} \in \mathcal{C}, E_n \uparrow E \implies \lim \mu(E_n) = \mu(E)$.
- 2) μ is said to be *continuous from above* at $E \in \mathcal{C}$ if $\forall (E_n)_{n \geq 1} \in \mathcal{C}, E_n \downarrow E$ and $\exists n_0$ such that $\mu(E_{n_0}) < \infty$ implies that $\lim \mu(E_n) = \mu(E)$.

 If the condition of existence of n_0 such that $\mu(E_{n_0}) < \infty$ is removed then some unwanted cases arise. For example consider a measure on some class of \mathbb{R} . Consider the sequence of intervals $I_n = [n, \infty)$. Clearly $\bigcup_{n \geq 1} [n, \infty) = \emptyset$, but $\mu(\emptyset) = 0$ while $\mu(I_n) = \infty, \forall n$. This shows that no measure can be continuous from above on \mathbb{R} . This leads us to add the condition of existence of some set in the sequence which has finite measure.

LEMMA 2.17 Let \mathcal{A} be an algebra and let $\mu : \mathcal{A} \rightarrow \mathbb{R}^+$ be an additive measure, then:

- 1) μ is σ -additive $\implies \mu$ is continuous.
- 2) μ is continuous from below $\implies \mu$ is σ -additive.
- 3) μ is continuous from above at \emptyset and μ is a finite measure $\implies \mu$ is σ -additive.

Proof | 1) Assume μ is σ -additive, let $E \in \mathcal{C}$, and let $(E_n)_{n \geq 1} \in \mathcal{C}$ such that $E_n \uparrow E$. Let $F_1 = E_1$ and $F_n = E_n - E_{n-1}$. Clearly by this definition $\bigcup_{j \geq 1} F_j = \bigcup_{j \geq 1} E_j = E$. Then,

$$\mu \left(\bigcup_{j \geq 1} F_j \right) = \sum_{j \geq 1} \mu(F_j) = \lim_{n \rightarrow \infty} \sum_{j \geq 2}^n (\mu(E_j) - \mu(E_{j-1})) + \mu(E_1) = \lim \mu(E_n)$$

Hence μ is continuous from below.

For proving continuity from above, let $(E_n)_{n \geq 1} \in \mathcal{C}$ such that some $\mu(E_{n_0}) < \infty$ and $E_n \downarrow E$. Let $G_m = E_{n_0} - E_{n_0+m}$ be a sequence of sets, $\bigcup_{m \geq n_0} G_m = E_{n_0} - E$. Using the fact the μ is continuous from below,

$$\lim_{m \rightarrow \infty} \mu(G_m) = \mu(E_{n_0}) - \mu(E)$$

hence,

$$\begin{aligned} \lim_{m \rightarrow \infty} \mu(E_{n_0}) - \lim_{m \rightarrow \infty} \mu(E_{n_0+m}) &= \mu(E_{n_0}) - \mu(E) \\ \lim_{m \rightarrow \infty} \mu(E_{n_0+m}) &= \mu(E) \end{aligned}$$

This is the same as $\lim \mu(E_n) = \mu(E)$.

- 2) Assume that μ is continuous from below. Let $E \in \mathcal{C}$ be represented as the union of pairwise disjoint sets E_1, E_2, \dots . Let F_1, F_2, \dots be a sequence defined as:

$$F_k = \bigcup_{j=1}^k E_j$$

Clearly F_k is a sequence that converges to E from below. Using the fact that μ is additive and continuous from below:

$$\mu(E) = \lim_{n \rightarrow \infty} \mu(F_n) = \lim_{n \rightarrow \infty} \mu\left(\bigcup_{j=1}^n E_j\right) = \lim_{n \rightarrow \infty} \sum_{j=1}^n \mu(E_j) = \sum_{j \geq 1} \mu(E_j)$$

Hence μ is σ -additive.

- 3) Assume that μ is continuous from above at \emptyset . Let $A \in \mathcal{C}$ and let A_1, A_2, \dots be pairwise disjoint sets whose union is A . Define the sets E_1, E_2, \dots as

$$E_n = A - \bigcup_{j=1}^n A_j$$

Clearly $E_n \downarrow \emptyset$. Using finiteness, additivity, and continuity from above at \emptyset ,

$$\begin{aligned} \lim_{n \rightarrow \infty} \mu(E_n) &= 0 \\ \implies \lim_{n \rightarrow \infty} \mu\left(A - \bigcup_{j=1}^n A_j\right) &= 0 \\ \implies \mu(A) &= \sum_{j \geq 1} \mu(A_j) \end{aligned}$$

This completes the proof. ■

THEOREM 2.18 (Extension Theorem) Let \mathcal{S} be a semi-algebra, $\mu : \mathcal{S} \rightarrow \mathbb{R}^+$ be an additive measure, and let \mathcal{A} be the algebra generated by \mathcal{S} . Then there exists a $\nu : \mathcal{A} \rightarrow \mathbb{R}^+$, called the *extension* of μ , such that:

- 1) $\nu(A) = \mu(A)$, $\forall A \in \mathcal{S}$.
- 2) ν is additive.

In addition such a measure on \mathcal{A} is unique.

Proof | Let $\nu : \mathcal{A} \rightarrow \mathbb{R}^+$ be a function defined in the following way. Using lemma 2.8 we know that for any $A \in \mathcal{A}$ we can find disjoint $E_1, \dots, E_n \in \mathcal{S}$ such that $A = \bigcup_{j=1}^n E_j$; then define ν as,

$$\nu(A) = \sum_{j=1}^n \mu(E_j)$$

First we must show that ν is well defined, since there can be more than one sequence of pairwise disjoint sets whose union is A . Let E_1, \dots, E_n and F_1, \dots, F_m be two sequences of pairwise disjoint sets in \mathcal{S} whose union is A . Then,

$$\nu(A) = \sum_{j=1}^n \mu(E_j)$$

and

$$\nu(A) = \sum_{j=1}^m \mu(F_j).$$

Since $A = \bigcup_{k=1}^m F_k$

$$\begin{aligned} \implies E_j &= \bigcup_{k=1}^m F_k \cap E_j \\ \implies \mu(E_j) &= \sum_{k=1}^m \mu(F_k \cap E_j) \end{aligned}$$

Hence,

$$\nu(A) = \sum_{j=1}^n \sum_{k=1}^m \mu(F_k \cap E_j)$$

Similarly it can shown that,

$$\mu(F_j) = \sum_{k=1}^n \mu(E_k \cap F_j)$$

and therefore

$$\sum_{j=1}^m \mu(F_j) = \sum_{j=1}^n \mu(E_j).$$

Hence ν is well defined.

Clearly for $A \in \mathcal{S}$ we have $\nu(A) = \mu(A)$. For additivity, let A_1, A_2, \dots, A_n be pairwise disjoint sets in \mathcal{A} whose union is A . Again from lemma 2.8 for each $A_j = \bigcup_{k=1}^{n_j} E_{jk}$ where $E_{j1}, \dots, E_{jn_j} \in \mathcal{S}$ are pairwise disjoint. Let F_1, \dots, F_N , where $N = \sum_{j=1}^n n_j$, be defined as $F_1 = E_{11}, F_2 = E_{12}$ and so on. Then,

$$\begin{aligned} A &= \bigcup_{j=1}^N F_j \\ \implies \nu(A) &= \sum_{j=1}^N \mu(F_j) = \sum_{j=1}^n \sum_{k=1}^{n_j} \mu(E_{jk}) \end{aligned}$$

since

$$\nu(A_j) = \sum_{k=1}^{n_j} \mu(E_{jk}),$$

$$\implies \nu(A) = \sum_{j=1}^n \nu(A_j)$$

Therefore ν is additive.

For uniqueness, let's assume that two such functions ν_1 and ν_2 exist. From property 1 we know that $\nu_1(A) = \nu_2(A) \forall A \in \mathcal{S}$. Let $A \in \mathcal{A}$ and let $A_1, \dots, A_n \in \mathcal{S}$ be pairwise disjoint with union A . Then using additivity

$$\nu_1(A) = \sum_{j=1}^n \nu_1(A_j) = \sum_{j=1}^n \nu_2(A_j) = \nu_2(A)$$

This completes the proof. ■



This is theorem can be easily generalised for σ -additive measures. The only change in the proof would be considering countably many A_j in the proof of additivity.

3 CARATHEODORY THEOREM

Until now we have shown that extension ν of σ -additive measure μ on semi-algebra \mathcal{S} is also σ -additive on algebra \mathcal{A} generated by \mathcal{S} . The goal of this section is to show that the extension $\pi : \mathcal{F} \rightarrow \mathbb{R}^+$, where \mathcal{F} is the σ -algebra generated by \mathcal{S} is σ -additive and unique. In order to do this we follow the following steps:

- 1) Define a $\pi^* : \mathfrak{P}(\Omega) \rightarrow \mathbb{R}^+$ and show that it is something called an *outer measure*.
- 2) Define a class $\mathcal{M} \subset \mathfrak{P}(\Omega)$, and show that it is a σ -algebra.
- 3) Show that $\mathcal{A} \subset \mathcal{M}$. This has the implication that $\mathcal{F} \subset \mathcal{M}$.
- 4) Show that $\pi^*|_{\mathcal{M}}$ is σ -additive and $\pi^*|_{\mathcal{A}} = \nu$. Hence $\pi^*|_{\mathcal{M}}$ is an extension.
- 5) Finally show that this extension is unique.

