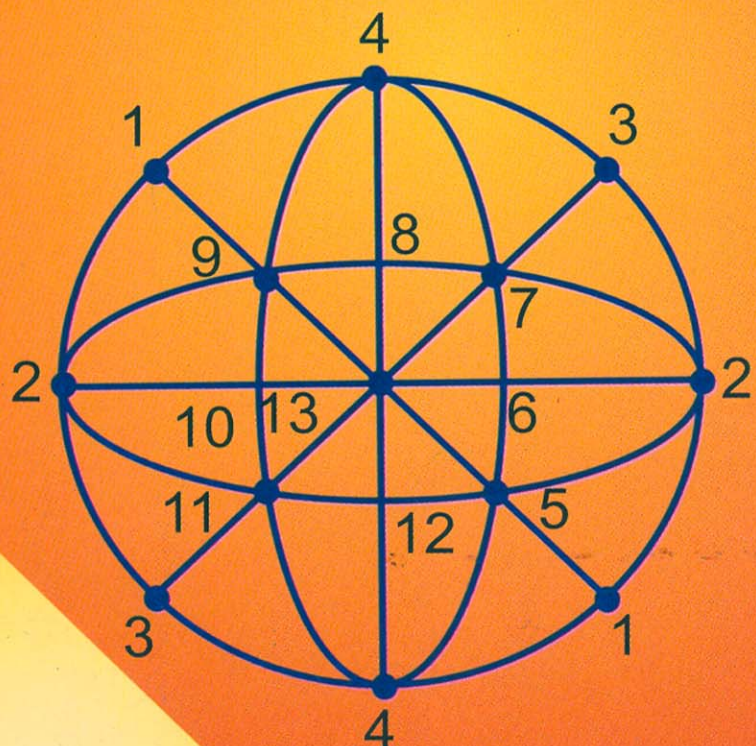
NEW AGE

TOPOLOGY

General & Algebraic



D. Chatterjee

NEW AGE INTERNATIONAL PUBLISHERS

TOPOLOGY

General & Algebraic

This page intentionally left blank

TOPOLOGY

General & Algebraic

D. Chatterjee

NEW AGE

NEW AGE INTERNATIONAL (P) LIMITED, PUBLISHERS

New Delhi • Bangalore • Chennai • Cochin • Guwahati • Hyderabad
Jalandhar • Kolkata • Lucknow • Mumbai • Ranchi

Visit us at www.newagepublishers.com

Published by New Age International (P) Ltd., Publishers

ISBN (13) : 978-81-224-2704-2

PUBLISHING FOR ONE WORLD

NEW AGE INTERNATIONAL (P) LIMITED, PUBLISHERS
4835/24, Ansari Road, Daryaganj, New Delhi - 110002
Visit us at **www.newagepublishers.com**

Preface

Some areas of human knowledge ever since its origin had shaken our understanding of the universe from time to time. While this is more true about physics, it is true about mathematics as well. The birth of topology as analysis *situs* meaning rubbersheet geometry had a similar impact on our traditional knowledge of analysis. Indeed, topology had enough energy and vigour to give birth to a new culture of mathematical approach. Algebraic topology added a new dimension to that. Because quantum physicists and applied mathematicians had noted wonderful interpretations of many physical phenomena through algebraic topology, they took immense interest in the study of topology in the twentieth century. Indian physicists too did not lag behind their counterparts in this respect. Some physicists of Kolkata and around invited me in 1978 to deliver a series of lectures on the subject in the Calcutta University under the auspices of Satyendra Nath Bose Institute of Physical Sciences. The same lecture was delivered earlier to the working physicists of the Indian Statistical Institute in 1976. The present manuscript is a slightly organized version of those lectures delivered at the said places.

To facilitate the readers distinguish the two approaches to the study of topology, matters have been divided into two parts, viz., general topology and algebraic topology. The general topology introduces the classical notions of topology such as compactness, completeness, connectedness etc. and the algebraic topology brings to light the purely algebraic aspects of them. In general, the treatment is sketchy but motivating and helpful for physicists to grasp quickly the basic ideas. The matters have been tested for presentation in Shibaji University and Mosul University. The author will feel rewarded if any one studying this monograph become interested in the subject.

In the preparation of this manuscript I got generous help from many–in particular from Prof A.B. Raha and Prof H. Sarbadhikari who opted to write a part of the manuscript from lectures. I owe a lot to both of them. I will be failing in my duty if I do not acknowledge my debt to Prof K. Sikdar, Prof T. Chandra, Prof S.M. Srivastava, Prof S. Roy, all of Indian Statistical Institute, Prof M. Datta, Director, SNBIPS, Prof B.K. Datta of the University of Trieste, Prof M.K. Das of the University of Nairobi, Prof S. Mukhopadhyay of the City University of New York, USA. My last words of gratitude must go to my wife, Suparna, sons Anandarup and Raju for what they did to see this project complete.

D. Chatterjee

This page intentionally left blank

Contents

Part I

This page intentionally left blank

CHAPTER 1

Sets, Relations and Functions

1.1 SYMBOLS AND NOTATIONS

Mathematics are full of symbols and notations. Symbols have their special meanings. So unless the symbols are understood, no mathematical expression can be interpreted. So we begin with some symbols, some of which mathematics has hired from logic. Students should always carry this collection as a tool box in their memory.

$\forall$:	for all
$\exists$:	there exists
$\nexists$:	there does not exist
$\exists!$:	there exists a unique
$\in$:	belongs to
$\notin$:	does not belong to
$\vee$:	or (disjunction)
$\wedge$:	and (conjunction)
$\Rightarrow$:	implies
$\Leftarrow$:	is implied by
$\Leftrightarrow$:	implies and is implied by
iff :	if and only if
i.e. :	that is
viz. :	namely
$\ni$:	such that
e.g. :	for example
$=$:	is equal to
$\neq$:	is not equal to
$\parallel$:	is parallel to
$\perp$:	is perpendicular to.

1.2 SETS AND SET OPERATIONS

The notion of a set is basic in mathematics. We can express our ideas very precisely and concisely by using this notion. We begin this notion here.

Definition: A *set* is a *well-defined* collection of distinct objects.

The words 'well-defined' and 'distinct' are to be carefully noted here. These words were not originally in the definition given by G. Cantor [1845 – 1918], who is known as the father of set theory. Bertrand Russell pointed out some logical faults in the original definition and mended it as above. The word 'well-defined' means unambiguously defined. Given any object, one should be able to determine whether the object is within the collection or not. The collection of intelligent students of a school is not well defined as opinions may differ in concluding who is intelligent and who is not.

The word 'distinct' implies distinguishable with respect to some features or characteristics. Thus the collection {1, 2, 3, 2} is not a set as the objects are not distinct.

Sets are usually denoted by capital letters or by putting the objects within the second brackets or curly brackets. This is a universally accepted convention and therefore by no means the convention should be disregarded. Thus $X = \{a, e, i, o, u\}$ is a set but {1, 2, 3} is not a set. A set may be expressed also by a property, e.g., $\{x;\ x \text{ is a prime number } \leq 10\}$. This is the same as {2, 3, 5, 7}. If an element belongs to a set, the fact of belonging is expressed symbolically by $2 \in \{1, 2, 3\}$. The fact of not belonging similarly is expressed symbolically by $4 \notin \{1, 2, 3\}$.

Thus if we write $x \in X$, it shall mean the object x is a member of the set X. Thus $x \in X$ can be read many ways as

(*i*) x belongs to X

(*ii*) x is a point of X

(*iii*) x is a member of X

(*iv*) x is an element of X

(*v*) x is an object of X

(*vi*) x is contained in X.

The symbol $y \notin X$ can be accordingly interpreted.

For the sake of precision and consistency the notion of a set which has no object in its collection is accepted in mathematics. This is known as the empty set and is usually denoted by the greek letter ϕ (phi).

Thus the empty set is a set having no element.

Definition: A set A is called a *subset* of another set B if either A is the empty set or every element of A is also an element of B.

Thus (*i*) The empty set is a subset of every set,

(*ii*) Every set is a subset of itself.

The fact that A is a subset of B is symbolically expressed as $A \subset B$.

A set B is called a *superset* of A if A is a subset of B. Note the notion of superset is just the opposite of the notion of subset.

Observe, if $A \subset B$ and $B \subset C$, then $A \subset C$.

Two sets A and B are said to be *equal*, denoted by $A = B$, if $A \subset B$ and $B \subset A$, i.e., every element of A is contained in B and every element of B is contained in A. If two sets A and B are not equal, we express that symbolically by $A \neq B$.

Thus the sets $A = \{a, b, c\}$ and $B = \{c, a, b\}$ are equal; but the sets $P = \{1, 2, 3, 4\}$, $Q = \{2, 3, 5, 6\}$ are not equal since $1 \in P$ but $1 \notin Q$.

Definition: To avoid logical contradictions in the long run all sets considered in any work are supposed to be subsets of a large set. This large set is called the *universal set* of the system and kept fixed through out the entire work.

Thus, when working with the sets $A = \{1, 2, 3\}$, $B = \{2, 4, 5\}$, $C = \{1, 3, 5\}$, the set $\{1, 2, 3, 4, 5\}$ or any superset of this may be taken as the universal set. But whichever is taken as the universal set must be kept unchanged throughout the work.

Set Operations

There are three basic operations with sets; one is called the *unary operation* because it requires only one set for its performance and two are *binary operations* because they require two sets for their performance. We define them as follows:

$A^c = \{x \in \Omega;\ x \notin A\}$. This set A^c is called the *complement* of A and the operation is sometimes referred to as *complementation*.

$A \cup B = \{x \in \Omega;\ x \in A \vee x \in B\}$. This set is called the union of the sets A and B and evidently consists of those elements of the universal set, which belong to any one of A and B.

$A \cap B = \{x \in \Omega;\ x \in A \wedge x \in B\}$. This set is called the *intersection* of the sets A and B. Evidently this consists of those elements of the universal set, which are common to both A and B. Thus if $A = \{1, 2, 3, 4\}$, $B = \{2, 3, 5\}$, $\Omega = \{1, 2, 3, 4, 5, 6\}$ then

$$A^c = \{5, 6\},\ B^c = \{1, 4, 6\}$$

$$A \cup B = \{1, 2, 3, 4, 5\}$$

$$A \cap B = \{2, 3\}.$$

Note that $\Omega^c = \phi, \phi^c = \Omega$ and $(A^c)^c = A$. Further note that $A \cup B = B \cup A$, $A \cap B = B \cap A$.

It is easy to see that

(*i*) $A \subset B \Rightarrow B^c \subset A^c$.

(*ii*) $A \cap \Omega = A,\ A \cup \Omega = \Omega$.

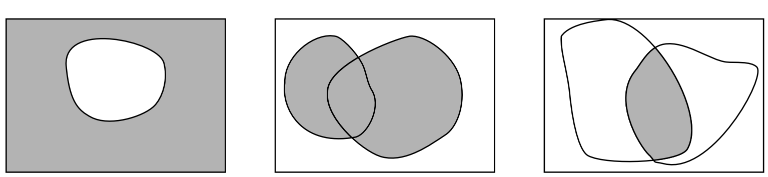
(*iii*) $A \cap \phi = \phi,\ A \cup \phi = A$.

(*iv*) $A \cap A = A,\ A \cup A = A$.

It is quite interesting and helpful to observe that the set operations through diagrams.

Because Venn used diagrams for the first time to visualize sets operations through diagrams such diagrams are called Venn diagrams.

To visualize diagrammatically these set operations we take all points inside a rectangle as the universal set and the region enclosed by a closed curve denotes a set. Thus we have



A^c $\qquad$ $A \cup B$ $\qquad$ $A \cap B$

The notion of set union and set intersections can be extended to finitely many or even infinitely many sets.

Thus if $A_1, A_2, \dots, A_n$ are n sets, then their union and intersection are denoted symbolically by

$$\underset{i=1}{\cup} A_i = \{x \in \Omega;\ x \in A_i \text{ for some i} = 1, 2, \dots, n\}$$

$$\underset{i=1}{\cap} A_i = \{x \in \Omega;\ x \in A_i \text{ for each } i = 1, 2, \dots, n\}$$

Therefore if $A_1 = \{1\}, A_2 = \{1, 2\}, \dots, A_n = \{1, 2, \dots, n\}$, then

$$\overset{n}{\underset{i=1}{\cup}} A_i = \{1, 2, \dots, n\}$$

$$\overset{n}{\underset{i=1}{\cup}} A_i = \{1\}$$

We now state some properties of set operations. Let A, B, C be three sets. Then

1. $(A \cup B) \cup C = A \cup (B \cup C)$ [*Associative property of union*]
2. $(A \cap B) \cap C = A \cap (B \cap C)$ [*Associative property of intersection*]
3. $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ [*Left distributive property*]
4. $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ [*Left distributive property*]
5. $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$ [*Right distributive property*]
6. $(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$ [*Right distributive property*]
7. $(A \cup B)^c = A^c \cap B^c$ [*De Morgan's Law*]
8. $(A \cap B)^c = A^c \cup B^c$ [*De Morgan's Law*]

The De Morgan's law is one of the finest properties of set operations and connects unions with intersection. This property can be generalized as

$$(\cup A_i)^c = \cap A_i^c \text{ and } (\cap A_i)^c = \cup A_i^c$$

From the three basic operations we can define two more frequently used operation, called *difference* and *symmetric difference* as follows:

$$A - B = \{x \in A;\ x \notin B\} \quad A \Delta B = (A - B) \cup (B - A).$$

Observe that $A - B = A \cap B^c$ and $A \Delta B = A \cup B - A \cap B$.

Thus if $A = \{1, 2, 3, 4\}$, $B = \{2, 4, 5, 6\}$,

then $A - B = \{1, 3\}$, $B - A = \{5, 6\}$, $A \Delta B = \{1, 3, 5, 6\}$.

Note from the definitions above it follows that

$$(i)\ A - B \neq B - A, \qquad (ii)\ A \Delta B = B \Delta A.$$

Two sets are said to be *disjoint* if their intersection is empty. Thus A and B are disjoint if $A \cap B = \phi$. Clearly the set $\{1, 2, 4\}$ and $\{3, 5\}$ are disjoint.

The *power set* of a set is the set of all subsets of the given set. The power set of the set X is usually denoted by $\wp(X)$.

Thus if $X = \{1, 2, 3\}$, then

$$\wp(X) = \{\phi, \{1\}, \{2\}, \{3\}, \{1,2\}, \{2,3\}, \{1,3\}, X\}.$$

Note that $X \subset Y$ implies $\wp(X) \subset \wp(Y)$ and conversely.

The *Cartesian product* of two sets *A* and *B*, written as $A \times B$ is defined as a set whose elements are ordered pairs (*a*, *b*) where $a \in A\, b \in B$.

Thus $A \times B = \{(a, b); a \in A, b \in B\}$.

One can similarly define

$$A \times B \times C = \{(a, b, c); a \in A, b \in B, c \in C\}.$$

As a convention we write $A \times A = A^2$, $A \times A \times A = A^3$ etc.

Note $\mathbf{R}^2$ is the set of order pairs of real numbers.

Finite and Infinite Sets

A set is called *finite* if counting can exhaust its collection.

A set is *infinite* if it is not finite.

Thus the set {*a*, *e*, *i*, *o*, *u*} is finite but the set of natural numbers is infinite.

The set of rational numbers and the set of integers are examples of infinite sets.

The number of elements in a finite set is called the *cardinality* of the set and is usually denoted by *n*(*A*) or card (*A*).

The following theorem, known as the *Cardinality theorem*, is of much practical importance.

Theorem 1.2.1: If *A* and *B* are two finite sets, then

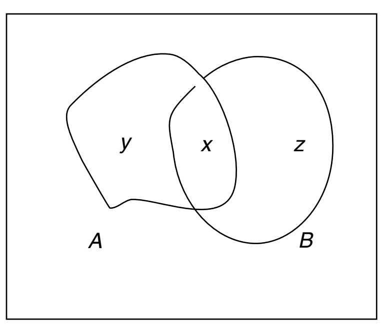
$$n(A \cup B) = n(A) + n(B) - n(A \cap B).$$

Proof: Let *x* be the number of elements common to both *A* and *B*, *y* the number of elements which are in *A* but not in *B* and *z* the number of elements which are in *B* but not in *A*.

Then $n(A) = x + z$.

$n(B) = y + z$.

$n(A \cup B) = x + y + z$.



Evidently, $n(A \cup B) = (x + y) + (x + z) - x$

$$= n(A) + n(B) - n(A \cap B).$$

Corollary 1: If A, B and C are finite sets, then

$$n(A \cup B \cup C) = n(A) + n(B) + n(C) - n(A \cap B) - n(B \cap C) - n(C \cap A) + n(A \cap B \cap C).$$

Corollary 2: If $\{A_i\}$ is a family of m finite sets, then

$$n(\bigcup_{i=1} A_i) = \sum_{i=1} n(A_i) - \sum_{\substack{i,j=1\\ i<j}} n(A_i \cap A_j) + \sum_{\substack{i,j,k=1\\ i<j<k}} n(A_i \cap A_j \cap A_k) - \ldots + (-1)^{n-1}\, n(A_1 \cap A_2 \cap \ldots \cap A_m)$$

Example 1: In a town 70% population use Colgate and 80% use Promise and 95% use any of the tooth-pastes. Determine what percent of people uses both brands.

Solution: Let C denotes the percentage of people who uses Colgate, P denotes the percentage of people who use Promise.

Hence by the given condition,

$$n(C \cup P) = 95,\ n(C) = 70,\ n(P) = 80.$$

By the cardinality theorem, we get $95 = 70 + 80 - n(C \cap P) = 55.$

Thus the percentage of people who use both the brands is 55.

Example 2: If in a class, of the students attending optional English, 45 like Cricket, 40 Football and 30 like Hockey, 15 like both Football and Hockey, 10 like both Hockey and Cricket, 5 like both Football and Cricket. How many like all three? How many like Cricket only? How many students like Cricket and Football but not Hockey?

Solution: Let F denotes the set of Football lovers, C denotes the set of Cricket lovers and H denotes the set of Hockey lovers. Then by the given condition, we get

$$n(F) = 40,\ n(C) = 45,\ n(H) = 30,\ n(F \cap C) = 5,\ n(C \cap H) = 10,\ n(F \cap H) = 15$$

By the cardinality theorem, we see

$$n(C \cup F \cup H) = n(C) + n(F) + n(H) - n(C \cap F) - n(F \cap H) - n(C \cap H) + n(C \cap F \cap H)$$

or

$$85 = 40 + 45 + 30 - 15 - 10 - 6 + n(C \cap F \cap H)$$

$$\therefore\ n(C \cap F \cap H) = 1.$$

Presenting these informations in a diagram, we see –

Since 6 students like both Cricket and Football and only 1 like all three, 6–1 or 5 like Cricket and Football but not Hockey.

Similarly, since 10 students like both Cricket and Hockey, 10–1 or 9 like Cricket and Hockey but not Football. Likewise we get 14 like Football and Hockey but not Cricket.

So the number of students who like Cricket only is 45–5–1–9 or 30.

Some Special Sets: In mathematics and statistics some sets are of much practical use. In order to apply mathematics effectively the following sets are to be well understood.

N = The set of natural numbers.
= {1, 2, 3, 4, ...}

= The set of positive integers

P = The set of prime numbers

= {2, 3, 5, 7, 11, 13, 17, 19, ...}

Z = The set of integers

= {..., –3, –2, –1, 0, 1, 2, 3, 4, ...}

Q = The set of rational numbers

= $\{p/q;\ p, q \in \mathbf{Z}, q \neq 0\}$

I = The set of irrational numbers

= $\{\sqrt{2}, 3\sqrt[7]{5}, 3 + 5\sqrt[3]{7}, 2 - 3\ 11\sqrt[5]{15}, \ldots\}$

R = The set of real numbers

= The set of rational and irrational numbers

= $\mathbf{Q} \cup \mathbf{I}$.

C = The set of complex numbers

= $\{x + iy, x,\ y \in \mathbf{R}, i^2 = -1\}$

= $\{2 + 3i, -5i, 6/7, \ldots\}$

Note every real number is necessarily is a complex number or $x \in \mathbf{R}$ can be written as $x + i\,0$, but every complex number need not be a real number, e.g., $3i$ is a complex number which is not a real. Numbers $x + iy$ where $y \neq 0$ is sometimes referred to as imaginary numbers. Thus $2 - 3i$ is an imaginary number.

Natural numbers are also counting numbers because we count with them.

Evidently the set of natural numbers is infinite. Further every natural number has a successor and every natural number except the first one, viz. 1, has a predecessor.

The set **Z** of integers is also an infinite set. Every integer has a successor and also a predecessor.

The set of rational numbers is an infinite set. The sum of two rational numbers is a rational number, the difference of two rational numbers is a rational number; the product of two rational numbers is a rational number and the quotient of two rational numbers is a rational number, provided the denominator is not zero. In fact, *division by zero is undefined in the real number system* but if we extend the real number system by adjoining $\propto$ and $-\propto$ to it, the set $\mathbf{R} \cup \{-\propto, \propto\}$ is called the *extended real number system* and is denoted by $\overline{\mathbf{R}}$. In the extended real number system we shall not say 5/0 is undefined rather we shall denote this by $\propto$.

The set of rational numbers has two very important properties, viz., the density property and the property of trichotomy. The first one suggests that between any two distinct rational numbers lie infinitely many rational numbers and the second one suggests for any two rational numbers x and y, exactly one of the three relations

$x = y$, $x < y$ and $x > y$

must hold.

Where as the sum, difference, product and quotient of two rationals are rational, the same can not be said about irrational numbers. For example, $2 + \sqrt{3}$ and $4 - \sqrt{3}$ when added gives 6 which is not irrational. Thus we can show that the sum, difference, product and quotient of two irrational numbers need not be irrational always.

Like the set of rationals, the set **I** of irrationals too have the density property and the trichotomy property, i.e., between any two distinct irrational numbers there lie infinitely many irrational numbers and for any two irrational numbers one of the three relations:

$$x = y, \quad x < y \quad \text{and} \quad x > y$$

must hold.

The set of real numbers because it is made of rational and irrational numbers only has many properties analogous to the set **Q** and **I**. For example, the density property and the law of trichotomy do hold for real numbers also. In fact, a stronger statement may be made about real numbers, i.e., between any two real numbers there lie infinitely many rational and irrational numbers.

A very useful identity of a real number is that every real number whether rational or irrational can be expressed by a decimal expression. For example,

$$13/5 = 2.6$$

$$41/7 = 5.857142857142....$$

$$\sqrt{2} = 1.414213562$$

$$2 + 3\sqrt{5} = 3.709975947$$

The question that arises naturally in this context is how to identify a real number as rational or irrational from its decimal expression. The answer is also simple. If the decimal expression is terminating or recurring it is a rational number, if neither it is irrational. Thus the number 3.14285 is a rational number and the number 1.213213213... is rational but the number 0.1010010001 ... is irrational.

A set of real numbers is said to be *bounded above* if all members of the set are less than or at most equal to a fixed real number. This fixed real number is called *an upper bound* of the set. If a real number is an upper bound, any number greater then that is also an upper bound. Thus for a set bounded above there are infinitely many upper bounds. The least of these upper bounds is called the *least upper bound* or *supremum*.

Exactly in the same way we get the *greatest lower bound* or *infimum* of a set bounded below. Interestingly, the supremum of a set of rational numbers, bounded above need not be rational always. Just as the infimum of a set of irrational numbers bounded below need not be irrational always, but the supremum of a set of real numbers bounded above is always a real number and the infimum of a set of real numbers bounded below is always a real number. This is what we call the *completeness property* of **R** For convention we write for any set A of real numbers

$$\sup A \in \mathbf{R} \text{ if } A \text{ is bounded above}$$

$$= \propto \text{ if } A \text{ is not bounded above}$$

and

$$\inf A \in \mathbf{R} \text{ if } A \text{ is bounded below}$$

$$= -\propto \text{ if } A \text{ is not bounded below.}$$

Another very useful property of **R** is the following:

For every real number x, there exist an integer n such that $n \le x < n + 1$.

This n is called the greatest integer n, which is not greater than x, i.e., less than or equal to x. This n is usually denoted by $[x]$.

Thus [3.17] = 3 and [– 2.54] = – 3.

Finally it is worthwhile to remember

(*i*) $x < y \Rightarrow x \pm z < y \pm z$,

i.e., an inequality remains unchanged if the same quantity is added to or subtracted from both the sides.

(*ii*) $x < y \Rightarrow zx < zy$ if $z > 0$.

$zx = zy$ if $z = 0$

$zx > zy$ if $z < 0$.

i.e., an inequality remains unchanged if both sides are multiplied by a positive real number, but changed if multiplied by a negative real number and the inequality is reduced to an equality if multiplied by zero.

(*iii*) $0 < x < y \Rightarrow 1/x > 1/y > 0$.

Finally we conclude this section with another definition:

For a real number *x*, we define

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

The notation $|x|$ is read as the modulus of *x*, or simply mod *x*. From the definition it readily follows:

(*a*) $|x| = |-x|$.

(*b*) $|x| \geq 0$.

(*c*) $|xy| = |x||y|$.

(*d*) $|x/y| = |x|/|y|$.

(*e*) $|x \pm y| \leq |x| + |y|$.

(*f*) $||x| - |y|| \leq |x + y|$.

Note that the expression $f(x) = 2|x - 1| + 3$ can be rewritten as

$$f(x) = \begin{cases} 2x + 1 & \text{if } x \geq 1 \\ 5 - 2x & \text{if } x < 1 \end{cases}$$

as $|x - 1| = x - 1$ if $x \geq 1$ and so $2|x - 1| + 3 =$

$2(x - 1) + 3 = 2x + 1$ but $|x - 1| = -(x - 1) = 1 - x$ if $x < 1$ and so $2|x - 1| + 3 = 2(1 - x) + 3 = 5 - 2x$.

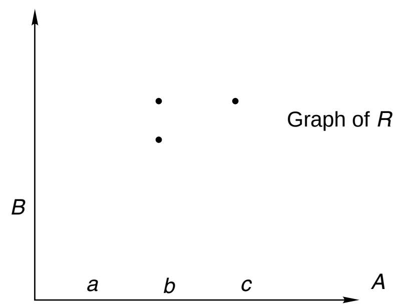
Similarly $|x - a| < \delta$ can be rewritten as $a - \delta < x < a + \delta$, since $|x - a| = x - a$ or $a - x$ and then $x - a < \delta$ gives $x < a + \delta$ but $a - x < \delta$ implies $a - \delta < x$.

1.3 RELATIONS

A basic concept in set theory is that of relations, which has tremendous applications in almost every sphere of academic pursuit including economics, sociology, engineering and technology. The notion of relation is apparently intuitive but it can be defined very precisely. We begin with a few definitions.

Definition: A *relation R* from a set *A* to a set *B* is defined to be a subset of the cartesian product $A \times B$. Thus $R \subset A \times B$. That an element of *A* is related to an element of *B* is expressed symbolically by $x\,Ry$. Clearly $x\,Ry$ holds if $(x, y) \in R \subset A \times B$. Note in a special situation *A* may be the same set as *B*. The set of elements which are related to some elements of *B* is called the *domain of the relation R* and the set of elements of *B* which correspond to some elements of *A* related to the them is called the *range of the relation R*. As two extreme cases, the set ϕ is called the *empty relation* and the set $R \times R$ is called the *universal relation*. The diagrammatic presentation of the set is called the graph of the relation.

For example, let $\{a, b, c\}$ and $B = \{1, 2, 3, 4\}$, then $R = \{(b, 3), (b, 4), (c, 4)\}$ is a relation from *A* to *B* as *R* is a subset of $A \times B$. Clearly $\{b, c\}$ is the domain of *R* and the set $\{3, 4\}$ is the range of the relation *R*.



It is to be noted all elements of *A* may not be related to all elements of *B* in general. Thus the statements *xRy* holds and *uRv* does not hold are quite meaningful. Note that in the above example ${}_aR_2$ does not hold but ${}_bR_4$ does hold. Since a relation hold between two elements only, this is often referred to as a binary relation. A binary relation may be defined on a set *A* as a subset of A^2 or in other words, a binary relation *R* is the set $\{(x, y) \in A^2;\ xRy \text{ holds}\}$.

In reality, that a relation can be defined from a set into another set can be seen from the following example:

Let *A* denotes the set of all novels written in English and *B* denotes the set of all British novelists and *R* denotes the relation of being written by. Then evidently *R* is a subset of $A \times B$. Again a binary relation defined on a set *A* can be given by *R* denoting the relation of 'being a own brother or cousin brother of' in a tribal group *A*.

Definition: If *R* is a relation from *A* to *B*, then its inverse relation R^{-1} is defined as a subset of $B \times A$ given by

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

It is easy to observe that the domain of *R* is the range of R^{-1} and the range of *R* is the domain of R^{-1}.

Though relations have been defined from one set to another, we restrict our further discussions to relations defined on a set only.

Definition: A relation *R* defined on a set *A* is called *reflexive* if *aRa* holds for every $a \in A.$ For example, if *aRb* is defined as '*a* divides *b*' on the set **N**, then evidently **R** is reflexive since every natural number divides itself. But if *aRb* is defined as '*a* is less than *b*' on **N** then evidently **R** is not reflexive.

A relation *R* is called *anti-reflexive* if *aRa* does not hold for every $a \in A.$ Note that the above relation '*a* is less than *b*' is anti-reflexive.

A relation *R* defined on a set *A* is called *symmetric* if '*aRb* holds' implies '*bRa* hold', i.e., (*a*, *b*) $(a, b) \in R \Rightarrow (b, a) \in R.$

A relation is called *asymmetric* if it is not symmetric, i.e., there exists $(a, b) \in R$ such that $(b, a) \notin R.$

A relation is said to be *anti-symmetric* if '*aRb* and *bRa* hold' implies that *a* = *b*.

The relation *R* defined on *R* by '*aRb* holds if *a* is less than or equal to *b*' is clearly anti-symmetric but the relation *R* defined on the set of all straight lines in space by '*a* is perpendicular to *b*' is not anti-symmetric since '*a* is perpendicular to *b* and *b* is perpendicular to *a*' does not imply that *a* = *b*.

A relation *R* is said to be *transitive* if '*aRb* and *bRc* hold' implies that '*aRc* hold'.

The relation *R* defined on the set of all straight lines in space by '*a* is parallel to *b*' is a transitive relation since the parallelity of *a* and *b* and that of *b* and *c* implies the parallelity of *a* and *c*, but the relation *R* defined on the set of all individuals residing in a locality by '*a* likes *b*' is not transitive since '*a* likes *b* and *b* likes *c*' does not necessary imply that *a* likes *c*.

A relation that is reflexive, symmetric and transitive is called an *equivalence relation.* Equivalence relations play an important role in the whole of mathematics.

The relation *R* defined on **R** by *xRy* holds if *x* = *y* is a trivial example of an equivalence relation. The equality of sets is another equivalence relation defined on *P*(*X*) for a non-empty *X*.

Example 1: If *xRy* holds on **Z** if *x* – *y* is divisible by 5, show that *R* is an equivalence relation.

Solution: Clearly *xRx* holds for every *x* in **Z** as *x* – *x*, i.e., 0 is always divisible by 5. If *xRy* holds, then *x* – *y* is divisible by 5, i.e., *x* – *y* = 5*k* for some integer *k*. This implies that *y* – *x* = 5(– *k*), i.e., *y* – *x* is divisible by 5. Thus *R* is symmetric.

Finally, if *xRy* and *yRz* hold, i.e., *x* – *y* = 5*k* and *y* – *z* = 5*k*, then *x* – *z* = 5(*k* + *k*), i.e., *xRz* holds. This implies that *R* is transitive. Hence *R* is an equivalence relation on **Z**.

Remark: The above relation defined on **Z** will be referred to as the *congruence relation* modulo 5 and will be denoted by $x \equiv y$ (mod 5).

One of the most fundamental theorems about equivalence relation is the following:

Theorem 1.3.1: An equivalence relation *R* defined on a set *X* partitions the set into equivalence classes so that every pair of elements in any class is *R*-related and elements of two different classes are not.

Conversely, every partition of a set *X* defines an equivalence relation which generates that partition.

Proof: Let *R* be an equivalence relation. Let $x \in X.$ Put all *s* of *X* which are *R*-equivalent to *x* into one class and denote that by [*x*]. If *X* = [*x*], we stop there. If $[x] \subset X,$ there exists $y \in X, \; y \notin [x].$ Put all *t* of *X* which are *R*-equivalent to *y* in one class and denote it by [*y*].

If $X = [x] \cup [y],$ we stop there, otherwise continue the process till all the elements of *X* are exhausted. Thus we shall finally get $X = [x] \cup [y] \cup [z] \ldots.$ the classes making a partition of *X*. Clearly this partition satisfies all the requirements of the theorem.

Converesely, for any partition of *X* we can define a binary relation as follows:

xRy holds if x and y belong to the same class of the partition. Then it is straightforward to prove that R is an equivalence relation generating the same partition.

For example, the congruence relation $x \equiv y \pmod 5$ defined on Z partitions it into five classes [0], [1], [2], [3] and [4] where

[0] = The set of integers divisible by 5
= {..., –10, –5, 0, 4, 10, ...}

[1] = The set of integers, which leave 1 as remainder when divided by 5
= {..., –9, –4, 1, 6, 11, ...}

[2] = The set of integers, which leave 2 as remainder when divided by 5
= {..., –8, –3, 2, 7, 12, ...}

[3] = The set of integers, which leave 3 as remainder when divided by 5
= {..., –7, –2, 3, 8, 13, ...}

[4] = The set of integers, which leave 4 as remainder when divided by 5
= {..., –6, –1, 4, 9, 14, ...}

Quotient Set

If R is an equivalence relation defined on a set X, then the set of equivalence classes is called the quotient set of X by R and is denoted by X/R.

Thus if ρ denotes the congruence relation on **Z** modulo 5, then

$$\mathbf{Z}/\rho = \{[0], [1], [2], [3], [4]\}$$

Composition of Relations

If R is a relation from A to B and S is another relation from B to C, then a relation can be defined from A to C, called the composition of R and S and denoted by $S_0 R$ in the following way:

$$RS = \{(a, c) \in A \times C;\ \text{There exists } b \in B \text{ such that } (a, b) \in R \text{ and } (b, c) \in S\}$$

Note that if there does not exist any such element as b in B, then RS is the empty relation.

For example if $A = \{a, b, c\}$, $B = \{1, 2\}$, $C = \{5, 6, 7\}$ and $R = \{(a, 1), (b, 1), (b, 2), (c, 1)\}$ and $S = \{(1, 5), (1, 6), (2, 6), (2, 7)\}$, then

$RS = \{(a, 5), (a, 6), (b, 5), (b, 6), (b, 7), (c, 5), (c, 6)\}$

The following result follows readily from the definition.

Proposition 1.3.2: If R_1 is a relation from A to B, R_2 is a relation from B to C and R_3 is a relation from C to D, then

$$(R_1\ R_2)\ R_3 = R_1\ (R_2\ R_3)$$

Proposition 1.3.3: If R_1 and R_2 are two relations from A to B and R_3 and R_4 are two relations from B to C and if $R_1\ R_2$ and $R_3\ R_4$ exist, then

(*i*) $R_1\ R_3 \subset R_2\ R_4$, (*ii*) $(R_1 \cup R_2)\ R_3 = R_1\ R_3 \cup R_2\ R_3$

Relation Matrix

A relation R defined on a finite set A can be expressed as a matrix $[m_{ij}]$, called the *relation matrix* of R, as follows:

$$m_{ij} = \begin{cases} 1 & \text{if } a_i \; R a_j \text{ holds} \\ 0 & \text{if } a_i \; R a_j \text{ does not hold} \end{cases}$$

Thus if $A = \{1, 2, 3, 4\}$ and $R = \{(1, 4), (2, 3), (2, 4), (3, 3), (4, 2)\}$, then

$$M(R) = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

Clearly a relation matrix is a square matrix of order same as the order of the basic set.

Note a reflexive relation has its relation matrix whose all the diagonal elements are ones. The relation matrix of a symmetric relation is symmetric. The relation matrix of an anti-reflexive relation has all its diagonal elements zero.

Note: The notion of relation matrix can be extended to relations from a set to another set in which case this matrix is a rectangular matrix.

If $A = \{a, b\}$, $B = \{p, q, r\}$ and $C = \{u, v, w, x\}$ and $R= \{(a, q), (b, r)\}$ and $S = \{(p, u), (q, w), (r, x)]$, then evidently $RS = \{(a, w), (b, x)\}$ which has a 2×4 relation matrix given by

$$M(RS) = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \text{ where } M(R) = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},\; M(S) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Interestingly, $M(RS) = M(R)\, M(S)$.

1.4 ORDER RELATIONS AND POSETS

A relation of special importance is the partial order. Correspondingly partially ordered sets too play an important role in mathematics.

Definition: A relation defined on a set X is called a *partial order* relation if it is reflexive, anti-symmetric and transitive. We shall use generally $\leq$ to denote a partial order relation.

A set on which a partial order is defined is called a *partially ordered set* or simply *poset*.

An example of a poset is the set **R** when equipped with the usual order $\leq$. The power set $P(X)$ of a nonempty set X is also a poset when equipped with the inclusion relation.

An interesting example of a poset is the set of positive divisors of 12 when equipped with the partial order $\leq$ defined by $a \leq b$ iff a is *a* divisor of *b*.

A relation $\leq$ defined on X is called a *total order* or *linear order* if it is a partial order and for every pair of elements *a* and *b* of X either $a \leq b$ or $b \leq a$.

A set equipped with a total order is called a totally ordered set or linearly ordered set or simply a chain.

The natural order relation $\leq$ defined on **R** is a total order but the relation '*m* is a divisor of *n*' defined on **N** is not a partial order as 2 is not a divisor of 5 nor 5 is a divisor of 2.

An element x in a partially ordered set A is called a *first element* if $x \le a$ for every $a \in A$. Similarly an element y of A is called a *last element* if $a \le y$ for every $a \in A$.

Note that a partially ordered set may have no first element and no last element or may have one of the two but not the other. For example, the set **N** equipped with the natural order $\le$ has a first element 1 but no last element, the set **R** equipped with the natural order has neither a first element nor a last element. The set of all positive divisors of 12 equipped with the divisibility order defined above has a first element 1 and a last element 12.

An element u of a partially ordered set A is called a *maximal element* of A if $u \le a$ for any $a \in A$ implies $a = u$, i.e., no element succeeds u except itself. Similarly an element v is called a minimal element if $a \le v$ implies $a = v$, i.e., no element precedes v except itself.

Note the set **R** equipped with the natural order has neither a *maximal element* nor a minimal element. Every finite set of real numbers equipped with the natural order has a maximal element and a minimal element. The set non-negative real numbers with the usual ordering has a minimal element but no maximal element.