DEFINITION 3.1 Let $A \subset \Omega$ for some set Ω . Then the collection $\{E_i \subset \Omega \mid i \geq 1\}$ is said to be a covering of A if $A \subset \bigcup_{i \geq 1} E_i$. Note that at least one covering exists for every subset and that is $\{\Omega\}$.

DEFINITION 3.2 Let $\pi^* : \mathfrak{P}(\Omega) \rightarrow \mathbb{R}^+$ for some set Ω defined in the following way: let $A \subset \Omega$ and let $\{E_i \in \mathcal{A} \mid i \geq 1\}$ be a covering of A then

$$\pi^*(A) = \inf_{\{E_i\}} \sum_{i \geq 1} \nu(E_i)$$

This is to be read as infimum of $\sum_{i \geq 1} \nu(E_i)$ over all coverings of A which are in the algebra \mathcal{A} .

DEFINITION 3.3 Let \mathcal{C} be a class of subsets of Ω such that $\emptyset \in \mathcal{C}$, and let $\mu : \mathcal{C} \rightarrow \mathbb{R}^+$ be a function such that:

- 1) $\mu(\emptyset) = 0$,
- 2) μ is monotone, i.e. $E \subset F$ where $E, F \in \mathcal{C} \implies \mu(E) \leq \mu(F)$,
- 3) μ is sub-additive, i.e. $E \in \mathcal{C}$ and $\{E_i \in \mathcal{C} \mid i \geq 1\}$ is a covering of E then $\mu(E) \leq \sum_{i \geq 1} \mu(E_i)$.

Then μ is said to be an outer measure.

PROPOSITION 3.4 The function π^* as defined above is an outer measure.

Proof | Since $\emptyset \subset \Omega$, and it is a subset of every possible covering, clearly for the covering $\{E_i = \emptyset \mid \forall i \geq 1\}$,

$$\sum_{i \geq 1} \nu(E_i) = 0$$

and hence $\pi^*(\emptyset) = 0$.

Let $E \subset F$ where $E, F \in \mathcal{C}$. Let $\{F_j \mid j \geq 1\}$ be a covering of F . Since $E \subset F$ any covering of F is also a covering of E . If $E_j = F \cap F_j$ then $E_j \subset F_j$ and $\bigcup_{j \geq 1} E_j = F \cap E = E$. Hence $\{E_j\}$ is a covering of E . Since ν is a σ -additive measure,

$$\nu(E_j) \leq \nu(F_j) \text{ and hence, } \sum_{i \geq 1} \nu(E_i) \leq \sum_{i \geq 1} \nu(F_i).$$

Since for every covering of F a covering of E can be constructed in the above manner such that the above inequality is true, hence π^* is monotone.

For sub-additivity let $E \subset \Omega$ and let $\{E_i \in \mathcal{A} \mid i \geq 1\}$ be a covering of E . In the case when $\pi^*(E_i) = \infty$, clearly $\pi^*(E) \leq \pi^*(E_i)$. In the case when $\pi^*(E_i) < \infty \forall i \geq 1$, for each $\epsilon > 0$ we can find a covering of E_i , say $\{F_{ij} \in \mathcal{A} \mid j \geq 1\}$ such that,

$$\pi^*(E_i) \leq \sum_{j \geq 1} \nu(F_{ij}) \leq \pi^*(E_i) + \frac{\epsilon}{2^i}$$

Hence,

$$\sum_{i \geq 1} \pi^*(E_i) \leq \sum_{j \geq 1} \nu\left(\bigcup_{i \geq 1} F_{ij}\right) = \sum_{i \geq 1} \nu(E_i) \leq \sum_{i \geq 1} \pi^*(E_i) + \epsilon$$

Using $\pi^*(E) \leq \sum_{i \geq 1} \nu(E_i)$, we get

$$\pi^*(E) \leq \sum_{i \geq 1} \nu(E_i) \leq \sum_{i \geq 1} \pi^*(E_i) + \epsilon$$

Since this is true for arbitrary ϵ we conclude that π^* is sub-additive. ■

DEFINITION 3.5 A set $A \subset \Omega$ is said to be measurable if $\forall E \subset \Omega$,

$$\pi^*(E) = \pi^*(E \cap A) + \pi^*(E \cap A^c)$$

Define \mathcal{M} to be the set of all measurable subsets of Ω .



Using sub-additivity of π^* it is possible to prove that

$$\pi^*(E) \leq \pi^*(E \cap A) + \pi^*(E \cap A^c)$$

since $E = (E \cap A) \cup (E \cap A^c)$. Hence showing that A is measurable just boils down to showing

$$\pi^*(E) \geq \pi^*(E \cap A) + \pi^*(E \cap A^c)$$

PROPOSITION 3.7 The algebra \mathcal{A} of subsets of Ω is a subset of \mathcal{M} .

Proof | Let $A \in \mathcal{A}$ and let $E \in \Omega$. If we can show that A is measurable, we prove the proposition. Let $\{E_i \in \mathcal{A}\}$ be a covering of E , and let $\epsilon > 0$. In the case when $\pi^*(E_i) = \infty$ even for a single i , it is clear that the inequality

$$\pi^*(E) \geq \pi^*(E \cap A) + \pi^*(E \cap A^c) \tag{3.1}$$

holds. In the case when $\pi^*(E_i) < \infty$ for all $i \geq 1$ let $\epsilon > 0$. Then

$$\pi^*(E) \leq \sum_{i \geq 1} \nu(E_i) \leq \pi^*(E) + \epsilon$$

Since $E \cap A \subset \bigcup_{i \geq 1} E_i \cap A$,

$$\pi^*(E \cap A) \leq \sum_{i \geq 1} \nu(E_i \cap A)$$

Using similar arguments for A^c

$$\pi^*(E \cap A^c) \leq \sum_{i \geq 1} \nu(E_i \cap A^c)$$

Adding these two inequalities

$$\pi^*(E \cap A^c) + \pi^*(E \cap A) \leq \sum_{i \geq 1} \nu(E_i)$$

Here I have used the additivity of ν since $E_i \cap A$ and $E_i \cap A^c$ are in the algebra. Using inequality (1):

$$\pi^*(E \cap A^c) + \pi^*(E \cap A) \leq \sum_{i \geq 1} \nu(E_i) \leq \pi^*(E) + \epsilon$$

Since this is true for arbitrary ϵ , we have

$$\pi^*(E \cap A^c) + \pi^*(E \cap A) \leq \pi^*(E)$$

Hence we have shown that A is measurable, completing the proof. ■

PROPOSITION 3.8 \mathcal{M} is a σ -algebra.

Proof | Since every algebra is a subset of \mathcal{M} clearly $\Omega \in \mathcal{M}$. Also it is easy to see that if $A \in \mathcal{M}$ then $A^c \in \mathcal{M}$ since replcaing A by A^c in the condition of measurable set does not change the inequality. The only condition that remains to be checked is closure under countable union. First consider the finite case. Let $A, B \in \mathcal{M}$. We are required to show that $\forall E \subset \Omega$,

$$\pi^*(E) \geq \pi^*(E \cap (A \cup B)) + \pi^*(E \cap (A \cup B)^c).$$

Since

$$\begin{aligned} \pi^*(E) &= \pi^*(E \cap A) + \pi^*(E - A), \\ \pi^*(E) &= \pi^*(E \cap B) + \pi^*(E - B) \end{aligned}$$

Thus

$$\begin{aligned} \pi^*(E - A) &= \pi^*((E - A) \cap B) + \pi^*((E - A) - B) \\ &= \pi^*(E \cap A^c \cap B) + \pi^*(E - (A \cup B)^c), \end{aligned}$$

implying that

$$\begin{aligned} \pi^*(E) &= \pi^*(E \cap A) + \pi^*(E \cap A^c \cap B) + \pi^*(E - (A \cup B)^c) \\ &\geq \pi^*(E \cap A \cup B) + \pi^*(E - (A \cup B)^c). \end{aligned}$$