A partially ordered set is said to be *well ordered* if every subset of it has a first element.

The set **N** is well ordered as every subset of **N** has a first element but the set **R** is not well ordered.

Note that every subset of a well ordered set is well ordered.

Well-ordering Principle: Every set can be well-ordered.

Definition: Let A be a subset of a poset X. An element $u \in X$ is called an *upper bound* of A if $a \le u$ for every $a \in A$. Similarly an element l is called a *lower bound* of A if $l \le a$ for every $a \in A$.

An element g of X is called a *least upper bound* or *supremum* of A if g is an upper bound of A and it is the smallest of all upper bounds, i.e., $h \le g$ and h is an upper bound of A implies $g = h$.

An element t of X us called a *greatest lower bound* or *infimum* of A if t is a lower bound of A and it is the greatest of all lower bounds, i.e., $t \le s$ and s is a lower bound implies $s = t$. Note that a supremum may not be an element of A. A similar statement holds for infimum.

For the subset $\{x \in \mathbf{R}; 1 < x \le 5\}$ of the linearly ordered set **R**, 10 is an upper bound just as 5 or 25 are upper bounds, 0 is a lower bound and so also are –7 and 1. Neither the subset **N** nor **R** has an upper bound but has a lower bound 0; the subset **Z** has neither any upper bound nor any lower bound. The element 5 is the supremum and 1 is the infimum of the set.

One of the most powerful tools of modern mathematics is the Zorn's lemma, which asserts the existence of certain type of elements, but there exists no constructive proof of obtaining them.

Zorn's Lemma: Every partially ordered set in which every totally ordered subset has an upper bound contains at least one maximal element.

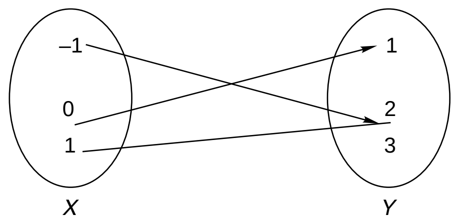
It is one ofthe finest results of mathematics that the Zorn's lemma is equivalent to the well ordering principle.

1.5 FUNCTIONS AND THEIR GRAPHS

A function relates the elements of one set to the elements of another set. Thus a function is of basic importance in mathematics.

To define a function we need two sets and a rule. Thus a function from a set X to a set Y is a rule which assigns to every element of X one or more elements of Y. But if the rule is such that to every element of X a unique element of Y is corresponded, then such a rule defines a single-valued function. If the rule is otherwise, we call the corresponding function many valued or multi-valued or set-valued.

Thus the rule $f(x) = x^2 + 1$ defines a single-valued function from the set $X = \{-1, 0, 1\}$ to the set $Y = \{1, 2, 3\}$, because the rule assigns to –1 of X the element 2 of Y. [note $(-1)^2 + 1 = 2$], to 0 of X the element 1 of Y [note $0^2 + 1 = 1$] and to 1 of X the element 2 of Y [note $1^2 + 1 = 2$]. Diagrammatically this function can be presented as



In the above function, note

(*i*) To every element of X, there corresponds a unique element of Y,

(*ii*) There are more elements in Y than the corresponded elements,

(*iii*) Two distinct element of X correspond to a single element of Y.

Evidently the above function is single-valued. But consider the rule $f(x) = \sin^{-1} x$. Evidently this rule defines a multi-valued function from the set $X = \{0, 1\}$ to the set $\{0, \pi/2, 5\pi/2, 9\pi/2\}$. Since $f(0) = 0$, but $f(1) = \sin^{-1} 1 = \{\pi/2, 5\pi/2, 9\pi/2\}$.

Though multi valued functions have a very important rule to play in mathematics, we shall refrain from discussing such functions and concentrate on single-valued functions only. So the reference to a function anywhere hereafter shall mean only a single-valued function.

Let f be a (single valued) function from X to Y. We shall denote this symbolically by $f: X \to Y$. If x corresponds the element y of Y, then y is called the *image* of x under f or the *value* of f at x. The element x is called the *pre-image* of y. The set X is called the *domain* of f and will be denoted by $D(f)$. The *range* of f, written as $R(f)$, is defined to be the set of images of all elements of X. Thus $R(f) = \{f(x), x \in X\}$. The range or f will be sometimes referred to as $f(X)$. Note an element of Y may have no pre-image, one pre-image or several pre-images. For example, in the above function the element 3 has no pre-image, the element 1 has just one pre-image and the element 2 has two pre-images. Denoting the set of pre-images of y by $f^{1}(y)$, we thus see: $f^{1}(3) = \phi$, $f^{-1}(1) = \{0\}$, $f^{-1}(2) = \{-1, 1\}$

In the same example, $D(f) = \{-1, 0, 1\}$ and $R(f) = \{1, 2\}$. The set Y is sometimes called the *codomain* of Y. If the range of a function is a set of real numbers, the function will be called *real-valued* or **R**-*valued*, if the range is a set of integers, the function is called *integer-valued*, if the range is a set of matrices, the funciton is called *matrix-valued* and so on. As an example of integer-valued function, consider the set of all students in a class and their ages in completed years. Thus to every student we can associate a positive integer and hence this function (may be called age-function) is integer valued.

The function $f(x) = [x]$, called the *greatest integer function*, is also integer valued, i.e., **Z**-valued with domain **R**.

A function $f: X \to Y$ is called *injective* or *one-one* if different elements of X have distinct images, i.e., $x_1 \neq x_2,\ x_1, x_2 \in X \Rightarrow f(x_1) \neq f(x_2)$.

From logical point of view this is as good as saying $f(x_1) = f(x_2) \Rightarrow x_1 = x_2$.

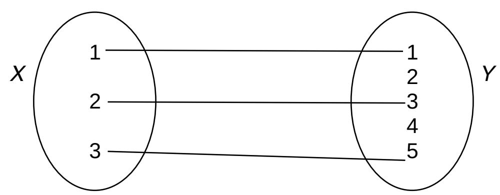
A function $f: X \to Y$ is called *surjective* or *onto* if every element of Y has a pre-image in X, i.e., $\forall\ y \in Y, \exists\ x \in X$ such that $y = f(x)$.

This amounts to saying that there are no (excess) elements in Y, having no pre-image, i.e., $Y = f(X) = R(f)$. If a function is not onto, it is sometime called *an into function*.

A function $f: X \to Y$ is called *bijective* if it is both injective and surjective, i.e., one-one and onto. We now consider examples of such functions.

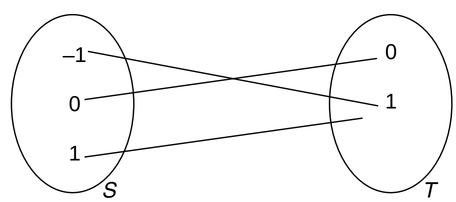
Example 1: Let $X = \{1, 2, 3\}$, $Y = \{12, 2, 3, 4, 5\}$, $f(x) = 2x - 1$. Clearly the function $f: X \to Y$ defined by the give rule is injective but not surjective. The diagram below amply explains the situation.

Note 2 and 4 have no pre-images.



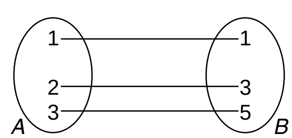
2 and 4 are the (excess) elements of Y, having no pre-image.

Example 2: Let $S = \{-1, 0, 1\}$, $Y = \{0, 1\}$, $g(x) = x^2$. Clearly the function $g: S \to T$ defined by the above rule is surjective but not injective.



Note 1 and –1 have the same image.

Example 3: $A = \{1, 2, 3\}$, $B = \{1, 3, 5\}$ and $h(x) =$ evidently $h: A \to B$ defined by the above rule is both injective and surjective.



Note A and B have the same number of elements.

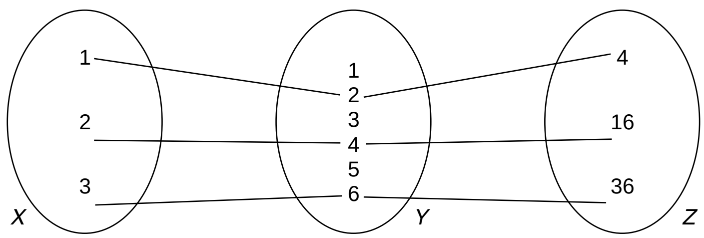
A function $f: X \to Y$ is called a *constant function* if all the elements of X are carried by f to one and the same element of Y.

As for example, if $X = \{1, 2, 3\}$, $Y = \{5, 6, 7, 8\}$, $f(x) = 5$, then evidently $f(x)$ is a constant function.

A function $i: X \to X$ is called the *identity function* of X if $i(x) = x$ for every x in X. Note here an identity function is defined from a set onto itself and it is always a bijective function. Instead of i, i_x or i_x are also used to denote the identity function of X.

If $f: X \to Y$ and $g: Y \to Z$ are two functions, then a function can be defined from $X \to Z$ making use of the rules of f and g. This function is called the *composition function* and is denoted by $g_0\ f$. Indeed, $g_0\ f$ is defined by $(g_0\ f)(x) = g(f(x))$ fore every $x \in X$.

For example, if $f: X \to Y$ and $g: Y \to Z$ be defined by the rules $f(x) = 2x$ and $g(x) = x^2$, where $X = \{1, 2, 3\}$, $Y = \{1, 2, 3, 4, 5, 6\}$ and $Z = \{1, 2, 3, ... 36\}$, then $(g_0\ f)(x) = 4x^2$. This is evident as $(g\ f)(x) = g(f(x)) = \{f(x)\}^2 = (2x)^2 = 4x^2$. Diagrammatically this composition will appear as



One can compose three or more functions in the same manner. In fact, if $f: X \to Y$, $g: Y \to Z$, $h: Z \to W$ be three functions, one can define $h_o\,(g_0 f)$ as $(h_0\,(g_0 f))(x) = h\,\{(g_0 f)(x)\} = (\{g\,(f(x))\}$. It is easy to verify $h_0\ (g_0\ f) = (h_0\ g)\ f$. Recall that two functions f and g are equal if both have the same domain and $f(x) = g(x)$ for every x in their common domain.

To every function $f: X \to Y$ we can associate another function f^{-1}, called the inverse of f provided f is bijective. The inverse function f^{-1} will have its domain Y and its range X.

Thus the inverse of $f: X \to Y$ is a function $f^{-1}: Y \to X$ such that

$f^{-1}\{f(x)\} = x\ \forall\ x \in X$, i.e., f^{-1} carries the image $f(x)$ of x back to x. Thus $f^{-1}{}_0\ f = i_X$. One can verify also $f_0\ f^{-1} = i_y$. Remember the condition of bijectivity of f is a prerequisite for the existence of f^{-1} as a function.

For example, if $f: X \to Y$ is a function defined by the relation $f(x) = 2x - 1$ where $X = \{1, 2, 3\}$ and $Y = \{1, 3, 5\}$, then f^{-1} is a function from Y to X defined by the rule $f^{-1}(y) = \frac{1}{2}(y + 1)$. This is evident because $f^{-1}\{f(x)\} = 1/2\{f(x) + 1\} = (1/2)\{2x - 1 + 1\} = x$ for every $x \in X$. The question arises as to how can one find the inverse of a function. The following example clarifies the method.

Example: Find the inverses of the following functions:

(*i*) $f: \mathbf{Z} \to \mathbf{Z}$ defined by $f(n) = n + 1$.

(*ii*) $f: \mathbf{R} \to \mathbf{R}^+$ defined by $f(x) = e^x$.

(iii) $f : \mathbf{N} \to \mathbf{N}$ defined by $f(n) = n^2$

Solution: (i) Here f is clearly bijective and hence it has an inverse. To find the inverse we write $m = f(n) = n + 1$. So $n = m - 1$. Thus $f^{-1}(m) = m - 1$ is the inverse function.

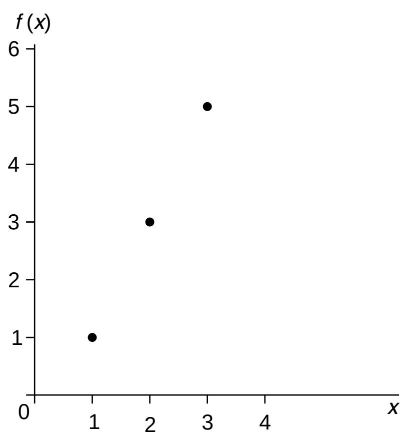
(ii) Clearly the function $f(x) = e^x$ is bijective. Hence its inverse exists. To find the inverse we write $y = f(x) = e^x$. So $x = ln\, y$. Therefore the inverse of f is given by $f^{-1}(y) = ln\, y$.

(iii) The function $f(n) = n^2$ is not bijective. In fact it is injective but not surjective. Hence its inverse does not exist.

Graphs of Functions

The diagrammatic representation of a function is usually referred to as the graph of a function. Set theoretically, the graph of a function $f : X \to Y$ is a set, denoted by graph f and defined by graph $f = \{(x, f(x)); x \in X\}$.

Thus if $f : X \to Y$ is defined by $f(x) = 2x - 1$ where $X = \{1, 2, 3\}$, $Y = \{1, 2, 3, 4, 5\}$, its graph is given by



Set theoretically, graph $f = \{(1, 1), (2, 3), (2, 5)\}$.

Example: Draw the graphs of the following functions:

(i) $f(x) = 2x + 1$ defined ou [–3, 3].

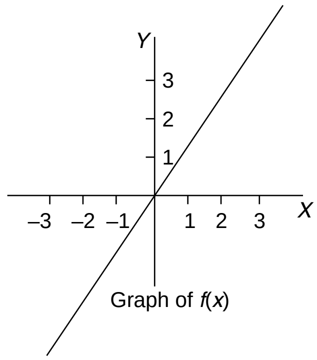
(ii) $f(x) = x + 1$ if $x \le 0$, $= 1 - x$ if $x > 0$.

(iii) $f(x) = x^2$ if $x > 1$, $= 2x - 1$ if $x \le 1$.

(iv) $f(x) = |x - 1|$ defined on $\mathbf{R}$

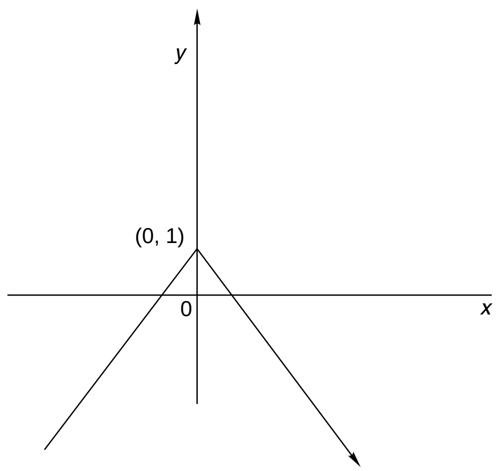
(v) $f(x) = [x + 1]$ defined on [–2, 2].

Solution: (i) Since $y = f(x) = 2x + 1$ is linear in x and y, if represents a straight line.Two points on this line can be found out by putting $x = 0$ and 1 and obtaining $y = 1$ and 3 respectively. So the points are (0, 1) and (1, 3). Thus the graph is as follows:

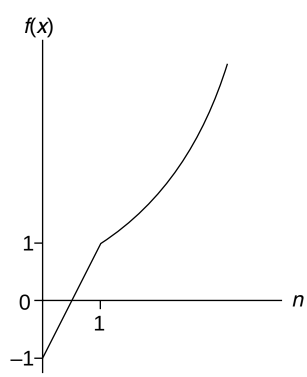


Graph of $f(x)$

(*ii*) The expressions $y = x + 1$ and $y = 1 - x$ are linear and therefore represent parts of straight lines respectively on i $x \leq 0$ and $x > 0$. For $y = x + 1$, two points are (0, 1) and (–1, 0) and for $y = 1 - x$, two points are (1, 0) and (2, 1). Thus the graph will be as follows:

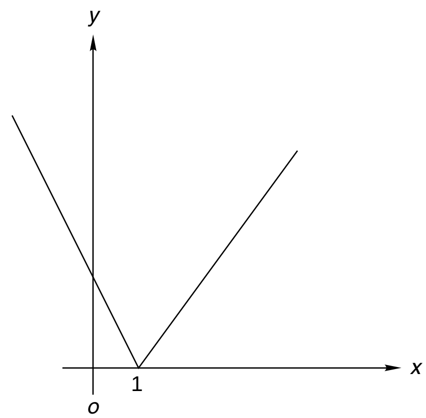


(*iii*) The expression $y = x^2$ represents a parabola passing through the origin and the expression $y = 2x - 1$ represents a straight line. So the given function is made of two parts, one is a segment of a parabola defined for $x > 1$, the other is a segment of a straight line for $x \leq 1$. Thus the graph will be as follows:



(*iv*) The function $f(x) = |x - 1|$ can be rewritten as $f(x) = x - 1$ if $x \geq 1$, $= 1 - x$ if $x < 1$. For $x \geq 1$ the straight line segment will be determined by the points (0, –1) and (1, 0) and for $x < 1$, the straight line segment will be determined by the points (0, 1) and (–1, 2).

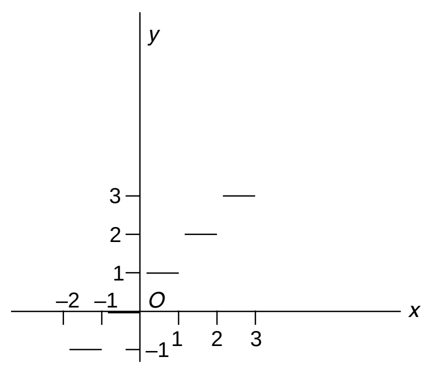
(*v*)



(*vi*) Here we observe that

$$\begin{aligned} f(x) &= -1 \text{ if } -2 \leq x < -1 \\ &= 0 \quad \text{if } -1 \leq x < 0. \\ &= 1 \quad \text{if } 0 \leq x < 1. \\ &= 2 \quad \text{if } 1 \leq x < 2. \\ &= 3 \quad \text{if } x = 2. \end{aligned}$$

So the graph will be as follows:



Remark: If the graph of a function can be drawn without lifting the pencil or pen from the plane of it is casually referred to as a continuous function. In the above graphs, note the functions in (*i*), (*iii*) and (*iv*) are continuous.

We conclude this section with the statement of an apparently obvious and candid hypothesis but for which a large number of mathematical conclusions would have been impossible.

Axiom of Choice: From a family of non-empty sets, a new set can be formed choosing single element from each member of the family.

At this moment the deep impact of this axiom cannot be understood but soon the application of this axiom in proving some basic results will manifest its profundity.

1.6 COUNTABILITY

The notion of countability is an important tool in topology and analysis and needs to be grasped carefully.

Definition: A set X is said to be equivalent (or equipollent or equipotent) to the set Y if there exists a bijective mapping from X onto Y.

Since the inverse of a bijective mapping is also bijective, it follows readily that if X is equivalent to Y, then Y is equivalent to X.

This induces one to conclude that X and Y are equivalent if there exists a bijective mapping from one to the other. The fact that X is equivalent to Y is expressed symbolically by $X \sim Y$.

It follows readily that (*i*) $X \sim X$, (*ii*) $X \sim Y$ implies $Y \sim X$, (*iii*) $X \sim Y$ and $Y \sim Z$ imply $X \sim Z$.

A rigorous definition of finite set is the following:

Definition: A set is said to be finite if it is equivalent to the set $\{1, 2, ..., n\}$ for some natural number n.

Analogously we define the following notions.

Definition: A set is said to be *denumberable* if it is equivalent to **N**. A set is called *countable* if it is finite or *denumberable*.

The cardinality of **N** is defined to be d or $\aleph_0$ (aleph null). Any set equivalent to **N** is said to have the same cardinality.

The cartesian product of two denumerable sets is also denumberable. Even the cartesian product of a denumberable number of denumberable. The first result implies that the set **Q** of rational numbers is denumberable and therefore has the cardinality d or $\aleph_0$. The second result above induces the result that the set of algebraic numbers is denumberable.

We shall see that a simple diagonal argument will prove that the set **R** of real numbers is not denumberable. The cardinal number of **R** is defined to be c or $\aleph_1$ (aleph one). Any set equivalent to **R** also has the cardinality $\aleph_1$. Thus the set **I** of irrational numbers has cardinarlity $\aleph_1$.

An interesting result of much use is the following:

Schroeder-Bernstein Theorem 1.6.1: If X is equivalent to a subset of Y and Y is equivalent to a subset of X, then X and Y are equivalent.

Proof: Let f be an injective mapping from X to Y and g be an injective mapping from Y to X. Let $x \in X$. $g^{-1}(x)$, if it exiusts, is called the first ancestor of x. For convention x will be called the zeroth ancestor of x. The element $f^{-1}\{g^{-1}(x)\}$, if it exists, will be called the second ancestor or x; the element $g^{-1}[f^{-1}\{g^{-1}(x)\}]$ will be called the third ancestor of x, In this way the succesive ancestors of x can be defined. Clearly there are three posibilities: (1) x has infinitely many ancestors, (2) x has even number of ancestors, (3) x has odd number of ancestors. Let X_i denote the set of elements of X which have infinitely many ancestors, X_e the set of elements of X having even number of ancestors and X_o the set of elements of X having odd number of ancestors. Then $X = X_i \cup X_e \cup X_o$ and X_i, X_e, X_o are mutually disjoint. We now define a function $F: X \to Y$ as follows:

$$F(x) = \begin{cases} f(x) & \text{if } x \in X_i \cup X_e, \\ g^{-1}(x) & \text{if } x \in X_o. \end{cases}$$

It is straight forward to prove now that F is bijective. This proves the theorem.

With regard to cardinalities a natural question is whether there are (infinite) cardinals other than $\aleph_0$ and $\aleph_1$. Cantor has provided the answer to this question.

Cantor's Theorem: The power set of any set has cardinality greater than the cardinality of the set itself.

Proof: A mapping f from a set X into $P(X)$ defined by $f(x) = \{x\}$ proves that card $(X) \leq$ card $P(X)$. We need to show that the inequality is strict. If possible suppose it is an equality that is, f is a bijective mapping from X onto $P(X)$. We shall show then that a contradiction arises. Call $a \in X$ a bad element if $a \notin f(a)$. Let B denote the set of all bad elements of X. Note $B \in P(X)$. Since f is surjective, there exists $b \in X$ such that $f(b) = B$. If $b \notin B$, then b is a bad element, but then $b \in f(b) = B$ which is a contradiction. Again if $b \notin B$, then, $b \in f(b) = B$, again a contradiction. So the assumption of the existence of a bijective mapping is not tenable. This completes the proof.

Another relevant question in this regard is whether there is any cardinal between $\aleph_0$ and $\aleph_1$. Interestingly assumption of the absence of any such cardinal has been proved to be consistent with the other axioms of set theory just as the negation has been proved to be consistent with the others as well. This assumption is known to be the Continuum Hypothesis.

Continuum Hypothesis: There is no set X such that $\aleph_0 <$ card $(X) < \aleph_1$.

CHAPTER 2

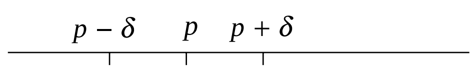
Topologies of R and R^2

In order to have a motivation for the main results of topology in an abstract space or in R^n in particular, it is convenient to have the related notions studied in the set of **R** and $\mathbf{R}^2$.

2.1 TOPOLOGY OF R

Let **R** denote the set of real numbers equipped with its natural order. The best characterization of **R** is that it is a complete Archimedean ordered field. Familiarity of the reader with the field properties, density property and order completeness property are assumed.

Definition: Let δ be a small positive number. The δ-neighbourhood of any point p of **R**, denoted by $N_\delta(p)$, is defined to be set $\{x \in \mathbf{R}; |x - p| < \delta\}$, i.e., the interval $(p - \delta,\ p + \delta)$ centered at p.



The set $N_\delta(p) - \{p\}$ is called the *deleted* δ*-neighbourhood* of p and will be denoted by $\hat{N}_\delta(p)$.

Let G be a subset of **R**. A point p is called an *interior point* of G if there exists a $\delta > 0$ such that $N_\delta(p) \subset G$. The set of all interior points of G is called the *interior* of G and will be denoted by G^o or int (G).

A set G of **R** is called *open* if $G^o = G$, i.e., every point of G is an interior point of G.

The following are simple examples of open sets of **R**:

(i) **R**, (ii) ϕ, (iii) any open interval (a, b), (iv) $(-\infty, b)$, (v) (a, ∞)

(vi) $(a, b) \cup (c, d)$.

The collection of all open sets of **R** is called the usual topology of **R** and will be denoted by U. Thus (0, 1) $\in U$ but [–1, 1] $\notin U$. Note that a nontrivial open set of **R** is either an open interval or union of countably many open intervals (which includes sets like (a, ∞) and $(-\infty, b)$.

It is straightforward to prove that

(i) ϕ, **R** $\in U$

(ii) Arbitrary union of open sets is open

(iii) Finite intersection of open sets is open

Definition: A set F of **R** is called *closed* if F^c is open.

The following are examples of closed sets:

(*i*) $\phi, \mathbf{R}$.

(*ii*) Any finite set, e.g., {1, 3/2, 2}.

(*iii*) Any closed interval, e.g., [*a*, *b*].

(*iv*) Sets like $[a, b] \cup [c, d]$.

The above observations lead to the following results:

(*a*) ϕ and **R** are closed sets.

(*b*) Arbitrary intersection of closed sets are closed.

(*c*) Finite union of closed sets is closed.

The notion of closed sets can be arrived at yet through another approach.

Definition: A point *p* of **R** is called a *limit point* (or *accumulation point*) of a set *F* if every δ-neighbourhood of *p* intersects *F* at least at one point other than *p*, i.e., the deleted δ-neighbourhood of *p* intersects *F*. A limit point of a set may or may not belong to the set.

For example, the set {1, ½, 1/3, ...} has a limit point 0 which does not belong to the set. On the other hand, every limit point of [0, 1] belongs to the set (every point of the set is also a limit point). The set {1, 2, 3, 4, 5} has no limit point.

The set of limit points of a set *F* is called the *derived set* of *F* and is denoted by *F** or *D* (*F*).

A set *F* of **R** is called *closed* if $F' \subset F$, i.e., every limit point of *F* is contained in *F*.

Thus [0, 1] is a closed set since every limit point of [0, 1] is a point of [0, 1].

The set $F \cup F'$ is called the *closure* of *F* and is denoted by $\overline{F}$.

It is straightforward to prove that *F* is closed iff $F = \overline{F}$.

Note that we could have started with the definition of closed sets in this way and thereby arrive at the usual topology through open sets defined as complements of closed sets.

The question whether an infinite set will always have a limit point is settled by the following result:

Bolzano Weirstrass Theorem: Every bounded infinite set has a limit point within or outside the set.

The proof can be seen in any book of real analysis or in [6].

Definition: A real valued function whose domain is **N** is called a *real sequence* or simply *sequence*. Often the range of such a function will be referred to as a sequence by convention and will be denoted by $\{x_n\}$.

A sequence $\{x_n\}$ is said to be bounded if its range set is bounded, i.e., if there exists a real number *k* such that $|x_n| \leq k$ for all *n*.

Thus $\{1/n\}, \{(n+1)/n\}$ and $\{(-1)^n\}$ are examples of bounded sequences.

A sequence $\{x_n\}$ is said to be convergent if there exists a real number *l* satisfying the condition:

For every $\varepsilon > 0,$ there exists a positive integer n_o such that $|x_n - l| < \varepsilon$ for all $n \geq n_o$.

The number *l* is called the limit of the sequence $\{x_n\}$.

Note that the above definition asserts that every ε-neighbourhood of *l* contains all but finitely many terms of the sequence.

Definition: If $\{i_n\}$ is a sequence of positive integers such that $i_1 < i_2 < i_3 < \ldots < i_n < \ldots$, then $\{x_{i_n}\}$ is called a *subsequence* of $\{x_n\}$.

For example the sequence {1/*n*} has a subsequence $\{1/2^n\}$ but the sequence {1/2, 1, 1/4, 1/3, 1/6, 1/5, ...} is not a subsequence of {1/*n*}. A sequence may a have a subsequence which is convergent and another subsequence which is not convergent. For example consider the sequence $\{x_n\}$ defined by

$$x_n = \begin{cases} 1/2^{(n+2)/2} & \text{if } n \text{ is even.} \\ 1-1/2^{(n+1)/2} & \text{if } n \text{ is odd.} \end{cases}$$

Clearly the sequence $\{x_n\}$ is not convergent but a has a convergent subsequence{1/2, 1/4, 1/8, 1/6, ...}. The sequence {*n*} has no convergent subsequence.

A condition that assures the existence of a convergent subsequence is contained in the following result:

Theorem 2.1.1: Every bounded sequence has a convergent subsequence.

The proof of this can be seen any book of analysis or in [6].

Definition: A sequence $\{x_n\}$ is called *Cauchy* if for every $\varepsilon > 0$, there exists a positive integer n_o such that $|x_m - x_n| < \varepsilon$ for all $m,\ n \geq n_o$, i.e., if the terms of the sequence become arbitrarily close to each other as *n* gets larger.

A simple argument proves that every convergent sequence is Cauchy. The converse is also true for real sequences.

Note if $\{x_n\}$ is a Cauchy sequence of integers, then it is of the form $\{x_1, x_2, x_3, \ldots, x_{n_o}, k, k, k, \ldots\}$, i.e., the sequence is constant after the n_o th term.

Completeness

Definition: A set *K* or real numbers is said to be *complete* if every Cauchy sequence $\{x_n\}$ of points in *K* converges to a point in *K*.

Clearly **Z** is complete since every Cauchy sequence in **Z** is convergent. The set **Q** is not complete since the sequence {1, 1.4, 1.41, 1.412, ... } is convergent and therefore Cauchy with the limit $\sqrt{2}$ which is not a point of **Q**. The set **R** is complete.

Compactness

Definition: A subset *K* of **R** is said to be *compact* if every family of open intervals of **R** whose union contains *K* (i.e., covers *K*) has a finite subfamily which also covers *K*.

A result that follows readily from the definition is that every compact set is closed and bounded.

A natural question that arises therefore whether the converse is true because then all compact sets of **R** will be fully characterized. Heine and Borel provided the answer to this question.

Heine Borel Theorem: Every closed and bounded set in **R** is compact.

For a nice proof see [6].

The above result provides us plently of compact sets of **R**.

For example, every finite set is compact. The infinite set {0, 1, 1/2, 1/3, ...} is compact. Every closed and bounded interval is compact but **R** itself is not compact.

Connectedness

Definition: A set K of **R** is said to be connected if the segment joining any two points of K lies in K.

It is immediate from the definition that a connected set is an interval, finite or infinite.

It must be of one of the forms: $[a, b]$, $(a, b]$, $[a, b)$, $(-\infty, b)$, $(\infty, b]$, (a, ∞), $[a, \infty)$ or $(-\infty, \infty)$.

Clearly **R** is connected but **Q** is not, neither is **I**. The set $(0, 1) \cup (2, 5)$ is also not connected.

Note that the union of two connected sets in **R** need not be connected but the intersection of two connected sets is connected.

2.2 CONTINUOUS FUNCTIONS AND HOMEOMORPHISMS

We begin with the local concept of continuity and then pass on to the global concept.

Definition: A function $f : \mathbf{R} \to \mathbf{R}$ is called *continuous at a point p* if for every $\varepsilon > 0$, there exists a $\delta > 0$ such that

$$|f(x) - f(p)| < \varepsilon \text{ for all } x \text{ satisfying } |x - p| < \delta$$

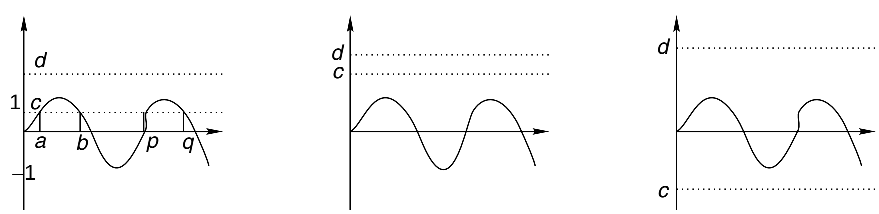
i.e., $x \in N_\delta(p)$ implies $f(x) \in N_\varepsilon(f(p))$.

Note the choice of δ depends on both p and ε. If in particular it does not depend upon p, then the continuity is known as *uniform continuity*. A function $f : \mathbf{R} \to \mathbf{R}$ is said to be *continuous* if it is continuous at every point of **R**.

An interesting characterization of continuous functions is the following:

The proof is an easy consequence of the definition and can be seen in [ ].

That the function $f(x) = \sin x$ is continuous can be visualized from the following diagram:



$f^{-1}(c, d) = (a, b) \cup (p, q) \cup \ldots,$ $f^{-1}(c, d) = \phi$ $f^{-1}(c, d) = \mathbf{R}$

From the above definition it follows that a function f is continuous on any subset D of **R** if it is continuous at every point of D but if D contains some isolated points (*a* point p is an isolated point of a set D if there exists $\delta > 0$ such that $N_\delta(p) \cap D = \phi$), then for continuity at these points it is enough to verify whether the function is defined at these points or not. From definition it follows that every uniformly continuous function is continuous.

Some important results worth noting are the following:

(*a*) The continuous image of a compact set is compact and therefore bounded and closed.

(*b*) The continuous image of a connected set is connected and therefore is an interval.

(*c*) If A is a dense subset of **R**, then every uniformly continuous function on A can be extended uniquely as uniformly continuous function on **R**.

Definition: A bijective function f from **R** onto **R** is called a *homeomorphism* if f is bicontinuous, i.e., both f and f^{-1} are continuous.

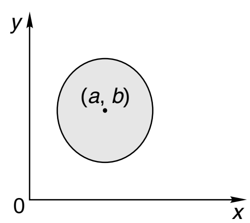
Clearly the function $f(x) = 3x - 2$ is a homeomorphism as f is bijective and both f and f^{-1} are continuous, but $g(x) = \sin x$ is not a homeomorphism as $g(x)$ is not surjective and hence is not bijective though it is continuous.

2.3 TOPOLOGY OF $\mathbf{R}^2$

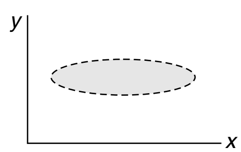
The usual topology of $\mathbf{R}^2$ can be introduced in many ways but the easiest would be to follow the steps as of **R**.

Definition: The δ-neighbourhood of any point (a, b) of $\mathbf{R}^2$ is the set $\{(x, y) \in \mathbf{R}^2;\ (x-a)^2 + (y-b)^2 < \delta^2\}$. This is a circular neighbourhood of (a, b) and is often referred to as the open disc centered at (a, b). A square neighbourhood or a rectangular neighbourhood or an elliptic neighbourhood can also be defined in a similar way. A δ-neighbourhood will be denoted by $N_\delta(a, b)$ or $S_\delta(a, b)$.

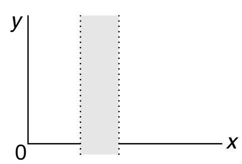
A point (a, b) is called an *interior point* of a set G in $\mathbf{R}^2$ if there exists $\delta > 0$ such that $N_\delta(a, b) \subset G$. The set of all interior points of G will be denoted by G^o or int (G) and will be known as the interior of G.



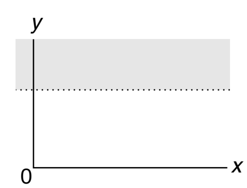
A set G in $\mathbf{R}^2$ is called open if $G = G^o$, i.e., every point of G is an interior point.



(bounded) open set



(unbounded) open set



(unbounded) open set

The complement of an open set in $\mathbf{R}^2$ is called a *closed* set.

The collection of all open sets in $\mathbf{R}^2$ is called the *usual topology* of $\mathbf{R}^2$ and will be denoted by **U**.

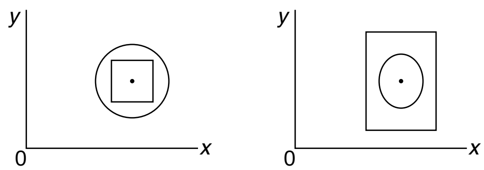
It is easy to verify that

(*i*) $\mathbf{R}^2$ is an open set

(*ii*) Union of any number of open sets is open (*iii*) Intersection of finite number of open sets is open.

Trivially the set ϕ is an open set.

Since every δ-neighbourhood of a point of $\mathbf{R}^2$ contains a square neighbourhood of the point and vice versa, it follows that the topology obtained by δ-neighbourhoods will be the same as the one obtained by square neighbourhoods.



Definition: A point (a, b) is called a *limit point* of a set K in $\mathbf{R}^2$ if every δ-neighbourhood of (a, b) intersects K at least at one point other than (a, b).

The limit point of a set may or may not belong to the set. The set of all limit points of K is called the *derived set* of K and is denoted by K'. The closure of a set K, denoted by $\overline{K}$, is defined as $K \cup K'$.

It is easy to verify that a set F in $\mathbf{R}^2$ is closed if $F = \overline{F}$.

For closed sets the following properties can be proved easily from similar properties of open sets:

(*i*) $\mathbf{R}^2$ is a closed set.

(*ii*) Union of finite number of closed sets is closed.

(*iii*) Intersection of any number of closed sets is closed.

The set ϕ can be proved trivially to be a closed set.

Observe that

(*a*) Every finite set is closed.

(*b*) Union of any number of closed sets need not be closed.

For example $\bigcup_{n=1}^{\infty} [0, 1 - 1/n]^2 = [0, 1)$ which is not closed.

(*c*) Interestingly the set $\{(0, y); 0 \le y \le 1\}$ is closed but the set $\{(0, y); 0 < y < 1\}$ is neither open nor closed.

Definition: A set K is $\mathbf{R}^2$ is called *bounded* if there exists a real number r such that $K \subset S_r\,(0, 0)$.

Bolzano Weirsstrass theorem like in $\mathbf{R}$ gives a condition for the existence of limit points of a set in $\mathbf{R}^2$.

Bolzano Weirsstrass Theorem: Every infinite bounded set in $\mathbf{R}^2$ has a limit point within or outside the set.

A proof of this can be seen in any book of analysis or in [6].

Definition: A function $f: \mathbf{N} \to \mathbf{R}^2$ is called a *sequence* in $\mathbf{R}^2$. A sequence can also be given by its image $\{(x_n, y_n); n \in \mathbf{N}\}$.

A sequence in $\mathbf{R}^2$ is bounded if its range set is bounded and this is equivalent to the boundedness of both the sequences $\{x_n\}$ and $\{y_n\}$ in $\mathbf{R}$.

A sequence (x_n, y_n) is said to be *convergent* if there exists $(x_o, y_o) \in \mathbf{R}^2$ satisfying the condition: For every $\varepsilon > 0,$ there exists a positive integer n_o such that

$$\sqrt{(x_n - x_o)^2 + (y_n - y_o)^2} < \varepsilon \text{ for all } n \geq n_o.$$

i.e., $$\| (x_n, y_n) - (x_o, y_o) \| < \varepsilon \text{ for all } n \geq n_o.$$

The point (x_o, y_o) is called the *limit of the sequence* (x_n, y_n). It is easy to prove that for a sequence in $\mathbf{R}^2$ there exists, if at all, exactly one such point. It can be further observed that the sequence (x_n, y_n) converges in $\mathbf{R}^2$ if and only if $\{x_n\}$ and $\{y_n\}$ both coverge in **R**.

A simple result worth noting is that every convergent sequence is bounded but the converse need not be true.

A sequence (x_n, y_n) is said to be Cauchy in $\mathbf{R}^2$ if for every $\varepsilon > 0,$ there exists a positive integer n_o such that

$$\| (x_m, y_m) - (x_n, y_n) \| < \varepsilon \text{ for all } m, n \geq n_o.$$

i.e., $$\sqrt{(x_m - x_n)^2 + (y_m - y_n)^2} < \varepsilon \text{ for all } m, n \geq n_o.$$

i.e., the terms of the sequence get closer to one another as n gets larger.

Note that every convergent sequence is Cauchy.

A set K in $\mathbf{R}^2$ is said to be *complete* if every Cauchy sequence in K converges to a point in K.

Note that $\mathbf{R}^2$ is complete but the set $\{(1, 1), (1/2, 1/2), (1/3, 1/3), \ldots\}$ is not complete since the sequence being convergent is Cauchy but the limit (0, 0) does not belong to the set.

Compactness

The notion of compactness is very much the same as in **R**.

A set in $\mathbf{R}^2$ is compact if every open cover has a finite subcover.

The argument needed to prove that every compact set is closed and bounded is also similar to that for **R**.

Heine Bortel Theorem: Every closed and bounded set $\mathbf{R}^2$ is compact.

As in **R**, we have the following results in $\mathbf{R}^2$.

(*i*) Every finite set is compact.

(*ii*) $\mathbf{R}^2$ is not compact since $\{(-n, n)^2; n \in \mathbf{N}\}$ is a cover of $\mathbf{R}^2$ which has no finite subcover.

(*iii*) $[a, b] \times [c, d]$ is compact.

Connectedness

A non-empty set K in $\mathbf{R}^2$ is *disconnected* if there exists open sets G and H of $\mathbf{R}^2$ such that $K = (K \cap G) \cup (K \cap H)$ and $(K \cap G) \cap (K \cap H) = \phi.$ A non-empty set is called *connected* if it is not disconnected.

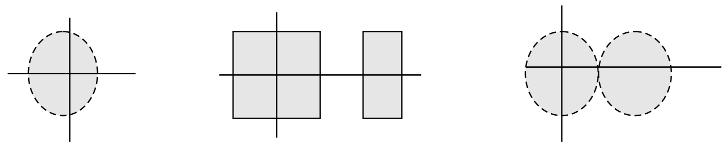
A connected open set in $\mathbf{R}^2$ is called a *region*.

A non-empty set K is said to be *locally connected* if every neighbourhood of a point of $\mathbf{R}^2$ has a connected neighbourhood.

The open unit disc is connected and hence is a region.

The set $K = \{(x, y) \in \mathbf{R}^2; -1 \le x \le 1, -1 \le y \le 1\} \cup \{(x, y) \in \mathbf{R}^2; 2 \le x \le 3, -1 \le y \le 1\}$ is disconnected, because the open sets $G = \{(x, y) \in \mathbf{R}^2; x < 3/2\}$ and $H = \{(x, y) \in \mathbf{R}^2; x > 3/2\}$ satisfy the requirements $K = (K \cap G) \cup (K \cap H)$ and $(K \cap G) \cap (K \cap H) = \phi$.

The set $S = \{(x, y); x^2 + y^2 < 1\} \cup \{(x, y); (x-2)^2 + y^2 < 1\}$ is also disconnected since the sets $U = \{(x, y); x < 1\}$ and $V = \{(x, y); x > 1\}$ satisfy the requirements for disconnection.



2.4 CONTINUOUS FUNCTION AND HOMEOMORPHISM

A function $f : \mathbf{R}^2 \to \mathbf{R}^2$ is called *continuous* if the inverse image of any open disc is open in $\mathbf{R}^2$.

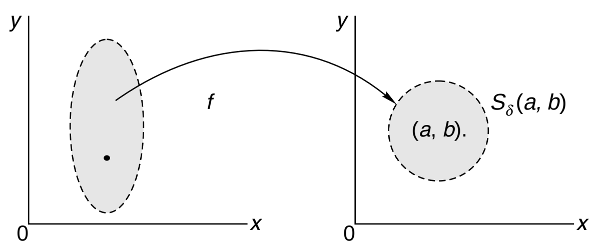
Clearly the function $f(x, y) = (2y + 1, 3x - 1)$ is continuous

$$f^{-1}(S_\delta(a, b)) = \{(x, y); (2y + 1 - a)^2 + (3x - 1 - b)^2 < \delta^2\}$$

$$= \{(x, y); 9(x - \{1 + b\}/3)^2 + 4(y - \{a - 1\}/2)^2 < \delta^2\}$$

$$= \{(x, y); \frac{(x - \{1 + b\}/3)^2}{\delta^2/9} + \frac{(y - \{a - 1\}/2)^2}{\delta^2/4} < 1\}$$

is an open set and in fact an open ellipse. (an elliptic region withou its boundary)



Definition: A bijective function $f : \mathbf{R}^2 \to \mathbf{R}^2$ is called a *homeomorphism* if *f* is bicontinuous, i.e., both *f* and f^{-1} are continuous.

The above example of a continuous fucntion is in fact a homeomorphism.

The results that play a significant role in analysis are the following:

(*a*) The continuous image of a compact set is compact and hence closed and bounded.

(*b*) The continuous image of a connected set is connected.

(*c*) If *D* be a dense set in $\mathbf{R}^2$, then every uniformly continuous function from *D* to $\mathbf{R}^2$ can be extended in a unique way to $\mathbf{R}^2$.

CHAPTER 3

Metric Space

Metric spaces have many interesting properties, which are analogous to those of **R**, yet metric spaces are more general than the Euclidean spaces $\mathbf{R}^n$. Thus metric spaces provide a generalization of the Euclidean spaces offering more flexibility. In this chapter we shall explore these generalization in the metric space set up.

3.1 SOME DEFINITIONS

We begin with the definition of a metric space.

Definition: A non-empty set X equipped with a function $d: X \times X \to \mathbf{R}^+ \cup \{0\}$ is called a *metric space* if the following conditions are satisfied:

(i) $d(x, y) = 0$ if and only if $x = y$ for $x, y \in X$

(ii) $d(x, y) = d(y, x)$ for $x, y \in X$ [*Symmetry*]

(iii) $d(x, z) \leq d(x, y) + d(y, z)$ for $x, y, z \in X$ [*Triangle inequality*]

The function d is called a *metric* on X.

Since on a set several metrics can be defined, a metric space X equipped with the metric d is usually referred to by writing (X, d).

Example 1: The set **R** is a metric space with respect to the metric d defined by

$$d(x, y) = |x - y| \text{ for } x, y \in \mathbf{R}$$

Solution: Clearly $d(x, y) \geq 0$ for $x, y \in \mathbf{R}$.

Since $|x - y| = 0$ iff $x = y$, the first condition follows easily.

Further, since $|x - y| = |y - x|$, the second condition is obvious.

Evidently, $|x - z| = |x - y + y - z| \leq |x - y| + |y - z|$, the triangle inequality follows.

Remark: The above metric of **R** is called the *usual metric* of **R**.

Example 2: The set **R** is a metric space with respect to the metric d' defined by

$$d'(x, y) = 1 \text{ if } x \neq y$$
$$= 0 \text{ if } x = y \text{ for } x, y \in \mathbf{R}$$

Solution: Evidently by definition $d'(x, y) \geq 0$ and $d'(x, y) = 0$ iff $x = y$.

Further, by definition $d'(x, y) = d'(y, x)$. The triangle inequality follows easily.

Hence $(\mathbf{R}, d')$ is a metric space.

Remark 1: This metric is known as the *discrete metric* of **R**.

Remark 2: Exactly the same way we can make any non-empty set a metric sapce, which will be referred to as a discrete metric space.

Example 3: The set $\mathbf{R}^n$ is a metric space with respect to the metric ρ defined by

$$\rho(x, y) = \|x - y\| \text{ for } x, y \in \mathbf{R}^n$$

where $\|x\| = (x_1{}^2 + x_2{}^2 + \ldots, x_n{}^2)^{1/2}$ and $x = (x_1, x_2, \ldots, x_n)$

Solution: Clearly $\rho(x, y) = 0$ iff $x = y$.

The symmetry is also obvious

$$\rho(x, z) = \{\sum_{i=1}^{n} (x_i - z_i)^2\}^{1/2} = \{\sum_{i=1} (x_i - y_i - y_i - z_i))^2\}^{1/2}$$

$$= \{\Sigma[(x_i - y_i)^2 + (y_i - z_i)^2 + 2(x_i - y_i)(y_i - z_i)]\}^{1/2}$$

$$= [\{[\Sigma(x_i - y_i)^2]^{1/2}\}^2 + \{[\Sigma(y_i - z_i)^2]^{1/2}\}^2 + 2[\Sigma(x_i - y_i)^2]^{1/2}[\Sigma(y_i - z_i)^2]^{1/2}\}]^{1/2}$$

$$= \{\Sigma(x_i - y_i)^2 + \Sigma(y_i - z_i)^2 + 2\Sigma(x_i - y_i)(y_i - z_i)]\}^{1/2}$$

$$\leq \{\Sigma(x_i - y_i)^2 + \Sigma(y_i - z_i)^2 + 2[\Sigma(x_i - y_i)^2 \Sigma(y_i - z_i)^2]^{1/2}\}^{1/2}$$

by Cauchy-Schwarz inequality

$$= [\Sigma(x_i - y_i)^2]^{1/2} + [\Sigma(y_i - z_i)^2]^{1/2} = \rho(x, y) + \rho(y, z).$$

Remark: The above metric is known as the *usual metric* of the space $\mathbf{R}^n$.

Example 4: The set $\mathbf{R}^n$ is a metric space when equipped with the metric d defined by

$$d(x, y) = \sum_{i=1}^{n} |x_i - y_i|$$

Solution: Evidently $d(x, y) \geq 0$ and $d(x, y) = 0$ iff $x = y$.

The symmetry $d(x, y) = d(y, x)$ is obvious.

For the triangle inequality we note that

$$d(x, z) = \sum_{i=1}^{n} |x_i - z_i| \leq \sum_{i=1}^{n} |x_i - y_i| + \sum_{i=1}^{n} |y_i - z_i|$$

$$= d(x, y) + d(y, z)$$

Remark: The above metric is called the *rectangular metric* of $\mathbf{R}^n$.

Example 5: The set $\mathbf{R}^n$ is a metric space when equipped with the metric

$$d^*(x, y) = \max_{1 \le i \le n} |x_i - y_i|$$

Solution: Evidently $d^*(x, y) \ge 0$ and $d^*(x, y) = 0$ iff $x = y$

The symmetry is obvious for triangle inequality we note that

$$d^*(x, z) = \max_{1 \le i \le n} |x_i - z_i| = \max_{1 \le i \le n} |x_i - y_i + y_i - z_i|$$

$$\le \max_{1 \le i \le n} |x - y| + \max_{1 \le i \le n} |y - z|$$

$$= d^*(x, y) + d^*(y, z)$$

Hence $(\mathbf{R}^n, d^*)$ is a metric space.

Example 6: The set $C_{\mathbf{R}}[0, 1]$ of all real-valued continuous functions defined on [0, 1] is a metric space with respect to the metric defined as follows:

$$d(f, g) = \sup \{| f(x) - g(x) |;\ x \in [0, 1]\} \text{ where } f, g \in C_{\mathbf{R}}[0, 1]$$

Solution: By definition, $d(f, g) \ge 0$ and $d(f, g) = 0$ iff $f(x) = g(x)$ for all $x \in [0, 1]$

The symmetry is obvious. For triangle inequality observe that

$$d(f, h) = \sup \{| f(x) - h(x) |;\ x \in [0, 1]\}$$

$$= \sup \{| f(x) - g(x) + g(x) - h(x) |;\ x \in [0, 1]\}$$

$$\le \sup \{| f(x) - g(x)|;\ x \in [0, 1]\} + \sup \{|g(x) - h(x)|;\ x \in [0, 1]\}$$

$$= d(f, g) + d(g, h)$$

Hence the result.

Remark: This metric measures the longest distance between the graphs of the two functions

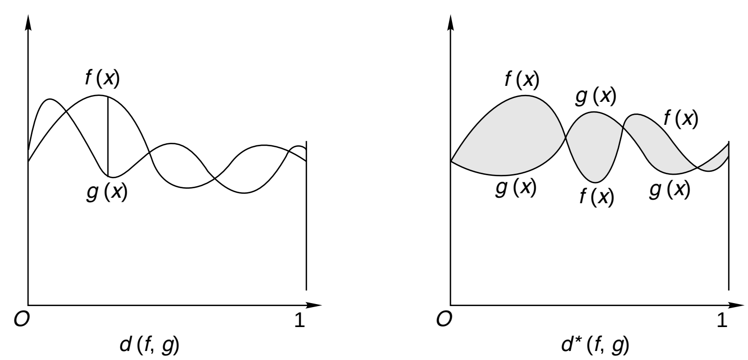


Fig. 1

Example 7: The set $C_{\mathbf{R}}[0, 1]$ of all real-valued continuous functions defined on [0, 1] is a metric space with respect to the metric defined as follows:

$$d^*(f, g) = \int_0^1 |f(x) - g(x)|\, dx \text{ where } f, g \in C_{\mathbf{R}}[0, 1]$$

Solution: The non-negativity of $d^*(f, g)$ is obvious. Further, $d^*(f, g) = 0$ iff $f \equiv g$. The symmetry follows trivially. For triangle inquality observe that

$$d^*(f, h) = \int_0^1 |f(x) - h(x)|dx \le \int_0^1 |f(x) - g(x)|dx + \int_0^1 |g(x) - h(x)|dx$$

$$= d^*(f, g) + d^*(g, h)$$

Hence the result.

Remark: This metric measures the area between the graphs of the two functions.

Definition: The *diameter of a set A* in a metric space (M, d), written as $d(A)$, is defined as follows:

$$d(A) = \sup \{d(x, y); x\ y \in M\}$$

Clearly the diameter of a circular region, called a circular disc, is the same as its normal diameter. The diameter of an elliptic disc is the length of its major axis. The diameter of a rectangular region is the length of its diagonal.

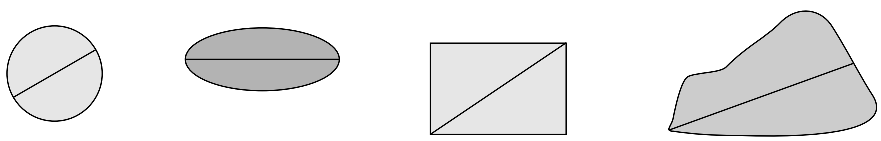


Fig. 2

Definition: The *distance between a point p and a set A* in a metric space (M, d), written as $d(p, A)$, is defined as follows:

$$d(p, A) = \inf \{d(p, x); x \in A\}$$

Intuitively it is the shortest distance between the point p and the points of A, i.e., the distance between p and the point of A nearest to p.

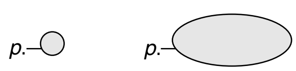


Fig. 3

Definition: The *distance between two sets A and B* in a metric space (M, d), written as $d(A, B)$, is defined as follows:

$$d(A, B) = \inf \{d(x, y); x \in A, y \in B\}$$

It is clear that the distance between two sets is the shortest distance between the points of one set from those of the other.

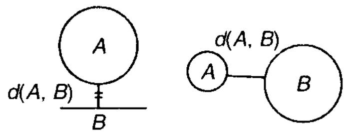


Fig. 4

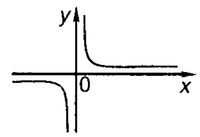


Fig. 4(a)

Example 8: If $A = \{(x, y) \in \mathbf{R}^2; xy = 1\}$ and $B = \{(x, y) \in \mathbf{R}^2; y = 0\}$, find $d(A, B)$.

Solution: Since the distance between the hyperbola $xy = 1$ and the x-axis can be made less than an arbitrary positive number ε, $\inf\{d(x, y); x \in A, y \in B\} = 0$ So $d(A, B) = 0$. [Fig. 4(a)]

Example 9: If $A = \{(x, y) \in \mathbf{R}^n; 2x + 3y = 5\}$ and $p = (1, -1)$, find $d(p, A)$.

Solution: By definition $d(p, A)$ is the shortest distance between the point and the straight line. Hence

$d(p, A) = |2(1) + 3(-1) - 5|/\sqrt{13} = 6/\sqrt{13}$.

Definition: A set A in a metric space (M, d) is said to be *bounded* if $d(A)$ is finite, i.e., $d(A) < \infty$. A set is called *unbounded* if it is not bounded.

Note that every finite set is bounded. A circle is a bounded set but a straight line is an unbounded set in $\mathbf{R}^2$.

3.2 TOPOLOGY OF METRIC SPACES

In this section we shall establish the fact that a metric space is a topological space in the usual sense of topology. To this end we begin with a few definitions.

Definition: A δ-neighbourhood of a point p in a metric space (M, d) is defined to be the set $\{x \in M; d(p, x) < \delta\}$ where $\delta > 0$. This is usually referred to as an open δ-ball or open δ-sphere with centre at p. Sometimes this is written as $N_\delta(p)$ or $S_\delta(p)$. The corresponding closed δ-ball is defined to be the set $\{x \in M; d(p, x) \le \delta\}$ and is denoted by $\overline{N}_\delta(p)$ or $\overline{S}_\delta(p)$.

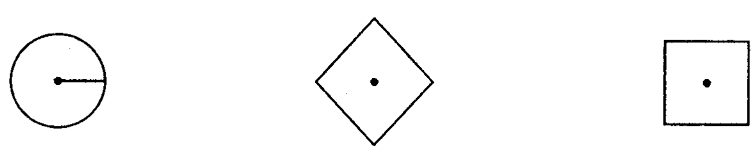
A δ-*nbhd* of a real number x in $\mathbf{R}$ equipped with the usual metric is $(x - \delta, x + \delta)$.

In $\mathbf{R}^2$, let us define three metrics d_1, d_2 and d_3 as follows:

$$d_1(x, y) = \{(x_1 - y_1)^2 + (x_2 - y_2)^2\}^{1/2}$$

$$d_2(x, y) = |x_1 - y_1| + |x_2 - y_2|$$

$$d_3(x, y) = \max\{|x_1 - y_1|, |x_2 - y_2|\}$$



δ-*nbhd* by d_1 δ-*nbhd* by d_2 δ-*nbhd* by d_3

Fig. 5

Definition: A point *p* is called an *interior point* of a set *G* is a metric space (*M*, *d*) if there exists a $\delta > 0$ such that $S_\delta(p) \subset G$.

Note that every interior point of *G* is a point of *G*.

The set of all interior points of *G* is called the *interior* of *G* and is usually denoted by int (*G*) or simply G^0. Note that $G^0 \subset G$ by definition.

Note that the interior of the set [0, 1] in **R** when equipped with the usual metric is (0, 1) and the interior (0,1) is (0, 1). The interior of any finite set is the empty set.

Definition: A set *G* in a metric space (*M*, *d*) is said to be *open* if $G^0 = G$, i.e., every point of *G* is an interior point.

Note that the set (0, 1) is open in **R**. The set $\{(x, y) \in \mathbf{R}^2;\ x^2 + y^2 < 1\}$ is open in $\mathbf{R}^2$ and is known as the open unit disc in $\mathbf{R}^2$.

Theorem 3.2.1: The open δ-sphere $S_\delta(p)$ is an open set in $\mathbf{R}^2$ when equipped with the usual topology.

Proof: Enough to prove that every point of $S_\delta(p)$ is an interior point. To this end let us take an arbitrary point $q \in S_\delta(p)$. So if we choose $\alpha < (1/2)\,\|p - q\|$, then $\alpha > 0$ and $S_\alpha(q) \subset S_\delta(p)$. This implies *q* is an interior point. Hence the result.

Theorem 3.2.2: In a metric space (*M*, *d*) the following are true:

(*i*) ϕ and *M* are open sets,

(*ii*) $\bigcup_{\alpha \in \Lambda} G_\alpha$ is open when each G_α is open,

(*iii*) $\bigcap_{i=1}^{n} H_i$ is open when each H_i is open,

Proof: (*i*) Trivial

(*ii*) Let $p \in \bigcup_{\alpha \in \Lambda} G_\alpha$ when each G_α is open. Then $p \in G_{\alpha 0}$ for some $\alpha_0 \in \Lambda$.

Since $G_{\alpha 0}$ is open, *p* is an interior point of $G_{\alpha 0}$. So there exists $\delta > 0$ such that $S_\delta(p) \subset G_{\alpha 0}$. Hence $S_\delta(p) \subset \bigcup_{\alpha \in \Lambda} G_\alpha$. This implies that $\bigcup_{\alpha \in \Lambda} G_\alpha$ is open.

(*iii*) Let $q \in \bigcap_{i=1}^{n} H_i$ when each H_i is open. Then $q \in H_i$ for each *i*.

Since each H_i is open, there exists $\delta_i > 0$ such that $S_{\delta i}(q) \subset H_i$ for each *i*.
Choose $\delta = \min\{\delta_1, \delta_2, \ldots, \delta_n\}$, then $S_\delta(q) \subset \cup H_i$.

Hence *q* is an interior point of $\bigcap_{i=1}^{n} H_i$. This implies that $\bigcap_{i=1}^{n} H_i$ is open.

Remark 1. The condition (*ii*) above implies that the union of any number of open sets, countable or uncountable, is open but the condition (*iii*) implies that the intersection of finitely many open sets is open in a metric space.

Remark 2. The collection of all open sets in a metric space satisfying the above four conditions is usually referred to as a *topology* of the metric space.

Definition: A point p is called a *limit point* of a set X in a metric space (M, d) if for every $\delta > 0,\ X \cap (S_\delta(p) - \{p\}) \neq \phi,$ i.e., every deleted δ-neighbourhood of p intersects X.

The set of all limit points of X is called the *derived set* of X and is usually denoted by X'.

Clearly the point 0 is a limit point of the set $X = \{1, 1/2, \ldots, 1/n, \ldots\}$ in **R** equipped with the usual metric as for every $\delta > 0,\ S_\delta(0)$ intersects the set. Here $X' = \{0\}$. The derived set of the set (0, 1) is [0, 1] as every point of (0, 1) is a limit point of (0, 1) and so also are 0 and 1.

The *closure of a set X*, written as $\overline{X}$, in a metric space (M, d) is defined as $X \cup X'$. Thus $\overline{X} = X \cup X'.$ Thus the closure of a set is obtained by adjoining to it all its limit points.

A set F in a metric space (M, d) is said to be *closed* if $F' \subset F$, i.e., every limit point of F is a point of the set.

Clearly every finite set is closed as the derived set of a finite set is the empty set and the empty set is a subset of every set. The set [0, 1] in **R** is closed when **R** is equipped with the usual metric.

Theorem 3.2.3: The following are equivalent:

(*i*) F is closed

(*ii*) F^c is open,

(*iii*) $F = \overline{F}$

Proof: (*i*) ⇔ (*ii*)

Let F be closed. Then to show that F^c is open, it is enough to show that every point of F^c is an interior point. To this end take an arbitrary point p in F^c. Since F is closed it cannot be a limit point of F. So there exists a $\delta > 0$ such that $S_\delta(p) \cap F = \phi,$ i.e., $S_\delta(q) \subset F^c$. But this implies that p is an interior point of F^c. Thus F^c is open.

Conversely, let F^c be open. To show that F is closed we shall show that every limit point of F must be in F, i.e., no point of F^c can be a limit point of F. To this end take an arbitrary point q of F^c. Since F^c is open, q must be an interior point of F^c. Hence there exists a $\delta > 0$ such that $S_\delta(q) \subset F^c$. But this implies that q cannot be a limit point of F. Thus no point of F^c can be a limit point of F. Hence F is closed.

(*i*) ⇔ (*ii*)

Let F be closed. Then by definition $F' \subset F$. Hence $\overline{F} = F \cup F' = F$ since $F' \subset F$.

Conversely, let $F = \overline{F}\ F \cup F'$. Hence $F' \subset F$. This implies F is closed.

Remark: The above theorem in no way conveys the message that an open set cannot be closed or vice versa. We have already seen that in the metric space (M, d), M is open and since it is the complement of ϕ in M, it is closed also. It should be noted a set may be neither open nor closed. For example the interval (0, 1] is neither open nor closed in **R** when equipped with the usual topology. Every subset of a discrete metric space is both open and closed.

The following theorem is an analogue of the corresponding results on open sets.

Theorem 3.2.4: The following are true in a metric space (M, d):

(i) ϕ and M are closed

(ii) The intersection of an arbirary family of closed sets is closed

(iii) The union of a finite family of closed sets is closed.

Proof: (i) Trivial.

(ii) Let $\{F_\alpha\}_{\alpha\in\Lambda}$ be an arbitrary family of closed sets.

Then each $F_\alpha{}^c$ is open. Hence $\bigcup\limits_{\alpha\in\Lambda} F_\alpha{}^c$ is open.

Therefore by De Morgan's law, $(\bigcup\limits_{\alpha\in\Lambda} F_\alpha{}^c)^c = (\bigcap\limits_{\alpha\in\Lambda} F_\alpha{}^c)^c = \bigcup\limits_{\alpha\in\Lambda} F_\alpha$ is closed.

(iii) Similar to (ii).

Definition: A set F is said to be *perfect* in a metric space (M, d) if $F = F'$.

From defintion it is clear that every perfect set is closed but the converse is not true. An example of a perfect set is [0, 1] in **R** equipped with the usual topology. One of the most beautiful examples of a perfect set is the Cantor set.

Definition: A set G is said to be dense in X in a metric space (M, d) if $\overline{G} \supset X$.

A set G is called *dense* (*everywhere dense*) if $\overline{G} = M$.

A set H is called *nowhere dense* if $(\overline{H})^0 = \phi$, i.e., the interior of the closure of H is empty.

As for example, the set **Q** of rational numbers is dense in **R** when equipped with the usual topology and any discrete set is nowhere dense in **R**. The set (0, 1) is not nowhere dense.

Proposition 3.2.5: If N is nowhere dense in a metric space (M, d), then $(\overline{N})^c$ is dense.

Proof: If possible let $(\overline{N})^c$ be not dense. Its closure is a paper subset of M. Hence there exists $p \in M$ and an open set G such that $p \in G$ and $G \cap (\overline{N})^c = \phi$.

Then $p \in G \subset \overline{N}$ and so $p \in (\overline{N})^0$ but this is impossible as N is nowhere dense. i.e., $(\overline{N})^0 = \phi$.

Hence $\overline{N}$ is dense.

Proposition 3.2.6: If G is open and N is nowhere dense in a metric space (M, d), then there exist $p \in M$ and $\delta > 0$ such that $S_\delta(p) \subset G$ and $S_\delta(p) \cap N = \phi$.

Proof: Let $H = G \cap N$. Then $H \subset G$ and $H \cap N = \phi$. Further, H is non-empty as G is open and $(\overline{N})^c$ is dense. Take $p \in H$. As H is open, there exists $\delta > 0$ such that $S_\delta(p) \subset G$ and $S_\delta(p) \cap N = \phi$.

Definition: A metric space (M, d) is said to be *separable* if it has a countable dense subset.

An example of a separable metric space is **R** equipped with the usual metric, as **R** has a countable dense subset **Q**. Analogously, $(\mathbf{R}^n, d)$ is separable when d is the usual metric in $\mathbf{R}^n$.

Note a discrete metric space is separable if and only if it is countable. Thus an uncountable set with its discrete topology is not separable.

Definition: A point p is called an *exterior point* of a set K in a metric space (M, d) if it is an interior point of K^c.

The set of exterior points of a set K is called the *exterior* of K and is usually denoted by *ext* (K).

A point q is called a *boundary point* of a set K in a metric space (M, d) if every δ-neighbourhood of q intersects both K and K^c i.e., q is a limit point of both K and K^c.

The set of all boundary points of a set K is called the (topological) *boundary* of K and is usually denoted by *bdry* (K) or ∂K.

Thus $\partial K = K \cap \overline{K}^c$.

As for example if $K = (0, 1)$, then its boundary $\partial K = \{0, 1\}$. The exterior of K is $(-\infty, 0) \cup (1, \infty)$ in **R**.

3.3 SUBSPACE

In this section we shall see how from a given metric space we can construct other metric spaces.

Definition: Let Y be a subset of a metric space (M, d).

Then the function $d: M \times M \to \mathbf{R}^+ \cup \{0\}$ restricted to the set $Y \times Y$ is also a metric, called the *induced metric* on Y. This metric is usually denoted by d_Y

The metric space (Y, d_Y) is called a subspace of (M, d).

Thus (0.1) is a subspace of the metric space **R** when induced by the usual metric of **R**.

It is very natural to inquire about the nature of the open sets and closed sets of a subspace in relation to the original space. The following theorem clarifies the position.

Theorem 3.3.1: Let (M, d) be a metric space and Y be a subset of M.

Then (i) a subset H of Y is open in the subspace (Y, d_Y) iff there exists an open set G in (M, d) such that $H = G \cap Y$

(ii) a subset F of Y is closed in the subspace (Y, d_Y) iff there exists a closed set K in (M, d) such that $F = K \cap Y$.

Proof: (i) Let H be an open set in Y. Then for every y in H there exists a sphere $S_\varepsilon{}^Y(y)$ in the metric d_Y such that $S_\varepsilon{}^Y(y) \subset Y$. So $H = \cup\{S_\varepsilon{}^Y(y);\ y \in Y\}$. But since $S_\varepsilon{}^Y(y) = S_\varepsilon(y) \cap Y$ where $S_\varepsilon(y)$ is the ε-sphere in (M, d), $H = \cup\{S_\varepsilon{}^Y(y) \cap Y\} = H \cap Y$ where $H = \cup\{S_\varepsilon(y);\ y \in Y\}$ is open in (M, d).

Conversely, let $H = Y \cap G$ where G is open in (M, d). To show that H is open in (Y, d_Y), let us take a point $p \in H$. Then as $p \in G$ and G is open, there exists an $\varepsilon > 0$ such that the sphere $S_\varepsilon(p)$ in the d metric is contained in H. But then $S_\varepsilon{}^Y(p) = S_\varepsilon(p) \cap Y \subset G \cap Y = H$. This proves that p is an interior point of H in the d^Y metric.

Hence H is open in (Y, d^Y).

(ii) Similar to (i).

Definition: The set of all open spheres of the metric space (M, d) together with M and the empty set ϕ is called a *base* of the topology of the metric space. It is easy to note

(i) every open set is a union of some open spheres of the metric space,

(ii) for every point x belonging to an open set G in M, there is an open sphere S such that $x \in S \subset G$.

Note that the topology of a metric space may have several bases.

Two bases are said to be *equivalent* if for every point x belonging to a member S of the first base there is a member S' of the second base such that $x \in S' \subset S$ and vice versa.

Evidently two bases generate the same topology.

3.4 COMPLETENESS

Convergence plays an interesting role in any metric space. In this section we discuss some of its special features.

Definition: A sequence x_n is said to be *convergent* in a metric space (M, d) if there exists a point x in M satisfying the condition that for every $\varepsilon > 0$, there exists a positive integer n_0 such that $d(x_n, x) < \varepsilon$ for all $n \geq n_0$. The point x is called a *limit* of the sequence. The fact that the sequence x_n converges to x is denoted by $\lim\limits_{n \to \infty} x_n = x$ or simply lim $x_n = x$.

It is to be noted that if a sequence x_n converges, it converges to a unique limit. In fact if it is supposed that x_n converges to x and also to y, then for an arbitrary $\varepsilon > 0$ there exists n_1 and n_2 such that $d(x_n, x) < \varepsilon/2$ for all $n \geq n_1$ and $d(x_n, y) < \varepsilon/2$ for all $n \geq n_2$ and hence by the triangle inequality we see

$$d(x, y) \leq d(x, x_n) + d(x_n, y) = d(x_n, x) + d(x_n, y) < \varepsilon/2 + \varepsilon/2 = \varepsilon \text{ for } n > \max\{n_1, n_2\}.$$

But the arbitrariness of ε implies that $d(x, y) = 0$ and therefore $x = y$.

The sequence $x_n = 1/n$ is evidently convergent and converges to the limit 0.

Definition: A sequence x_n is said to be *Cauchy* if for every $\varepsilon > 0$, there exists a positive integer no such that $d(x_m, x_n) < \varepsilon$ for all $m, n > n_0$.

For example the sequence $\{1/n\}$ is Cauchy.

A very natural question that arises in this connection is what relation exists between convergent sequences and Cauchy sequences. The following theorem gives an answer to this question.

Theorem 3.4.1: Every convergent sequence is Cauchy but not conversely.

Proof: Let x_n be convergent with limit x. Let $\varepsilon > 0$ be arbitrary.

Then there exists $n_o \in \mathbf{N}$ such that $d(x_n, x) < \varepsilon/2$ for all $n \geq n_o$.

Now let $m, n > n_o$. Then

$$\begin{aligned} d(x_m, x_n) &\leq d(x_m, x) + d(x, x_n) \\ &= d(x_m, x) + d(x_n, x) < \varepsilon/2 + \varepsilon/2 = \varepsilon. \end{aligned}$$

Hence x_n is Cauchy.

For the converse consider the metric space (0, 1) equipped with the induced usual metric and the sequence $\{1/n\}$. Clearly the sequence is Cauchy but not convergent.

Theorem 3.4.2: Let (M, d) be a metric space and $K \subset M$. Then $x \in K'$ iff there exists a sequence x_n in K such that $\lim x_n = x$.

Proof: Let $x \in K'$. Then for every $\delta > 0, S_\delta(x) \cap K \neq \phi$. So if we choose $\delta = 1/n$, we get a sequence x_n in $S_\delta(x) \cap K$ such that $d(x_n, x) < 1/n$. This is our desired x_n which lies in K and converges to x.

Conversely, let x_n be a sequence in K, converging to x. We shall show that $x \in K'$.

By definition, for every $\varepsilon > 0$, there exists a positive integer n_0 such that $d(x_n, x) < \varepsilon$ for all $n \geq n_0$. Thus for every $\varepsilon > 0, S_\varepsilon(x) \cap K \neq \phi$. This implies that $x \in K'$.

Definition: A metric space (M, d) is said to be *complete* if every Cauchy sequence in it is convergent.

Note that the metric space **R** is complete when equipped with the usual metric but the metric space (0, 1) is not complete with respect to the induced metric.

Example 1: Prove that the metric space $C_{\mathbf{R}}[0, 1]$ is complete when equipped with the metric $d(f, g)$ defined by:

$$d(f, g) = \max_{x \in [0,1]} |f(x) - g(x)| \text{ for } f, g \in C_{\mathbf{R}}[0, 1]$$

Solution: Let f_n be a Cauchy sequence in $C_{\mathbf{R}}[0, 1]$. Then for every $\varepsilon < 0$, there exists a positive integer no such that $d(f_m, f_n) < \varepsilon$ for all $m, n \geq n_0$.

That is, $\max_{x \in [0,1]} |f_m(x) - f_n(x)| < \varepsilon$ for all $m, n \geq n_0$

That is, $|f_m(x) - f_n(x)| < \varepsilon$ for all $m, n \geq n_0$ and for all $x \in [0, 1]$.

By Cauchy's criterion the sequence $f_n(x)$ converges uniformly to $f(x)$, say.

This means every Cauchy sequence is convergent in $C_{\mathbf{R}}[0, 1]$. Hence the result.

Example 2: The metric space $C_{\mathbf{R}}[0, 1]$ is not complete with respect to the metric $d_1(f, g)$ defined by

$$d_1(f, g) = \int_0^1 |f(x) - g(x)|\, dx \text{ where } f, g \in C_R[0, 1].$$

Solution: We shall present Cauchy sequence here which will not be convergent in the above metric.

Consider the sequence

$$f_n(x) = n \quad \text{when } 0 \leq x \leq 1/n^2$$
$$= 1/\sqrt{x} \quad \text{when } 1/n^2 < x \leq 1$$

We shall show now that f_n is Cauchy but not convergent.

For $m > n$, we see

$$d_1(f_m, f_n) = \int_0^1 | f_m(x) - f_n | \, dx$$

$$= \int_0^{1/m^2} |m - n| \, dx + \int_{1/m^2}^{1/n^2} |1/\sqrt{x} - n| + \int_{1/n^2}^{1} |1/\sqrt{x} - 1/\sqrt{x}| \, dx$$

$$= (m - n)/m^2 + \left[2\sqrt{x} - nx\right]_{1/m^2}^{1/n^2}$$

$= 1/n - 1/m$ which tends to 0 as m and n tend to infinity.

Now we show that the limit of this sequence does not belong to $C_R[0, 1]$, i.e., not continuous.

Let f_n converge to f. Then observe that

$$d_1(f_n, f) = \int_0 | f_n(x) - f(x) | \, dx$$

$$= \int_0^{1/n^2} |n - f(x)| \, dx + \int_{1/n^2}^{1} |1/x - f(x)| \, dx$$

Since the integrals are non-negative, lim $d(f_n, f) = 0$ will imply that each integral on the right must approach zero.

This gives

$$f(x) = 1/\sqrt{x} \text{ if } 0 < x \leq 1$$
$$= 0 \quad \text{if } x = 0$$

Evidently f is discontinuous. Hence $(C_{\mathbf{R}}[0, 1], d_1)$ is not complete.

Example 3: The metric space (l_∞, d) is complete.

Solution: Let x_n be a Cauchy sequence in l_∞ and let $x_n = (a_1{}^n, a_2{}^n, a_3{}^n, \ldots, a_i{}^n, \ldots)$.

Since x_n is bounded, there exists $k > 0$ such that $|a_i{}^n| < k$ for all $i = 1, 2, \ldots$ to ∞.

Since x_n is Cauchy, for every $\varepsilon > 0$, there exists a positive integer n_0 such that $d(x_m, x_n) < \varepsilon$ for all $m, n \geq n_0$.

i.e., $\sup |a_i{}^m - a_i{}^n| < \varepsilon$ for all $m, n \geq n_0$.

Hence $|a_i{}^m - a_i{}^n| < \varepsilon$ for all $m, n \geq n_0$ and all $i = 1, 2, \ldots$ to ∞.