The final inequality comes from the sub-additivity of π^* and the fact that $(E \cap A) \cup (E \cap A^c \cap B) = E \cap A \cup B$. Hence $A \cup B \in \mathcal{M}$. Now extending this to the countable case, let $A_j \in \mathcal{M}$, $A = \bigcup_{j \geq 1} A_j$, and $B_n = \bigcup_{j=1}^n A_j$. Using closure under finite unions we can say that

$$\pi^*(E) = \pi^*(E \cap B_n) + \pi^*(E - B_n)$$

Since $B_n \subset A \implies E - B_n \supset E - A$. Hence,

$$\pi^*(E) \geq \pi^*(E \cap B_n) + \pi^*(E - A)$$

Define the sets $F_1 = A_1, \dots, F_j = A_j - B_{j-1}, \dots$; and observe that $F_j \in \mathcal{M}$, $A = \bigcup_{j \geq 1} F_j$, and that these sets are pairwise disjoint. If we define $G_n = \bigcup_{j=1}^n F_j$, then using a similar logic as B_n

$$\pi^*(E) \geq \pi^*(E \cap G_n) + \pi^*(E - A).$$

Using induction one can show that

$$\pi^*(E \cap \bigcup_{j=1}^n F_j) = \sum_{j=1}^n \pi^*(E \cap F_j).$$

For $n = 1$ it is obviously true. Assuming it to be true for some n ,

$$\begin{aligned} \pi^*(E \cap \bigcup_{j=1}^{n+1} F_j) &= \pi^*(E \cap \bigcup_{j=1}^{n+1} F_j \cap F_{n+1}) + \pi^*(E \cap \bigcup_{j=1}^{n+1} F_j \cap F_{n+1}^c) \\ &= \pi^*(E \cap F_{n+1}) + \pi^*(E \cap \bigcup_{j=1}^n F_j) \\ &= \pi^*(E \cap F_{n+1}) + \sum_{j=1}^n \pi^*(E \cap F_j) \\ &= \sum_{j=1}^{n+1} \pi^*(E \cap F_j). \end{aligned}$$

Hence using this property,

$$\pi^*(E) \geq \pi^*(E \cap G_n) + \pi^*(E - A) = \sum_{j=1}^n \pi^*(E \cap F_j) + \pi^*(E - A).$$

Taking the limit $n \rightarrow \infty$,

$$\pi^*(E) \geq \sum_{j \geq 1} \pi^*(E \cap F_j) + \pi^*(E - A) \geq \pi^*(E \cap A) + \pi^*(E - A).$$

The final inequality comes from sub-additivity of π^* . This completes the proof that \mathcal{M} is a σ -algebra. ■



Since the algebra \mathcal{A} is a subset of \mathcal{M} and \mathcal{M} is a σ -algebra it follows that \mathcal{M} contains the σ -algebra generated by \mathcal{A} .

PROPOSITION 3.10 $\pi^*(A) = \nu(A) \forall A \in \mathcal{A}$.

Proof | Let $A \in \mathcal{A}$. Consider the covering $\{A_1 = A, A_j = \emptyset \ j \geq 2\}$. Then $\pi^*(A) \leq \nu(A)$ by definition. The opposite inequality can be proved by constructing sets $F_1 = E_1, F_n = E_n - \bigcup_{j=1}^{n-1} E_j$, where $\{E_n \in \mathcal{A}\}$ is some covering of A . As discussed in the previous proof F_j are pairwise disjoint. Since,

$$\begin{aligned} A &\subset \bigcup_{j \geq 1} F_j \\ \implies A &= \bigcup_{j \geq 1} F_j \cap A \\ \implies \nu(A) &= \nu\left(\bigcup_{j \geq 1} F_j \cap A\right) \\ \implies \nu(A) &= \sum_{j \geq 1} \nu(F_j \cap A) \leq \sum_{j \geq 1} \nu(E_j) \end{aligned}$$

The last inequality comes from the fact that $F_j \cap A \subset E_j$. This inequality shows that $\nu(A)$ is infact the infimum of the sum over all coverings of A . Hence $\pi^*(A) = \nu(A)$. ■

PROPOSITION 3.11 $\pi^*|_{\mathcal{M}}$ is σ -additive.

Proof | It is clear that $\pi(\emptyset) = 0$. Let $A_1, A_2, \dots \in \mathcal{M}$ be pairwise disjoint sets and let their union be A . Since \mathcal{M} is a σ -algebra $A \in \mathcal{M}$. Since we have already shown that for any pairwise disjoint sets $F_1, F_2, \dots \in \mathcal{M}$ and any $E \subset \Omega$

$$\pi^*(E \cap \bigcup_{j=1}^n F_j) = \sum_{j=1}^n \pi^*(E \cap F_j).$$

Letting $E = A$ and $F_j = A_j$,

$$\pi^*(\bigcup_{j=1}^n A_j) = \pi^*(A \cap \bigcup_{j=1}^n A_j) = \sum_{j=1}^n \pi^*(A \cap A_j) = \sum_{j=1}^n \pi^*(A_j).$$

Since $\bigcup_{j=1}^n A_j \subset \bigcup_{j \geq 1} A_j$, using monotonicity of π^*

$$\pi^*(A) \geq \sum_{j=1}^n \pi^*(A_j).$$

Taking the limit

$$\pi^*(A) \geq \sum_{j \geq 1} \pi^*(A_j).$$

Since using sub-additivity we already know that

$$\pi^*(A) \leq \sum_{j \geq 1} \pi^*(A_j)$$

it follows that π^* acting on \mathcal{M} is σ -additive. ■

DEFINITION 3.12 A set Ω is said to be σ -finite with respect to a function μ if there exists a sequence $E_1, E_2, \dots \subset \Omega$, such that $E_j \uparrow \Omega \implies \mu(E_j) < \infty$.

DEFINITION 3.13 A class $\mathcal{G} \subset \mathfrak{P}(\Omega)$ is said to be a monotone class if all monotonic sequences of sets converge in \mathcal{G} .

PROPOSITION 3.14 If \mathcal{G}_α where $\alpha \in I \subset \mathbb{R}$ are monotone classes then the intersection $\bigcap_{\alpha \in I} \mathcal{G}_\alpha$ is also a monotone class.

Proof | If A_1, A_2, \dots is any monotone sequence in $\bigcap_{\alpha \in I} \mathcal{G}_\alpha$ then it is in all \mathcal{G}_α and hence converge in all \mathcal{G}_α . ■



Using this proposition one can define the smallest monotone class generated by some class \mathcal{C} as the intersection of all the monotone classes containing \mathcal{C} .

LEMMA 3.16 Let \mathcal{A} be any algebra of subsets of Ω , \mathcal{G} be the monotone class generated by \mathcal{A} , and let \mathcal{F} be the σ -algebra generated by \mathcal{A} . Then $\mathcal{G} = \mathcal{F}$.

Proof | Let $A_j \in \mathcal{F}$ monotonically increase (decrease) to A ; since the countable union (intersection) of A_j is in \mathcal{A} so A must also be in \mathcal{F} . Since $\mathcal{A} \subset \mathcal{F}$ it follows that $\mathcal{G} \subset \mathcal{F}$.

All that remains to be shown is $\mathcal{F} \subset \mathcal{G}$. If we can show that \mathcal{G} is an algebra, then for any $A_1, A_2, \dots \in \mathcal{G}$ let $B_i = \bigcup_{j=1}^i A_j \in \mathcal{G}$ (since if \mathcal{G} is an algebra it will be closed under finite union), it follows that $\bigcup_{j \geq 1} B_j = \bigcup_{j \geq 1} A_j \in \mathcal{G}$ (using the fact that \mathcal{G} is monotone class). To prove that \mathcal{G} is an algebra define $\mathcal{M}(A) \in \mathcal{G}$ the class:

$$\mathcal{M}(A) = \{M \in \mathcal{G} \mid A - M, M - A, A \cap M \in \mathcal{G}\}.$$

Clearly $\mathcal{M}(A) \subset \mathcal{G}$. Let $E_1 \subset E_2 \subset \dots \in \mathcal{M}(A)$ converge to some E . Then $E - A = \bigcup_{i \geq 1} (E_i - A)$, but since by definition $E_i - A \in \mathcal{G}$ and $E_i - A \subset E_{i+1} - A$ it follows that $E - A \in \mathcal{G}$ (since \mathcal{G} is monotone class). Similarly it can be shown that $A - E$ and $A \cap E$ are in \mathcal{G} , and thus $E \in \mathcal{M}(A)$. The same argument can be used to show that if $E_1 \supset E_2 \supset \dots \in \mathcal{M}(A)$ converges to E then $E \in \mathcal{M}(A)$, hence concluding that $\mathcal{M}(A)$ is a monotone class. $\mathcal{M}(A)$ is also symmetric in the sense that if $A \in \mathcal{M}(B) \iff B \in \mathcal{M}(A)$, because if $A - B, B - A, A \cap B \in \mathcal{G}$ then both $A \in \mathcal{M}(B)$ and $B \in \mathcal{M}(A)$.

Let $A, B \in \mathcal{A}$ then we know that $A - B, B - A, A \cap B \in \mathcal{A}$. Hence $B \in \mathcal{M}(A)$ for all $A, B \in \mathcal{A}$, implying that $\mathcal{A} \subset \mathcal{M}(A) \forall A \in \mathcal{A}$. Since \mathcal{G} is the smallest monotone class containing \mathcal{A} , $\mathcal{G} \in \mathcal{M}(A) \forall A \in \mathcal{A}$. Since $M \in \mathcal{A}$ for all $M \in \mathcal{G}$ using the symmetry it implies that $A \in \mathcal{M}(M)$. Hence $\mathcal{A} \in \mathcal{M}(M)$. Therefore $\mathcal{G} = \mathcal{M}(M)$ for all $M \in \mathcal{G}$. This means that \mathcal{G} is closed under finite difference and intersection, proving that it is an algebra. This completes the proof. ■

THEOREM 3.17 (Uniqueness of Extension) Let $\mu_1, \mu_2 : \mathcal{F} \rightarrow \mathbb{R}^+$ be σ -additive functions, where \mathcal{F} is the σ -algebra generated by algebra \mathcal{A} of a set Ω which is σ -finite with respect to μ_1 and μ_2 (with the additional condition that the finite sequence exists in \mathcal{A}), be such that $\mu_1|_{\mathcal{A}} = \mu_2|_{\mathcal{A}}$ then $\mu_1 = \mu_2$.

Proof | Let $E_1, E_2, \dots \in \mathcal{A}$ be the sequence such that $\mu_1(E_n) < \infty$ and $\mu_2(E_n) < \infty$ for all n and $E_n \uparrow \Omega$. This sequence is guaranteed by the σ -finiteness of Ω . Define $\mathcal{B}_n = \{E \in \mathcal{F} \mid \mu_1(E \cap E_n) = \mu_2(E \cap E_n)\}$. Clearly $\mathcal{B}_n \subset \mathcal{F}$. If $E \in \mathcal{A}$ then $E \cap E_n \in \mathcal{A}$ and since $\mu_1|_{\mathcal{A}} = \mu_2|_{\mathcal{A}}$ it follows that $\mathcal{A} \subset \mathcal{B}_n$. Let $A_1, A_2, \dots \in \mathcal{B}$ be a sequence monotonically converging to some A . Since

$$\begin{aligned} \mu_1(A_j \cap E_n) &= \mu_2(A_j \cap E_n) \\ \implies \mu_1(A \cap E_n) &= \mu_2(A \cap E_n) \end{aligned}$$

where we have used lemma 2.17 and the finiteness of E_n in case of continuity from above. It follows that \mathcal{B}_n is a monotone class. Since it contains the algebra \mathcal{A} as well it must contain the monotone class generated by \mathcal{A} . Using lemma 3.16 we can conclude that $\mathcal{F} \subset \mathcal{B}_n$ and hence $\mathcal{B}_n = \mathcal{F}$. Let $A \in \mathcal{F}$ then

$$\begin{aligned} \lim_{n \rightarrow \infty} \mu_1(A \cap E_n) &= \lim_{n \rightarrow \infty} \mu_2(A \cap E_n) \\ \implies \mu_1(A) &= \mu_2(A) \end{aligned}$$

Therefore the extension is unique. ■

As a result of this theorem we can conclude that the function $\pi^* : \mathcal{F} \rightarrow \mathbb{R}^+$ is a σ -additive extension of the σ -additive measure $\nu : \mathcal{A} \rightarrow \mathbb{R}^+$ on the σ -algebra \mathcal{F} generated by the algebra \mathcal{A} and is uniquely determined. This is known as Caratheodory theorem. The formal statement of this theorem is:

THEOREM 3.18 (Caratheodory Theorem) Let \mathcal{A} be an algebra, $\nu : \mathcal{A} \rightarrow \mathbb{R}^+$ be a σ -additive measure, and \mathcal{F} be the σ -algebra generated by \mathcal{A} . Then there exists a unique σ -additive measure $\pi : \mathcal{M} \rightarrow \mathbb{R}^*$ such that $\pi|_{\mathcal{A}} = \nu$. Explicitly this measure is given by restricting the outer measure $\pi^* : \mathfrak{P}(\Omega) \rightarrow \mathbb{R}^+$,

$$\pi^*(A) = \inf_{\{E_j \in \mathcal{A}\}} \sum_{j \geq 1} \nu(E_j), \text{ where } \{E_j\} \text{ is a covering of } A$$

on \mathcal{M} ; i.e. $\pi = \pi^*|_{\mathcal{M}}$.


4 LEBESGUE MEASURE

In this section we define a σ -additive measure on a class of subsets \mathbb{R} and formalise the notion of length of subsets of \mathbb{R} . The procedure to do this is as follows:

- 1) Construct a semi-algebra \mathcal{S} and a σ -additive measure $\mu : \mathcal{S} \rightarrow \mathbb{R}^+$.
- 2) Use theorem 2.18 to construct a σ -additive measure ν on the algebra \mathcal{A} generated by \mathcal{S} .
- 3) Use the Carathéodory theorem to determine the σ -measure on \mathcal{F} , the σ -algebra generated by \mathcal{A} .

DEFINITION 4.1 Let $\mathcal{S} = \{\emptyset, \mathbb{R}, (a, b], (a, \infty), (-\infty, b]\}$. It is easy to check that \mathcal{S} is a semi-algebra of subsets of \mathbb{R} . Let $F : \mathbb{R} \rightarrow \mathbb{R}$ be a non-decreasing function. Then define $\mu_F : \mathcal{S} \rightarrow \mathbb{R}^+$ as:

$$\begin{aligned}\mu_F(\emptyset) &= 0, & \mu_F(\mathbb{R}) &= F(\infty), & \mu_F((a, b]) &= F(b) - F(a), \\ \mu_F((a, \infty)) &= F(\infty) - F(a), & \mu_F((-\infty, b]) &= F(b) - F(-\infty).\end{aligned}$$

 Observe that if we construct a function $G : \mathbb{R} \rightarrow \mathbb{R}$ given by $G(x) = \lim_{n \rightarrow \infty} F(x_n)$ when $-\infty < x < \infty$ and $G(\pm\infty) = F(\pm\infty)$, where $x_n \downarrow x$. It is easy to verify that G is non-decreasing, right continuous and $\mu_G(A) = \mu_F(A)$, $\forall A \in \mathcal{S}$. Hence without loss of generalization it is fair to assume that F is always right continuous. Also observe that μ_F is monotone.

PROPOSITION 4.3 μ_F is a σ -additive measure.

Proof | By definition we have that $\mu_F(\emptyset) = 0$. Consider the interval $(a, b] = \bigcup_{j=1}^n (a_j, b_j]$, where $(a_j, b_j]$ are pairwise disjoint. Then it is always possible to reindex the intervals such that $b_j = a_{j+1}$ when $j < n$, $a_{n+1} \equiv b_n = b$, and $a_1 = a$ (this uses both the fact that the intervals are disjoint and their union is $(a, b]$). Since,

$$\begin{aligned}F(b) - F(a) &= F(a_{n+1}) - F(a_1) \\ &= F(a_{n+1}) - F(a_2) + F(a_2) - F(a_1) \\ &= (F(a_{n+1}) - F(a_n)) + \dots + (F(a_2) - F(a_1)) \\ &= \sum_{j=1}^n F(a_{j+1}) - F(a_j)\end{aligned}$$

This sum can again be reindexed such that

$$F(b) - F(a) = \sum_{j=1}^n F(b_j) - F(a_j)$$

This implies that

$$\mu_F((a, b]) = \sum_{j=1}^n \mu_F((a_j, b_j])$$

Hence μ_F is additive. Now consider the case when $(a, b] = \bigcup_{j=1}^k (a_j, b_j]$ where $(a_j, b_j]$ are pairwise disjoint. Using monotonicity and additivity of μ_F

$$\mu_F((a, b]) \geq \mu_F\left(\bigcup_{j=1}^k (a_j, b_j]\right) = \sum_{j=1}^k \mu_F((a_j, b_j]).$$

Taking the limit $k \rightarrow \infty$,

$$\mu((a, b]) \geq \sum_{j \geq 1} \mu_F((a_j, b_j]).$$

All that remains to be proven is that the \leq inequality. To prove this fix an $\epsilon > 0$. Choose a $c > a$ such that $F(c) - F(a) < \epsilon$, choose $d_j > b_j$ such that $F(d_j) - F(b_j) < \epsilon/2^j$ and $[c, b] \subset \bigcup_{j \geq 1} (a_j, d_j]$ (such choices are possible since F is continuous from the right). Using Heine-Borel theorem, since $[c, b]$ is closed and bounded and $\{(a_j, d_j)\}$ forms an open cover of $[c, b]$, there exists a finite subcover $\{(a_j, d_j) \mid j \leq k\}$ of $[c, b]$. Without loss of generality we can assume that $c \in (a_1, d_1)$ and $b \in (a_k, d_k)$. Since,

$$\mu_F((a, b]) = F(b) - F(a) < F(b) - F(c) + \epsilon = \mu_F((b, c]) + \epsilon$$

Then using the monotonicity of μ_F ,

$$\begin{aligned} \mu_F((c, b]) &\leq \mu_F\left(\bigcup_{j=1}^k (a_j, d_j]\right) \\ &\leq \sum_{j=1}^k \mu_F((a_j, d_j]) \\ &\leq \sum_{j=1}^k F(d_j) - F(a_j) \\ &< \sum_{j=1}^k F(b_j) - F(a_j) + \epsilon \\ &< \sum_{j \geq 1} F(b_j) - F(a_j) + \epsilon. \end{aligned}$$

Thus for any $\epsilon > 0$

$$\sum_{j \geq 1} \mu_F((a_j, b_j]) \leq \mu_F((a, b]) < \sum_{j \geq 1} \mu_F((a_j, b_j]) + 2\epsilon$$

It is easy to show that this is true also for intervals $(-\infty, b]$ and (a, ∞) . This proves σ -additivity. ■

Using the extension theorem and then Carathéodory theorem the function $\mu_F^* : \mathcal{F} \rightarrow \mathbb{R}^+$

$$\mu_F^*(A) = \inf \left\{ \sum_j \mu_F(A_j) \mid A_j \in \mathcal{A} \text{ \& } A \subset \bigcup_{j \geq 1} A_j \right\}$$

is a unique σ -additive extension of μ on the σ -algebra \mathcal{M}_{μ^*} (which is the set of measurable functions w.r.t. μ^*). The measure space $(\mathbb{R}, \mathcal{M}_{\mu^*}, \mu_F^*)$ is called Lebesgue-Stieltjes measure space. In the case when $F(a) = a$ and hence $\mu((a, b]) \equiv \mu_F((a, b]) = b - a$, $(\mathbb{R}, \mathcal{M}_{\mu^*}, \mu^*)$ is called the Lebesgue measure space.