We now observe that for each i, $\{a_i{}^1, a_i{}^2, \ldots\}$ is a Cauchy sequence and hence converges to a_i, say.

Hence making $m \to \infty$, we get

$$|a_i - a_i{}^n| \leq \varepsilon \text{ for all } n \geq n_0.$$

Thus $|a_i| \leq |a_i - a_i{}^n| + |a_i{}^n| < \varepsilon + k$ for all i.

This implies that the sequence $\{a_i\}$ is bounded. Hence $x = \{a_i\} \in l_\infty$.

Hence l_∞ is complete.

The following theorem is of immense importance.

Theorem 3.4.5: Let (M, d) be a complete metric space and Y be a subspace of M. Then Y is complete iff Y is closed in (M, d).

Proof: Let Y be closed in (M, d) and x_n be a Cauchy sequence in Y. Therefore x_n is Cauchy in M and hence convergent in M. Let it converge to $x \in M$. Then either all but finitely many $x_n = x$ or all x_n's are distinct and x is a limit point of the set $\{x_n\}$. In the first case x is in Y and in the second case $x \in Y$ as Y is closed. Hence Y is complete.

Conversely, let Y be complete and y be a limit point of Y. For each positive integer $n, S_{1/n}(y)$ contains a point y_n in Y. Clearly y_n converges to y and hence y_n is Cauchy. Since Y is complete, y belongs to Y. Thus Y is closed.

Definition: A sequence $\{A_n\}$ of subsets of a metric space (M, d) is called *decreasing* if $A_1 \supset A_2 \supset A_3 \supset \ldots \supset A_n \supset \ldots$

In complete metric spaces such sequences of non-empty closed sets have an interesting feature given by the following theorem:

Cantor Intersection Theorem: If $\{F_n\}$ be a decreasing sequence of non-empty closed subsets of a complete metric space (M, d) such that $d(F_n)$ tends to zero, then $F = \bigcap_{n=1}^{\infty} F_n$ is a singleton set.

Proof: First we note that F cannot contain more that one point as in that case $d(F_n)$ will fail to tend to zero.

Next we choose x_n from each F_n. Since $d(F_n)$ tends to zero, x_n must be Cauchy.

Therefore it must be convergent as M is complete. We shall show that its limit x, say, belongs to F.

If possible, let $x \notin F$. Then $x \notin F_{no}$ for some $n_o \in \mathbf{N}$.

Hence $d(x, F_{no}) = r > 0$. But as $d(x, F_{no}) = \inf\,\{d(x, y);\, y \in F_{no}\}$ and F_{no} is closed, $S_{r/2}(x) \cap F_{no} = \phi$. Therefore $x_n \in F_{no}$ for all $n \geq n_o$. This implies $x_n \notin S_{r/2}(x)$ which is impossible as x is a limit point of x_n. Hence x must belong to F.

Definition: A metric space (M, d) is said to be the first category if M can be expressed as a union of countably many nowhere dense sets.

A metric space (M, d) is said to be of the *second category* if it is not of the first category.

We shall now prove one of the most beautiful and useful theorems of topology.

Baire's Category Theorem: A complete metric space is of the second category.

Proof: If possible let (M, d) be a complete metric space which is not of the second category. So it must be of the first category and therefore there exists a countable family of nowhere dense sets $\{N_i\}$ such that $M = \cup N_i$.

Since N_1 is nowhere dense in *M*, there exists $p_1 \in M$ and $\delta_1 > 0$ such that $S_{\delta 1}(p_1) \subset G$ and $S_\delta(p) \cap N_1 = \phi$. Choose $\varepsilon_1 = \delta_1/2$. Then $\bar{S}_{\varepsilon 1}(p_1) \cap N_1 = \phi$.

Again, $S_{\varepsilon 1}(p_1)$ is open and N_2 is nowhere dense in *M*. Hence there exits $p_2 \in M$ and $\delta_2 > 0$ such that $S_{\delta 2}(p_2) \subset S_{\varepsilon 1}(p_1) \subset \bar{S}_{\varepsilon 1}(p_1)$ and $S_{\delta 2}(p_2) \cap N_2 = \phi$.

Choose $\varepsilon_2 = \delta/2 \le \varepsilon_1/2 = \delta_1/4$. Then $S_{\varepsilon 2}(p_2) \subset S_{\varepsilon 1}(p_1)$ and $\bar{S}_{\varepsilon 2}(p_2) \cap N_2 = \phi$.

Continuing in this manner we get a sequence of non-empty closed sets

$$\ldots \quad \ldots \bar{S}_{\varepsilon 4}(p_4) \subset \bar{S}_{\varepsilon 3}(p_3) \subset \bar{S}_{\varepsilon 2}(p_2) \subset \bar{S}_{\varepsilon 1}(p_1)$$

such that for every positive integer *n*, $\bar{S}_{\varepsilon n}(p_n) \cap N_n = \phi$ and $\varepsilon_n \le \delta_1/2^n$.

As $d(\bar{S}_{\varepsilon n}(p_n)) = \varepsilon_n \to \mathbf{0}$ as $n \to \propto$, by Cantor Intersection Theorem it follows that $\cap \bar{S}_{\varepsilon n}(p_n) \ne \phi$. Let $p \in \cap \bar{S}_{\varepsilon n}(p_n)$. Clearly $p \notin N_n$ for each *n*. But this is a contradiction. Hence *M* must be of the second category.

3.5 CONTINUITY AND UNIFORM CONTINUITY

So long we had been busy discussing the various structures of a metric space. Now we concentrate on relations between metric spaces. To this end we begin with some definitions.

Definition: Let (X, d_1) and (Y, d_2) be two metric spaces. A function $f: X \to Y$ is said to be *continuous at a point a* of *X* if for every $\varepsilon > \mathbf{0}$, there exists a $\delta > 0$, depending on ε and a such that

$$d_2(f(x), f(a)) < \varepsilon \quad \text{whenever } d_1(x, a) < \delta.$$

i.e., $$f(x) \in S_\varepsilon(f(a)) < \varepsilon \quad \text{whenever } x \in S_\delta(a)$$

i.e., $$f(S_\delta(a)) \subset S_\varepsilon(f(a))$$

A function $f: X \to Y$ is said to be *continuous* if it is continuous at every point of *X*.

A function $f: X \to Y$ is said to be *sequentially continuous* if $f(x_n)$ converges to $f(x)$ in *Y* whenever x_n converges to *x* in *X*.

The following theorem gives useful characterizations of continuous functions.

Theorem 3.5.1: Let $f: (X, d_1) \to (Y, d_2)$ be a function. Then the following are equivalent:

(*i*) *f* is continuous

(*ii*) $f^{-1}(G)$ is open for every open set *G* of *Y*

(*iii*) $f^{-1}(F)$ is closed for every closed set *F* of *Y*

(*iv*) *f* is sequentially continuous.

Proof: (*i*) $\Rightarrow$ (*ii*)

Let $f: (X, d_1) \to (Y, d_2)$ be continuous and *G* is open in *Y*. To show that $f^{-1}(G)$ is open in *X*, it is enough to prove that every point of $f^{-1}(G)$ is an interior point. Let $p \in f^{-1}(G)$ and $\varepsilon > 0$ be arbitrary. Then $f(p) \in G$ and because *G* is open, there exists $\varepsilon > 0$ such that $S_\varepsilon(f(p)) \subset G$. As *f* is continuous at p, there exists a $\delta(\varepsilon, p) > 0$ such that $f(S_\delta(p)) \subset S_\varepsilon(f(p)) \subset G.$ Therefore $S_\delta(p) \subset f^{-1}(G).$ Hence p is an interior point of $f^{-1}(G).$

(*ii*) $\Rightarrow$ (*i*)

To prove that $f : (X, d_1) \to (Y, d_2)$ is continuous it is enough to prove that f is continuous at every point of X. To this end let $p \in X$ be arbitrary. Then for every $\varepsilon > 0,$ the sphere $S_\varepsilon(f(p))$ is open in Y and hence by hypothesis, $f^{-1}(S_\varepsilon(f(p))$ is open in X. The point p being an interior point of this, there exists a $\delta > 0$ such that $S_\delta(p) \subset f^{-1}(S_\varepsilon(f(p)).$ Thus if $x \in S_\delta(p),$ then $f(p) \in S_\delta(f(p)).$ This implies f is continuous at p. Hence the result.

(*ii*) $\Rightarrow$ (*iii*)

Let F be a closed set in Y. Then F^c is an open set in Y. So by (*ii*), $f^{-1}(F^c)$ is open in X. Hence $[f^{-1}(F^c)]^c$ is closed. But $f^{-1}(F) = [f^{-1}(F^c)]^c.$ Therefore $f^{-1}(F)$ is closed.

(*iii*) $\Rightarrow$ (*ii*)

Similar to above.

(*i*) $\Rightarrow$ (*ii*)

Let f be continuous and let x_n converge to x in X. We must show that $f(x_n)$ converge to $f(x).$ Then for an $\varepsilon > 0$ and $S_\varepsilon(f(x)),$ there exists $S_\delta(x)$ such that $f(S_\delta(x)) \subset S_\varepsilon(f(x)).$ Since x_n converges to x, there exists $n_o \in \mathbf{N}$ such that $x_n \in S_\delta(x)$ for all $n \geq n_o.$ This implies that $f(x_n) \in S_\varepsilon(f(x))$ for all $n \geq n_o.$ Hence $f(x_n)$ converges to $f(x).$

If possible, let f be not continuous. So there is a point x where it fails to be continuous. We shall show that x_n converges to x does not imply that $f(x_n)$ converges to $f(x).$ Since f is not continuous, there exists an open sphere $S_\varepsilon(f(x))$ with the property that the image of every sphere centered at x is not contained in $S_\varepsilon(f(x)).$ Consider the sequence of open sphere $\{S_1(x), S_{1/2}(x), ..., S_{1/n}(x), ...\}$ and form a sequence $x_n \in S_{1/n}(x)$ and $f(x_n) \notin S_\varepsilon(f(x)).$ Clearly x_n converges to x but $f(x_n)$ does not converge to $f(x).$

Definition: A function $f : (X, d_1) \to (Y, d_2)$ is said to be *uniformly continuous* if for every $\varepsilon > 0,$ there exists $\delta > 0$ such that $d_1(x, y) < \delta$ implies $d_2(f(x), f(y)) < \varepsilon.$

A direct consequence of the definition of uniform continuity is the following theorem:

Theorem 3.5.2: Every uniformly continuous function is continuous.

Proof: Straightforward and left therefore.

The converse of the theorem is in general not true. For example the function $f(x) = x$ is uniformly continuous on **R** but the function $g(x) = x^2$ is not uniformly continuous on **R**. The function $h(x) = 1/x$ is also not uniformly continuous on (0, 1).

Theorem 3.5.3: The image of a Cauchy sequence under a uniformly continuous function is also Cauchy.

Proof: Let x_n be a Cauchy sequence in *X* and let $f:(X, d) \to (Y, d^*)$ be uniformly continuous. We are to show that $f(x_n)$ is Cauchy in *Y*. Let $\varepsilon > 0$ be arbitrary. Since *f* is uniformly continuous, there exists a $\delta > 0$ such that $d^*(f(x), f(y)) < \varepsilon$ whenever $d(x, y) < \delta$. Since x_n is Cauchy, there exists $n_o \in \mathbf{N}$ such that $d(x_m, x_n) < \delta$ for all $m, n \geq n_o$. Thus for $\varepsilon > 0$, there exists $n_o \in \mathbf{N}$ such that $d^*(f(x_n), f(x_n)) < \varepsilon$ since $d(x_m, x_n) < \delta$ for all $m, n \geq n_o$. Hence the result.

Theorem 3.5.4: If *X* be a dense subspace of a complete metric space (M, d_1) and $f:(x, d_1) \to (Y, d_2)$ be uniformly continuous and *Y* is complete, then *f* admits of a unique uniformly continuous extension $\bar{f}$ to *M*.

Proof: We shall first define $\bar{f}: M \to Y$ as follows:

$$\bar{f}(x) = f(x) \text{ if } x \in X$$
$$= \lim f(x_n) \text{ where } x_n \to x, x_n \in X.$$

Note since *X* is dense in *M*, such a sequence can be chosen and the convergence of x_n implies that x_n is Cauchy and therefore $f(x_n)$ is Cauchy which is convergent as *Y* is complete.

Clearly $\bar{f}$ is an extension of *f* to *M*. So it is enough to prove that $\bar{f}$ is uniformly continuous.

To this end, let $\varepsilon > 0$ be arbitrary. Since *f* is uniformly continuous there exists a $\delta > 0$ such that $d_1(a, b) < \delta$ implies $d_2(f(a), f(b)) < \varepsilon/2$ for $a, b \in X$. We shall show that for $p, q \in M$, $d_1(p, q) < \delta$ implies $d_2(f(p), f(q)) < \varepsilon$. Since *X* is dense in *M*, there exist sequences p_n and q_n such that p_n converges to *p* and q_n converges *q*. By the triangle inequality we see now

$$d_1(p_n, q_n) \leq d_1(p_n, p) + d_1(p, q) + d_1(q, q_n)$$

Since p_n tends to *p* and q_n tends to *q*, then there exists n_o such that $d_1(p_n, p) < \delta/3$, $d_1(q_n, q) < \delta/3$ for all $n \geq n_o$. So if $d_1(p, q) < \delta$, then $d_1(p_n, q_n) < \delta$ and hence $d_2(f(p_n), f(q_n)) < \varepsilon/2$.

So $d_2(p, q) = \lim d_2(p_n, q_n) \leq \varepsilon/2 < \varepsilon$.

This proves that $\bar{f}$ is uniformly continuous. The uniqueness $\bar{f}$ is obvious.

Definition: A function $f:(X, d) \to (X, d)$ is called a *contraction mapping* or simply a *contraction* if there exists $\alpha \in [0, 1)$ such that

$$d(f(a), f(b)) \leq \alpha d(a, b) \text{ for every } a, b \in X.$$

It is easy to see that every contraction mapping is continuous.

Definition: A point *x* of the metric space (*X*, *d*) is called a *fixed point* of the function $f:(X, d) \to (X, d)$ if $f(x) = x$.

The following theorem is of basic importance in many branches of mathematics.

Banach Fixed Point Theorem: Every contraction mapping from a complete metric space into itself has a fixed point.

Proof: Let $f : (M, d) \to (M, d)$ be a contraction mapping.

First note that every contraction mapping is (uniformly) continuous. This is clear as for every $\varepsilon > 0$, if we choose $\delta = \varepsilon/\alpha$, then evidently $d(x, y) < \delta$ implies $d(f(x), f(y)) < \varepsilon$. Now we shall define a sequence in *M*. Let x_o be any point of *M*. Define $x_1 = f(x_o)$, $x_2 = f(x_1)$, $x_3 = f(x_2)$, ..., $x_n = f(x_{n-1})$, ... so that $x_n = f^n(x_o)$.

Obviously $d(x_n, x_{n-1}) = d(f(x_{n-1}), f(x_{n-2}))$

$$\leq \alpha d(x_{n-1}, x_{n-2})$$
$$= \alpha d(f(x_{n-2}), f(x_{n-3}))$$
$$\leq \alpha^2 d(x_{n-2}, x_{n-3}) = ...$$
$$\leq \alpha^n d(x_o, x_1).$$

Now we shall prove that the sequence $\{x_n\}$ is Cauchy. To this end note that for $m < n$,

$$d(x_m, x_n) = d(x_m, x_{m+1}) + d(x_{m+1}, x_{m+2}) + ... + d(x_{n-1}, x_n)$$
$$\leq \alpha^m d(x_o, x_1) + \alpha^{m+1} d(x_o, x_1) + ... + \alpha^{n-1} d(x_o, x_1)$$
$$= \alpha^m \{1 + \alpha + \alpha^2 + ... + \alpha^{n-m-1}\} d(x_o, x_1)$$
$$= [\alpha^m (1 - \alpha^{n-m})/(1 - \alpha)]\ d(x_o, x_1)$$

which tends to zero as *m* and hence *n* tend to zero.

Since *M* is complete, x_n converges to a point x_o in *M*. We claim that $f(x_o) = x_o$, i.e., x_o is a fixed point. For this observe only

$f(x_o) = f(\lim x_n) = \lim f(x_n) = \lim x_{n+1} = x_o$ as *f* is continuous.

This completes the proof.

3.6 EQUIVALENCE, HOMEOMORPHISM AND ISOMETRY

Definition: Two metrics d_1 and d_2 defined on a non-empty set are said to be equivalent if they generate the same topology, i.e., the set of open sets obtained by the metric d_1 is same as the set of open sets obtained by the metric d_2 and vice versa.

Thus two metrics are equivalent if for an open sphere centered at a point defined by one metric there is an open sphere centered at the same defined by the other metric contained inside the first sphere and vice versa.

Example 1: The metrics d_1, d_2 and d_3 defined on $\mathbf{R}^2$ as follows are equivalent

$$d_1(x, y) = \{(x_1 - y_1)^2 + (x_2 - y_2)^2\}^{1/2}$$

$d_2(x, y) = \max\{|x_1 - y_1|, +|x_2 - y_2|\}$

$d_3(x, y) = |x_1 - y_1| + |x_2 - y_2|$ where $x = (x_1, x_2)$, $y = (y_1, y_2)$

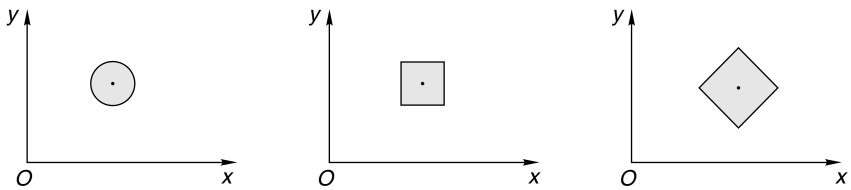


Fig. 6

The following theorem gives a necessary and sufficient condition for the equivalence of two metrics.

Theorem 3.6.1: Two metrics d_1 and d_2 defined of a metric space *M* are equivalent if and only if there exists two non-zero constants α and β such that

$$\alpha\, d_2(x, y) \le d_1(x, y) \le \beta\, d_2(x, y)$$

Proof: Let $S_\varepsilon(x)$ be a sphere in the d_1 metric. Then $S_\varepsilon(x) = \{p \in M; d_1(p, x) < \varepsilon\}$. Since $\alpha\, d_2(p, x) \le d_1(p, x) < \varepsilon$, then $d_2(p, x) \le (1/\alpha)\, d_1(p, x) < \varepsilon/\alpha$. So $S_{\varepsilon/\alpha}(x)$ is sphere in the d_2 metric contained in $S_\varepsilon(x)$.

A similar argument using the second inequality proves that every open sphere in the d_2 metric contains a sphere in the d_1 metric.

Hence the two metrics are equivalent.

It is immediate from the above theorem that

Theorem 3.6.2: Two metrics d_1 and d_2 are equivalent if and only if lim $x_n = x$ in the d_1 metric implies and is implied by lim $x_n = x$ in the d_2 metric.

Proof: It is enough to note

$$d_1(x_n, x) \to 0 \text{ as } n \to 0 \text{ when } d_2(x_n, x) \to 0 \text{ as } n \to 0 \text{ if } d_1(x_n, x) \le k\, d_2(x_n, x).$$

Example 2: Show that the metrics $d^*(x, y) = 1/\{1 + d(x, y)\}$ and $d(x, y)$ defined on *M* are euqivalent.

Solution: Let $x_n \to 0$ as $n \to 0$ in the metric *d*. Then evidently $x_n \to 0$ as $n \to 0$ in the metric *d**.

Conversely let $x_n \to 0$ as $n \to 0$ in the metric *d**. Then note $d^*(x_n, x) > d(x_n, x)$ and hence the result follows.

Example 3: Let *c* denote the set of all convergent real sequences and $x = \{x_n\}, y = \{y_n\} \in c$.

Define $d(x, y) = \sup|x_n - y_n|$ and $d^*(x, y) = \Sigma(1/2^n)\,|x_n - y_n|/\{1 + |x_n - y_n|\}$

Show that the metrics *d* and *d** defined above are not equivalent.

Solution: To prove the result exhibit a sequence which converges in the metric *d** but not in the metric *d*.

Consider the sequence $\{e_n\}$ where $e_0 = (0, 0, \ldots, 0, ...)$, $e_1 = (1, 0, \ldots, 0, ...)$, $e_n = (0, 0, \ldots, 1, ...)$

Note $d(e_n, e_0) = 1$ for each n but $d^*(e_n, e_0) = (1/2^n)\ \{1/(1+1)\} = 1/2^{n+1}$ which tend to zero.

Hence the metrics *d* and *d** are not equivalent.

Definition: A function $f : (X, d_1) \to (Y, d_2)$ is called a *homeomorphism* if

(*i*) *f* is bijective,

(*ii*) *f* and f^{-1} are both continuous.

A metric space (X, d_1) is said to be *homeomorphic* to another metric space (Y, d_2) if there exists a homeomorphism from (X, d_1) to (*Y*, *d*).

Since the bijectivity of *f* implies the bijectivity of f^{-1}, it readily follows that (X, d_1) is homeomorphic to (Y, d_2), then (Y, d_2) is also homeomorphic to (X, d_1). This justifies the often made statement that two spaces are homeomorphic.

Definition: A function $f : (X, d_1) \to (Y, d_2)$ is called an *isometry* if

$$d_2(f(a),\ f(b)) = d_1(a, b) \text{ for every } a, b \in X.$$

Evidently every isometry is injective since $f(x_1) = f(x_2)$ implies $d_2(f(x_1),\ f(x_2)) = 0$ and hence $d(x_1,\ x_2) = 0$, i.e., $x_1 = x_2$. Further, every isometry is uniformly continuous and hence continuous. This is clear as for every $\varepsilon > 0$ there exists a $\delta > 0$ (in fact $\delta = \varepsilon$) such that $d_1(a, b) < \delta$ implies $d_2(f(a),\ f(b)) = d_1(a, b) < \delta = \varepsilon$.

Two metric spaces are said to be *isometric* if there exists a surjective isometry from one of the metric spaces to the other. Note that the surjectivity of the isometry ensures the existence of an isometry from the second metric space to the first.

It is straightforward to verify that isometry of two metric spaces implies homeomorphism but not conversely. In fact it should be borne in mind that isometry preserves the metric properties only whereas homeomorphism preserves the topological properties. Thus all isometric metric spaces have the same metric properties but homeomorphic spaces have the same topological properties. Note completeness is only a metric property and not a topological property.

3.7 COMPACTNESS

Compactness is one of the most important features of any topological space and plays the most significant role in all of topology. Though in the general topological spaces compactness is somewhat intangible, in metric spaces it is very much tangible and more so in the Euclidean spaces.

Definition: A family of sets $\{G_\alpha\}$ is called a cover of a metric space (*M*, *d*) if $M = \cup\, G_\alpha$. If the family is finite, the cover is called a *finite cover*, if it is countable or uncountable, the cover is accordingly called *countable* or *uncountable*. If all the members of the cover are open sets, the cover is called an *open cover*; likewise if all the members are closed sets, the cover is called a *closed cover*. A subfamily of $\{G_{\alpha i}\}$ which also covers *M* is called a subcover of *M*.

The metric sapce (**R**, d) equipped with the usual metric d has a finite cover $\{(-\infty, 1), (0, \infty)\}$, a countable cover $\{(n, n+2); n \in \mathbf{N}\}$ and an uncountable cover $\{(x-1, x+1); x \in I\}$.

A family of sets $\{G\}$ is called a *cover* of a set K in a metric space (M, d) if $K \subset \cup G_\alpha$. The finiteness, countability and uncountability of the set are defined as above.

Definition: A metric space (M, d) is said to be *compact* if every open cover of it has a finite subcover.

A metric space which is not compact is called *non-compact*.

A subset K of a metric space (M, d) is called *compact* if every open cover of K has a finite subcover.

It is easy to see that a set K is compact if and only it is compact as a subspace with respect to the induced metric.

As examples of compact sets one can see that every finite set of a metric sapce is compact. We have already seen that a closed and bounded interval in **R** is a compact set with respect to the usual metric. An example of a set which is not compact is the open interval (0, 1) when equipped with the induced usual metric. The set **R** equipped with the usual metric is also not compact. In fact, any infinite set equipped with the discrete metric is not compact.

Definition: A metric space is called *sequentially compact* if every sequence in it has a convergent subsequence.

A set in a metric space is called *sequentially compact* if every sequence in it has a convergent subsequence.

As an example one should note every finite metric space is sequentially compact. If we take any sequence in the metric space at least one term of the sequence will be repeated infinitely many times. The constant sequence made of that element is a subsequence of the original sequence and evidently that converges to the same point. The same argument applies to a sequentially compact metric space.

The metric space (0, 1) equipped with the induced usual metric is not sequentially compact since the sequence $\{1/n\}$ has all its subsequences converging to 0 which is not a point of the metric space. An analogous argument proves that (0, 1) is not a sequentially compact set in **R**.

Definition: A metric space is called *countably compact* if every countable cover of it has a finite subcover.

A set in a metric space is *countably compact* if every countable cover of it has a finite subcover.

Clearly every closed and bounded interval equipped with the induced usual topology of **R** is countably compact since it being compact by Heine Borel theorem, every cover of it, whether countable or uncountable, has a finite subcover. The argument applies for the set [0, 1] also.

The interval (0, 1) equipped with the induced usual metric of R is not countably compact as the open cover $\{(1/(n+1), 1/n); n \in N\}$ has no finite subcover.

Definition: A metric space is called *locally compact* if each of its points has a neighbourhood whose closure is compact.

The set **R** with the usual metric is locally compact but not compact. This is clear because every point x of R has a neighbourhood $(x-\delta, x+\delta)$ whose closure $[x-\delta, x+\delta]$ is compact by the classical Heine Borel theorem. The space **Q** equipped with the induced metric is not locally compact.

The first thing that we can say about a compact set is the following:

Theorem 3.7.1: Every compact set in a metric space is closed.

Proof: We first prove a simple result:

For every pair of distinct points x and y in a metric space (M, d), there exist two disjoint open spheres centered at the point x and y respectively.

Since $x \neq y, d(x, y) > 0.$ So if we choose $\varepsilon < (1/2)\ d(x, y)$, Clearly $S_\varepsilon(x) \cap S_\varepsilon(y) = \phi$.

Now let K be compact in M. To prove that K is closed, it is enough to prove that K^c is open, i.e., every point of K^c is an interior point.

Let x be an arbitrary point of K^c. Then for every $y \in K$, there exist spheres $S_\varepsilon(x)$ and $S_\varepsilon(y)$ such that $S_\varepsilon(x) \cap S_\varepsilon(y) = \phi$. Evidently $\{S_\varepsilon(y);\ y \in K\}$ is a cover of K. Since K is compact, this cover has a finite subcover, say, $\{S_{\varepsilon_1}(y_1), S_{\varepsilon_2}(y_2), \ldots, S_{\varepsilon_n}(y_n)\}$, i.e., $K \subset \cup\, S_{\varepsilon i}(y_i)$. Then $\cap\, S_{\varepsilon i}(x)$ is open and disjoint from K. Hence x is an interior point. This implies that K is closed.

A natural question that arises in this connection is whether closed subsets of a metric space are also compact or not. The answer to this question is 'no'. But what can be said in this regard is contained in the following result:

Theorem 3.7.2: Every closed subset of a compact metric space is compact.

Proof: Let the metric space (M, d) be compact and let K be a closed subset of M. To prove that K is compact we take a cover $\{G_\alpha\}$ of K, i.e., $K \subset \cup\, G_\alpha$. Then the family $\{K^c, G_\alpha\}$ is a cover of M. Since M is compact, this family must have a finite subfamily which also covers M and this finite subfamily must contain K^c also. If K^c is taken out from this subfamily, the remaining sets shall cover K. Hence we get a finite subfamily of G which covers K. This means K is compact.

Definition: A family of sets $\{F_i\}$ is said to have *finite intersection property* (FIP) if every finite number of members of the family $\{F_i\}$ has non-empty intersection.

For example the family $\{(-n, n);\ n \in \mathbf{N}\}$ has the finite intersection property but the family $\{(n, n+1);\ n \in \mathbf{N}\}$ does not have the finite intersection property.

The finite intersection property has an intimate relation with compactness as will be evident from the following theorem:

Theorem 3.7.3: A metric space (M, d) is compact if and only if every family of closed subsets of M having finite intersection property has the entire intersection non-empty.

Proof: Let (M, d) be compact and let $\{F_i\}_{i \in I}$ be an arbitrary family of closed subsets of M having finite intersection property. If possible, let $\cap\, Fi = \phi$.

Then $\bigcup_{i \in I} Fi^c = M$, i.e., $\{Fi^c\}_{i \in I}$ is an open cover of M. Since M is compact, there is a finite subcover of this cover, say, $\{F_1{}^c, F_2{}^c, \ldots, Fm^c\}$, i.e., $M = F_1{}^c \cup F_2{}^c \cup \ldots \cup Fm^c$. But this means $F_1 \cap F_2 \cap \ldots \cap F_m = \phi$ which contradicts the fact that $\{F_i\}_{i \in I}$ has the fintie intersection property.

Hence $\bigcap_{i \in I} F_i \neq \phi$.

Conversely let every family of closed subsets of M having finite intersection property has the entire intersection non-empty. Let $\{G_\alpha\}$ be an open cover of M, i.e., $M = \cup G_\alpha$.

Then $\cap G_\alpha^c = \phi$. So the family $\{G_\alpha^c\}$ of closed sets do not have the finite intersection property since otherwise the intersection of the entire family would have been non-empty. So there exists a finite subfamily of $\{G_\alpha^c\}$ with empty intersection i.e., $G_1 \cap G_2 \cap \ldots \cap G_n = \phi$. Hence $M = \cup G_i$. This means M is compact.

Definition: A finite set F of points of K is called an ε*-net* for K if for every point $x \in K$, there exists a point $y \in F$ such that $d(x, y) < \varepsilon$.

A set K in a metric space (M, d) is called *totally bounded* if K possesses an ε*-net* for every $\varepsilon > 0$.

For example, the set $F = \{ (1, -1), (1, 0), (1, 1), (0, -1), (0, 0), (0, 1), (-1, -1), (-1, 0), (-1, 1) \}$ is an 3/2-net for $K = \{(x, y) \in R^2; x^2 + y^2 < 4\}$.

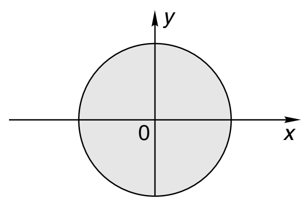


Fig. 7

It should be noted in this connection that every totally bounded set is bounded but not conversely. For example, the set $A = \{\varepsilon_1, \varepsilon_2, \varepsilon_3, \ldots, \varepsilon_n, \ldots\}$ where $\varepsilon_1 = (1, 0, 0, \ldots)$, $\varepsilon_2 = (0, 1, 0, \ldots), \ldots$, $\varepsilon_n = (0, 0, 0, \ldots, 1, \ldots), \ldots$ is a bounded set in l^2, since $d(\varepsilon_i, \varepsilon_j) = \sqrt{2}$ and therefore $d(A) = \sup \{d(\varepsilon_i, \varepsilon_j); \varepsilon_i, \varepsilon_j \in A\} = \sqrt{2}$. But A is not totally bounded, since for $\varepsilon = 1/2$, there is no ε-net. In fact the only non-empty sets with diameter less than 1/2 are the singleton sets.

Theorem 3.7.4: If a set K in a metric space (M, d) is compact, then it is totally bounded.

Proof: To prove the result, we must show that for every $\varepsilon > 0$, there exists an ε*-net* for K. Consider the open cover $\{S_\varepsilon(x); x \in K\}$ of K. Because K is compact, the above cover has a finite subcover, say, $\{S_\varepsilon(x_1), S_\varepsilon(x_2), \ldots, S_\varepsilon(x_n)\}$. Evidently the set $\{x_1, x_2, \ldots, x_n\}$ is an ε*-net* for K.

Definition: A metric sapce (M, d) is said to have *Bolzano-Weirstrass property* (BWP) if every infinite set in (M, d) has a limit point.

As for example **R** equipped with the usual metric does not have the Bolzano-Weirstrass property as the infinite set {1, 2, 3, ... } has no limit point.

Theorem 3.7.5: A metric space is sequentially compact if and only if it has the Bolzano-Weirstrass property.

Proof: Let (M, d) be a metric space and let it be sequentially compact. Let A be an infinite subset of M. We shall show that A has a limit point. Since A is infinite, a sequence $\{x_n\}$ of distinct points can be extracted from A. Since M is sequentially compact, there exists a subsequence $\{x_{n_i}\}$ which converges to a point x, say, and x is defacto a limit point of the set of points of the subsequence and hence that of the set of points of the sequence also. This proves the Bolzano-Weirstrass property.

Conversely, let (M, d) have tha Bolzano-Weirstrass property. Let $\{x_n\}$ be an arbitrary sequence in M. If x_n has a point repeated infinitely many times, then it has a constant subsequence which converges to the point repeated. If no point of $\{x_n\}$ is repeated infinitely many times. Then the set of points of the sequence is infinite and therefore by the Bolzano-Weirstrass property, it has a limit point. Now it is easy to extract from $\{x_n\}$ a subsequence which converges to x.

To prove the next important result we need the following results.

Theorem 3.7.6: If a metric space (M, d) is compact, then it has the Bolzano-Weirstrass property.

Proof: Let A be an infinite set in M. We shall show it has a limit point. If possible, let A has no limit point in M, i.e., no point of M is a limit point of A. Then for every point of M, there exists an open sphere centered at that point which does not contain any point of A other than its centre. The family of all such spheres constitute an open cover of M. Now M being compact, this cover has a finite sub-cover. Since A is a subset of M, A is not only covered by the finite subcover but also are the centres of these finitely many spheres of the subcover. Hence A must be finite which is a contradiction. Thus A must have a limit point. This implies that M has the BW property.

Theorem 3.7.7: Every sequentially compact metric space is totally bounded.

Proof: Let (M, d) be sequentially compact. To show M is totally bounded, let $\varepsilon > 0$ be arbitrary. Choose a point a_1 in M and the sphere $S_\varepsilon(a_1)$. If $S_\varepsilon(a_1)$ covers M, then $\{a_1\}$ is the ε*-net*. If not, choose another point a_2 and the sphere $S_\varepsilon(a_2)$. If $S_\varepsilon(a_1) \cup S_\varepsilon(a_2) = M$, then $\{a_1, a_2\}$ is the ε*-net*. Continuing this process we arrive at a set $\{a_1, a_2, \ldots,\}$ which will be finite otherwise this will be an infinite sequence which no convergent subsequence–contary to our assumption of sequential compactness of M. Hence the set $\{a_1, a_2, \ldots,\}$ and therefore also an ε*-net*. This implies that M is totally bounded.

Definition: A positive real number a is called a Lebesgue number for an open cover $\{G_\alpha\}$ of a metric space (M, d), if each subset S of M with $d(S) < a$ is contained in one of the open sets of the cover $\{G_\alpha\}$.

Theorem 3.7.8: In a sequentially compact metric space each open cover of the space has a Lebesgue number. [*Lebesgue Covering Lemma*]

Proof: (M, d) be sequentially compact and let $\cup G_\alpha$ be a cover of M. Assume if possible M has no Lebesgue number. Hence for every $n \in N$, there exists a non-empty set A_n with $d(A_n) < 1/n$ such that A_n in not contained in G_α for each α. So we can generate a sequence $\{x_n\}$ choosing x_n from A_n for each n. This sequence $\{x_n\}$ has a convergent susequence $\{x_{n_i}\}$ as M is sequentially compact. Let lim $x_{n_i} = x$. Clearly $x \in G_{\alpha 0}$ for some α as $\{G_\alpha\}$ is a cover of M and $x \in M$. Now $G_{\alpha 0}$ being open there exists an $\varepsilon > 0$ such that $S_\varepsilon(x) \subset G_{\alpha 0}$. For this ε we thus see that there exists $i_0 \in \mathbf{N}$ such that $d(x_{ni_o}, x) < \varepsilon/2$ and $d(A_{nio}) < \varepsilon/2$.

Thus for any element $y \in An_{i_o}$, we see

$$d(y, x) < d(y, xn_{i_o}) + d(xn_{i_o}, x) < \varepsilon/2 + \varepsilon/2 = \varepsilon.$$

This implies that $y \in S_\varepsilon(x) \subset G_\alpha$. But this contradicts the fact that for each natural number n, A_n is not contained in any of the open sets of the cover. Hence $\{G_\alpha\}$ must have a Lebesgue number.

We are now in a position to prove the main result.

Theorem 3.7.9: Let (M, d) be a metric space. Then the following are equivalent:

(i) M is compact

(ii) M is sequentially compact

(iii) M has the Bolzano-Weirstrass property

Proof: $(i) \Rightarrow (ii)$.

Let M be compact. Then by Theorem above it has the BW propety and by Theorem then it must be sequentially compact.

$(ii) \Rightarrow (i)$

Let M be sequentially compact. Let $\{G_\alpha\}$ be an open cover of M. So it must have a Lebesgue number a, say. If we take $\varepsilon = a/3$, then by the total boundedness of the sequentially compact space M we get ε-*net* $\{a_1, a_2, \ldots, a_n\}$. For each i = 1, 2, ..., n, we have $d(S_\varepsilon(a_i)) \le 2\varepsilon = 2a/3 < a$.

Now by the definition of Lebesgue number, for each I, there exists a $G_{\alpha i}$ such that $S_\varepsilon(a_i) < G_{\alpha i}$. Since every point of M belongs to one of the $S_\varepsilon(a_i)'$s, the family $\{G_{\alpha_1}, G_{\alpha_2}, \ldots, G_{\alpha_n}\}$ is a finite subcover of M. Hence M is compact.

$(ii) \Leftrightarrow (iii)$

It is already proved in theorem.

It is natural to ask the question now whether every closed and totally bounded set is compact. The answer to this question is assertive in a complete metric space.

We now prove two very useful theorems.

Theorem 3.7.10: The continuous image of a compact set is compact.

Proof: Let $f:(M, d) \to (N, d')$ be continuous and let K be compact in M. Then to show $f(K)$ is compact in (N, d'), assume $\{G_\alpha\}$ to be an open cover of $f(K)$, i.e., $f(K) \subset \cup\, G_\alpha$. Since f is continuous, $f^{-1}(G_\alpha)$ is open for each α. Further $K \subset \cup\, f^{-1}(G_\alpha)$, i.e., $\{f^{-1}(G_\alpha)\}$ is an open cover of K. Since K is compact it must have a subcover, say, $\{f^{-1}(G_1), f^{-1}(G_2), \ldots, f^{-1}(G_n)\}$. The clearly $f(K) \subset G_1 \cup G_2 \cup \ldots \cup G_n$. Hence $f(K)$ is compact.