CONVENTION From now on we refer to $(\Omega, \mathcal{F}, \mu)$ a measure space if Ω is some set, \mathcal{F} is some σ -algebra containing Ω , and μ is a σ -additive measure. From now we also adopt the convention of calling σ -additive measures as just measures.

5 COMPLETE MEASURES

DEFINITION 5.1 A measure space $(\Omega, \mathcal{F}, \mu)$ is said to be *complete* if $A \in \mathcal{F}$, $\mu(A) = 0$ and $E \subset A$ imply that $E \in \mathcal{F}$.

DEFINITION 5.2 Let $(\Omega, \mathcal{F}, \mu)$ measure space and $A \in \mathcal{F}$ such that $\mu(A) = 0$. Then subsets of A are said to be *negligible sets*.

PROPOSITION 5.3 If $(\Omega, \mathcal{F}, \mu)$ is a measure space and \mathcal{F}' is defined as

$$\mathcal{F}' = \{A \cup N \mid A \in \mathcal{F} \text{ \& } N \subset E \in \mathcal{F} \text{ where } \mu(E) = 0\}.$$

Then \mathcal{F}' is a σ -algebra.

Proof | Let $A \in \mathcal{F}$ and $E = \emptyset$ (implying that $\mu(E) = 0$) then $A \cup N = A$ where $N \subset E$, hence $A \in \mathcal{F}'$. It further follows that $\mathcal{F} \subset \mathcal{F}'$. This means that $\Omega \in \mathcal{F}'$.

Let $A \in \mathcal{F}'$. Then $A = E \cup N$ where $E \in \mathcal{F}$ and $N \subset H$ such that $\mu(H) = 0$. One can then write $A^c = E^c \cap N^c = (E^c \cap H^c) \cup (E^c \cap (H - N))$. Clearly $E^c \cap H^c \in \mathcal{F}$ and $E^c \cap (H - N) \subset H - N \subset H$. Hence $A^c \in \mathcal{F}'$.

Let $A_1, A_2, \dots \in \mathcal{F}'$. Let $A_j = E_j \cup N_j$ where $E_j \in \mathcal{F}$, $N_j \subset H_j$ and $H_j \in \mathcal{F}$ such that $\mu(H_j) = 0$. Then

$$\bigcup_{j \geq 1} A_j = \left(\bigcup_{j \geq 1} E_j \right) \cup \left(\bigcup_{j \geq 1} N_j \right)$$

Since $\bigcup_{j \geq 1} E_j \in \mathcal{F}$, $\bigcup_{j \geq 1} N_j \subset \bigcup_{j \geq 1} H_j$ and $\mu(\bigcup_{j \geq 1} H_j) = \sum_{j \geq 1} \mu(H_j) = 0$, it follows that $\bigcup_{j \geq 1} A_j \in \mathcal{F}'$. ■

DEFINITION 5.4 Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $\mathcal{F}' \supset \mathcal{F}$ be a σ -algebra as defined in proposition 5.3. Then define $\mu' : \mathcal{F}' \rightarrow \mathbb{R}^+$ as follows. If $A \in \mathcal{F}'$ and $A = E \cup N$ where $A \in \mathcal{F}$ and N is a negligible set then

$$\mu'(A) = \mu(E)$$

PROPOSITION 5.5 μ' is a unique, σ -additive extension of μ .

Proof | Let $E \cup N = F \cup M$ where $E, F \in \mathcal{F}$ and $N \subset H, M \subset H'$ where $H, H' \in \mathcal{F}$ and $\mu(H) = \mu(H') = 0$. Clearly $E \subset E \cup N = F \cup M \subset F \cup H'$. Using monotonicity we have $\mu(E) \leq \mu(F)$. Similarly it can be shown that $\mu(F) \leq \mu(E)$, implying that $\mu(E) = \mu(F)$. This shows that $\mu'(E \cup N) = \mu(E) = \mu(F) = \mu'(F \cup M)$. Hence μ' is well defined.

Clearly $\mu'(\emptyset) = 0$. Let $A_1, A_2, \dots \in \mathcal{F}'$ be pairwise disjoint, and let their representation be $A_j = E_j \cup N_j$ where $E_j \in \mathcal{F}$ and N_j are negligible. By definition $\mu'(A_j) = \mu(E_j)$. Also since A_j are pairwise disjoint so will be E_j . Thus:

$$\mu'\left(\bigcup_{j \geq 1} A_j\right) = \mu'\left(\bigcup_{j \geq 1} E_j \cup \bigcup_{j \geq 1} N_j\right)$$

$$\begin{aligned}
&= \mu\left(\bigcup_{j \geq 1} E_j\right) \\
&= \sum_{j \geq 1} \mu(E_j) \\
&= \sum_{j \geq 1} \mu'(A_j)
\end{aligned}$$

Hence μ' is σ -additive.

Let $A \in \mathcal{F}$. Then clearly $\mu'(A) = \mu'(A \cup \emptyset) = \mu(A)$. Thus μ' is an extension of μ . To prove that this is a unique extension let $\mu_1, \mu_2 : \mathcal{F}' \rightarrow \mathbb{R}^+$ be σ -additive functions such that $\mu_1(A) = \mu_2(A) = \mu(A) \forall A \in \mathcal{F}$. Then for some $E \cup N$, where $E, H \in \mathcal{F}$, $N \subset H$, and $\mu(H) = 0$:

$$\mu_1(E \cup N) \leq \mu_2(E \cup H) = \mu_2(E) \leq \mu_2(E \cup N)$$

Similarly it can be shown that $\mu_1(E \cup N) \geq \mu_2(E \cup N)$. Hence μ' is also unique. ■

PROPOSITION 5.6 The measure space $(\Omega, \mathcal{F}', \mu')$ is complete.

Proof | If $A \in \mathcal{F}'$, $\mu'(A) = 0$, $A = F \cup M$ where $F \in \mathcal{F}$ and M is negligible. Then $\mu(F) = \mu'(A) = 0$. Thus we could simply make the choice $M \subset F$ and hence $A = F$. Therefore shown that if $\mu'(A) = 0$ then $A \in \mathcal{F}$. Let $E \subset F$. Then we can simply represent E as $\emptyset \cup E$. Since $\emptyset \in \mathcal{F}$ and $E \subset A \in \mathcal{F}$ where $\mu(A) = 0$, it follows that $E \in \mathcal{F}'$. ■

PROPOSITION 5.7 Let $(\Omega, \mathcal{M}, \pi^*|_{\mathcal{M}})$ be the measure space as defined in theorem 3.18. This is a complete measure space.

Proof | Let $B \in \mathcal{M}$, $\pi^*(B) = 0$, and $A \subset B$. For any $F \subset \Omega$,

$$F \cap A \subset A \subset B,$$

hence $\pi^*(F \cap A) \leq \pi^*(B) = 0$. Since $F \cup A^c \subset B \implies \pi^*(F \cup A^c) \leq \pi^*(B)$. Adding these two inequalities we get:

$$\pi^*(F) \geq \pi^*(F \cap A) + \pi^*(F \cap A^c).$$

Thus $A \in \mathcal{M}$. ■

6 INTEGRATION

In this section we would like to formulate the concept of integral in the context of measure spaces. In the Riemann integral we first partition the domain and then approximate then integral to be:

$$\int f \approx \sum_{k \geq 1} y_k (x_k - x_{k-1})$$

Using a similar concept we would later define an integral, called the Lebesgue integral, where we partition the y -axis, take the inverse of that interval to get a set in the domain, and then use a measure to find the "length" of this interval. Then we approximate the integral as follows:

$$\int f \approx \sum_{k \geq 1} y_k \mu(f^{-1}(A_k))$$

See fig. 2 to understand this better. But in order to define this integral the function must have the property that its inverse belongs to the σ -algebra on which the measure μ is defined. For this a new class of functions known as measurable functions is defined which hold this property.

DEFINITION 6.1 Let (X, \mathcal{T}) be a topological space. Then the σ -algebra generated by the open sets of this space is called the *Borel σ -algebra*, and represented $\mathcal{B}(X, \mathcal{T})$.

In the case when $X = \mathbb{R}^n$ and \mathcal{T} is the usual topology on \mathbb{R} , the Borel σ -algebra is simply denoted $\mathcal{B}(\mathbb{R}^n)$.

DEFINITION 6.2 Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, and $f : \Omega \rightarrow \mathbb{R}$ be a function Ω then f is said to be \mathcal{F} -*measurable*, or simply measurable, if $B \in \mathcal{B}(\mathbb{R}) \implies f^{-1}(B) \in \mathcal{F}$.

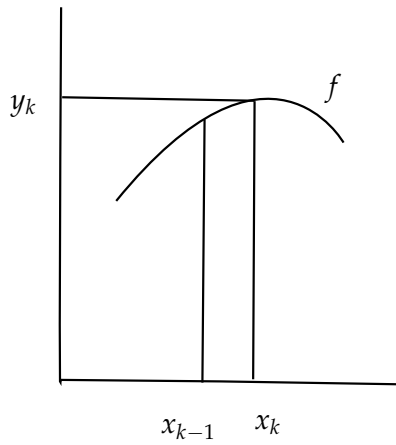
In general if $(\Omega_1, \mathcal{F}_1, \mu_1)$ and $(\Omega_2, \mathcal{F}_2, \mu_2)$ are measure spaces then $f : \Omega_1 \rightarrow \Omega_2$ is said to be $\langle \mathcal{F}_1, \mathcal{F}_2 \rangle$ -measurable if $A \in \mathcal{F}_2 \implies f^{-1}(A) \in \mathcal{F}_1$. Hence \mathcal{F} -measurable functions are just $\langle \mathcal{F}, \mathcal{B}(\mathbb{R}) \rangle$ -measurable.

LEMMA 6.3 Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f : \Omega \rightarrow \mathbb{R}$. Then f is measurable $\iff f^{-1}((-\infty, x]) \in \mathcal{F}$.

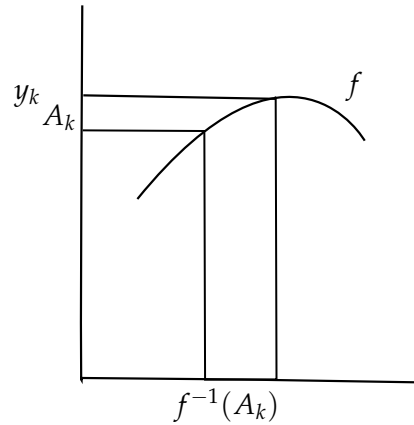
Proof | Before we begin proving this lemma, note that $\mathcal{B}(\mathbb{R})$ is the σ -algebra generated by the usual topology \mathcal{T} on \mathbb{R} . Let $\mathcal{C} = \{(-\infty, x] \mid x < \infty\}$ and \mathcal{G} be the σ -algebra generated by \mathcal{C} . Since for any $(-\infty, x]$ the sequence of elements $(-\infty, x_n) \in \mathcal{B}(\mathbb{R})$ where $x_n \downarrow x$ satisfies $(-\infty, x] = \bigcap_{j \geq 1} (-\infty, x_j]$. This shows that $(-\infty, x] \in \mathcal{B}(\mathbb{R})$ (using closure under countable intersection), further implying that $\mathcal{G} \subset \mathcal{B}(\mathbb{R})$. Similarly it can be shown that any open set (x, y) can be expressed as countable unions and intersections of elements of \mathcal{G} , implying that $\mathcal{B}(\mathbb{R}) \subset \mathcal{G}$. Hence $\mathcal{B}(\mathbb{R}) = \mathcal{G}$.

(\implies) If f is measurable then $A \in \mathcal{B}(\mathbb{R}) \implies f^{-1}(A) \in \mathcal{F}$, and since $(-\infty, x] \in \mathcal{B}(\mathbb{R})$ this implies that $f^{-1}((-\infty, x]) \in \mathcal{F}$.

(\impliedby) Let $\mathcal{M} = \{A \in \mathcal{B}(\mathbb{R}) \mid f^{-1}(A) \in \mathcal{F}\}$. Since $\mathbb{R} \in \mathcal{B}(\mathbb{R})$ and $f^{-1}(\mathbb{R}) = \{\omega \mid f(\omega) \in \mathbb{R}\} = \Omega \implies \Omega \in \mathcal{M}$. Also observing that f^{-1} preserves countable unions and compliments, if $A, A_1, \dots \in \mathcal{M}$ then $f^{-1}(A^c) = (f^{-1}(A))^c \in \mathcal{F}$ and $f^{-1}(\bigcup_{j \geq 1} A_j) = \bigcup_{j \geq 1} f^{-1}(A_j)$. This shows that \mathcal{M} is closed under compliments and countable unions. Hence \mathcal{M} is a σ -algebra. Since we are showing the backward implication we assume that $\mathcal{C} \subset \mathcal{M}$. Since \mathcal{M} is a σ -algebra containing \mathcal{C} it follows that $\mathcal{B}(\mathbb{R}) \subset \mathcal{M}$. Since by definition of \mathcal{M} it is a subset of $\mathcal{B}(\mathbb{R})$ further follows that $\mathcal{M} = \mathcal{B}(\mathbb{R})$. Thus proving the lemma. ■



(a) Riemann integral



(b) Lebesgue integral

Figure 2: Visualization of the integral

💡 Note that the key to proving the lemma really was showing that $\mathcal{B}(\mathbb{R}) \subset \mathcal{M}$, which required the fact that $\mathcal{B}(\mathbb{R})$ was the σ -algebra generated by \mathcal{C} . Hence this theorem can be easily extended to any class \mathcal{C} which generates the σ -algebra $\mathcal{B}(\mathbb{R})$, in the following sense:

LEMMA 6.5 Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, \mathcal{C} be a class of subsets of \mathbb{R} which generates the σ -algebra $\mathcal{B}(\mathbb{R})$, and $f : \Omega \rightarrow \mathbb{R}$. Then f is measurable $\iff f^{-1}(A) \in \mathcal{F}$ where $A \in \mathcal{C}$.

Hence lemma 6.3 also holds if $(-\infty, x]$ is replaced with one of $(-\infty, x)$, (x, ∞) , or $[x, \infty)$.

DEFINITION 6.6 Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, $E_1, \dots, E_n \in \mathcal{F}$ be pairwise disjoint sets such that $\Omega = \bigcup_{j=1}^n E_j$, and 1_{E_j} be the indicator function of E_j . Then a *simple function* $f : \Omega \rightarrow \mathbb{R}$ is a function which can be written as:

$$f(\omega) = \sum_{j=1}^n c_j 1_{E_j}(\omega)$$

where $c_j \in \mathbb{R}$.

PROPOSITION 6.7 Simple functions are measurable.

Proof | Let $f : \Omega \rightarrow \mathbb{R}$ be a simple function expressed as:

$$f(\omega) = \sum_{j=1}^n c_j 1_{E_j}(\omega)$$

The set $f^{-1}((-\infty, x]) = \{\omega \mid f(\omega) \leq x\}$ should belong to \mathcal{F} for f to be measurable by lemma 6.3. Notice that $f(\omega) \leq x$ only when $\omega \in E_j$ such that the corresponding $c_j \leq x$. Hence $f^{-1}((-\infty, x]) = \bigcup_{j \mid c_j \leq x} E_j$. Since each $E_j \in \mathcal{F}$ so will be any finite union. Hence $f^{-1}((-\infty, x]) \in \mathcal{F}$. ■

DEFINITION 6.8 (Integral of Non-negative Simple functions) Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and let f be a non-negative simple function of the form

$$f(\omega) = \sum_{j=1}^n c_j 1_{E_j}(\omega), \quad c_j \geq 0$$

Then we define:

$$\int f = \sum_{j=1}^n c_j \mu(E_j)$$

The non-negative condition was applied to avoid cases like the following: $\mu(E_1) = \mu(E_2) = \infty$ and $c_1 = -c_2$. Then the sum on the RHS would have $\infty - \infty$, which is not well defined.

PROPOSITION 6.9 The integral of non-negative simple function is well defined.

Proof | Let $f : \Omega \rightarrow \mathbb{R}$, $\{E_1, \dots, E_n\}, \{F_1, \dots, F_2\} \in \mathcal{F}$ be two partitions of Ω , and f be represented as:

$$f(\omega) = \sum_{j=1}^n c_j 1_{E_j} = \sum_{j=1}^n d_j 1_{F_j}$$

where $c_j, d_j \geq 0$. Consider the case when $E_{j_0} \cap F_{k_0} \neq \emptyset$. If $\omega \in E_{j_0} \cap F_{k_0}$ then $f(\omega) = c_{j_0} = d_{k_0}$. Since

$$\begin{aligned} \mu(E_j) &= \mu(E_j \cap \Omega) \\ &= \mu(E_j \cap \bigcup_{k=1}^n F_k) \\ &= \sum_{k=1}^n \mu(E_j \cap F_k), \end{aligned}$$

it follows that

$$\int f = \sum_{j=1}^n \sum_{k=1}^n c_j \mu(E_j \cap F_k).$$

Similarly in case of F_j ,

$$\int f = \sum_{j=1}^n \sum_{k=1}^n d_k \mu(E_j \cap F_k).$$

When $E_j \cap F_k = \emptyset$ the corresponding term in the sum is 0, and when $E_j \cap F_k \neq \emptyset$ then $d_k = c_j$. Thus both the sums are equal. ■

LEMMA 6.10 Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, $f, g : \Omega \rightarrow \mathbb{R}$ be measurable functions, and α be some constant; then

- 1) αf ,
- 2) $f + \alpha$,
- 3) $f + g$,
- 4) f^2 ,
- 5) $1/f$,

- 6) f^\pm , where $f^\pm(\omega) = \max(\pm f(\omega), 0)$,
- 7) $|f|$,
- 8) fg

are measurable functions.