Corollary: A real valued continuous function with compact domain is bounded.

This follows from the classical Heine Borel theorem.

Theorem 3.7.11: A continuous funtion with compact domain is uniformly continuous.

Proof: Let $f:(M, d) \to (N, d')$ be a continuous function and let M be compact. Since f is continuous at every point of M, for every $p \in M$ and every $\varepsilon > 0$, there exists a $\delta(p) > 0$ such that $x \in S_{\delta(p)}(p)$ implies $f(x) \in f^{-1}(S_{\varepsilon/2}(f(p)))$. Thus $\{(S_{\delta(p)}(p));\ p \in M\}$ form a cover of M. Since M is compact and hence sequentially compact, there is a Lebesgue number $\delta > 0$ corresponding to this open cover.

Now let $x, y \in M$ and $d(x, y) < \delta$. Since $d(\{x, y\}) = d(x, y) < \delta$, there exists an open sphere $S_{\delta(p_0)}(p_0)$ of the cover such that $\{x, y\} \subset S_{\delta(p_0)}(p_0)$ and hence $f(x), f(y) \in S_{\varepsilon/2}(f(p_0))$. But as $d(S_{\varepsilon/2}(f(p_0))) < \varepsilon$, we conclude.

$d(x, y) < \delta$ impliese $d'(f(x), f(y)) < \varepsilon$.

This implies *f* is uniformly continuous.

3.8 CONNECTEDNESS

Intuitively connected means joined and thus a connected space is one, which consists of a single piece. This is one of the simplest properties of a space yet one of the most important so far as applications to real and complex analyses is concerned. Spaces that are not connected are also quite interesting. At the opposite end of the connectivity spectrum lies a totally disconnected space and the Cantor set is one such example. The space or rational numbers equipped with the induced metric from *R* is also a space with this opposite feature.

Definition: A metric space (*M*, *d*) is said to be *connected* if it cannot be expressed as the union of two non-empty disjoint open sets. Thus if $M = A \cup B$ where *A* and *B* are non-empty and $A \cap B = \phi$, the space *M* is disconnected. If a space is not disconnected, it is called connected. Evidently *A* and *B* are closed sets also as $A = B^c$ and $B = A^c$.

Note the connectedness of a space *M* assures that the set *M* and ϕ are the only clopen sets (sets which are both open and closed).

A subspace *X* of *M* is called *connected* if *X* is connected in its own right. An expression of a space as the union of two disjoint open sets is sometimes referred to as a disconnection of the space.

Examples of connected spaces are **R** and **C** equipped with the usual metric. An interval, closed or open or semi-closed, equipped with the relative topology by the induced metric from **R** is also a connected space. The unit disc, the elliptic region, the square region with side of length one, with or without the (topological) boundary are examples of connected spaces when equipped with the induced metric from $\mathbf{R}^2$.

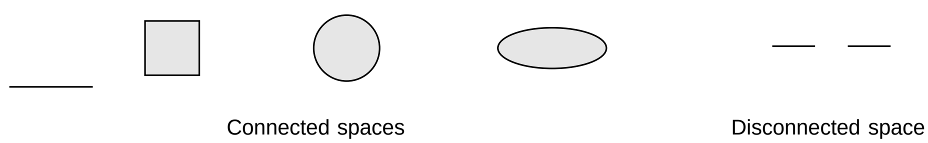


Fig. 8

Analogously, a set *X* in metric space (*M*, *d*) is called *disconnected* if there exists open subsets *G* and *H* of *M* such that $X \cap G$ and $X \cap H$ are disjoint non-empty sets whose union is *X*. In this case, $G \cup H$ is called a disconnection of *X*.

A set *X* is called *connected* if it is not disconnected.

Note that $X = (X \cap G) \cup (X \cap H)$ if and only if $X \subset G \cap H$.

And $(X \cap G) \cup (X \cap H) = \phi$ if and only if $G \cap H \subset X^c$.

So $G \cup H$ is a disconnection of *X* if and only if $X \cap G \neq \phi$, $X \cap H \neq \phi$, $X \subset G \cup H$ and $G \cap H \subset X^c$.

Note that the set ϕ is connected and so is any singleton set.

Clearly the set $X = (1, 2) \cup (2, 3)$ is disconnected in **R** as $X \cap (-\infty, 2)$ and $X \cap (2, \infty)$ are two non-empty disjoint open subsets of *R* whose union is *X*.

Similarly the set $S = \{(x, y); x^2 - y^2 \geq 4\}$ is disconnected in $\mathbf{R}^2$ as the open sets $G = \{(x, y); x < -1\}$ and $H = \{(x, y); x > 1\}$ are such that $S = (S \cap G) \cup (S \cap H)$ and $(S \cap G) \cap (S \cap H) = \phi$.

It readily follows that a set in a metric space is connected if and only if it is connected as a subspace of the metric space.

We now prove an impressive result.

Theorem 3.8.1: A subspace of **R** is connected if and only if it is an interval of **R**.

Proof: Let *X* be a subspace of **R**. We shall show that if *X* is not an interval, it is not connected as a subspace of **R**. The fact that *X* is not an interval ensures the existence of real numbers *x*, *y* and *z* such that *x* and *z* are in *X* but *y* is not in *X*. Clearly the $X = [X \cap (-\infty, y)] \cup [(y, \infty) \cap X]$ is a disconnection of *X*.

Conversely, let *X* be an interval of *X*. We show then that *X* is connected. Let assume the contrary and arrive at a contradiction. Since by assumption *X* is not connected, there exists a disconnection of *X*, say, $X = A \cup B$, where *A* and *B* are disjoint non-empty relatively open subsets of *X*. Choose *x* in *A* and *z* in *B* and without loss of generality assume *x* < *z*. But since *X* is an interval, $[x, z] \subset X$. Again as $X = A \cup B, [x, z]$ is either contained entirely within *A* or within *B* as *A* and *B* are disjoint. Define $y = \sup ([x, z] \cap A)$. Then $x \leq y \leq z$ and so *y* is in *X*.

But as *A* is closed, $y \in A$. From this it follows *y* < *z*. But the definition of *y* we get $y + \varepsilon$ is in *B* for every $\varepsilon > 0$ such that $y + \varepsilon \leq z$. Since *B* is closed, $y \in B$. But this contradicts the fact that *A* and *B* are disjoint. Hence *X* must be connected.

Remark: A similar statement can be made about a connected set.

Corollary: The real line **R** is connected.

The next theorem asserts that connectedness is a topological property.

Theorem 3.8.2: The continuous image of a connected space is connected.

Proof: Let (*X*, *d*) be connected and let $f : (X, d) \to (Y, d^*)$ be continuous. To show that *f*(*X*) is a connected subspace of *Y*, let us assume the contrary, i.e., there is a disconnection of *f*(*X*) as $f(X) = A \cup B$ where $A = G \cap f(X), B = H \cap f(X)$, *G* and *H* open in *Y* and $A \cap B = \phi$. Then evidently $x = f^{-1}(G) \cup f^{-1}(H)$, $f^{-1}(G)$ and $f^{-1}(H)$ are nonempty open sets in *x*, $f^{-1}(G) \cap f^{-1}(H) = \phi$. Thus *X* is not connected–a contradiction. Hence *f*(*X*) must be connected.

Remark: A similar statement can be made about a connected set.

Corollary: For a real valued continuous function, the range is an interval.

When a space is not connected, it is quite useful often to see what are the largest connected subsets of that space. Such subsets are known as components. The formal definition is as follows:

Definition: A *component* of a metric space is defined to be a maximal connected subspace of it. Thus a component is not a proper subset of any connected subspace.

As for example, if $X = [0, 1] \cup [2, 5]$ is a metric space equipped with the induced usual metric, then [0, 1] and [2, 5] are two components of X but [3, 4] is not a component.

Definition: A metric space (M, d) is called *totally disconnected* if for every pair of distinct points x and y in M, there is a disconnection $X = A \cup B$ such that $x \in A$ and $y \in B$.

The discrete spaces are examples of totally disconnected spaces. The set **Q** of rationals and the set **I** of irrationals when euipped with the induced metric from **R** are also totally disconnected. A compact space which is totally disconnected is the Cantor set.

A result of interest is the following:

Theorem 3.8.3: The components of a totally disconnected space are its points.

Proof: Let (M, d) be a totally disconnected space. It suffices to prove that every subspace of M containing more than one point is disconnected. To this end let x and y be two points in a subspace X of M and let $X = A \cup B$ be a disconnection of X with $x \in A$ and $y \in B$. Then evidently $X = (X \cap A) \cup (X \cap B)$ is a disconnection of X. Hence the result.

Definition: A metric space every point of which has a connected neighbourhood is called *locally conntected*.

Note that every discrete space is locally connected. The set **R** equipped with the usual metric is locally connected.

Definition: A *path* from a point a to a point b in metric space (M, d) is by definition a continuous function $f : [0, 1] \to M$ with $f(0) = a$ and $f(1) = b$. The point a is called the *initial point* of the path and the point b is called the *terminal point* of the path.

For any point p, the constant function $e_p : [0, 1] \to M$ defined by $e_p(x) = p$ is called the *constant path* at p.

A path $f : [0, 1] \to M$ is called a *closed path* if the initial point coincides with the terminal point, i.e., $f(0) = f(1)$.

A paths $f : [0, 1] \to M$ is said to be *homotopic* to $g : [0, 1] \to M$ if there exists a continuous function $H : I^2 \to M$ where $I = [0, 1]$ such that

$$\begin{aligned} H(s, 0) &= f(s), \quad H(0, t) = p \\ H(s, 1) &= g(s), \quad H(1, t) = q. \end{aligned} \qquad \text{where } p, q \in M$$

The function H is called a *homotopy* from f to g.

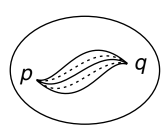


Fig. 9

A closed path $f : [0, 1] \to M$ is said to be *contractible* to a point p if it is homotopic to the constant path e_p.

A metric space (M, d) is called *simply connected* if every closed path in M is contractible to a point. For example an open disc in $\mathbf{R}^2$ is simply connected but an annular region is not.

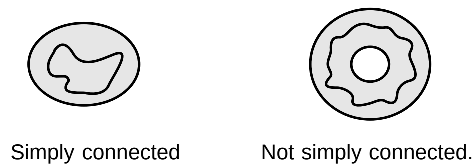


Fig. 10

Note that the image of the connected space [0, 1] is connected and therefore is a curve without any break. As for example the function $f(x) = (x^2, 2x)$ defined on [0, 1] gives a path in $\mathbf{R}^2$.

Definition: A metric space (M, d) is called *path connected* or *arcwise connected* if for every pair of points a and b in M there is a path from a to b.

A set X in a metric space (M, d) is called *path connected* if for every pair of points a and b in X there is a path from a to b contained entirely within X.

It easily follows that every path-connected space is connected. So also is the case of a set.

Since connectedness is a topological notion and is easily extendable to topological spaces in general, we refrain from proving many other results on connectedness here which can be seen in any book of topology.

CHAPTER 4

Topological Spaces

The peculiar properties enjoyed by metric spaces and the topology of **R** and $\mathbf{R}^2$ admit of far more generality demostrating a beauty of its own and free from the specter of real numbers.

4.1 SOME DEFINITIONS

We begin with a non-empty set *X*.

Definition: A family *T* of subsets of *X* is called a *topology* of *X* if the following conditions are satisfied:

(*i*) $\phi, X \in T$.

(*ii*) Union of any arbitrary number of sets from *T* is also in *T*.

(*iii*) Intersection of finite number of sets from *T* is also in *T*.

The pair (*X*, *T*) is called a *topological space*. The members of *T* are called *open sets* of *X*.

The complements of open sets in *X* are called the *closed sets* of *X*.

Note that many topologies can be defined on the same set. Referring to the topology is necessary when referring to a topological space unless one works with only one and fixed topology.

The topology consisting of only two subsets, namely ϕ and *X*, is called the *indiscrete topology* of *X* and will be referred to by *I*.

The topology consisting of all subsets of *X* is called the *discrete topology* of *X* and will be referred to by *D*.

Clearly every singleton set is an open set in a discrete space.

If T_1 and T_2 are two topologies of *X* and $T_1 \subset T_2$ i.e., every open set of T_1 is an open set in T_2, then T_1 is said to be *weaker than* T_2 or T_2 is said to be *stronger than* T_1. Clearly *I* is the weakest of all topologies on *X* and *D* is the strongest of all topologies on *X*.

Example 1: Let $X = \{a, b, c\}$. Then $T_1 = \{\phi, \{a\}, \{a, b\}, X\}$ is a non-trivial topology of *X*. $T_2 = \{\phi, \{a\}, \{a, b\}, \{a, c\}, X\}$ is another non-trivial topology. $I = \{\phi, X\}$ is indiscrete topology and $D = \{\phi, \{a\}, \{b\}, \{c\}, \{a, b\}, \{b, c\}, \{c, a\}, X\}$ is the discrete topology.

Example 2: Let *X* be an infinite set. Let *T* consists of the empty set ϕ and all subsets of *X* whose complements are finite. Then it is easy to verify that *T* is a topology, called the *cofinite topology*.

Example 3: Let *X* be an uncountable set, e.g., **R** and let *T* consist of the empty set ϕ and all subsets of *X* whose complements are countable. It readily follows that *T* is a topology, called the *cocountable topology* of *X*.

It is to be noted that a set may be both open and closed simultaneously or may be neither. For example, the sets X and ϕ are both open and closed. A proper subset of X may be open and closed simultaneously, e.g., in the discrete topology every set is closed and open. Such sets will be sometimes referred to as *clopen* sets of X.

Note further that the intersection of any two topologies is also a topology of X but the union need not be. For example, consider $X = \{a, b, c\}$ and two topologies $T_1 = \{\phi, \{a\}, X\}$ and $T_2 = \{\phi, \{b\}, X\}$. Clearly $T_1 \cup T_2$ is not a topology of X.

A result that readily follows from the properties of open sets is

Theorem 4.1.1: For a topological space (X, T), the following are true:

(*i*) ϕ and X are closed sets

(*ii*) Intersection of any number of closed sets is closed

(*iii*) Union of finitely many closed sets is closed.

4.2 NEIGHBOURHOOD, INTERIOR, EXTERIOR AND BOUNDARY

Let (X, T) be a topological space and $p \in X$.

Definition: A subset N_p of X is called a *neighbourhood* of p if there exists an open set G of X such that $p \in G \subset N_p$.

If N_p itself is open, N_p will be called an *open neighbourhood*. Similarly if N_p is closed, then N_p will be called a closed neighbourhood. The set $N_p - \{p\}$ will be called the deleted neighbourhood of p and will be denoted sometimes by N_p.

For example, if U denotes the usual topology of **R**, then (0, 1] is a neighbourhood of 1/3, (0, 1) is an open neighbourhood of 1/3, [0, 1] is a closed neighbourhood of 1/3 and (0, 1] – {1/3} is a deleted neighbourhood of 1/3.

Definition: A point p of X is called an *interior point* of a set K of X if there exists a neighbourhood N_p of p such that $p \in N_p \subset K$. The set of all interior points of K is called the *interior* of K and is denoted by K^0 or int(K). Note the interior of a set may be empty. For example, **Q** has empty interior in the usual topology of **R**. The point 1/3 is an interior of point of (0, 1], so also is the point 0 but 1 is not an interior point. The interior of (0, 1] is (0, 1). In the topological space (X, T) with $X = \{a, b, c\}$ and $T = \{\phi, \{a\}, \{a, b\}, X\}$, the point a is an interior point and the interior of $\{a, c\}$ is $\{a\}$.

Theorem 4.2.1: In a topological space the following are euqivalent:

(*i*) G is open

(*ii*) $G = G^0$

Theorem 4.2.2: In a topological space (X, T), the following are equivalent:

(*i*) $K^0 = \cup\{G \subset K; G \in T\}$

(*ii*) K^0 is the largest open set contained in K.

Definition: A point p of X is called an *exterior point* of a set K in X if p is an interior point of K^c. The set of all exterior points of K is called the *exterior* of K and will be denoted by ext(K). Note the exterior of a set may be empty also. For example the exterior of **Q** in **R** is empty since its complement **I** has also an empty interior.

The exterior of (0, 1] is $(-\infty, 0) \cup (1, \infty)$. It should be remebered that the complement of the interior of *K* is not the exterior of *K* nor is the vice versa.

Definition: A point *p* of *X* is called a *boundary point* of a set *K* if every neighbourhood of *p* intersects both *K* and K^c, i.e., $K \cap N_p \neq \phi$ and $K^c \cap N_p \neq \phi$. The set of all boundary points of *K* is called the *boundary* of *K* and will be denoted by bdry (*K*). Since there are other concepts of boundary, this boundary will be sometimes be referred to as the topological boundary of *K*. Note the boundary of a set may be empty.

For example, if *K* = (0, 1), then bdry(*K*) = {0, 1} in (**R**, *U*), Observe that a boundary point of a set may or may not belong to the set. In the space (*X*, *T*) with *X* = {*a*, *b*, *c*} and $T = \{\phi, \{a\}, \{a, b\}, X\}$, the boundary of {*a*, *b*} is empty.

Definition: A point *p* of *X* is called a *limit point* or *accumulation point* of a set *K* if every deleted neighbourhood of *p* intersects *K*, i.e., every neighbourhood of *p* intersects *K* in at least one point other than *p*. The limit point may or may not belong to the set. The set of limit points of *A* is called the *derived set* of *K* and will be denoted by K'. A point *p* is called an *isolated point* of *K* if there exists a neighbourhood of *p* disjoint from *K*, i.e., $\exists\, N_p \ni K \cap N_p = \phi$. The set $K \cup K'$ is called the *closure* of *K* and is denoted by $\overline{K}$. A set *K* is said to be *closed* if it contains all its limit points, i.e., $K' \subset K$.

In (**R**, *U*), the set $K = \{1, 1/2, 1/3, \ldots\}$ has only one limit point 0, which does not belong to the set. The point 1/2 is an isolated point of the set. Hence $K' = \{0\}$. Note the derived set a set may be empty. For example, every finite set in (**R**, *U*) has an empty derived set.

The closure of (0, 1) is [0, 1].

Theorem 4.2.4: The following are equivalent in a space (*X*, *T*)

(*i*) $\overline{K} = \cap \{F \supset K;\ F$ is closed in $X\}$

(*ii*) $\overline{K}$ is smallest closed set containing *K*.

A result characterizing closed sets is the following.

Theorem 4.2.5: In (*X*, *T*) the following are equivalent

(*i*) *F* is closed

(*ii*) *F* = *F*

(*iii*) $F' \subset F$.

The proofs of the above results are straightforward and are left here.

Definition: A set *K* in a topological space (*X*, *T*) is called *dense* (everywhere dense) if $\overline{K} = X$.

For example, the set **Q** of rational is dense in (**R**, *U*) since $\mathbf{Q}' = \mathbf{R}$ and then $\overline{\mathbf{Q}} = \mathbf{R}$. Note the set **I** of irrational is dense in **R** also.

The following properties of neighbourhoods of a point of a topological space can be easily verified:

Theorem 4.2.6: In a topological space (*X*, *T*), let $\mathcal{N}_p$ denote the family of all neighbourhood of a point *p* of *X*, then the following are true (*i*) *p* belongs to each member of $\mathcal{N}_p$.

(*ii*) Every superset of a member of $\mathcal{N}_p$ belongs to $\mathcal{N}_p$.

(*iii*) For every member *N* of $\mathcal{N}_p$ there is another member *G* of $\mathcal{N}_p$, which is the neighbourhood of each of its points.

The family $\mathcal{N}_p$ mentioned here is sometimes referred to as the neighbourhood system of *p*. Note $\mathcal{N}_p$ is always non-empty as $X \in \mathcal{N}_p$.

Definition: A topological space (*X*, *T*) is called *separable* if *X* has a countable dense subset.

Clearly *R* is separable since **Q** is a countable dense subset of **R**.

4.3 RELATIVE TOPOLOGY AND SUBSPACE

Let (*x*, *T*) be a topological space and let $Y \subset X$. A natural topology can be defined on *Y* from *T* as follows:

Definition: The relative topology T_Y of *Y* is the family of those subsets of *Y* which can be obtained as the intersection of a member of *T* and *Y*, i.e.,

$$T_Y = \{G \cap Y;\, G \in T\}.$$

Thus if $X = \{a, b, c\}$, $T = \{\phi, \{a\}, \{a, b\}, X\}$ and $Y = \{b, c\}$, then $T_Y = \{\phi, \{b\}, Y\}$ is the relative topology of *Y*. Similarly for (**R**, *U*) and $Y = (0, 1], (0, 1/3) \in T_Y$ since $(0, 1/3) = Y \cap (-1, 1/3)$ and $(-1, 1/3) \in U$. Also $(1/2, 1] \in T_Y$ since $(1/2, 1] = Y \cap (1/2, 2)$ and $(1/2, 2) \in U$. Note (1/2, 1) is an open set in *Y* but it is not open in *u*.

It should be remembered that whenever we refer to a subset *Y* of **R** as a space, unless otherwise specifically mentioned, we shall understand that *Y* is equipped with the relative topology obtained from *u*.

4.4 BASE AND SUBBASE OF A TOPOLOGY

Let (*X*, *T*) be a topological space.

Definition: A class *B* of open sets of *X* is called a **base** for the topology *T* if every open set of *X* is the union of members of *B*.

Note that *B* is a base for *T* iff for every point *p* belonging to an open set *G* of *X*, there exists $B \in B$ with $p \in B \in B$. Further, if $B \subset C$, then *C* is also a base for *T*.

Example 1: The (finite) open intervals of **R** form a base for its usual topology *U*, i.e., $U = \{(a, b);\, a, b \in \mathbf{R}\}$. It is easy to see that for any point *p* belonging to an open set *G* of *R*, there exists an open interval (*a*, *b*) such that $p \in (a, b) \subset G$. Similarly the open discs of $\mathbf{R}^2$ form a base for the usual topology of $\mathbf{R}^2$. The open rectangles with sides parallel to the coordinate axes also form a base for the usual topology of $\mathbf{R}^2$.

Example 2: For any discrete topological space (*X*, *D*), the family of all singleton sets form a base for *D*.

A natural question that arises in this connection is 'under what conditions can a family of subsets of a non-empty set be a base for a topology of X ?'. The following result gives an answer to this question.

Theorem 4.4.1: A non-empty class of subsets of a set X is a base for a topology of X iff

(*i*) $X = \{B;\ B \in B\}$

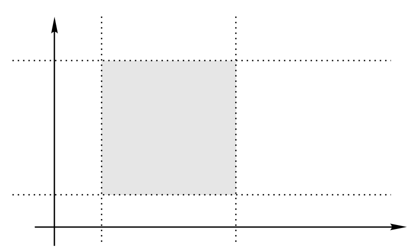
(*ii*) For any $B_1, B_2 \in B$, $B_1 \cap B_2$ is the union of members of B, i.e., if $p \in B_1 \cap B_2$, then there exists $B' \in B$ such that $p \in B' \subset B_1 \cap B_2$.

Example 1: If B denotes the class of all open-closed intervals of **R**, i.e., $B = \{(a, b];\ a, b \in \mathbf{R}\}$, then clearly $\mathbf{R} = \cup \{B;\ B \in B\}$ as for any $p \in \mathbf{R}$, there exists $(a, b] \in R$ such that $p \in (a, b]$. Further the intersection of any two such intervals is either empty or a set of the same form. Hence B is a base of a topology of **R**. This topology is called the *upper limit topology* of **R**, which is different from the usual topology of **R**. Taking intervals of the form $[a, b)$ one can similarly generate another topology of **R**, called the lower limit topology of **R**.

Definition: Let (X, T) be a topological space with a base B. A class of open sets of X is called a subbase for T if finite intersections of sets from S form the base B, i.e., every basic open set can be obtained as the intersection of some members of S.

Example 1: The class of intervals of the form $(-\infty, b)$ and (a, ∞) is a subbase for the usual topology of R, since intersection of sets of the above form gives a basic open set of **R**.

Example 2: The class of infinite open strips of the form $(a, b) \times \mathbf{R}$ and $\mathbf{R} \times (c, d)$ is a subbase for the usual topology of $\mathbf{R}^2$, since finite intersections of such sets give open rectangles of the form $(a, b) \times (c, d)$.



Definition: Let $\mathscr{A}$ be a family of subsets of a topological space X, then the topology generated by $\mathscr{A}$, denoted by $T(\mathscr{A})$, is defined to be the smallest topology containing $\mathscr{A}$. Indeed this topology can be obtained from $\mathscr{A}$ by taking finite intersections of members of $\mathscr{A}$ as the base of a topology which is unique. Evidently $T(\mathscr{A})$ is the intersection of all topologies containing $\mathscr{A}$.

Example: If $X = \{a, b, c, d, e\}$ and $\mathscr{A} = \{\{c, d\}, \{d, e\}, \{a, b, c\}\}$, then the topology generated by A is the topology whose bas is given by

$$B = \{\phi, \{c\}, \{d\}, \{d, e\}, \{a, b, c\},\ X\}$$

and hence $T(\mathscr{A}) = \{\phi, \{c\}, \{d\}, \{c, d\}, \{d, e\}, \{a, b, c\}, \{c, d, e\}, \{a, b, c, d\},\ X\}$.

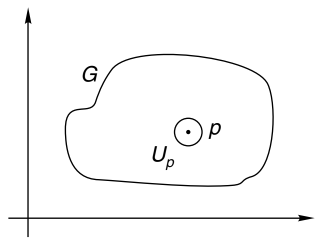
Definition: Let (X, T) be a topological space and p be a point of X. A family B_p of open sets containing p is called a *local base* at p if for each open set G containing p, there exists $U_p \in B_p$ such that $p \in U_p \subset G$.

Example: The family $\{(p-\delta,\ p+\delta),\ \mathbf{R}^+\}$ is a local base at p for the usual topology of R. This is obvious as any open set containing p contains a set of the form $(p-\delta,\ p+\delta)$.

Example: The family of δ - neighbourhoods of a point $p=(x_0,\ y_0)$,

i.e., $\{(x,\ y);\ (x-x_0)^2+(y-y_0)^2<\delta^2\}$

is a local base at p for the usual topology of $\mathbf{R}^2$.



The following two axioms will turn out extremely useful in many situations.

First Axiom of Countability. A topological space has a countable local base at each of its points.

Second Axiom of Countability. A topological space has a countable base.

Definition: A topological space is called *first countable* (*respectively second countable*) if it satisfies the first (respectively second) axiom of countability.

It readily follows from the observation made above that every second countable space is first countable. Further, any subsapce of a second countable space is second countable as the class of all intersections with the subspace is an open base for the subspace. This observation leads to the following result:

Lindeloff's Theorem: Every non-empty open set in a second countable space can be represented by a countable union of basic open sets.

An immediate corollary of the above is the following:

Corollary: Every open base of a second countable space has a countable subclass which is also an open base.

An interesting result about second countable space is sated below:

Theorem 4.4.2: Every second countable space is separable.

Proof: The proof of this result follows from the observation that the set formed by choosing a point from each basic open set of the countable basis is not only countable but also dense.

The converse of the above theorem is however not true as is clear from the classic example of the space **R** equipped with the upper limit topology which satisfies the second axiom of countability but is not separable.

However the following special case in true.

Theorem 4.4.3: Every separable metric space is second countable.

4.5 CONTINUOUS FUNCTIONS

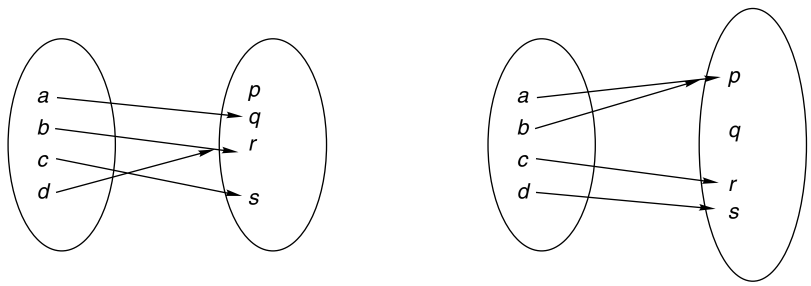
Of basic importance in the whole of the subject is the concept of continuous functions.

Definition: Let (X, T) and (Y, S) be two topological spaces. A function $f: X \to Y$ is *continuous* if the inverse image of an open set in Y is an open set in X, i.e., $f^{-1}(U) \in T$ whenever $U \in S$.

Example 1: Let $X = \{a, b, c, d\}$, $Y = \{p, q, r, s\}$, $T = \{\phi, \{a\}, \{a, b\}, \{a, b, c\}, X\}$ and

$$S = \{\phi, \{p\}, \{q\}, \{p, q\}, \{q, r, s\}, Y\}.$$

Let two functions f and g be defined by the diagrams



Clearly f is continuous, since

$$f^{-1}(\phi) = \phi \in T,\ f^{-1}(\{p\}) = \phi \in T,\ f^{-1}(\{q\}) \in T = \{a\},$$

$$f^{-1}(\{p, q\}) = \{a\} \in T,\ f^{-1}(\{q, r, s\}) \in T = \{a, b, c, d\} \in T.$$

But g is not continuous, since $g^{-1}(\{q, r, s\}) = \{c, d\} \notin S$.

Note any function from a discrete space to any topological space is continuous.

An interesting result about continuous funtions is the following:

Theorem 4.5.1: Let (X, T) and (Y, S) be two topological spaces. The following are equivalent:

(i) $f: X \to Y$ is continuous,

(ii) The inverse of each number of a base B of Y is an open set of X

(iii) The inverse image of any subbasic open set is open in X.

Proof: Since every open set H in Y is a union of a family of basic open sets of Y, i.e., $H = \cup \{B_i, B_i \in B\}$ and $f^{-1}(\cup B_i) = \cup f^{-1}(B_i)$, the result follows directly. The result (ii) is a consequence of the result $f^{-1}(G \cap H) = f^{-1}(G) \cap f^{-1}(H)$.

A similar result of much use is the following:

Theorem 4.5.2: A function $f: X \to Y$ is continuous iff the inverse image of a closed set of Y is closed in X.

The proof follows from the observation $f^{-1}(Y - F) = X - f^{-1}(F)$.

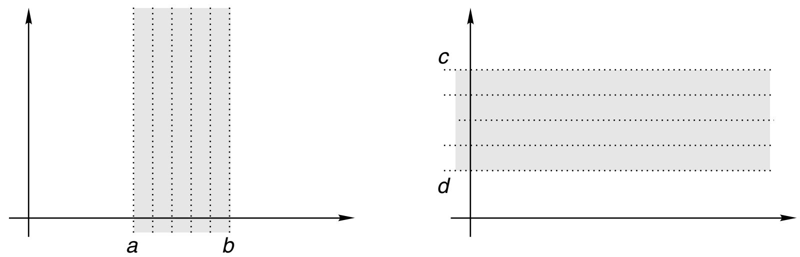
The following example has a bearing upon many results of topology:

Example 2: The projection mappings from $\mathbf{R}^2$ to $\mathbf{R}$ are continuous with respect to their usual topologies.

Recall there are two projection mappings here

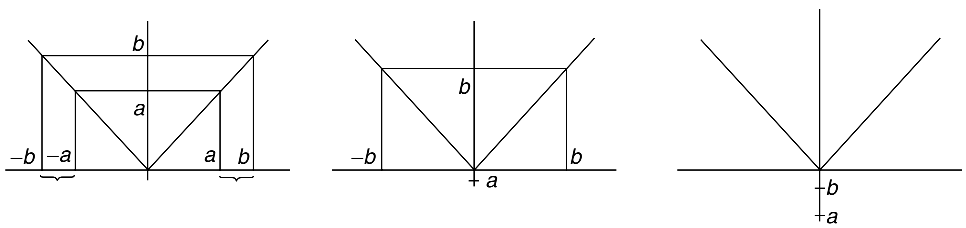
$$\pi: \mathbf{R}^2 \to \mathbf{R} \text{ and } \pi: \mathbf{R}^2 \to \mathbf{R} \text{ defined by } \pi_1(x, y) = x \text{ and } \pi_2(x, y) = y.$$

Clearly the inverse image of any open interval (i.e., any basic open set) under a projection is an open strip as illustrated in the diagram below and hence is open.



The result holds good also for projections from $\mathbf{R}^n$ to $\mathbf{R}$.

Example 3: The function $f: \mathbf{R} \to \mathbf{R}$ defined by $f(x) = |x|$ is continuous with respect to the usual topology of $\mathbf{R}$. This is clear from the following diagram.



$$\text{Clearly, } f^{-1}(a, b) = \begin{cases} (a, b) \cup (-b, -a) & \text{if } \quad 0 \le a \le b \\ (-b, b) & \text{if } \quad a \le 0 \le b \\ \phi & \end{cases}$$

The above definition of continuity is of global character. The following definition concerns the local character of continuity.

Definition: A function $f: X \to Y$ is said to be *continuous at a point p* of X if the image of every neighbourhood of $f(p)$ is a neighbourhood of p.

Note the ε-δ definition of real valued functions of a real variable concerns nothing but the local continuity.

Theorem 4.5.3: A function $f: X \to Y$ is continuous iff it is continuous at every point of X.

Proof: Let f be continuous. Let p X be an arbitrary. Let $N_{f(p)}$ be a neighbourhood of $f(p)$. Then there exists an open set G of Y such that $f(p) \in G \subset N_{f(p)}$. Since f is continuous, $f^{-1}(G)$ is open and $p \in f^{-1}(G)$. Further, $f^{-1}(G) \subset f^{-1}(N_{f(p)})$. Hence $f^{-1}(N_{f(p)})$ is a neighbourhood of p. This proves the continuity of f at every point of X.

Conversely, let f be continuous at every point of X and let G be an open set of Y. We shall show that $f^{-1}(G)$ is open in X, i.e., for every point p in $f^{-1}(G)$, there exists an open set U containing p and entirely contained in $f^{-1}(G)$. To this end we note $p \in f^{-1}(G)$ implies that $f(p) \in G$. As G is an open neighourhood of $f(p)$, there exists an open set U of X such that $p \in U \subset f^{-1}(G)$. This proves that f is continuous.

Open and Closed Functions

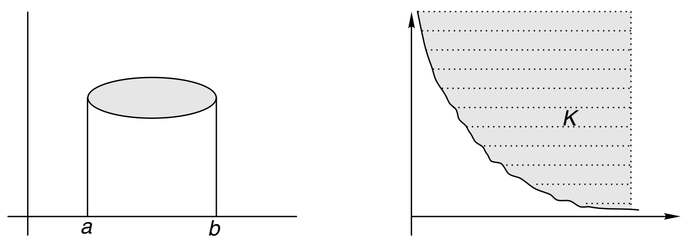
The notion of continuity of a function leads to the question whether the image of an open set by a function can be open and whether the image of a closed set is also closed.

Definition: A function $f: X \to Y$ is called an *open function* (mapping) if the image of an open set of X is open in Y.

A function $f: X \to Y$ is called a *closed function* (mapping) if the image of a closed set in X is closed in Y.

In general a function which is not closed need not be open and not also the vice versa. Further a continuous function may not be open also.

The simplest example of an open mapping is a projection mapping. For example, the mapping $\pi_1: \mathbf{R}^2 \to \mathbf{R}$ defined by $\pi_1(x, y) = x$ is an open mapping, but it is not a closed mapping, since $\pi_1(K) = (0, \infty)$ is not closed in $\mathbf{R}$ where $K = \{(x, y);\ xy \geq 1\}$.



Homeomorphism

Of fundamental importance in topology is the notion of homeomorphisms – mappings of very special character, between topological spaces.

Definition: A topological space (X, T) is said to be *homeomorphic* to another topological space (Y, S) if there exists a bijective mapping f from X onto Y which is bicontinuous also (i.e., both f and f^{-1} are continuous).

Indeed, the bijectivity of f implies that the function f is open iff f^{-1} is continuous. Thus f is bicontinuous iff f is both open and continuous.

4.6 INDUCED TOPOLOGY

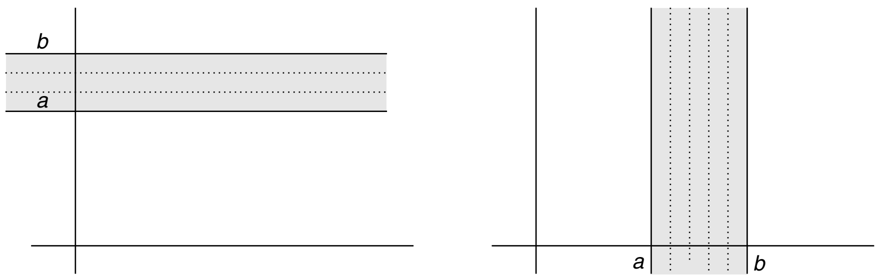
Let X be a set and $\{Y_i\}_{i \in I}$ be a finite or infinite family of topological spaces.

Let $f_i: X \to Y_I$ be an arbitrary function for each i. The weakest (also called coarsest) topology of X with respect to which each f_i is continuous is called the *induced topology* of X, induced by the family $\{f_i\}$. It is easy to observe that this topology is the intersection of all topologies with respect to which the functions $\{f_i;\ i \in I\}$ are continuous. Indeed $S = \{f_i^{-1}(G);\ G \in T_i\}$ is the subbase of the induced topology.

As a direct application of the above we get the product topology as follows.

Definition: If $\{X_i, T_i\}$ is a family of topological spaces, then the topology induced by the projections $\pi_i \colon \pi X_i \to X_i$ defined by $\pi_i((x_i)) = x_i$ is the *product topology* of πX_i.

Example 1: The product topology of $\mathbf{R}^2$ is the weakest topology of $\mathbf{R}^2$ with respect to which the two projections $\pi_1 \colon \mathbf{R}^2 \to \mathbf{R}$ and $\pi_2 \colon \mathbf{R}^2 \to \mathbf{R}$ defined by $\pi_1(x, y) = x$ and $\pi_2(x, t) = y$ are continuous. Thus the subbasic open sets of the product topology of $\mathbf{R}^2$ are as shown in the following diagram:



The space $S^1 \times S^1$ equipped with the product topology is called *torus*.

4.7 IDENTIFICATION TOPOLOGY

Let (X, T) be a topological space, Y be a non-empty set and $p \colon X \to Y$ be a surjection. The strongest topology of Y with respect to which p is continuous is called the *identification topology* of Y and is denoted by $T(p)$.

Note that p need not be open or closed with respect to $T(p)$.

As a direct consequence of the above we get:

Definition: If ρ be an equivalence relation defined on a topological space (X, T), then the strongest topology of X/ρ with respect to which the quotient mapping $p \colon X \to X/\rho$ defined by $p(x) = [x]$ is continuous, is called the *quotient topology* of X/ρ and the pair $(X/\rho, T(p))$ is called the *quotient space* of X by ρ.