Proof | From lemma 6.3, in each case we only have to show that the inverse map of $(-\infty, x]$ belongs to \mathcal{F} , given that $A \in \mathcal{B}(\mathbb{R}) \implies f^{-1}(A), g^{-1}(A) \in \mathcal{F}$.

- 1) In this case we need to show that $\{\omega \mid \alpha f(\omega) \leq x\} \in \mathcal{F}$. When $\alpha = 0$, for all $x \geq 0$ the set in question is Ω and when $x < 0$ it is \emptyset . Both of these are in \mathcal{F} . When $\alpha > 0$, the set $\{\omega \mid \alpha f(\omega) \leq x\} = \{\omega \mid f(\omega) \leq x/\alpha\} \in \mathcal{F}$. Similarly for $\alpha < 0$, $\{\omega \mid \alpha f(\omega) \leq x\} = \{\omega \mid f(\omega) \geq x/\alpha\} \in \mathcal{F}$.
- 2) Using similar logic as above $\{\omega \mid -\infty < f(\omega) + \alpha \leq x\} = \{\omega \mid -\infty < f(\omega) \leq x - \alpha\} \in \mathcal{F}$.
- 3) Consider the set $\{\omega \mid f(\omega) + g(\omega) \leq x\}$. Using density of \mathbb{Q} in \mathbb{R} we know that it is always possible to find $r \in \mathbb{Q}$ such that $f(\omega) \leq r$ and hence $g(\omega) \leq x - r$. Thus $\{\omega \mid f(\omega) + g(\omega) \leq x\} = \bigcup_{r \in \mathbb{Q}} \{\omega \mid f(\omega) \leq r\} \cap \{\omega \mid g(\omega) \leq x - r\}$. Since each $\{\omega \mid f(\omega) \leq r\}$ and $\{\omega \mid g(\omega) \leq x - r\}$ is in \mathcal{F} , by closure under countable unions and intersections $\{\omega \mid f(\omega) + g(\omega) \leq x\} \in \mathcal{F}$.
- 4) Consider the set $\{\omega \mid f^2(\omega) \leq x\}$. In the case when $x < 0$, the set $\{\omega \mid f^2(\omega) \leq x\} = \emptyset \in \mathcal{F}$. In the case when $x \geq 0$, $\{\omega \mid f^2(\omega) \leq x\} = \{\omega \mid -\sqrt{x} \leq f(\omega) \leq \sqrt{x}\} \in \mathcal{F}$.
- 5) Consider the set $\{\omega \mid 1/f(\omega) < x\}$. In the case when $x > 0$,

$$\begin{aligned} \{\omega \mid 1/f(\omega) < x\} &= \{\omega \mid 1/f(\omega) < 0\} \cup \{\omega \mid 0 < 1/f(\omega) < x\} \\ &= \{\omega \mid f(\omega) \leq 0\} \cup \{\omega \mid 0 < f(\omega) \leq 1/x\} \end{aligned}$$

Since each set in the RHS is in \mathcal{F} it follows that $\{\omega \mid 1/f(\omega) < x\} \in \mathcal{F}$. When $x = 0$, $\{\omega \mid 1/f(\omega) < 0\} = \{\omega \mid f(\omega) < 0\} \in \mathcal{F}$. When $x < 0$ then

$$\{\omega \mid 1/f(\omega) < x\} = \{\omega \mid 0 > f(\omega) > 1/x\}$$

which is clearly in \mathcal{F} .

- 6) The set $\{\omega \mid f^+(\omega) \leq x\} = \{\omega \mid f(\omega) \leq x\}$ when $x \geq 0$, and $\{\omega \mid f^+(\omega) \leq x\} = \emptyset$ when $x < 0$. Thus $\{\omega \mid f^+(\omega) \leq x\} \in \mathcal{F}$. Similarly since $\{\omega \mid f^-(\omega) \leq x\} = \{\omega \mid f(\omega) \geq x\}$ when $x \geq 0$ (and \emptyset otherwise). Hence $\{\omega \mid f^-(\omega) \leq x\} \in \mathcal{F}$.
- 7) Since $|f| = f^+ + f^-$, it is obvious that $|f|$ is also measurable.
- 8) In the case of product of measurable functions, we simply use the identity:

$$fg = \frac{1}{2}((f+g)^2 - f^2 - g^2)$$

Since $f+g$, f^2 , and g^2 are measurable (by points 3,4) it follows that fg is also measurable (again by point 3).



PROPOSITION 6.11 Let $(\Omega, \mathcal{F}, \mu)$ be measure space, $f : \Omega \rightarrow \mathbb{R}$ be some function, and $f_j : \Omega \rightarrow \mathbb{R}$ are a sequence of measurable functions. Then

- 1) $\sup f_n$,
- 2) $\inf f_n$,
- 3) $\limsup f_n$,
- 4) $\liminf f_n$, and
- 5) $\lim f_n$

are measurable functions.

Proof | 1) Consider the set $\{\omega \mid \sup f_n < x\}$. Since $f_n \leq \sup f_n$ for all n , hence $f_n < x$. Thus

$$\{\omega \mid \sup f_n < x\} = \bigcup_{n \geq 1} \{\omega \mid f_n < x\} \in \mathcal{F}$$

2) Since $\inf f_n = -\sup\{-f_n\}$, it is clear that $\inf f$ is also measurable.

3) Since

$$\limsup f_n = \inf_n \sup_{m \geq n} f_m$$

and both the infimum and supremum of sequence of measurable functions is measurable, it follows that $\limsup f_n$ is also measurable.

4) Similar argument as above for $\liminf f_n$.

5) For converging sequences $\limsup f_n = \liminf f_n = \lim f_n$. Hence the limit of a converging sequence of functions is converging. ■

DEFINITION 6.12 A property P is said to be true almost everywhere w.r.t. μ , also written as a.e. (μ) , if $\mu(\{\omega \mid P \text{ is false}\}) = 0$. In other words P is true everywhere except in a set with zero measure.

PROPOSITION 6.13 Let f and g be simple functions, then:

1) Integral is linear, i.e.

$$\int af = a \int f \quad \& \quad \int f + g = \int f + \int g.$$

2) Integral is monotonic, i.e.

$$f \leq g \implies \int f \leq \int g.$$

3) $\int f = 0 \iff f = 0$, a.e. (μ) .

4) If $f = g$ a.e. (μ) , then $\int f = \int g$.

Proof | Let $\{E_j\}, \{F_j\}$ be a partition of Ω and let the representation of f, g be:

$$f(\omega) = \sum_{j \geq 1} c_j 1_{E_j}(\omega)$$

$$g(\omega) = \sum_{j \geq 1} d_j 1_{F_j}(\omega)$$

then:

1) The representation of af would be:

$$f(\omega) = \sum_{j \geq 1} (ac_j) 1_{E_j}(\omega)$$

Thus its integration would be:

$$\begin{aligned} \int af &= \sum_{j \geq 1} (ac_j) \mu(E_j) \\ &= a \sum_{j \geq 1} c_j \mu(E_j) \\ &= a \int f. \end{aligned}$$

The representation of $f + g$ would be:

$$\begin{aligned} (f + g)(\omega) &= \sum_{j \geq 1} c_j 1_{E_j}(\omega) + \sum_{j \geq 1} d_j 1_{F_j}(\omega) \\ &= \sum_{j \geq 1} \sum_{k \geq 1} (c_j + d_k) 1_{E_j \cap F_k}(\omega) \end{aligned}$$

The final inequality is due to the fact that if $\omega \in E_j \cap F_k \neq \emptyset$ in which case we have c_j from the first sum and d_k from the second sum in the LHS and a $c_j + d_k$ in the RHS. Since both the summations are countable it is possible to write it out as a single sum (using the fact that product of countable sets is countable). Thus

$$\begin{aligned} \int (f + g) &= \sum_{j \geq 1} \sum_{k \geq 1} (c_j + d_k) \mu(E_j \cup F_k) \\ &= \sum_{j \geq 1} \sum_{k \geq 1} c_j \mu(E_j \cup F_k) + \sum_{j \geq 1} \sum_{k \geq 1} d_j \mu(E_j \cup F_k) \\ &= \int f + \int g \end{aligned}$$

2) If $\omega \in E_i \cap F_j \neq \emptyset$ then

$$c_i = f(\omega) \leq g(\omega) = d_j,$$

and since

$$\int f = \sum_{i \geq 1} \sum_{j \geq 1} c_i \mu(E_i \cap F_j)$$

Either $E_i \cap F_j = \emptyset$, or $E_i \cap F_j \neq \emptyset$ in which case $c_i \leq d_j$. Thus

$$\int f \leq \sum_{i \geq 1} \sum_{j \geq 1} d_j \mu(E_i \cap F_j) = \int g$$

3) Its clear that when $f = 0$, then $\int f = 0$. For the forward implication let $D = \{\omega \mid f > 0\}$ and $D_n = \{\omega \mid f > 1/n\}$. Clearly $D_n \uparrow D$. Since


$$\begin{aligned} f &\geq f 1_{D_n} \geq \frac{1}{n} 1_{D_n} \\ 0 &= \int f \geq \int \frac{1}{n} 1_{D_n} = \frac{1}{n} \mu(D_n) \end{aligned}$$

$$\implies \mu(D_n) \leq 0 \implies \mu(D_n) = 0$$

Using monotone continuity from below, we get that $\mu(D) = 0$. Thus $f = 0$ a.e. (μ) .

- 4) Let $h = f - g$. Thus $h = 0$, a.e. (μ) . From the previous property we know that $\int h = 0$. Then using linearity $\int f = \int g$.

Thus completing the proof. ■

 The second property can be further generalised to $f \leq g$ a.e. (μ) implies that $\int f \leq \int g$ using the fourth property.

LEMMA 6.15 Let $f : \Omega \rightarrow \mathbb{R}$ be a non-negative function, then there exists a sequence of non-negative simple functions $f_n : \Omega \rightarrow \mathbb{R}$ such that $f_n \uparrow f$.

Proof | Consider the sequence of simple functions $(f_n)_{n \geq 1}$ given by:

$$f_n(\omega) = \begin{cases} n, & \text{if } f(\omega) > n \\ \frac{k}{2^n}, & \text{if } \frac{k}{2^n} \leq f(\omega) \leq \frac{k+1}{2^n}, 0 \leq k \leq n2^n - 1 \end{cases}$$

If $f(\omega) = \infty$, then $f_n(\omega) = n$ implying that $f_n(\omega) \rightarrow f(\omega)$. When $f(\omega) < \infty$, then $\exists n_0$ such that $f(\omega) < n_0$. If $n > n_0$ then

$$f_n(\omega) = \frac{[2^n f(\omega)]}{2^n} \leq f(\omega).$$

Since $f \geq 0$ it follows that:

$$\begin{aligned} \frac{2^n f(\omega) - 1}{2^n} &\leq f_n(\omega) = \frac{[2^n f(\omega)]}{2^n} \leq f(\omega) \\ \implies f(\omega) &\leq \lim_{n \rightarrow \infty} f_n(\omega) \leq f(\omega) \end{aligned}$$

Thus as $n \rightarrow \infty$ we get $f_n \rightarrow f$.

Now it is required to show that $f_{n+1} > f_n$. In the case when $f(\omega) = \infty$ clearly $f_n(\omega) = n < f_{n+1}(\omega) = n + 1$. When $f(\omega) > n + 1$, then $f_n(\omega) < f_{n+1}(\omega)$. When $n < f(\omega) < n + 1$, we know that

$$f_{n+1}(\omega) = \frac{[2^{n+1} f(\omega)]}{2^{n+1}} \geq \frac{[n2^{n+1}]}{2^{n+1}} \geq n = f_n(\omega)$$

And finally when $f_n(\omega) < n < n + 1$ then:

$$f_{n+1}(\omega) = \frac{[2^{n+1} f(\omega)]}{2^{n+1}}$$

Since for any $x > 0$ we have:

$$[2x] = \begin{cases} 2[x] + 1, & \text{if } \{x\} \geq 0.5 \\ 2[x], & \text{if } \{x\} < 0.5 \end{cases}$$

Thus:

$$\frac{[2^{n+1} f(\omega)]}{2^{n+1}} = \begin{cases} \frac{2[2^n f(\omega)] + 1}{2^{n+1}}, \\ \frac{[2^n f(\omega)]}{2^n} \end{cases}$$

$$\implies f_{n+1}(\omega) \geq f_n(\omega)$$

Hence we have proved that $f_n \uparrow f$. ■

DEFINITION 6.16 (Definition of integral of non-negative functions) Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f : \Omega \rightarrow \mathbb{R}$ be a non-negative function, and f_n be a sequence of non-negative simple functions such that $f_n \uparrow f$. Then

$$\int f := \lim_{n \rightarrow \infty} \int f_n$$

PROPOSITION 6.17 Integral in definition 6.16 is well defined.

Proof | Let f_n, g_n be sequences of non-negative simple functions such that $f_n, g_n \uparrow f$. Let the representation of f_n, g_n be as follows,

$$\begin{aligned} f_n &= \sum_{i \geq 1} c_{ni} 1_{E_{ni}} \\ g_n &= \sum_{i \geq 1} d_{ni} 1_{F_{ni}}, \end{aligned}$$

let

$$\begin{aligned} I_n &= \int f_n = \sum_{i \geq 1} c_{ni} \mu(E_{ni}) = \sum_{i \geq 1} \sum_{j \geq 1} c_{ni} \mu(E_{ni} \cap F_{nj}) \\ J_n &= \int g_n = \sum_{j \geq 1} d_{nj} \mu(F_{nj}) = \sum_{j \geq 1} \sum_{i \geq 1} d_{nj} \mu(F_{nj} \cap E_{ni}), \end{aligned}$$

and let $\epsilon > 0$. Since we know that f_n and g_n converge to the same function, f , assuming that $\omega \in E_{nk} \cap F_{nl} \neq \emptyset$ it follows that $\exists N$ such that $n > N$ implies

$$\begin{aligned} |f_n(\omega) - g_n(\omega)| &< \frac{\epsilon}{2^{k+l} \mu(E_{nk} \cap F_{nl})} \\ \implies \left| \sum_{i \geq 1} c_{ni} 1_{E_{ni}}(\omega) - \sum_{i \geq 1} d_{ni} 1_{F_{ni}}(\omega) \right| &< \frac{\epsilon}{2^{k+l} \mu(E_{nk} \cap F_{nl})} \\ \implies |c_{nk} - d_{nl}| &< \frac{\epsilon}{2^{k+l} \mu(E_{nk} \cap F_{nl})}. \end{aligned}$$

Also,

$$|I_n - J_n| = \left| \sum_{j \geq 1} \sum_{i \geq 1} (c_{ni} - d_{nj}) \mu(F_{nj} \cap E_{ni}) \right|$$

In this sum we either have that $E_{ni} \cap F_{nj} = \emptyset$, or $E_{ni} \cap F_{nj} \neq \emptyset$ in which case we know that $|c_{ni} - d_{nj}|$ is arbitrarily close to zero. Hence we get

$$\begin{aligned} |I_n - J_n| &\leq \sum_{j \geq 1} \sum_{i \geq 1} |c_{ni} - d_{nj}| \mu(F_{nj} \cap E_{ni}) \\ &< \sum_{j \geq 1} \sum_{i \geq 1} \frac{\epsilon}{2^{i+j} \mu(E_{ni} \cap F_{nj})} \mu(F_{nj} \cap E_{ni}) = \epsilon \end{aligned}$$

Thus $n > N \implies |I_n - J_n| < \epsilon$. Hence

$$\lim I_n = \lim J_n = \int f$$

proving that the integral is well defined. ■

PROPOSITION 6.18 The properties in proposition 6.13 extend to non-negative measurable functions.

Proof | Let f, g be non-negative measurable functions, let $f_n \uparrow f$, and $g_n \uparrow g$ where f_n, g_n are non-negative simple functions. Then:

1) Since $f_n \uparrow f \implies af_n \uparrow af$. Thus

$$\begin{aligned} \int af &= \lim_{n \rightarrow \infty} \int af_n \\ &= a \lim_{n \rightarrow \infty} \int f_n \\ &= a \int f. \end{aligned}$$

2) If $f \leq g$ then there exists N such that $n \geq N \implies f_n \leq g_n$. Thus

$$\begin{aligned} \int f_n &\leq \int g_n \\ \implies \lim_{n \rightarrow \infty} \int f_n &\leq \lim_{n \rightarrow \infty} \int g_n \\ \implies \int f &\leq \int g. \end{aligned}$$

3) The proof for the third property did not assume anything about the function f thus it is true for non-negative functions too.

4) Since $f = g$ a.e. (μ) , let $\epsilon > 0$, then there exists N such that $n > N$ implies that

$$\begin{aligned} g_n - \epsilon &< f_n < g_n + \epsilon, \text{ a.e. } (\mu) \\ \implies \int g_n - \epsilon\mu(\Omega) &< \int f_n < \int g_n + \epsilon\mu(\Omega), \text{ everywhere} \\ \implies \lim_{n \rightarrow \infty} \int g_n - \epsilon\mu(\Omega) &< \lim_{n \rightarrow \infty} \int f_n < \lim_{n \rightarrow \infty} \int g_n + \epsilon\mu(\Omega) \\ \implies \int g &\leq \int f \leq \int g \\ \implies \int f &= \int g \end{aligned}$$

Thus completing the proof. ■

THEOREM 6.19 (Monotone Convergence Theorem) If $(f_n)_{n \geq 1}$, f be non-negative measurable functions such that $f_n \uparrow f$ a.e. (μ) . Then,

$$\int f = \lim_{n \rightarrow \infty} \int f_n.$$

Proof |