Example 1: Let $A \subset X$ and (X, T) be a topological space. Let ρ be defined by $\rho = (A \times A) \cup \{(x, x); x \; X\}$. Clearly ρ is an equivalence relation. The quotient space $(X/\rho, T(p))$ where $p \colon X \to X/\rho$ is the quotient mapping is obtained from X by collapsing the set A to a point (note any two points of A are ρ-equivalent and thus all points of A form a class).

4.8 FREE UNION OF SPACES AND ATTACHMENTS

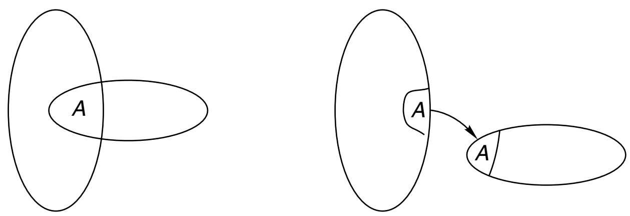
The attachment of a space X to another space Y by a mapping f has great importance in topology. The construction of cone and suspension follows as a consequence.

Definition: Let (X, T_X) and (Y, T_Y) be two disjoint topological spaces. The *free union* of X and Y, denoted by $X + Y$, is the set $X \cup Y$ equipped with the topology T_{X+Y} defined as follows:

$U \in T_{X+Y}$ iff $U \cap X \in T_X$ and $U \cap Y \in T_Y$

It readily follows that $G \in T_X$ implies $G \in T_{X+Y}$ and $H \in T_Y$ implies $H \in T_{X+Y}$.

Let (X, T_X) and (Y, T_Y) be disjoint topological spaces and let A be a closed subset of X and $f : A \to Y$ be continuous, A being equipped with the relative topology. The relation ρ defined on $X + Y$ by '$a \sim f(a)$ for each $a \in A'$ is an equivalence relation. The quotient space $(X + Y)/\rho$ is referred to as "X attached to Y by f" and will be expressed symbolically by $X \cup Y$, f being called the *attaching function*.



4.9 TOPOLOGICAL INVARIANT

Any property that remains unaltered by homeomorphisms is known as a topological invariant. Such a property will be referred to as a topological property in general.

As for example, the length of a segment is not a topological property since the mapping $f : \mathbf{R} \to \mathbf{R}$ defined by $f(x) = 3x$ is a homeomorphism and $f((3, 1)) = (9, 3)$. Similarly boundedness of a set is not a topological property. The property of a sequence being Cauchy is also not a topological property, but the openness of an interval is a topological property. So also is the closedness of a set. The compactness and connectedness are important topological invariants which will be studied in the chapters to follow.

4.10 METRIZATION PROBLEM

We have seen in an earlier chapter that the metric space (X, d) defines a topology on X, namely the topology generated by all open spheres. This topology is known as the topology induced by d. This process of obtaining a topology from a metric evolves a pertinent question:

Given a topological space (X, T), is it possible to define a metric d on X such that the topology induced by d will be the same as T ?

This problem is known as the metrization problem of topology. The solution of this problem will appear in due course.

CHAPTER 5

Separation Axioms

The character, in particular richness of a topological space, depends much upon the abundance of open sets in its topology. In fact, the more is the number of open sets, the stronger is the topology and then higher is the chance of any function defined on it being continuous. The study of separation axioms initiated by Alexandroff and Hopf provides us a tool of analyzing the strength of various topological spaces.

5.1 THE AXIOMS

Let (X, T) be a topological space.

The separation axioms as per Alexandroff and Hopf are as follows:

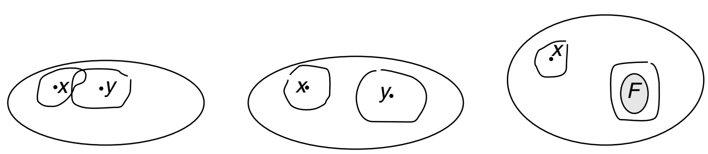
T_1*-axiom*: For every pair of distinct points x, y of X, there exist open sets G and H of X such that $x \in G,\ y \notin G$ and $y \in H,\ x \notin H$.

T_2*-axiom*: For every pair of distinct points x, y of X, there exist open sets G and H of X such that $x \in G$ and $y \in H,\ G \cap H = \phi$.

T_3*-axiom*: For every point x of X and every closed set F of X not containing x, there exist open sets G and H of X such that $x \in G$ and $F \subset H,\ G \cap H = \phi$.

T_4*-axiom*: For every pair of disjoint closed sets F and K of X, there exist open sets G and H of X such that $F \subset G$ and $K \subset H,\ G \cap H = \phi$.

Definition: A topological space (X, T) is called a T_1*-space* if it satisfies the T_1-axiom.



T_1-axiom T_2-axiom T_3-axiom

First observe the following:

Theorem 5.1.1: A topological space (X, T) is a T_1-space iff every singleton subset of X is closed.

Proof: Let (X, T) be a T_1-space and let p be an arbitrary point of X. We shall show that $\{p\}^c$ is open. To this end let $x \in \{p\}^c$. Now x and p being distinct points there exist open sets G_x and H_x such that $x \in G_x$ and $p \notin G_x$. Let $G = \cup \{G_x;\ x \in \{p\}^c\}$. Clearly G is open and $G = \{p\}^c$. So $\{p\}$ is closed.

Conversely, let x and y be two distinct points of X. Then $\{x\}$ and $\{y\}$ are two closed sets. So $\{x\}^c$ and $\{y\}^c$ are two open sets satisfying the conditions that $y \notin \{x\}^c$ and $x \notin \{y\}^c$. Hence (X, T) satisfies the T_1-axiom, i.e., X is a T_1-space.

Corollary: Every finite subset of a T_1-space is closed.

Definition: A topological space (X, T) is a called a T_2-*space* or a *Hausdorff space* if it satisfies the T_2-axiom of separation.

Note the following then:

(*a*) Every Hausdorff space is a T_1-space.
(*b*) Every metric space is a Hausdorff space.
(*c*) Every subspace of Hausdorff space is Hausdorff.
(*d*) The cartesian product of Hausdorff spaces is Hausdorff.
(*e*) The set $\{(x, x);\ x \in X\}$ is closed in $X \times X$ if X is Hausdorff.
(*f*) Every convergent sequence in a Hausdorff space converges to a unique limit.

An interesting at the same time important result about Hausdorff spaces is the following:

Theorem 5.1.2: If (X, T_X) be a topological space and (Y, T_Y) be a Hausdorff space and $f, g: X \to Y$ are continuous functions, then

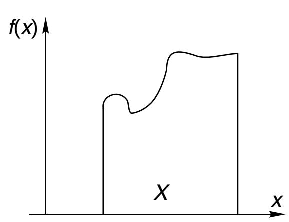
(*i*) The set $\{x, \in X;\ f(x) = g(x)\}$ is closed in X
(*ii*) $f(x) = g(x)$ for all x in D, D is dense in X implies $f(x) = g(x)$ for all x in X.
(*iii*) The set graph $f = \{(x, f(x));\ x \in X\}$ is closed in $X \times Y$.

Proof: (*i*) Since $\{x \in X;\ f(x) = g(x)\} = \psi^{-1}(\Delta)$ where $\psi: X \to Y \times Y$ defined by

$\psi(x) = (f(x), g(x))$ is continuous and $\Delta = \{(y, y);\ y \in Y\}$, the result follows:

(*ii*) Since by (*i*) D is closed and by the given condition D is dense, $X = \overline{D} = D$. Hence the result.

(*iii*) Note graph $f = \varphi^{-1}(\Delta)$ where $\varphi: X \times Y \to Y \times Y$ defined by $\varphi(x, y) = (f(x), y)$ is continuous. Hence the result follows:



Definition: A topological space (X, T) is called a *regular space* if it satisfies the T_3-space.

First note that a regular space need not that be a T_1-space and hence need not be a Hausdorff space. This is clear as the space (X, T) with $X = \{a, b, c\},\ T = \{\phi, \{a\}, \{b, c\},\ X\}$ is regular but not T_1-space (note the singleton set is $\{b\}$ is closed}.

Definition: A topological space (X, T) is called a T_3 *-space* if it is regular and T_1.

Observe that every T_3-space is T_2-space *b* but not conversely. Since a T_3-space is also T_1-space, every singleton set is closed and therefore for any two distinct points *x* and *y* of *X*, $\{y\}$ being closed, there exist two open sets *G* and *H* separating *x* and $\{y\}$, i.e., $x \in G,\ \{y\} \subset H,\ G \cap H,$ but this is exactly the requirement of a Hausdorff space.

A Hausdorff space which is not regular is the space $(\mathbf{R}, U)$ where *T* is the topology generated by all open intervals and **Q**. Evidently $T \supset U$ and hence **R** is Hausdorff but it is not regular since the closed set $\mathbf{Q}^c$ and the point $1 \notin \mathbf{Q}^c$ can not be separated by disjoint open sets of *T*.

Other important results about regular spaces are

(*i*) Every subspace of a regular space is regular,

(*ii*) The cartesian product of regular spaces is regular.

Let $C(X, \mathbf{R})$ denote the set of all bounded continuous real functions defined on $X \in.$ A family *A* of functions from $C(X, \mathbf{R})$ is said to separate points of *X* if for every pair of distinct points *x*, *y* of *X*, there exist $f \in A$ such that $f(x) \neq f(y)$.

It readily follows that (X, T) is Hausdorff if $C(X, \mathbf{R})$ separates points of *X*. This is clear, as for $x < y$, there exist $f \in C(X, \mathbf{R})$ and $r \in \mathbf{R}$ such that $f(x) < r < f(y)$ or $f(y) < r < f(x)$. Evidently in the first case $\{x \in X;\ f(x) < r\}$ and $\{x \in X;\ f(x) > r\}$ are disjoint open sets of *X* containing *x* and *y* respectively. Similarly for the second case we can find disjoint open sets separating *x* and *y*.

Definition: A topological space (X, T) is called a *completely regular space* if it satisfies the property that for every *x* in *X* and closed set *F* of *X*, not containing *x*, there exist $f \in C(X, [0, 1])$ such that $f(x) = 0$ and $f(F) = 1$.

A completely regular space is sometimes referred to as a $T_{3\frac{1}{2}}$ *space* or *Tychonoff* space.

From the definition it follows that

(*i*) Every completely regular space is a regular space. This is evident as for a point *x* of *X* and a closed set not containing *x*, there exists $f \in C(X, [0, 1])$ such that $f(x) = 0$ and $f(F) = 1$. Now consider $G = \{x \in X;\ f(x) < 1/2\}$ and $H = \{x \in X;\ f(x) > 1/2\}$. Clearly *G* and *H* are disjoint open sets containing *x* and *F* respectively.

(*ii*) For every *x* and a closed set *F* of *X*, there exists a function $g \in C(X, [0, 1])$ such that $g(x) = 1$ and $f(F) = 0$.

(*iii*) $C(X,\ [0, 1]$ separates points of *X* if *X* is completely regular.

(*iv*) Every Tychonoff space is a T_3-space.

Definition: A topological space is called *normal* if it satisfies the T_4-axiom and is a T_1-space, i.e., for every pair of disjoint closed sets *F* and *G*, there exist disjoint open sets *G* and *H* of *X* such that $F \subset G, K \subset H$.

A normal T_1-space is called a T_4-*space*.

The following results directly from the definition:

(*i*) Every metric space is normal

(*ii*) A normal sapce need not be a T_1-space. Consider the space (*X*, *T*) where *X* = {*a*, *b*, *c*} and $T = \{\phi, \{a\}, \{b\}, \{a, b\}, X\}$. Clearly *X* is normal but it is not a T_1-space as {*a*} is not closed.

(*iii*) Every T_4-space is a T_3-space and hence Hausdorff also.

5.2 URYSHON'S LEMMA AND TIETZE'S EXTENSION THEOREM

The fact that a topological space is rich in open sets ensures that it is rich in continuous functions. This is assured by the following results:

Urysohn's Lemma: If (*X*, *T*) is a normal space and *F* and *K* are two closed subsets of *X*, then there exists $f \in C(X, [0, 1])$ such that *f*(*F*) = 0 and *f*(*K*) = 1.

The proof of this theorem is bit lengthy and is therefore left here but can be seen in [9].

This lemma however leads to the following very useful result.

Tietze's Extension Theorem: If *X* be a normal space and *F* be a closed subset of *X* equipped with the relative topology, then every continuous function from *F* to [*a*, *b*] admits of a continuous extension to *X*.

For a proof the reader is referred to [9].

Note the condition of closedness of *F* is an unavoidable condition for the conclusion as can be seen from the following example:

Let *X* = [0, 1], *F* = (0, 1] and *f*(*x*) = (1/*x*). Clearly *X* is normal, *F* is not closed, *f* is continuous on *F*, but *f* does not admit of a continuous extension to [0, 1].

We conclude this section with the statement of a remarkable result on metrization.

Urysohn's Metrization Theorem: Every second countable normal T_1-space is homeomorphic a subset of the Hilbert Cube I^{∞}.

Recall the Hilbert cube $I^{\infty} = \{(x_1, x_2, x_3, \ldots) \in \mathbf{R}^{\infty}; |x_n| < 1/n\}$ endowed with the relativized usual topology.

CHAPTER 6

Compactness

One of the most useful topological properties is compactness. In what follows we shall explore this feature as much as is necessary for our purpose.

6.1 SOME BASIC NOTIONS

The definition of compactness of a topological space requires the notion of a cover of a set. So we begin with this notion here.

Definition: A family $\{G_i\}_{i\in I}$ of open sets of a topological space X is called an *open cover* of X if $X = \bigcup_{i\in I} G_i$.

If the index set I is finite, the cover is called a *finite cover*. On the other hand if I is infinite, it is called a *infinite cover*. If in particular if I is countable, it is called a *countable cover*. If I is uncountable, the cover is known as an *uncountable cover*.

For example, the family $\{(m, m+1); m \in \mathbf{Z}\}$ is a countable (infinite) cover of **R**. Note $\mathbf{R} = \cup\{(m, m+1); m \in \mathbf{Z}\}$. Similarly the family $\{(r-1, r+1); r \in I\}$ is an uncountable cover of **R**.

Definition: A family $\mathscr{C} = \{G_i\}_{i\in I}$ of open sets is called an *open cover* of a subset K of X if $K \subset \cup G_i$. The cover is *finite* if the index set I is finite. The cover is infinite, if I is infinite. The cover is *countable*, if I is countable.

Clearly the family $\{(r-1, r+1); r \in \mathbf{Q} \cap [0, 1]\}$ is a countable cover of [0, 1].

Note $[0, 1] \subset \cup\{(r-1, r+1); r \in \mathbf{Q} \cap [0, 1]\}$. Similarly the family of open unit discs centered at (p, q) where $p, q \in \mathbf{Q} \cap [0, 1]$ is a cover of $[0, 1]^2$.

If $S \subset C$ and S is a cover of K, then S is called a subcover of K.

Definition: A topological space (X, T) is said to be *compact* if every cover of X has a finite subcover.

A subset K of a topological space X is said to be *compact* if every cover of K has a finite subcover.

It is easy to see that a subset K of (X, T) is compact iff it is compact as a subspace of X in the relative topology.

First note that **R** is not compact since the cover $\{\{-n, n); n \in \mathbf{N}\}$ of **R** has no finite subcover. So also is the set (0, 1) as the cover $\{(1/n+1, 1/n); n \in \mathbf{N}\}$ has no finite subcover. The set $\{0, 1, 1/n, 1/2, \ldots, 1/n, \ldots\}$ is compact.

A notion that is closely related to the compactness criterion is the following:

Definition: A family $\mathscr{A}$ of subsets of a topological space (X, T) is said to have *finite intersection property* (*FIP*) if every finite subfamily of A has non-empty intersection.

For example, the family $\{(-n, n); n \in \mathbf{N}$ has the finite intersection property but the family $\{(n, n+1); n \in \mathbf{N}\}$ does not have the finite intersection property.

The following characterization of compactness in terms of finite intersection property is worth noting:

Theorem 6.1.1: A topological space (X, T) is compact iff every family $\mathscr{F}$ of closed sets of X having finite intersection property has itself a non-empty intersection.

Proof: Let X be compact. Since proving that a family of closed sets having finite intersection property has entire intersection non-empty is equivalent to proving that a family of closed sets with the entire intersection empty has a subfamily with empty intersection, we shall prove the second proposition only. T this end let $\{Fi\}$ be family of closed sets of X having entire intersection empty, i.e., $\cap \{F_i ; i \in I\} = \phi$. Hence $\cup F_i^c = X$, is an open cover of X is compact, $\{F_i{}^c\}$ has a finite subcover, i.e., $F_{i_1}{}^c \cup F_{i_2}{}^c \cup F_{i_n}{}^c = X$. This implies that $F_{i_1} \cap F_{i_2} \cap ... \cap F_{i_n} = \phi$. Thus every family of closed sets having finite intersection property has the entire intersection non-empty.

Conversely, $\{Gi\}$ be an open cover of X. Then $\{G_i^c\}$ is a family of closed sets with $\cap G_i^c = \phi$. But by the equivalence of finite intersection property, there exists a finite family $G_{i_1}{}^c \cap G_{i_2}{}^c \cap ... \cap G_{i_n}{}^c = \phi$, or $G_{i1} \cup G_{i2} \cup ... \cup G_{in} = X$. This implies the compactness of X.

The following results have important bearings over many results of analysis:

(i) Every compact set in a Hausdorff space is closed.

(ii) Every closed subset of a compact space is compact.

(iii) Every compact set in a metric space is closed and totally bounded.

(iv) Continuous image of a compact set is compact.

(v) For every pair of disjoint compact sets in a Hausdorff space, there exist two disjoint open sets, which separate them.

(vi) Every compact Hausdorff space is normal.

An interesting result on the impact of compactness on homeomorphism is demonstrated below:

Theorem 6.1.2: Every bijective continuous function from a compact space onto a Hausdorff space is a homeomorphism.

Proof: Let $f : X \to Y$ be a bijective continuous mapping and X be a compact space and Y be a Hausdorff space. To prove that f is a homeomorphism, it is enough to prove that f is open. To this end let G be open in X. Then G^c is closed. Sine X is compact G^c is compact. As f is continuous, $f(G^c)$ is compact. Y being Hausdorff, $f(G^c)$ is closed. Since f is a bijection, $f(G^c) = Y = f(G)$. Hence $f(G)$ is open. This completes the proof.

Corollary: If $f: X \to Y$ is an injective continuous mapping, X is compact, Y is Hausdorff, then f is a homeomorphism from X onto $f(X)$, since $f(X)$ is a Hausdorff space in the relative topology.

6.2 OTHER NOTIONS OF COMPACTNESS

There are a few other notions of compactness which are also related to the above notion of compactness in a nice way.

Sequentially Compact Sets

A subset of topological space (X, T) is said to be *sequentially compact* if every sequence in K has a subsequence which also converges to a point of K.

Note every finite subset of a topological space is sequentially compact as every sequence must have at least one term repeated infinitely many times in a sequence which constitutes the convergent subsequence. The set (0, 1) is not sequentially compact as there exists a sequence $\{x_n\}$ which has no convergent subsequence with limit within (0, 1).

We have already seen that in a metric space a set is compact iff it is sequentially compact.

Countably Compact Sets

A subset of a topological space (X, T) is called *countably compact* if every infinite subset A of K has a limit point in K.

As for example, the set [0, 1] is countably compact, since every infinite subset of [0, 1] is bounded and therefore by Bolzano Weisstrass theorem has a limit point in [0, 1], the set [0, 1] being closed. Observe the set (0, 1) is not countably compactas the infinite subset {1/2, 1/3, 1/4, …} has a limit point not in (0, 1).

The following observations can be made easily

1. Every compact set is countably compact but not conversely.
2. Every sequentially compact set is countably compact but not conversely.
3. Every countable open cover of a sequentially compact set has a finite subcover.
4. In a metric space the following are equivalent.
   (*i*) K is compact
   (*ii*) K is sequentially compact
   (*iii*) K is countably compact.
   (*iv*) The product of compact spaces is compact.

6.3 COMPACTIFICATION

A topological space may not be compact but is can be made compact by adjoining one or more points to it and redefining topology in a natural way. For example the real line R is not compact but if a single point is added to R, then it becomes compact in a naturally extended topology. This phenomenon is referred to as compactification of the original space. The topologization of the expanded set is done as follows:

Definition: A topological space X is said to be embedded in another topological space Y if X is homeomorphic to a subspace of Y. If Y is compact, then Y is called a *compactification* of X.

The *one-point compactification* (also called *Alexandroff compactification*) of a topological space (X, T) is achieved in the following way:

Step 1. Adjoin a point $\{\propto\}$ to the set *X* is denote the expanded set by $X_\propto$, i.e., $X_\propto = X \cup \{\propto\}$ The point $\propto$ is called the point at infinity.

Step 2. The topology $T_\propto$ of $X_\propto$ consists of all the members of T together with the complements in $X_\propto$ of all closed and compact subsets of *X*.

It is easy to verify that $T_\propto$ is a topology and $(X_\propto, T_\propto)$ is a compactification of (*X*, *T*).

Though the topology $T_\propto$ is obtained by extending *T*, it need not possess the same properties as those of *T*. But interestingly the following is true:

Theorem 6.3.1: If (*X*, *T*) is locally compact Hausdorff, then its one-point compactification $(X_\propto, T_\propto)$ is also compact Hausdorff.

Another observation that leads to a distinct compactification of a topological space is this that if a topological space *X* is a subspace of a compact Hausdorff space, then it is completely regular. This observations raises a question as to whether every completely regular space T_1-s[pace admits of a compactification. The following theorem gives an answer to this question:

Stone-Cech Theorem: Every completely regular T_1-space *X* can be embedded as a dense subspace of a compact Hausdorff space $\beta(X)$, called the *Stone-Cech Compactification* of *X* with the property that every bounded continuous real function has a unique extension to a bounded continuous real function on $\beta(X)$.

The proof is much involved and left here but can be seen in [9].

CHAPTER 7

Connectedness

Connectedness is another topological property which has a significant bearing upon many results of analysis and topology. We begin with some basic notions.

7.1 SOME BASIC NOTIONS

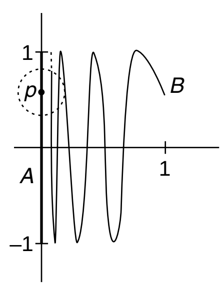
The notion of connectedness is best explained by the notion of separated sets.

Definition: Two subsets of a topological space (*X*, *T*) is said to be separated if $\bar{A} \cap B = \phi$ and $A \cap \bar{B} = \phi$, i.e., none of *A* and *B* contains a limit point of the other.

For example, the sets *A* = (0, 1) and *B* = (1, 2) are separated, since *A* = [0, 1], *B* = [1, 2], $[0, 1] \cap (1, 2) = \phi$ and $(0, 1) \cap [1, 2] = \phi$

For examples of sets not separated, consider, $A = \{(0, y) \in \mathbf{R}^2; -1 \le y \le 1\}$ and $B = \{(x, y) \in \mathbf{R}^2,$ $y = \sin(1/x), 0 \le x \le 1\}$.

Clearly *A* and *B* are not separated as A contains many limit points of *B*.



Definition: A topological space (*X*, *T*) is said to be *disconnected* if there exist two disjoint open sets *G* and *H* of *X* such that $X = G \cup H$. Note that both *G* and *H* are closed also. The expression $G \cup H$ is called a disconnection of *X*.

A topological space *x* is called *connected* if it is not disconnected.

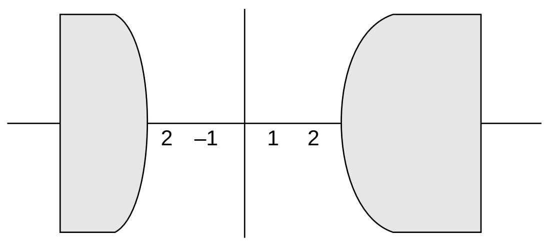
A subset K of a topological space (*X*, *T*) is said to be disconnected if there exists two open set *G* and *H* of *X* such that $K = (G \cap K) \cup (G \cap K), (G \cap K) \cap (G \cap K) = \phi$.

It is easy to note that a subset of *X* is disconnected iff it is disconnected as a subspace of *X* endowed with the relative topology.

Further note that any singleton set is always connected.

Example 1: Consider the topological space (X, T) where $X = \{a, b, c, d\}$ and $T = \{\phi, \{a, b\}, \{c, d\}, X\}$. Clearly the set $\{a, c\}$ is disconnected, as the set $G = \{a, b\}$ and $H = \{c, d\}$ are disjoint open sets and $K = \{a, c\} = (G \cap K) \cup (G \cap K)$.

Example 2: The set $K = \{(x, y) \in \mathbf{R}^2; x^2 - y^2 \geq 2\}$ is disconnected in $(\mathbf{R}^2, U)$ since $G = \{(x, y) \in \mathbf{R}^2; x > 1\}$ and $H = \{(x, y) \in \mathbf{R}^2; x < 1\}$ are open in R^2 and $K = (G \cap K) \cup (G \cap K)$.



The relation of separated sets with connectedness is exhibited in the following result:

Theorem 7.1.1: A set K is connected in a topological space (X, T) iff K is not the union of two separated sets.

Proof: To prove the result it is enough to prove that a set K is disconnected in X iff it is the union of two separated sets of X,

To this end assume K to be disconnected. Then there exit $G, H \in T$ such that $K = (G \cap K) \cup (H \cap K)$, $(G \cap K) \cap (H \cap K) = \phi$. If possible let $(G \cap K$ and $H \cap K)$ be not separated.

Then without loss of generality let, $\overline{(G \cap K)} \cap (H \cap K) \neq \phi$. This implies that a limit point p of $G \cap K$ belongs to $H \cap K$. But since H is open containing p, it must intersect $G \cap K$ at a point other than p. But this implies $G \cap K$ and $H \cap K$ are not disjoint. This is a contradiction. Hence $G \cap K$ and $H \cap K$ must be separated. The converse follows by a similar argument. This proves the result.

Theorem 7.1.2: If A and B are connected and not separated in (X, T), then $A \cup B$ is connected.

Proof: Suppose $A \cup B$ is disconnected. Then there exist open sets G and H such that $A \cup B = ((A \cup B) \cap G) \cup ((A \cup B) \cap H)$ and $((A \cup B) \cap G) \cap ((A \cup B \cap H) = \phi$. Since A is connected, either $A \subset G$ or $A \subset H$ (Recall if $G \cup H$ is a disconnection of a set K and F is a connected subset of K, then either $F \subset G$ or $F \subset H$) and similarly $B \subset G$ or $B \subset H$. If $A \subset G$ and $B \subset H$, then $(A \cup B) \cap G = A$ and $(A \cup B) \cap H = B$ are separated sets. But this contradicts the hypothesis. Hence $A \cup B \subset G$ or $A \cup B \subset H$ and so $G \cup H$ is not a disconnection of $A \cup B$, i.e., $A \cup B$ is connected.

An interesting characterization of connectedness is contained in the following result whose proof follows directly.

Theorem 7.1.3: A topological space X is connected iff X and ϕ are the only closed subsets of X.

As a direct consequence of the above we observe that $\mathbf{R}$ is connected.

Example 3: The space (X, T) is disconnected where $X = \{a, b, c\}$ and $T = \{\phi, \{a\}, \{b, c\}, X\}$

The following result is very useful in real and complex analysis.

Theorem 7.1.4: The continuous image of a connected set is connected.

Proof: Let A be a connected set in a topological space (X, T) and let $f : X \to Y$ be a continuous function where (Y, T') is another topological space. If possible let $f(A)$ be not connected in Y. Then there exist $G, H \in T'$ such that $(f(A) \cap G) \cup (f(A) \cap H)$ and $(f(A) \cap G) \cap (f(A) \cap H) = \phi$. Clearly $f^{-1}(G)$ and $f^{-1}(H)$ are open sets of X and $A = f^{-1}(G) \cup f^{-1}(H)$. Further $(A \cap f^{-1}(G)) \cap (A \cap f^{-1}(H)) = \phi$. This means A is disconnected, contrary to our hypothesis. Hence $f(A)$ must be connected.

Definition: A maximal connected subset of a topological space is called a *component* of X. Thus a component K is connected and if F is also a connected subset of X and $K \subset F$, then $K = F$.

A central fact about components of a topological space is the following:

Theorem 7.1.5: Every topological space is the disjoint union of all its components, i.e., the components of a space constitute a partition of the space.

A standard argument proves the theorem.

Another result of interest about connectedness is given below.

Theorem 7.1.6: A product of connected spaces is connected.

The proof is a consequence of the fact that the projections are continuous functions.

Corollary: The Euclidean space $\mathbf{R}^n$ is connected.

7.2 OTHER NOTIONS OF CONNECTEDNESS

We begin with local connectedness.

Locally Connected Space

A topological space (X, T) is said to be *locally connected at a point p* of X if every open neighbourhood of p contains a connected open set containing p.

A space (X, T) is called locally connected if it is locally connected at each point of X.

A subset K of a topological space is called *locally connected* if it is locally connected as a subspace of X when endowed with the relative topology.

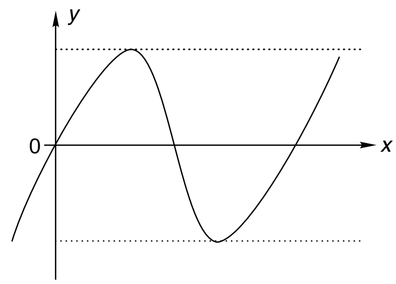
For example, $\mathbf{R}$ is locally connected since it is locally connected at each point of it.

Again consider $A = \{(0, y); -1 \le y \le 1\}$ and $B = \{(x, y); y = \sin(1/x), 0 < x \le 1\}$

Clearly AB is not locally connected since any neighbourhood of p does not contain a connected neighbourhood.

Path-connected Space

A path from a point a to the point b in a topological space (X, T) is a continuous function $f : I \to X$ with $f(0) = a$ and $f(1) = b$ where $I = [0, 1]$ The point a is called the *initial point* and b is called the *terminal point* of the path.



It should be noted that though the function has been referred to as a path, it is actually the image of I in X. Further if there is a path from a to b, there is also a path from b to a given by $g(s) = f(1 - s)$.

The constant function $f : I \to X$ defined by $f(s) = p \in X$ defines a constant path at p. If there is a path from a to b and another path from b to c, then there is a path from a to c.

Indeed, if $f : I \to X$ is a path from a to b and $g : I \to X$ be a path from b to c, then the path from a to c is given by $f * g : I \to X$ where $f * g$ is defined as

$$(f * g)\ (s) = f(2s) \text{ when } 0 \le s \le 1/2.$$

$$g\ (2s - 1) \text{ when } 1/2 \le x \le 1.$$

Definition: A topological space (X, T) is said to be *path-connected* if every pair of points of X can be connected by a path.

A subset K of a space X is said to be *path-connected* if every pair of points of K can be connected by a path.

The maximal path connected subset of a space is called a *path-connected component.*

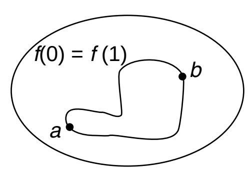
Simply Connected Space

We need to begin with the notion of closed path and contractibility.

A path $f : I \to X$ is said to be *closed* if $f(0) = f(1)$.

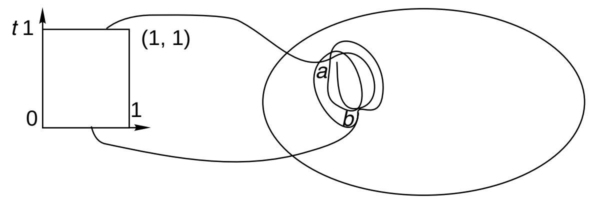
A path $f : I \to X$ is said to be *homotopic* to another path $g : I \to X$ both having the same initial point a and the same terminal point b if there exist a continuous function $H : I^2 \to X$ such that

$H(s, 0) = f(s), H(0, b) = a$
$H(s, 1) = g(s), H(1, b) = b.$



The fact that f is homotopic to g is denoted by $f \cong g$.

It can be easily verified that homotopy relation is an equivalence relation in the collection of all paths from a to b in X.

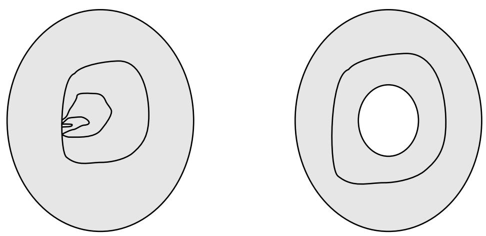


Definition: A closed path $f : I \to X$ is said to be *contractible* to a point if it is homotopic to the constant path $e_p: I \to X$ where $e_p(s) = p$ for all $s \in I$.

Definition: A topological space I said to be simply-connected if every closed path in X is contractible to a point.

For example an open connected disc in $\mathbf{R}^2$ is in simply connected but an annular region in $\mathbf{R}^2$ is not simply connected.

The diagram below explains the difference geometrically.



Part II

This page intentionally left blank

CHAPTER 1

Algebraic Preliminaries

We assume familiarity with the following concepts: [a] group [b] abelian group [c] subgroup [d] ring [e] subring [f] field [g] homomorphism (group, ring and field) [h] isomorphism (group, ring and field). We denote by $\cdot$ the binary operation in a group and let 1 denote its multiplicative identity. If a is an element of a group, write a^{-1} to denote its inverse element, $a \cdot b$ will be sometimes written as ab. For an additive group, we use +, 0 and $-a$ instead of ., 1 and a.

1.1 SOME BASIC NOTIONS

Definition: Let A be an *additive abelian* group and B, C two subsets of A. We write $B + C$ for the set $\{b + c: b \in B, c \in C\}$. $B + C$ is called the *sum* of B and C. If $\{B_\alpha : \alpha \in I\}$ is an arbitrary collection of subsets of A, then the sum of $\{B_\alpha : \alpha \in I\}$ is defined to be the set of all finite sums $b_{\alpha_1} + \ldots + b_{\alpha_2} : b_{\alpha_i} \in B$, $i = 1 \ldots, n, n \geq 1$.

Proposition 1.1.1: If for each $\alpha \in I$, B_α is a subgroup of A, then so is their sum. Proof is easy.

Definition: If A is an abelian group of which A_0 is a subgroup, then a *coset* of A by A_0 is a set of the form $[a] = \{a + a_0 : a_0 \in A_0)$, where a is a fixed element of A. A/A_0 denotes the set of all cosets of A by A_0. Define $[a_1] + [a_2] = [a_1 + a_2]$. With this operation, it is easy to see that A/A_0 is an abelian group. This is called the *factor group* or *quotient group* of A by A_0.

Proposition 1.1.2: If A_1 is a subgroup of A_0 and A_0 is a subgroup of A, then A_0/A_1 is a subgroup of A/A_0 and $(A/A_0)/(A_0/A_1)$ is isomorphic to A/A_1.

Proposition 1.1.3: If B, C are subgroups of A, then so is $B \cap C$ and $(B + C)/C$ is isomorphic to $B/B \cap C$.

The proofs are straightforward.

If A, B are abelian groups and $\phi: A \to B$ is a homomorphism, then $\phi^{-1}(o)$ is a subgroup of B. This subgroup is called the *kernel* of ϕ.

Definition: A sequence $\cdots \to A_{n+1} \xrightarrow{\phi_{n+1}} A_n \xrightarrow{\phi_n} A_{n-1} \to \cdots$ of abelian groups and homomorphisms [finite or infinite] is called *exact* at A_n, if *kernel* of ϕ_n = image of ϕ_{n+1}. The sequence is called *exact* if it is exact at every A_n.

Proposition 1.1.4: If $\{0\}$ is the group consisting of 0 only, A_o is a subgroup of A, $i: A_o \to A$ is the inclusion map given by $i(a) = a$ and $p: A \to A/A_o$ is the projection map given by $p(a) = [a]$, then $\{0\} \xrightarrow{j_1} A_o \xrightarrow{i} A \xrightarrow{p} A/A_o \xrightarrow{j_2} (o)$ is *exact* where $j_1(o) = 0$ and $j_2([a]) = 0$ for all $[a] \in A/A_o$.

Let A and B be abelian groups and $\phi, \phi': A \to B$ be a homomorphism.

Define $\phi, \phi': A \to B$ by $(\phi + \phi')(a) = \phi(a) + \phi'(a)$. Clearly $\phi + \phi'$ is a homomorphism. With this operation, the set of all homomorphisms from A into B is an abelian group. This group is denoted by Hom (A, B).

If ϕ, ϕ' are homomorphism from A into B and ψ, ψ' are homomorphisms from B into C, then $\phi\psi: A \to C$ defined by $\phi\psi(a) = \phi(\psi(a))$ is again a homomorphism and

$$(\phi + \phi')\psi = \phi\psi + \phi'\psi$$
$$\phi(\psi + \psi') = \phi\psi + \phi\psi'$$

Let $\{A_\alpha\}$ be an indexed family of abelian groups. Consider all elements (a_α) in their Cartesian product for which $a_\alpha = 0$ except possible for finitely many values of α.

Define $[a] + [a'] = [a + a']$. Then we obtain an abelian group called the *external direct sum* of $\{A_\alpha\}$ and denoted by $\sum A_\alpha$. If we withdraw the restriction "$\alpha_\alpha = 0$ except possibly for finitely many values of α" the resulting group is called the *direct product* of $\{A_\alpha\}$ and denoted by $\prod A_\alpha$. If $\{A_\alpha\}$ is a finite family, $\prod A_\alpha = \sum A_\alpha$. If $\{A_\alpha\}$ is a family of subgroups of an abelian group A whose sum is A such that $A_\alpha \cap A_{\alpha'} = \{0\}$ if $\alpha \neq \alpha'$, then A is called the *internal direct sum* of $\{A_\alpha\}$. In this case A is isomorphic to $\sum A_\alpha$. On the other hand, if $\{A_\alpha\}$ is any family of abelian groups, each A_α is isomorphic to a subgroup B_α of $\sum A_\alpha$ and $\sum A_\alpha$ is the *internal direct sum* of the family $\{A_\alpha\}$.

1.2 FREE ABELIAN GROUP

If A is an additively abelian group, let na denote $a + a \ldots + a$ [n times] for a positive integer n, $= o$ for $n = o$ and $(-a) + (-a) + \ldots + (-a)$ [n times] if n is a negative integer. For $B \subset A$, call the smallest subgroup containing B the *group generated by B*. This group generated by B consists of all finite sums $n_1 b_1 + \ldots + n_\gamma b_\gamma$ where $n_1, \ldots, n_\gamma$ are integers and $b_1, \ldots b_\gamma \in B$. The *order* of $a \in A$ is the cardinality [or order] of the subgroup generated by a.

An additively abelian group F is called *free* if it contains a subset B such that (i) F is generated by B and (ii) any element of F can be uniquely expressed as $n_1 b_1 + \ldots + n_\gamma b_\gamma$, n_i integers, $b_i \in B$. B is called a *basis* for F.

Proposition 1.2.1: The cardinalities of any two bases of F are equal.

The cardinality of a *basis* of F is called the *rank* of F.

Note: If the rank of F is infinite, it equals the order of F.

Proposition 1.2.2: If F is a free abelian group with a basis B and A is any abelian group, then a function from B into A can be uniquely extended to a homomorphism from F into A.

1.3 NORMAL SUBGROUPS

Let G be a multiplicative group [not necessarily abelian]. A subgroup G_o of G is called *normal* [or *self conjugate*] if $g^{-1} g_o\ g \in G$ for all $g \in G$, $g_o \in G_o$. If G_o is normal, the factor group G/G_o can be defined as in the abelian case. The *center* C of G is the set of elements c for which $cg = gc$ for any $g \in G$.

Proposition 1.3.1: The centre C of a group G is a normal subgroup of G.

Definition: An element of the form $g^{-1}g'^{-1}gg',\ g, g' \in G$ is called a *commutator* of G. The subgroup generated by the commutators of G is called the *commutator subgroup* and denoted by [G, G].

Proposition 1.3.2: [G, G] is normal in G and G/[G, G] is commutative.

If G_o is normal in G and G/G_o is commutative, then $[G, G] \subset G$.

Proof: The elements of [G, G] are finite products $h_i \ldots, h_n$ where $h_i = a_i^{-1} b_i^{-1} a_i\, b_i,\ a_i, b_i \in G$. Let $g \in G$, then

$$g^{-1} a_1^{-1} b_1^{-1}\ a_1 b_1\, g = g^{-1} a_1^{-1}\, ga_1\, a_1^{-1}\, g^{-1} b_1^{-1}\ a_1 b_1$$

$$g = g^{-1} a_1^{-1}\ ga_1 \cdot a_1^{-1} (b_1\, g)^{-1} a_1\ (b_1\, g) \in [G, G].$$

Suppose $g^{-1} h_1, \ldots h_m g \in [G, G]$ for $m \le n-1,\ n > 1$. We show $g^{-1} h_i \ldots h_n g \in [G, G]$.

$g^{-1} h_1, \ldots h_n g = g^{-1} h_1 \ldots h_{n-1} g\ g^{-1} h_n g\ g^{-1} h_1, \ldots h_n g\ \varepsilon\, [G, G]$ by induction hypothesis since $g^{-1} h_n g\ \varepsilon\, [G, G]$. Hence $g^{-1} h_1, \ldots h_n g \in (G, G)$. Thus [$G$, G] is normal.

To show G/[G, G] is commutative, enough to show ab[G, G] = ab for any a, b in G.

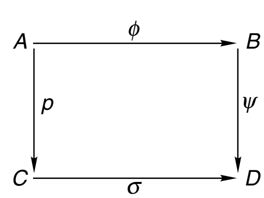
$(ab)\,[G, G] = a \cdot b \cdot (ab)^{-1} [G, G]\ (ab) = [G, G](ab)$. Let G_o be any normal subgroup of G such that G/G_o is commutative. To show $[G, G] \subset G$, enough to show $a^{-1} b^{-1} ab \in G$ for any a, b in G. Consider the element $a^{-1} b^{-1}\ ab$ in G/G_o. This is $(a)^{-1} (b)^{-1} (a)\ (b) = (1)$ by commutativity. Hence $a^{-1} b^{-1} ab \in G$.

1.4 IDEALS OF RINGS

Definition: Let R be a ring. A *two-sided ideal* in R is a subring I such that for $a \in I$, $r \in R$, $ra, ar \in I$. Considering R as an additive abelian group, R/I is well defined. Multiplication can be defined on R/I in a natural way. With this structure, R/I is a ring and it is called the *quotient ring* of R by I.

If R, S are two rings and $\phi\colon R \to S$ is a ring homomorphism, then the kernel K of $\phi[\,= \phi^{-1}(o)]$ is an ideal in R, $\phi(R)$ is a subring of S and ϕ induces an isomorphism between R/K and $\phi(R)$.

Definition: A diagram



where *A*, *B*, *C*, *D* are sets, ϕ, p, ψ, σ are maps $A \to B$, $A \to C$, $B \to D$, $C \to D$ respectively, is said to be commutative if $\psi\phi = \sigma p$.

1.5 *G*-SPACES

Definition: Let *G* be a group. A set *E* is called a *left G-space* if there exists a mapping from $G \times E$ into *E* [the image of [*g*, *x*] being denoted by $g \cdot x$] such that

(*a*) for any $x \in E$, $1 \cdot x = x$

(*b*) for any $x \in E$, $g_1, g_2 \in G$, $(g_1 g_2) \cdot x = g_1 \cdot (g_2 \cdot x)$.

A set *E* called a *right G-space* if there exists a mapping: $E \times G \to E$ such that

(*a*) $x \cdot 1 = x$

(*b*) $x \cdot (g_1 g_2) = (x \cdot g_1) \cdot g_2$

for all $x \in X$, $g_1, g_2 \in G$.

In what follows, we deal only with right *G*-spaces. A similar treatment of left *G* spaces is possible.

Theorem 1.5.1: Let *E* be a right *G*-space. Fix $g \in G$ and define $\phi_g : E \to E$ by $\phi_g(x) = x \cdot g$. Then ϕ_g is one-one onto.

Proof: Let $x \ \varepsilon \ E$ $\phi_g \phi_g^{-1}(x) = (x \cdot g^{-1}) \cdot g = x \cdot (g^{-1}g) = x \cdot 1 = x$. Similarly, $\phi_g^{-1}\phi_g(x) = x$.

Hence $\phi_g^{-1} = (\phi_g)^{-1}$

Hence ϕ is one-one onto.

Definition: A group *G* is said to *operate effectively* on *E* if ϕ_g is the dentity map implies *g* = 1. Let E_1, E_2 be right *G*-spaces, A mapping $f: E_1 \to E_2$ is called *G-equivariant* if.

$$f(x \cdot g) = f(x) \cdot g \text{ for all } x \in E, g \in G.$$

f is called an *isomorphism* if there is a *G*-equivariant map $f': E_2 \to E_1$ such that $f' = f^{-1}$.

It is enough to show that if $f: E_1 \to E_2$ is *G*-equivariant, one-one, and onto, then *f* is *G*-equivariant.

Let *f* be *G*-equivariant, one-one, onto. Let $y \ \varepsilon \ E$. Then $y = f(x)$ for some *x*. Let $g \ \varepsilon \ G$.

$$f(y \cdot g) = f^1(f(x) \cdot g) = f^1(f(x \cdot g)) = x \cdot g = f(y) \cdot g.$$

An *automorphism* of a right *G*-space is a self isomorphism.

Definition: A right *G*-space *E* is called *homogeneous* if for any *x*, *y* in *E*, there is a *g* in *G* such that $x \cdot g = y$. In such a case *G* is said to *act transitively* on *E*.

Theorem 1.5.2: Let *G* be a group. If *H* is a subgroup, then the set *G*/*H* of all right cosets of *G* by *H* form a homogeneous right *G*-space.

Conversely given any homogeneous right *G*-space *E*, there is a subgroup *H* of *G* such that *E* is isomorphic to the right *G*-space *G*/*H*.

Proof: Define on $G/H \times G$ by $[Hg_1] \cdot g_2 = Hg_1 g_2$. It is easy to verify that *G*/*H* is a right *G*-space.

Let E be homogenous G-space. Choose $x_o \in E$. Let $H = \{g \in G, x_o \cdot g = x_o\}$. Then H is a subgroup of G.

Define f on G/H into E by $f(Hg) = x_o \cdot g$. Then f is well-defined. For, let $Hg_1 = Hg_2$, then $H = Hg_2 g_1^{-1}$ so that $g_2 g_1^{-1} \varepsilon H$. Hence $x_o \, g_2 g_1^{-1} = x_o$ So that $x_o \cdot g_2 = x_o \cdot g_1$.

It is easy to see that f is one-one, onto G-equivariant.

Definition: A subgroup H of G as defined above is called the *isotropy subgroup of G* corresponding to x_o.

Remark: The subgroup H defined above depends on x_o. If we choose another point x_1, H is replaced by a conjugate subgroup of H.

Proof: There exists g_o such that $x_o \cdot g_o = x_1$. Then

$$\{g \in G: x_1 \cdot g = x_1\} = \{g \in G: x_o \cdot g_o \cdot g = x_o \cdot g_o\} = \{g \in G: x_o \cdot (g_o g g_o^{-1}) = x_o\} =$$
$$\{g: g_o g g_o^{-1} \in H\} = g_o^{-1} H g$$

Theorem 1.5.3: Let E be a homogeneous right G-space and $\phi: E \to E$ an automorphism. Then for any $x \in E$, x and $\phi(x)$ have the same isotropy subgroup. Conversely, if x, $y \in E$ have the same isotropy subgroup, then there is an automorphism on E such that $y = \phi(x)$.

Proof: If ϕ is an automorphism on E, $x \; \varepsilon \; E$, then

$$x \cdot g = x \leftrightarrow \phi(x \cdot g) = \phi(x) = \phi(x) \cdot g = \phi(x)$$

Thus x and $\phi(x)$ have the same isotropy subgroup.

Conversely, let x, y have the same isotropy subgroup. Let $z \; \varepsilon \; E$. As E is homogeneous, there is a $g \in G$ such that $z = x \cdot g$. Define $\phi(z) = y \cdot g$. Since x and y have the same isotropy subgroup, this definition is independent of the choice of g. It is easy to verify that ϕ is an automorphism.

Remark: If ϕ_1 and ϕ_2 are two automorphisms on a homogeneous right G-space E such that for some $x \in E, \phi_1(x) = \phi_2(x)$, then for every $y \in Y$, $\phi_1(y) = \phi_2(y)$.

This follows from the fact that $y = x \cdot g$ for some $g \in G$.

Theorem 1.5.3: A group A of automorphisms of a homogeneous right G-space E contains all automorphisms on E if and only if, for any x, $y \in E$ such that x and y have the same isotropy subgroup, there is a $\phi \; \varepsilon \; A$ such that $\phi(x) = y$.

Proof: If x, $y \in E$ have the same isotropy subgroup we have seen that there is an automorphism ϕ on E such that $\phi_1(x) = y$.

Conversely, suppose for any x, $y \in E$ there is a $\phi \; \varepsilon \; A$ such that $\phi(x) = y$.

Let Ψ be any automorphism on E. To show $\psi \in A$, it is enough to show that there is $\phi \in A$ such that ϕ and Ψ agree at some point. Let $x \in E$. $\psi(x)$ and x have the same isotropy subgroup. Hence there is a $\phi \in A$ such that $\phi(x) = \psi(x)$.

Definition: Let G be a group and H a subgroup of G.

Let $N(H) = \{g \in G: g H g^{-1} = H\}$.

Then $N(H)$ is called the *normalizer* of H. It is a subgroup of G containing H as normal subgroup and it is the largest such.

Theorem 1.5.4: Let E be a homogeneous right G-space and let H be the isotropy subgroup of G corresponding to any point $x_o \varepsilon E$. Then the group A of automorphisms of E is isomorphic to $N(H)/H$.

Proof: Let $S = \{x: x \in E, H$ is the isotropy subgroup of $x\}$.

Step 1: $N(H)$ acts transitively on S. Let $x \in S, g \in N(H), (x \cdot g) \cdot h = x \cdot g \cdot ((x \cdot g) \cdot h)g^{-1} = x \Leftrightarrow ((x \cdot g)h) \cdot g^{-1} = x \Leftrightarrow x \cdot g \cdot h \cdot g^{-1} = x \Leftrightarrow ghg^{-1} \in H \Leftrightarrow h \in H$. Thus $x \cdot g \in S$. Let $x, Y \in S$. G acts transitively on E. Hence there is $g \in G$ such that $y = x \cdot g$. We show that $g \in N(H) \cdot x \cdot g \in S \Leftrightarrow H$ is the isotropy subgroup of $x \cdot g$. Thus

$$h \,\varepsilon\, H \Leftrightarrow (x \cdot g) \cdot h = x \cdot g \Leftrightarrow x \,.\, ghg^{-1} = x \Leftrightarrow g \cdot h \cdot g^{-1} \in H.$$

Thus $g \cdot h \cdot g^{-1} = H$.

Step 2: If $x \in N(H)/H, x \in S$. Define $x \cdot \alpha = x \cdot g$ if $\alpha = Hg$. Note that this definition is unambiguous, for if $Hg_1 = Hg_2, g_1 \cdot g_2^{-1} \in H$ so that $x \cdot g_1 = x \cdot g_2$ for $x \in S$. Clearly $N(H)/H$ acts transitively on S. Further, if $x \cdot \alpha = x$ for any $x \in E, x \cdot \alpha \in N(H/H)$ then $\alpha = H$.

Step 3: Let $x, y \in S$. There is $\phi \in A$ such that $\phi(x) = y$. This is true as x and y have the same isotropy subgroup.

Step 4: Define F on A as follows: Let $\phi(x_o) = y$. then by *step 2*, there is a unique $\alpha \in N(H)/H$ such that $y = x_o$. Put $F(\phi) = \alpha$. By *step 3*, F is onto. F is one-one. For let $F(\phi) = F(\psi) = \alpha$. Then $\phi(x_o) = x_o \cdot \alpha = \psi(x_o)$. Hence $\phi = \psi$ as they agree at one point. Lastly, let $F(\phi) = \alpha, F(\psi) = \beta$. Then $\phi(\psi(x_o)) = \phi(x_o \cdot \beta) = \phi(x_o) \cdot \beta = (x_o \cdot \alpha) \cdot \beta = x_o \cdot (\alpha\beta)$. Hence $F(\phi\psi) = \alpha\beta = F(\phi) \cdot F(\psi)$.

Note: The definition of F depends on the choice of x.

1.6 CATEGORY AND FUNCTOR

Definition: A *category* C consists of:

[*i*] a class of objects, denoted by $Ob(C)$

[*ii*] For every pair, X, Y of objects, a set of morphisms from X to Y, denoted by $C(X, Y)$ or $[X, Y]$. If $\alpha \in C[X, Y]$ then X is called the domain of α and Y the range of α; one also writes $\alpha: X \to Y$ or $X \xrightarrow{\alpha} Y$, or simply $X \to Y$ to denote morphisms from X to Y.

[*iii*] For every ordered triple of objects X, Y, Z a map from $C(X, Y) \times C(Y, Z)$ to $C(X, Z)$, called composition; the image of (α, β) is denoted by $\beta_o \alpha$ or $\beta\alpha$, and is called the composite of β and α. These data have to satisfy the following two axioms:

[*iv*] $\gamma_o(\beta_o\alpha) = (\gamma_o\beta)_o\alpha$ [*associativity*] whenever

$$X \xrightarrow{\alpha} Y \xrightarrow{\beta} Z \xrightarrow{\gamma} W$$

[*v*] There exists an identity morphism $id: X \to Y$, for every object *X*, such that $\alpha_o\, id_x = \alpha$, $id_{yo}\, \alpha = \alpha$ whenever $\alpha: X \to Y$. These identities are easily seen to be unique.

Examples:

[*i*] The category of sets, *C*=Sets. The objects of the category are arbitrary sets (*Ob* [Sets] = The class of all sets), morphisms are maps ([*X*, *Y*]= set of all maps from *X* to *Y*), and composition has the usual meaning.

[*ii*] The category of abelian groups, Here *C* = *AG*. and *Ob*(*AG*) is the class of all abelian groups. [*X*, *Y*] = Hom (*X*, *Y*) is the set of all homomorphisms from *X* to *Y*, and composition has the usual meaning.

[*iii*] The category of topological spaces, *C* = Top. Here *Ob*(Top) is the class of all topological spaces, [*X*, *Y*] is the set of all continuous maps from *X* to *Y*, and composition has the usual meaning.

Definition: If C', *C* are categories, then C' is called a *subcategory* of *C* provided.

[*i*] $Ob(C') \subset Ob(C)$

[*ii*] $C'(X', Y') \subset C'(X', Y')$ for all $X', Y' \in Ob(C')$

[*iii*] The composites of $\alpha \in C'(X', Y')$, $\beta \in C'(Y', Z')$ in C' and *C* coincide.

[*iv*] The identity morphisms of $X \in Ob(C')$ in C' and *C* coincide.

If furthermore, $C'(X', Y') = C(X' \cdot Y')$ for all $X', Y' \in Ob(C)$, then C' is called a *full subcategory*.

Definition: Let *C* and *D* be categories. *A* [covariant] *functor T* from *C* to *D*, in symbols, $T: C \to D$, consists of

[*i*] a map $T: Ob(C) \to Ob(D)$, and

[*ii*] maps $T = T_{xy}: C(X, Y) \to D(TX, TY)$, for every $X, Y \in Ob(C)$, which preserve composition and identities, i.e., such that

[*iii*] $T(\beta \cdot \alpha) = (T\beta) \circ (T\alpha)$, for all morphisms $X \xrightarrow{\alpha} Y \xrightarrow{\beta} Z$ in *C*,

[*iv*] $T[id_x] = id_{TX}$ for all $X \in Ob(C)$.

Proposition 1.6.1: Let $T: C \to D$ be a functor. If $\alpha \in C(X, Y)$ is an isomorphism, then so is *T*, and $(T\alpha)^{-1} = (T\alpha^{-1})$.

Proof: Indeed $\alpha\alpha^{-1} = id_y \Rightarrow T(\alpha)\, T(\alpha^{-1}) = T(\alpha\alpha^{-1}) = T(id_y) = id_x$.

CHAPTER 2

Homotopy Theory

A main problem of topology is the classification of topological spaces: Given two spaces X and Y, are they homeomorphic? This is usually a very difficult question to answer without employing some fairly sophisticated machinery, and the idea of algebric topology is that in which one should transform such topological problems into algebraic problems in order to have a better chance of solution. It turns out, however, that the algebric techniques are usually not delicate enough to classify spaces up to homeomorphism. Hence we shall introduce the notion of *homotopy*, in order to achieve a somewhat *coarser classification*. For the sake of clarity, if we are presented with two spaces X and Y, the problem of deciding whether or not they are homeomorphic is formidable. We have either to construct a homeomorphism between X and Y or, worse still, to prove that no such homeomorphism exists. We therefore try to reflect the problem algebraically. We associate by some means [to be made precise later] a group $G(X)$ with a space X such that for every continuous map [not necessarily a homeomorphism] $f: X \to Y$, there is associated a homeomorphism $f_*: G(X) \to G(Y)$, in such a way that if $f: X \to Y$ happens to be a homeomorphism then $f_*: G(X) \to G(Y)$ is an isomorphism. Thus if X and Y are homeomorphic, $G(X)$ and $G(Y)$ are isomorphic. The converse of this result is not in general true, since $G(X)$ and $G(Y)$ can be isomorphic even though X and Y are not homeomorphic. Thus if $G(X)$ and $G(Y)$ are not isomorphic, X and Y are certainly non-homeomorphic. This method is coarser in the sense that it is more efficacious in declaring X and Y non-homeomorphic rather than determining whether they are actually homeomorphic.

2.1 BASIC NOTIONS

Definition: *A topological pair* (X, A) consists of a topological space X and a subspace $A \subset X$. If $A = \phi$, the pair (X, ϕ) stands for the space X.

A *subpair* $(X', A') \subset (X, A)$ consists of a pair with $X' \subset X$, $A' \subset A$.

A map $f: (X, A) \to (Y, B)$ between pairs is a *continuous function* $f: X \to Y$ such that $f(A) \subset B$.

If $A = \{p\}$, where $p \in X$, then the topological pair (X, A) is called a *pointed topological space* and is denoted by (X, p).

Result: The topological pairs and maps between them form a category, called the *category of topological pairs*. The category contains as full subcategories the category of topological spaces and continuous maps, *Top*, as well as the category of pointed topological spaces and continuous maps, $Top^{(2)}$.

Proof: Easy and, hence, left to the reader.

Definition: Given a pair (X, A), let us denote $(X \times I, A \times I)$ by $(X, A) \times I$ where $I = [0, I)$. Let $X' \subset X$. Suppose $f_o, f_1 : (X, A) \to (Y, B)$ such that $f_o \mid X' = f_1 \mid X'$ (i.e., $f_o = f_1$ on X'), then f is said to be *homotopic* to f relative to X', denoted by $f_o \cong f_1$ rel. X', if there exists a function $F: (X \times A) \times I \to (Y, B)$ such that

$$\left.\begin{array}{l} F(x, 0) = f_o(x) \\ F(x, 1) = f_1(x) \end{array}\right\} \text{for } x \in X$$

and $\quad F(x, t) = f_o(X)$ for $x \in X', t \in I$.

Such a map F is called a *homotopy* relative to X' from f_0 to f_1 and is denoted by $F: f_0 \cong f_1$ rel. X'. If $X' = \phi$, we omit the word relative to ϕ.

Remark: It is easy to see that if $X'' \subset X'$, $f_o \cong f_1$ rel. X'' when $f_o \cong f_1$ rel. X'.

Definition: A map from X to Y is said to be *null homotopic* or *inessential* if it is homotopic to some constant map.

Examples:

(1). Let $X = Y = \mathbf{R}^n \cdot f_o(X) = x \, f_1(x) = 0$ (i.e., $f = 1^n_{\mathbf{R}}$ and f is the constant map of $\mathbf{R}^n$ to its origin.)

If $F: R \times I \to \mathbf{R}^n$ is defined by $F(x, t) = (1-t)x$, then $F: f_o \cong f_1$ rel. $\{o\}$.

(2). Let $X = I = Y$ and define $f_o(t) = t$ and $f_1(t) = o$ for $t \in I$.

If $F: I \times I \to I$ is defined by $F(t, t') = (1-t')t$, then $F: f_o \cong f_1$ rel. $\{o\}$.

(3). Let X be an arbitrary space and Y a convex subset of $\mathbf{R}^n$. Let $f_0, f_1 : X \to Y$ agree on $X' \subset X$. Then $f_o \cong f_1$ rel. X', because $f: X \times I \to Y$ defined by $F(X, t) = tf_1(x) + (1-t)\, f_o(x)$ is a homotopy relative to X' from f_o to f_1.

Theorem 2.1.1: The homotopy relation is an equivalence relation.

Proof: For $f: (X, A) \to (Y, B)$, define

$F: f \cong f$ by $F(x, t) = f(x), x \in X$.

Symmetry: Given $f: f_o \cong f_1$, define F' as follows $F'(x, t) = F(x, 1-t)$.

Then $F': f_1 \cong f_o$.

Transitivity: Given $F: f_o \cong f_1, G: f_1 \cong f_2$,

Define $H: f_o \cong f_2$ as follows:

$$\begin{aligned} H(x, t) &= F(x, 2t) && \text{if } o \le t \le 1/2 \\ &= G(x, 2t-1) && \text{if } 1/2 \le t \le 1 \end{aligned}$$

Note that H is continuous, because its restriction to each of the closed sets $X \times [1/2, 1]$ is continuous and on he boundary also they agree

$$[H(x, 1/2) = F(x, 1) = f(x) = G(x, o)].$$

2.2 HOMOTOPY CLASS

Definition: The equivalance class under $\cong$ of f is denoted by $[f]$, and is called the *homotopy class of f.*

Proposition 2.2.1: The homotopy relation is compatible with composition, i.e., if $f_o, f_1: X \to Y$, $g_o, g_1: Y \to Z$ are maps such that $f_o \cong f_1$, $g_o \cong g_1$, then $g_o\ f_o \cong g_1\ f_1$.

Proof: Let $F: f_o \cong f_1$ and $G: g_o \cong g_1$.

Define $H: X \times I \to Z$ as follows

$$H(x, t) = G(F(x, t), t);\ x \in X, t \in I.$$

i.e., $H: X \times I \xrightarrow{F \times li} Y \times I \xrightarrow{G} Z$ where $F \times I(x, t) = F(x, t), t)$.

Hence H is continuous and $H: g_o\ f_o \cong g_1\ f_1$.

We can therefore define *composition of homotopy classes* by $[g]_o\ [f] = [g_o\ f]$.

This defines a new category Htp: its objects are topological spaces as in Top, *Ob* (Htp) = *Ob* (Top); *morphisms*, however, are *homotopy classes* of continuous maps, $\text{Htp}[X, Y] = \{[f]: f \in \text{Top}\ (X, Y)\}$.

If we assign to every continuous map $f: X \to Y$ its homotopy class $[f]$ we obtain a function π: Top $-$ Htp, $\pi(X) = x$ for $x \in Ob$ [Top], $\pi f = [f]$.

We can define another category, the *homotopy category of pairs* whose objects are topological pairs and whose morphisms are homotopy classes [relative to ϕ]. If Htp$^{(1)}$ denotes this category, Htp$^{(2)}$ contains as full subcategories the homotopy categories the topological spaces Htp [abbreviated as homotopy category] and the homotopy category of pointed topological spaces.

There is a covariant functor from the category of topological pairs and maps Top$^{(2)}$ to Htp$^{(2)}$ defined as follows:

π: Top$^{(2)} \to$ Htp$^{(2)}$ with $\pi(X, A) = (X, A)$ and $\pi(f) = [f]$ = homotopy classes of f.

2.3 HOMOTOPY EQUIVALENCE

Definition: Some of the main tools in algebraic topology are functors t: Top $\to A$ where A is some algebraic category (groups, rings, ...). In most cases these functors are homotopy-invariant i.e., $f_o \cong f_1 \Rightarrow tf_o = tf_1$. Equivalently t factors through π defined in 2.2, i.e., $t = t_o'\pi$ where Top $\xrightarrow{\pi}$ Htp $\xrightarrow{t'} A$. Due to this fact, algebraic topologists are often more interested in the category Htp than in Top. This leads to the following definitions:

Homotopy equivalence: A map $f: (X, A) \to (Y, B)$ is said to be a *homotopy equivalence* if $[f]$ is invertible in the homotopy category of pairs, i.e., there exists a continuous map $g: (Y, B) \to (X, A)$ such that $fg \cong I_Y$ and $gf \cong I_X$.

Define $(X, A)\ \beta\ (Y, B)$ holds if and only if there exists a homotopy equivalence f between (X, A) and (Y, B).

It is easy to check that ρ is, in fact, an equivalence relation among pairs. The equivalence classes are called *homotopy types.*

The simplest non-empty space is a one-point space (p, p). We characterise the *homotopy type* of such a space as follows:

Definition: A topological space X is said to be *contractible* if the identity map 1_X of X is homotopic to some constant map of X to itself.

A homotopy from 1_X to the constant map of X to $x_o \in X$ is called a *contraction* of X to x_o.

Examples 1 and 2 of Section 2.1 show that **R** and **I** are contractible and example 3 shows that any convex set of $\mathbf{R}^n$ is contractible.

Theorem 2.3.1: X is contractible if and only if for any space T, any two continuous maps $f, g: T \to X$ are homotopic.

Proof: Sufficiency is obtained by setting $T = X$ and letting f and g be respectively, the identity and constant maps.

For necessity, suppose X is contractible; say $1_X \cong c$, where c is a constant map from X to itself. Let $f, g: T \to X$ be any two continuous maps.

Now $f = 1_X \, o \, f$ and $g = 1_X \, o \, g$.

Since $1_X \cong c$ and $f \cong f$ and $g \cong g$ trivially, by composition $f = 1_X \, o \, f \cong c \, o \, f$ and $g = 1_X \, o \, g \cong c \, o \, g$. But $c \, o \, f = c \, o \, g$. So $f \cong g$.

Theorem 2.3.2: X is contractible if f it is homotopically equivalent to a one-point space.

Proof: Suppose X is contractible, say the identity map $1_X : X \to X$ is homotopic to the constant function $c(x) = x$.

Let $Y = \{x_o\}$, and $i: Y \to X$ be the inclusion map. Then $c \, o \, i = 1_Y$ and $i \, o \, c = c \cong 1_X$. Thus i is a homotopy equivalence from Y to X. Conversely, suppose, $f: X \to Y$ is a homotopy equivalence between X and Y where $Y = \{p\}$.

Let $g: Y \to X$ be a homotopy inverse of f.

Now $g_o \, f \cong 1_X$ but $g_o \, f$ is a constant function from X to X.

Hence X is contractible.

Corollary 1: If Y is contractible, any two continuous maps of Y into Y are homotopic.

Proof: Let X and Y be contractible and $f: X \to Y$ be a constant map. Since X is contractible, $1_X \cong c$ where c is a constant map on X into itself; consider $g: Y \to X$ as follows $g(y) = c$ for all $y \in Y$. Hence $g \, o \, f = c$ so that $g \, o \, f = c \cong 1_X$.

Again note that $f \, o \, g(y) = f \, o \, c = d$, a constant map of Y to Y. Because Y is contractible 1_X and d are homotopic by corollary 1, i.e., $f \circ g \cong 1_Y$.

Hence f is a homotopy equivalence.

Corollary 2.3.3: All maps $f: X \to Y$ are homotopic to a constant map when Y is contractible.

Proof: If ϕ is an automorphism on E, $x \varepsilon E$, then

$$x \cdot g = x \leftrightarrow \phi(x \cdot g) = \phi(x) = \phi(x) \cdot g = \phi(x).$$

Proof: Let $f: X \to Y$ be a map and Y contractible then $1_y \cong c$, where $c: Y \to Y$ a constant map. Since $f \cong f$ and $1_y \cong c$, we have $1_y f \cong cf$ i.e., $f \cong cf$. But cf is a constant map of X to Y. So f is homotopic to a constant map of Y.

Corollary 2.3.4: If X is contractible to a point $p \in X$, then all maps $f: X \to Y$ are homotopic to a constant map.

Proof: As X is contractible, $1_x \cong c$ where $c: X \to p$ the constant map. Since $f \cong f$, $f \cong f 1_X \cong fc = c'$. Now $c' = fc$ is constant map of X to Y. So $f \cong c'$.

The next result establishes an important relation between homotopy and the extendibility of maps.

Theorem 2.3.5: Let p_o be any point of S^n, the unit ball in R^{n+1}, and let $f: S^n \to Y$, The following are equivalent:

(*a*) f is null homotopic.

(*b*) f can be continuously extended over $E^{n+1} = \{x \in \mathbf{R}^{n+1}; \|x\| \leq 1\}$

(*c*) f is null homotopic relative to p_o.

Proof: $(a) \Rightarrow (b)$ Since f is null homotopic, let F: $f \cong c$, where c is the constant map of S^n to $y_0 \in Y$.

Define an extension f' of f over E^{n+1} by

$$\begin{aligned} f'(x) &= y_o && \text{if } 0 \leq \|x\| \leq 1/2, \\ &= F(x/\|x\|, 3 - \|x\|) && \text{if } 1/2 < \|x\| \leq 1, \end{aligned}$$

Since $F(x, 1) = y_o$ for all $x \in S^n$, the map f' is well defined. f' is continuous because its restriction to each of the closed sets $\{x \in E^{n+1}: 0 \leq \|x\| \leq 1/2\}$ and $\{x \in E^{n+1}; 1/2 \leq \|x\| \leq 1\}$ is continuous.

Since $F(x, o) = f(x)$ for $x \in S^n$, $f'/S^n = f$ and f' is a continuous extension of f to E^{n+1}.

$(b) \Rightarrow (c)$. If f has the continuous extension $f': E^{n+1} \to Y$, define $F: S \times I \to Y$ by $F(x, t) = f'((1-t)\, x + tp_o)$.

Then $F(x, 0) = f'(x) = f(x)$ and $F(x, 1) = f'(p_o)$ for $x \in S^n$. Since $F(p_o, t) = f'(p_o)$ for t in I, F is a homotopy relative to p_o from f to the constant map to $f'(p_o)$.

$(d) \Rightarrow (a)$: Obvious.

Corollary 2.3.6: Any continuous map from S^{n+1} to a contractible space has a continuous extension over E^{n+1}.

Proof: Combining Theorems 2.3.1 and 2.3.5. we get the result.

2.4 RETRACTION AND DEFORMATION

Definitions: A subset A of X is a *retract of X* if there exists a continuous map $r: X \to A,$ called a *retraction*, such that $r(a) = a$ for such $a \in A.$ We call A a *deformation retract* of X if there is a retraction $r: X \to A$ which is homotopic [as a map into X] to the identity function 1_X on x i.e., $1_X \cong ir$ where i is the inclusion map. $i: A \to X.$ If $F: 1_X \cong ir,$ F is called a *deformation retraction*.

Note: A retract need not be a deformation retract. In fact, the one-point subsets of any space are retracts, but no one-point subspace of S^1 is a deformation retract [To be proved later]

Theorem 2.4.1: If A is a deformation retract of X, then A is homotopically equivalent to X.

Proof: Let $i: A \to X$ be the inclusion and $r: X \to A$ be the retraction. Then $i \circ r \cong 1_x$ and $r \circ 1 = 1_A,$ hence i is a homotopy equivalence.

2.5 THE FUNDAMENTAL GROUP

The concept of path-connectedness, in which it is required that it be possible to reach any point in the space from any other point along a continuous path is necessary for the notion of fundamental group. This approach is especially useful in studying connectivity properties from an algebraic point of view, e.g., via homotopy theory.

Definition: A space X is *path-connected* [or *arcwise connected*] if for every pair of points x and y in X, there is a continuous function $f: I \to X$ such that $f(o) = x,$ $f(1) = y.$ Such a function f [as well as its range $f(I),$ when confusion is not possible] is called a *path* from x to y.

A space X is *locally path-connected* if each point has a neighbourhood base consisting of path connected sets [We should point out here that a subset A of X is path-connected if any two points in A can be joined by a path *lying in A*,]

Theorem 2.5.1: Every path connected space is connected.

Proof: Easy.

Addition of Paths: Paths can be 'added' in the following sense. If $a, b, c \in X,$ and $f_1: I \to X$ is a path from a to b, while $f_2: I \to X$ is a path from b to c, then the function $f: I \to X$ defined by

$$f(t) = \begin{cases} f_1(2t) & \text{if } 0 \le t \le 1/2 \\ f_1(2t-1) & \text{if } 1/2 \le t \le i \end{cases}$$

is a path from a to c, obtained by 'putting the paths f_1 and f_2 end-to-end'. [For example, f is continuous because it is continuous on each of the closed sets $[0, 1/2]$ and $[1/2, 1]$.

This path addition provides a way to associate with each path-connected space X a group $\pi_1(X)$ in such a way that homeomorphic spaces have isomorphic groups. The branch of algebraic topology which is concerned with relationships between X and $\pi_1(X)$ is known as *homotopy theory*. We shall use the addition of paths defined above to provide a partial converse to the Theorem 2.5.1.

Theorem 2.5.2: A connected, locally path-connected space X is path-connected.

Proof: Let $a \in X$ and let H be the set of all points of X which can be joined by a path to a. As $a \in H,$ H is non-empty. If H is closed and open, $H = X$.

But H is open. For $b \in H$, let U be a path-connected *neighbourhood* of b. Then any point $z \in U$ can be joined by a path to b and hence can be joined to a by adding the path from b to a.

Also H is closed. For if $b \in H$, let U be any path-connected *neighbourhood* of b. Then $U \cap H \neq \phi$; say $z \in U \cap H$. Now b can be joined to z by a path and z can be joined to a by a path, so by addition of paths again, $b \in H$.

Corollary: An open connected subset of **R** is path-connected. We are now in a position to use the 'addition' of paths to associate with any topological space a group [actually, several groups].

Definition: Let X be a topological space, x_o a fixed point in X. A continuous function $f: I \to X$ will be called a *loop based at* x_o if $f(0) = f(1) = x_o$.

Two loops f and g based at x_o will be called *loop homotopic* [or simply homotopic] if $f \cong g$ rel $\{0, 1\}$. Thus a loop homotopy between two loop S based at x_o must be a relative homotopy which at any stage carries the end points of 1 into x_o.

The relation $\cong$ between loops based at x_o is an equivalence relation and hence partitions the set $L(X, x_o)$ of loops based at x_o into equivalence classes. The equivalence class containing f will be denoted $[f]$, and the set of all such equivalence classes of loops based at x_o will be denoted by $\pi_1[X, x_o]$.

We can 'add' loops just as we 'added' paths earlier. If f_1 and f_2 are loops based at x_o, we define a new loops $f_1 * f_2$ as follows:

$$(f_1 * f_2)(t) = \begin{cases} f_1(2t) & \text{if } 0 \le t \le 1/2 \\ f_2(2t-1) & \text{if } 1/2 \le t \le 1 \end{cases}$$

Then we can lift the operation $*$ to the set $\pi_1(X, x_o)$ of equivalence classes of loops by defining. $[f_1] * [f_2] = [f_1 * f_2]$. We shall now show that $*$ is well defined in $\pi_1(X, x_o)$. That is, if $f_1 \cong g_1$ and $f_2 \cong g_2$, then $f_1 * f_2 \cong g_1 * g_2$ where $f_1, f_2, g_1, g_2 \in L(X, x)$. Let $F\colon f_1 \cong g_1$ rel $\{0, 1\}$ and $G\colon f_2 \cong g_2$ rel $\{0, 1\}$.

A homotopy $H = F * G\colon I \times I \to X$ is defined by the formula.

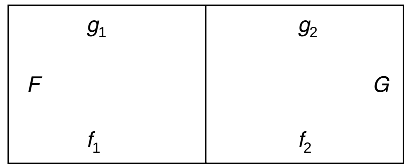
$$H(t, t') = \begin{cases} F(2t, t') & \text{when } 0 \le t \le 1/2 \\ G(2t-1, t') & \text{when } 1/2 \le t \le 1 \end{cases}$$

$$H(t, 0) = \begin{cases} F(2t, 0) & \text{when } 0 \le t \le 1/2 \\ G(2t-1, 0) & \text{when } 1/2 \le t \le 1 \end{cases} = f_1 * f_2(t)$$

$$H(t, 1) = \begin{cases} F(2t, 1) & \text{when } 0 \le t \le 1/2 \\ G(2t-1, 1) & \text{when } 1/2 \le t \le 1 \end{cases} = g_1 * g_2(t)$$

Then $H = F * G\colon f_1 * f_2 \cong g_1 * g_2$ rel $\{0, 1\}$.

This is illustrated in the diagram:



$F * G$

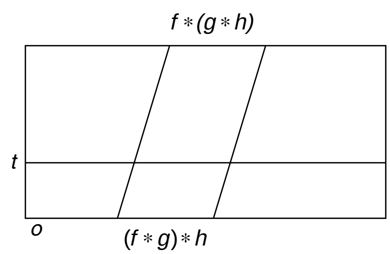
Thus $*$ is a binary operation on $\pi_1(X, x_o)$.

Theorem 2.5.3: $\pi_1(X, x_o)$ with the operation $*$ is a group.

Proof: We check associativity first. For this, it suffices to show that $(f * g) * h \cong f * (g * h)$ for loops *f*, *g* and *h* based at x_o. Such a homotopy $G: I \times I \to X$ is defined by the formula.

$$G: (t, t') = \begin{cases} f\left(\dfrac{4t}{t'+1}\right) & \text{when } 0 \le t \le \dfrac{t'+1}{4} \\ g(4t - t' - 1) & \text{when } \dfrac{t'+1}{4} \le t \le \dfrac{t'+2}{4} \\ h\left(\dfrac{4t - 2 - t'}{2 - t'}\right) & \text{when } \dfrac{t'+2}{4} \le t \le \dfrac{t'+1}{4} \end{cases}$$

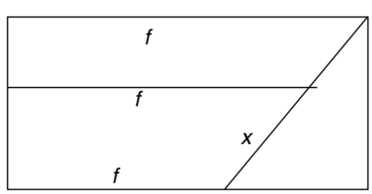
and pictured in the diagram:



where the bottom line represents $(f * g)*h$ and the top line $f * (g * h)$. Now let *a* denote the constant map $e(t) = x_o$ for all $t \in I$. We claim [*e*] serves as an indentity in $\pi_1[X, x_o]$. It suffices to show $f * e \cong f$ and $e * f \cong f$ for all $f \in L(X, x_o)$. To exhibit a homotopy for the first, define for each $t \in I$.

$$H(x, t) = \begin{cases} f\left(\dfrac{2x}{1+t}\right) & \text{when } 0 \le x \le (t+1)/2 \\ x_o & \text{when } (t+1)/2 \le x \le 1. \end{cases}$$

This is pictured in the diagram:



H

H is continuous on $I \times I$ since it is continuous on each of the closed sets $\{(x, t): x \leq (2 - t)/2\}$ and $\{(x, t): x \leq (2 - t)/2\}$ and it is checked that $H(x, 0) = f(x)$ and $H(x, 1) = (f * e)[x]$ for all $x \in I$. The relation $e * f \cong f$ is done in a similar fashion.

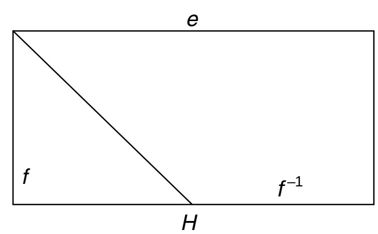
Finally, we show the existence of the inverses. For each loop at x_o, define f^{-1} to be the loop $f^1(x) = f(1 - x), 0 \leq x \leq 1.$ and let $[f] = [f^{-1}]$. This is well defined because $f \leq g \Rightarrow f^{-1} \cong g^{-1}$: For if $F: f \cong g$ rel $\{0, 1\}$ define $H: I \times I \to X$ as follows:

$H(x, t) = F(1 - x, t)$. Then $H: f^{-1} \cong g^{-1}$ rel $\{0, 1\}$.

To show $[f]^{-1}$ is the inverse of $[f]$, it suffices to check that $f * f^{-1} \cong e$ and $f^{-1} * f \cong e$. First let

$$H(x, t) = \begin{cases} f(x) & \text{when } o \leq x \leq (1 - t)/2 \\ f^{-1}(x + t) & \text{when } (1 - t)/2 \leq x \leq 1 - t \\ x_o & \text{when } 1 - t \leq x \leq 1. \end{cases}$$

The function H is continuous on each of three closed sets which cover the square and thus continuous, and clearly $H(x, o) = (f * f^{-1})(x)$ and $H(x, 1) = e(x)$ for all $x \in 1$. The diagram is given below



The homotopy showing $f^{-1} * f \cong e$ is similarly constructed.

Definition: The group $\pi_1[X, x_o]$ is called the *fundamental group of X at* x_o. The dependence on the base point x_o is not illusory in the general case. But for an important special class of spaces it can be ignored.

Theorem 2.5.4: If X is a path-connected space, then for any pair of points x_o and x_1 in X, $\pi_1(X, x_o)$ and $\pi_1(X, x_1)$ are isomorphic.

Proof: Let $h: I \to X$ be a path from x_o to x_1, h^{-1} the path h traversed in the opposite direction. For each loop f based at x_o, define $a(f)$ to be the following loop based at x_1: $a(f) = h^{-1} * f * h$

This induces a mapping $A[f] = [a(f) = [h^{-1} * f * h]$ of $\pi_1(X, x_o)$ to $\pi_1(X, x_1)$. We shall show that this is the desired isomorphism. First, a is a single-valued, i.e., if $f \cong g$, then $a(f) \cong a(g)$. For if $H: f \cong g$, then the function G defined by

$$G(x, t) = \begin{cases} h^{-1}(x) & \text{when } 0 \leq t \leq 1/3 \\ h(x, 3t - 1) & \text{when } 1/3 \leq t \leq 2/3 \\ h(x) & \text{when } 2/3 \leq t \leq 1. \end{cases}$$

is a homotopy between $h^{-1} * f * h$ and $h^{-1} * g * h$.

Secondly, *A* is a homeomorphism; that is $A([f] * [g]) = A(f) * A(g)$.

But $A(f) * A(g) = [a(f)] * [a(g)] = [a(f)] * [a(g)] = [h^{-1} * f * g * h^{-1} * g * h] =$

$[h^{-1} * f * g * h] = [a(f * g)] = A[f * g] = A([f] * [g])$.

A is onto: Let $[\sigma]$ be an element of $\pi_1[X, x_1]$. Consider a loop $f = h * \sigma * h$ based at *x*.

Now $a(f) = h^{-1} * f * h = \sigma$. Hence $A[f] = [\sigma]$.

A is 1–1: Let $[f] \neq [g]$, then *f* is not homotopic to *g*. We shall show that $a(f)$ is not homotopic to $a(g)$. Suppose $a(f) \cong a(g)$ i.e., there exists $G: a(f) \cong a(g)$. Note that $f = h * a(f) * h^{-1}$ and $g = h * a(g) * h^{-1}$. Define the function *H* as follows:

$$H(x, t) = \begin{cases} h(x) & \text{when } 0 \le t \le 1/3 \\ h(x, 3t - 1) & \text{when } 1/3 \le t \le 2/3 \\ h^{-1}(x) & \text{when } 2/3 \le t \le 1. \end{cases}$$

Then *H* is a homotopy between *f* and *g* – *a* contradiction. Hence A is 1–1, so that A is an isomorphism of $\pi_1(X, x_o)$ and $\pi_1(X, x_1)$.

Thus for a path-connected space *X*, we can speak of the fundamental group $\pi_1(X)$ of *x*. In fact $\pi_1(X)$ is a set of groups indexed by the points of *X*, any two of which are isomorphic.

Historical Note: The fundamental group was introduced by *H*. Poincare [Analysis Situs; Cinqueme complement a Analysis Situs] around 1900. This is the reason of calling fundamental group also Poincare Group. Hurewicz studied the fundamental group and introduced the higher homotopy groups in a series of four papers in the 1930's.

We have associated with each pointed space (X, x_o) an algebraic object $\pi_1(X, x_o)$. The power of the homotopy method in topology is largely traceable to the fact that mappings of pointed spaces induce homeomorphisms of the associated algebraic structures.

Theorem 2.5.5: Every continuous mapping $f: (X, x_o) \to (Y, y_o)$ induces a homeomorphism $f^*: \pi_1(X, x_o) \to (Y, y_o)$.

Proof: For each loop *g* at x_o in *X*, let $f'(g)$ be the loop at y_o in *Y* defined by $f'(g)(t) = f[g(t)]$. This defines a mapping f' from $L[X, x_o]$ to $L[Y, y_o]$ which in turn induces a mapping $f^*: \pi_1(X, x_o) \to \pi_1(Y, y_o)$ as follows:

$$f^*([g]) = [f'(g)]$$

To see that *f* is well-defined, not that if *H* is a homotopy between g_1 and g_2 in $L[X, x_o]$, then foH is a homotopy between $f'[g_1]$ and $f'[g_2]$ in $L[Y, y_o]$.

It remains to show that *f* is a homeomorphism, for which it suffices to establish.

The necessary algebraic property for *f*. But

$$f'(g * h) = \begin{cases} f[g\ (2x)] \quad = f'(g)\ (2x) & \text{when } 0 \le x \le 1/2 \\ f[h\ (2x-1)] = f'(h)\ (2x-1) & \text{when } 1/2 \le x \le 1. \end{cases}$$

$$= f'(g) * f'(h)$$

Hence the proof.

Theorem 2.5.6: (*a*) If *f* is the identify on *X*, i.e., $f = 1x, [1x]$ is identify on $\pi_1(X, x_o)$.

(*b*) If *f* and *g* are maps from (X, x_o) to (Y, y_o) such that $f \cong g$ rel x, then $f^* = g^*$.

(*c*) If $f: (X, x_o) \to (Y, y_o)$ and $g: (Y, y_o) \to (Z, z_o)$, then $(g_o f)^* = g^*{}_o f^*$.

(*d*) If $r: (X, x_o) \to (A, x_o)$ is a retraction and $i = (A, x_o)\ (X, x_o)$ is the inclusion map, then *i* is a homeomorphism and *r* is an epimorphism.

Proof: (*a*) Obvious.

(*b*) It suffices to show that if *h* is a loop based at x_o in *X*, then $f'(h)\ g'(h)$ based at y_o. But *f* and *g* are homotopic relative to x_o, then $f_o h$ and $g_o h$ are homotopic; that is, $f'(h)$ and $g'(h)$ are homotopic.

(*c*) If *h* is any loop based at x_o in *X*, then $(g_o f)\ (h) = [(g_o f)'\ (h)] = [g'(f'\ (h))]$.

(*d*) $r_o i$ is the identity map on (A, x_o), so $r *_o i * = (r_o i) *$ is the identify on $\pi_1(A, x_o)$

Theorem 2.5.7: If (X, x_o) and (Y, y_o) are homotopically equivalent as pointed topological spaces, then $\pi_1(X, x_o)$ and $\pi_1(Y, y_o)$ are isomorphic.

Proof: There are maps $f: (X, x_o) \to (Y, y_o)$ and $g: (Y, y_o) \to (X, x_o)$ such that *fog* is homotopic to the identity on *Y* and $g_o f$ is homotopic to the identity on *X*. Then from the previous theorem 2.5.7 $g *_o f * = (g_o f) *$ is the identity on $\pi_1[Y, y_o]$ and $f *_o g * = (fog) *$ is the identity on $\pi_1(X, x_o)$. Since $f *$ and $g *$ are homeomorphism, they are thus isomorphisms.

The importance of the theorem above is obvious; e.g., the question as to whether or not $\pi_1(X, x_o)$ has any given group theoretic property [e.g., it is abelian, finite, nilpotent, free, etc.] is independent of the point x_o, and thus depends only on the space *X*, provided *X* is connected. On the other hand, we must keep in mind that there is no *canonical* or *natural* isomorphism between $\pi_1(X, x_o)$ and $\pi_1(X, x_1)$; corresponding to each choice of a path from x_o to x_1 there will be an isomorphism from $\pi_1(X, x_o)$ to $\pi_1(X, x_1)$.

In view of the Theorem 2.5.6, a continuous map $f: X \to Y$ induces a homeomorphism $f*: \pi_1(X, x) \to \pi_1(Y, f(x))$; and if *f* is a homeomorphism, then $f *$ is an isomorphism. This induced homeomorphism will be extremely important in studying the fundamental group.

Note that the theorem 2.5.7[*b*] and the theorem 2.5.8 require relative homotopy. Unfortunately, the condition that the homotopy should be relative to the base point *x* is too restrictive for many purposes. The condition can be omitted and the following stronger result is true:

If X and Y are path-connected and homotopically equivalent, then $\pi_1(X)$ and $\pi_1(Y)$ are isomorphic. [The proof is difficult and is deferred to a later section].

We have seen that π_1 attaches a group $\pi_1(X, x)$ to each pointed topological space (X, x). Thus we can speak of the *fundamental group functor* from the category of pointed topological spaces top to the category of groups G.

Definition: A subset A of X is a *strong deformation retract* of X if there exists a retraction $r: X \to A$ such that $ir \cong 1_x$ rel. A where $i: A \to X$ is the inclusion map.

Theorem 2.5.8: If A is a strong deformation retract of X, then the inclusion map $i: A \to X$ induces an isomorphism of $\pi_1(A, a)$ onto $\pi_1(X, a)$ for every $a \in A$.

Proof: Since $ri = 1_A$, $r * i * =$ identity map on $\pi_1(A, a)$, for any $a \in A$. Take any $a \in A$, then $ir \cong 1_x$ rel. $A \Rightarrow ir \cong 1_x$ rel. $\{a\}$. Then by theorem. $2L(b)$, $(ir)* = (1_x)*: \pi_1(X, a) \to \pi_1(X, a)$. But $(1_x)*$ is the identity map on $\pi_1(X, a)$. Hence $i * \cdot r * = (ir)*$ is the identity map on $\pi_1(X, a)$. Consequently $i: \pi_1(A, a) \to \pi_1(X, a)$ is an isomorphism.

We shall use this theorem in two different ways. We shall use it to prove that two spaces have isomorphic fundamental groups. On the other hand, we can use it to prove that a subspace is not a strong deformation retract by proving the fundamental groups are non-isomorphic. A simple characterisation of a contractible space in terms of strong deformation retract is as follows:

Proposition 2.5.9: A topological space is contractible if and only if there exists a point $x_o \in X$ such that $\{x_o\}$ is a strong deformation retract of X.

Proof: Let X be contractible. There exists $x_o \in X$ such that 1_x is homotopic to the constant map $c: X \to x_o$. Since $1_X(x_o) = x_o = c(x_o)$ $1_X \cong c$ rel. $\{x_o\}$. That is, $\{x_o\}$ is a strong deformation retract of X. Converse also follows similarly.

Definition: A topological space X is *simply connected* if it is path connected and $\pi_1(X, x)$ is the trivial group for some [and hence any] $x \in X$.

Theorem 2.5.10: If X is contractible, then X is simply connected.

Proof: First we show that X is path connected. Let $x_o, x_1 \in X$, $x_o \neq x_1$. There exists $y \in X$ such that $1_X \cong c$ rel. $\{y\}$ where c is the constant map from X to $\{y\}$. Let $F: 1_X \cong c$. Put $f(x) = F(x_o, x)$.

Then $f(0) = f(x, 0) = x_o$ and $f(1) = f(x_o, 1) = y$.

Thus f is a path from x_o to y. Similarly there is a path from x_1 to y. Hence, by addition of paths, we get a path joining x_o and x_1.

Since $\{y\}$ is a strong deformation retract of X, $i*$ is an isomorphism from $\pi_1(y, y)$ onto $\pi_1(X, y)$ where $i*$ is as in the theorem 2.5.9. But $\pi_1(y, y)$ is trivial. Hence $\pi_1(X, y)$ is trivial. Thus X is simply connected by definition.

2.6 FUNDAMENTAL GROUP OF THE CIRCLE

We shall compute $\pi_1(\mathbf{S}^1, \underline{\mathbf{I}})$ following a method used by A.W. Tucker, where $\underline{I} = (1, 0)$.

The exponential map $e_X : \mathbf{R} \to \mathbf{S}^1$ is defined by $e_X[t] = e^{2\pi \text{it}}$. Then e_X is continuous. $e_X(t_1 + t_2) = e_X(t_1)\, e_X(t_2)$ [Where the right hand side is multiplication of complex numbers]. and $e_X(t_1) = e_X(t_2)$ if and only if $t_1 - t_2$ is an integer.

It follows that $e_x|(-1/2, 1/2)$ is a homeomorphism of the open interval $(-1/2, 1/2)$ onto $\mathbf{S}^1 - \{e^{i\pi}\} = S - \{-1\}$. We let $1 * g: \mathbf{S}^1 - (-1) \to (-1/2, 1/2)$ be the inverse of $e_X|(-1/2, 1/2)$.

We need two key lemmas.

2.7 LIFTING LEMMA

Lifting Lemma: If f is a path $\mathbf{S}^1$ with initial point I, there is a *unique* path f' in $\mathbf{R}$ with initial point 0 such that $e_X \ o \ f = f$.

Proof: Let $Y = I = [0, 1]$. Let $f: Y \to \mathbf{S}^1$ be a path in $\mathbf{S}^1$ starting at $\underline{1}$. Since Y is compact, f is uniformly continuous. Therefore, there exists $\varepsilon > 0$ such that $|y - y'| < \varepsilon$ implies $[f(y) - f(y')] < 1$.

2.8 COVERING HOMOTOPY LEMMA

Covering Homotopy Lemma: If f and g are paths in S with $f[0] = 1 = g[0]$ such that $F: f \cong g$ rel. $\{0, 1\}$, then there is a unique $F': \mathbf{I} \times \mathbf{I} \to \mathbf{R}$ such that $F': f' \cong g'$ rel. $\{0, 1\}$ and $e_X \ o \ F' = F$.

In particular, for such y and y', $f(y) \neq -f(y')$. Since Y is compact, Y is bounded, there exists a positive integer N such that $|y| < N \in$ for all $y \in Y$. Then for each $0 \leq j \leq n$ and all $y \in Y$.

$$\left| \frac{(j+1)y}{N} - \frac{jy}{N} \right| < \in \text{ and so } \left| \frac{f[(j+1)y]}{N} - \frac{f(jy)}{N} \right| < 1.$$

It follows that the quotient $f\left(\frac{j+1}{N} y\right) / f\left(\frac{j}{N} y\right)$ is a point of $\mathbf{S}^1 - \{-1\}$.

Let $h_j: \mathbf{I} \to \mathbf{S}^1 - \{-1\}$ for $0 \leq j \leq N$ be the map defined by

$$h_j(y) = f\left(\frac{j+1}{N} y\right) / f\left(\frac{j}{N} y\right)$$

Then, for all $y \in Y$, we see that $f(y) = f(0)\, h_o(y)\, h_1(y) \ldots h_{N-1}(y)$. We define $f': Y \to \mathbf{R}$ by $f^1(y) = \lg\{h_o(y)\} + \lg\{h_1(y) + \ldots + \lg(h_{N-1}{}^{(y)})$ f' is the sum of N continuous functions from Y to $\mathbf{R}$, so it is continuous.

Clearly, $f'(0) = 0$, $e_x \ o \ f' = f$.

To show the uniqueness let $\sigma':Y \to \mathbf{R}$ such that $\sigma'[\mathbf{0}] = \mathbf{0}$, $e_x\, o\, \sigma' = f$. Then $f' - \sigma'$ is a continuous map of Y into the kernel of e_x. Since kernel of $e_x = Z$, $f' - \sigma': Y \to Z$. But Y is connected, hence $f' - \sigma'$ must be a contract [which must be o] So $f' = \sigma'$.

Proof: Let $Y = \mathbf{I} \times \mathbf{I}$. Then $F: Y \to \mathbf{S}^1$ and $F: f \cong g$ rel. $\{0, 1\}$. Since $\mathbf{I} \times \mathbf{I}$ is compact, F is uniformally continuous. Hence there exists $\varepsilon > 0$ such that $\|y - y'\| < \varepsilon \Rightarrow |F(y) - F(y')| < 1$ so that $F(y) \neq -F(y')$ for such $y,\ y' \in Y$. As in the proof above choose $N > 0$ such that $\|y\| < \varepsilon$ for $y \in Y$.

Define $F': \mathbf{I} \times \mathbf{I} \to \mathbf{R}$ by

$$F'(y) = \log\left(F(y)/\left(\frac{N-1}{N}y\right)\right) + \log\left(F\left(\frac{N-1}{N}y\right)/F\left(\frac{N-2}{N}y\right)\right) + ... + \log\left(F\left(\frac{U}{N}\right)/F(o)\right).$$

Then (*i*) $F'(0) = 0$

(*ii*) $e_x\, o\, f' = F$

(*iii*) F' is continuous.

To check $F': f' \cong g'$ rel. $\{0,\ 1\}$: $e_x\, o\, F'(s,\ 0) = F(s,\ 0) = f[s]$.

By uniqueness part of the lemma above, $F'(s, 0) = f'(s)$. Similarly, $F'(s, 1) = g'(s)$. Since $F(0,\ t) = F(1,\ t) = 1$ for all $t \in\ f$, $ex\ o\ F'(0,\ t) = e_x\ o\ F'(1,\ t)$. Hence $F(0,\ t)$ and $F'(1,\ t)$ take on only integral values for all $t \in 1$. It follows that they must be constant. Because $F'(0,\ 0) = 0$, $F'(0,\ t) = 0$ for all $t \in 1$. Hence $F': f' \cong g'$ rel. $\{0, 1\}$. Arguments similar to that of the lifting lemma lead to the fact that F' is unique.

Corollary: The end point $f'(1)$ of f' depends only on the homotopy class of *f*.

Definition: Let $f: \mathbf{I} \to \mathbf{S}^1$ be a loop at $\underline{1}$. Since $\underline{1}$ is the initial point of *f*, by Sec. 2.7, there is a unique path f' in *R* with initial point $f'(0) = 0$ and $e_x\, o\, f' = f$.

Because $ex\ (f'(1)) = f(1) = 1$, $f'(1)$ is an integer. We define the degree of *f* by deg $f = f'[1]$. It follows that there is a well-defined function deg from $\pi_1(\mathbf{S}^i, \mathbf{I})$ to $\mathbf{Z}$ defined by $\deg [f] = \deg f$, where *f* is a loop in $\mathbf{S}^1$ at $\underline{1}$.

Theorem 2.8.1: The function $\deg: \pi_1(\mathbf{S}^1, 1) \to \mathbf{Z}$ is an isomorphism.

Proof: (*i*) deg is a homeomorphism: Let $[f], [g] \in \pi_1(\mathbf{S}^i, \mathbf{I})$. Let $m = \deg [f]$. $n = \deg [g]$. Then, $m = f'(1)$ and $n = g'(1)$. Let h' be the path from *m* to *m* + *n* in $\mathbf{R}$ defined by $h'(s) = g'(s) + m$. Then, $e_{x\ o}\ h' = e_{x\ o}\ g' = g$.

Now $e_{x\ o}\ (f' * h')\ (s) = f * g(s)$ i.e., $e_{x\ o}\ (f'h') = f * g$. Then $f' * h'$ is the lifting of $f * g$ with initial point 0, its end point is *m* + *n*. Hence

$$\deg\ ([f] * [g] = \deg\ ([f * g]) = f' * h'(1) = m + n = \deg [f] + \deg [g].$$

(*ii*) deg is one-one: Suppose $\deg [f] = 0$, i.e., $f'(1) = 0$. So, f' is a loop at 0 in $\mathbf{R}$; $\mathbf{R}$ being contractible $f' \cong 0$ rel. [0, 1]. Applying *ex* we get $f = e_x\, o\, f'\ e_x\, o\, 0$ rel. [0, 1], i.e., $f \cong 1$ rel. [0, 1]. Hence $[f] = 1$, the identity element of $\pi_1(\mathbf{S}^1, \mathbf{I})$

(*iii*) deg is onto: For if n is an integer, there is a path f_n' in **R** defined by f_n' $(t) = tn$. Let $f = e_x \, o \, f_n'$. Then clearly, deg $[f_n] = f_n'(1) = n$.

Consequently deg is an isomorphism onto the set **Z** of integers.

Remark: The basic idea involved in the above discussion of the degree of an element of $\pi_1[\mathbf{S}^1]$ will be refined and generalized in the discussion of the fundamental group of a covering space in this chapter.

As a corollary to theorem 2.3.1, we see that the fundamental group of any space with a circle as strong deformation retract is infinite cyclic. Examples of such spaces are the Mobius strip, a punctured disc, the punctured plane, a region in the plane bounded by two concentric circle, etc.

The only property of $\mathbf{S}^1$ used in the proof of Theorem 2.3.1 is that it is a topological group. The quotient of **R** by **Z**. The only property of **R** used in this proof is that it is a simply connected topological group. The only property of **Z** used is that it is a discrete subgroup of **R**. Thus exactly the same argument gives a more general result.

2.9 THE FUNDAMENTAL GROUP OF A PRODUCT SPACE

Theorem 2.9.1: The fundamental group of a product space, $\pi_1(X \times Y, (x, y))$ is naturally isomorphic to the direct product of fundamental groups, $[X, x] \times [Y, y]$. The isomorphism is defined by assigning to any element $[\delta] \in (X \times Y, (x, y))$ the ordered pair $(p*[\delta], q[\delta])$, where $p: X \times Y \to X$ and $q: X \times Y \to Y$ denote the projection of the product space onto its factors and $p*, q*$ their induced homeomorphisms on $\pi_1(X \times Y, (x, y)) \to \pi_1(X, x)$ and $\pi_1(X \times Y, (x, y)) \to \pi_1(Y, x)$ respectively.

Proof: Consider the projections $p: X \times Y \to X$, $q: X \times Y \to Y$, p and q induce homeomorphisms $p*$ and $q*$. defined as follows

$$p*: \pi_1(X \times Y, (x, y)) \to \pi_1(X, x)$$

$q*: \pi_1(X \times Y, (x, y)) \to \pi_1(Y, y)$ by $p*[\delta] = [p \, _o \, \delta]$ and $q*[\delta] = [q \, _o \, \delta]$

where $[\delta] \in \pi_1(X \times Y, (x, y))$.

Define $r*: \pi_1(X \times Y, (x, y)) \to \pi_1(X, x) \times \pi_1(Y, y)$ as $r*[\delta] = [p*[\delta], q*[\delta]$.

Clearly $r*$ is a homeomorphism as $p*$ and $q*$ are so,

(*i*) $r*$ is an epimorphism: Take $[\alpha], [\beta] \in \pi_1(X \;\; x) \; Y \times \pi_1(Y, y)$.

Let $\alpha: \mathbf{I} \to X$ represent $[\alpha]$ and $\beta: \mathbf{I} \to Y$ represent $[\beta]$.

Define: $\delta: \mathbf{I} \to X \times Y$ by $\delta[t] = (\alpha(t), \beta(t) \in X \times Y$.

$$r*([\delta]) = (p*[\delta], q*[\delta] = ([p \; o \; \delta]), [q \; o \; \delta]) = ([\alpha], [\beta]).$$

(*ii*) r is a monomorphism:

Let $r*([\delta]) =$ identity of the group $\pi_1(X, x) \times \pi_1(Y, y)$.

To show $[\delta] =$ identity of the group $\pi_1(X \times Y, (x, y))$.

Let $\delta: \mathbf{I} \to X \times Y$ represent $[\delta]$: since $r*([\delta]) =$ identity, $p*[\delta] =$ identity of $\pi_1(X, x)$ and $q*[\delta] =$ identity of $\pi_1(Y, y)$.

This implies, $p \circ \delta \cong x$ [the constant map at x] and $q \circ \delta \cong y$ [the constant map at y].

Let $F: p \circ \delta \cong x$ and $G: q \circ \delta \cong y$.

Define $H: \mathbf{I} \times \mathbf{I} \to X \times Y$ as follows:

$$H(s, t) = (F(s, t), G(s, t))$$

Clearly H is a homotopy of δ with the constant map at (x, y), i.e., $[\delta] =$ identity of the group $\pi_1(X \times Y, (x, y))$.

Therefore r_* is an isomorphism.

Corollary: $\pi_1(\mathbf{S}^1 \times \mathbf{S}^1) = \pi_1(S^1) \times \pi_1(S^1) = \mathbf{Z} \times \mathbf{Z}$.

Theorem 2.9.2: If G is a simply connected topological group, H is a discrete normal sub group, then

$$\pi_1(G/H, 1) \cong H$$

Proof of this needs the following lemma.

Lemma: If G is a topological group with a discrete normal subgroup H, then there exists an open set V in G containing I such that the quotient map. $f: G \to G/H$ restricted to V is a homeomorphism onto the range $f(V)$.

Proof: Since H is discrete, there is an open *nbhd* of I in G such that $U \cap H = \{1\}$. Let h be the continuous map given by

$$h: G \times G \to G,\ h(g_1, g_2) = g_1 g_2^{-1}$$

Consider $h^{-1}(U)$ in $G \times G$. It is clearly an open *nobhd* of $(1, 1) \in G \times G$. Hence there exists an open *nbhd* V of I in G such that $V \times V \in h^{-1}(U)$. Clearly $V \subset U$. Since f is always an open map, $f(V)$ is open in G/H and contains the identity $\underline{1}$ of G/H. We shall show f is 1–1 on V onto $f(V)$.

Since f is continuous, f is then a homeomorphism.

Let $f(g_1) = f(g_2)$: $(g_1), (g_2) \in V$. then $h(g_1, g_2) \in U$.

But $h(g_1, g_2) = g_1 g_2^{-1} \in H$. Consequently,

$$g_1 g_2^{-1} \in H \cap U = \{1\},\ \text{i.e., } g_1 = g_2.$$

Corollary: The fundamental group of $\mathbf{S}^1 \times \mathbf{S}^1$ [*torus*] is $\mathbf{Z} \times \mathbf{Z}$.

Proof: Since $\mathbf{S}^1 = \mathbf{R}/\mathbf{Z}$, $\mathbf{S}^1 \times \mathbf{S}^1 = \mathbf{R}/\mathbf{Z} \times \mathbf{R}/\mathbf{Z}$

which is a topological group isomorphic to $(\mathbf{R} \times \mathbf{R})/(\mathbf{Z} \times \mathbf{Z})$ and $\pi_1(\mathbf{R} \times \mathbf{R})/(\mathbf{Z} \times \mathbf{Z}) \cong \mathbf{Z} \times \mathbf{Z}$.

Therefore, $\pi_1(\mathbf{S}^1 \times \mathbf{S}^1) = \mathbf{Z} \times \mathbf{Z}$.

Lemma: A discrete normal subgroup of a connected topological group is central.

Proof: Let H be a discrete normal subgroup of a connected group G. We have to show that $hg = gh$ for all $g \in G$ and all $h \in G$. Fix $h \in H$. Look at the set $A = \{g\ h_o\ g^{-1};\ g \in G\}$. Since H is a normal subgroup, $A \subset H$, since G is a topological group the map $g \to g_o h_o\ g^{-1}$ is continuous. The range of this map is A and G is connected. Hence A is a connected subset of H and since $h_o \in A$ and H is discrete, $A = \{h_o\}$, i.e., $gh_o g^{-1} = h$ for all $g \in G$ i.e., $g\, h_o = h_o\, g$ for all $g \in G$.

We have proved that if G is a simply connected topological group and H is a discrete, normal subgroup, then $(G/H, I) \cong H$. Because of the lemma above $\pi_1(G/H, I)$ abelian.

CHAPTER 3

Compact Open Topology

3.1 COMPACT OPEN TOPOLOGY ON FUNCTION SPACES

Let X, Y be two topological spaces, and by Y^X we mean the set of all continuous maps from X into Y.

Definition: The *compact open topology* in Y^X is that having as sub-base all sets (A, V), where $A \subset X$ is compact and $V \subset Y$ is open and

$$(A, V) = \{ f \in Y^X : f(A) \subset V \}$$

Let X, Y, Z be three spaces. For $f \in Y^X$ and $g \in Z^Y$, we define a map $T : Y^X \times Z^Y \to Z^X$ by the composition $g_o\ f \in Z^X$ so that $T(f, g) = g_o\ f$.

We investigate the continuity of T.

Proposition 3.1.1: (i) $g \to g_o\ f_1$ is a continuous map $Z^Y \to Z^X$ for each fixed f_1.

(ii) $f \to g_1\, o\ f$ is a continuous map $Y^X \to Z^X$ for each fixed g_1, i.e., T is always continuous in each argument separately.

Proof: (i) Let (A, V) be any sub-basic *neighbourhood* of $g_o f_1$. Note that $g_o f_1 \in (A, V)$ $\Leftrightarrow g \in (f_1(A), V)$ and $(f_1(A), V)$ is a *neighbourhood* of g since $f_1\ (A)$ is compact. Thus, $T[f_1(f_1(A), V)] = (A, V)$ establishes the continuity.

(ii) This is proved similarly, noting that $g_1\, o\ f \in (A, V) \Leftrightarrow f(A) \subset g_1^{-1}(V) \Leftrightarrow f \in (A, g_1^{-1}(V))$.

Theorem 3.1.2: Let X, Z be Hausdorff and Y locally compact T_2. Then the map $T : Y^X \times Z^Y \to Z^X$ is continuous.

Proof: Let $f_1 \in Y^X$, $g_1 \in Z^Y$ and *neighbourhood* (A, W) of $g_{1o}\ f_1$ be given. Since $g_1^{-1}(W) \subset Y$ is open, $f_1(A) \subset g^{-1}(W)$ is compact and Y is locally compact Housdorff, there is an open set V such that ∇ is compact and $f_1(A) \subset V \subset \nabla \subset g_1^{-1}(W)$. Now $f_1 \in (A, V)$ and $g_1 \in (\nabla, W)$, and clearly $T[(A, V), (\nabla, W)] \subset (A, W)$.

Definition: For any two spaces Y, Z, the map $W : Z \times Y \to Z$ defined by $(f, y) \to f(y)$ is called the *evaluation map* of Z^Y.

Theorem 3.1.3: (i) For each fixed y_o, the map $Wy_o : Z \to Z$, given by $Wy_o(f) = f(y_o)$ is continuous.

(ii) If Y is locally compact T_2, then $W : Z^Y \times Y \to Z$ is continuous.

Proof: Whenever X is a single-point space, Y^X and Z can be looked upon as Y and Z respectively and, hence, W is precisely the composition map $T: Y^X \times Z^Y \to Z^X$.

Now the result follows from Theorems 3.1.1 and 3.1.2.

For notation, let $f: X \times Y \to Z$ be continuous in y for each fixed x. The formula $[f(x)](y) = f(x, y)$ defines, for each fixed x, and $f(x): Y \to Z$, and so $x \to f(x)$ is a map $f: X \to Z^Y$. Conversely, given an $f: X \to Z^Y$ the formula defines an $f: X \times Y \to Z$ continuous in y for each fixed x.

The most important feature of the compact-open topology is

Theorem 3.1.4: (i) If $f: X \times Y \to Z$ is continuous then $f: X \to Z^Y$ is also continuous.

(ii) If $f: X \to Z^Y$ is continuous, and if Y is locally compact T_2, then $f: X \times Y \to Z$ is also continuous.

Proof: (i) Let $x_o \in X$ and (A, V) be a sub-basic open set containing $f(x_o)$. Now $\hat{f}(x_o)(A) \subset V$ i.e., $f(x_o, A) \subset V$. Hence $\{x_o\} \times A \subset f^{-1}(V)$ and $f^{-1}(V)$ is open; since A is compact, there is a *neighbourhood*. U of x_o such that $U \times A \subset f^{1}(V)$ i.e., $f(U \times A) \subset V$. Hence f is continuous at x_o.

(ii) Define the map $\hat{f} \times 1: X \times Y \to Z^Y \times Y$ as follows: $f \times 1\,(x, y) = (f(x), y)$. Since $f: X \to Z^Y$ is continuous and $1: Y \to Y$ is the identity map, $f \times 1$ is a continuous map. Consider the evaluation map $W: Z^Y \times Y \to Z$. Since Y is locally compact Hausdorff by Theorem 3.1.3, W is continuous. Hence the composition $X \times Y \xrightarrow{\hat{f} \times 1} Z \times Y \xrightarrow{W} Z$ is continuous. But the composition is simply the function.

$f: X \times Y \to Z$.

A further advantage of the compact open topology is that homotopic maps induce homotopic maps of functions under mild restrictions.

Theorem 3.1.5: In all function spaces, use the compact open topology and let Z be an arbitrary space.

(i) If $f_o, f_1: X \to Y$ are homotopic and if Y is locally compact, then the induced map $q_o, q_1: Z \to Z$ are homotopic where $q_i(g) = g\ o\ f_i,\ g \in Z,\ I = [0, 1].$.

(ii) If $g_o, g_1: Y \to Z$ are homotopic, and if Y is locally compact, then the induced maps $p_o, p_1: Y^X \to Z^X$ are homotopic where

$p_i(f) = g_i\, o\ f,\ f \in Y^X,\ I = [0, 1].$

Proof: (i) Let $F: f \cong f$ i.e., $F: X \times I \to Y$ such that $F(., 0) = f_o$, $F(., 1) = f_1$. Hence $F: I \to Y^Z$ is continuous by 3.1.4. Also $F \times I: I \times Z^Y \to Z^X \times Z_Y$ is continuous. Since Y is locally compact, by 3.1.2., the composition $T: Y^X \times Z^Y \to Z^X$ is continuous from $I \times Z^Y \to Z^X$.

But $T\ o\,[f \times 1][0, g] = T(F(0), g) = g\ o\ F(0) = g\ o\ f_o = q_o(g)$

Similarly $T\ o\ (F \times I)(1, g) = q_1(g)$

$T\ o\ (F \times I): q_o \cong q_1$

(ii) Proof is similar to (1).

3.2 LOOP SPACES

Let Y^I denote the set of all continuous maps $f : \mathbf{I} \to Y$ where Y is any topological space and $I = [0, 1]$. Y is always equipped with the compact open topology. In this topology a typical sub-basic open set is of the form $(K, U) = \{f \in Y^I : f(K) \subset U\}$ where K is compact $\subset \mathbf{I}$ and U is open $\subset Y$.

Each element in Y is called a *path* in Y. By L (Y; a, b) we mean the sub-space of Y consisting of all paths starting at a and ending at b. a, b Y. In case $a = b$. L (Y; a, a) is written simply L (Y, a) and is called the *loop space* of Y based at a.

Result 3.2.1: The map $f \to f^{-1}$ of $L(Y; a, b) \to L(Y; b, a)$ is a homeomorphism.

Proof: The map is clearly bijective. Let $p: \mathbf{I} \to \mathbf{I}$ be the homeomorphism $p(t) = 1 - t$. Define $q: Y \to Y$ as follows: $q(f) = f \ o \ q, \ f \in Y^I$.

Now $q^{-1}(K, U) = \{f \in Y^I : q(f) \in (K, U)\} = \{f \in Y^I : f \ o \ q(K) \subset U\} = (q(K), U)\}$.

Hence q is continuous. But q is precisely the map $f^1 \to f^{-1}$. Similarly, the map $f^{-1} \to f$ is continuous. The following propositions are very useful.

Proposition 3.2.2: Let $p: \mathbf{I} \to \mathbf{I}$ be continuous and such that $p(0) = 0, p(1) = 1$. Then the map $f \to f \ o \ p$ of L [Y; a, b] to itself is homotopic to the identity map.

Proof: Note $p \cong 1_I$ rel. (0, 1) as $F(t, s) = 1\ (1 - s)\ p\ (t) + st$ $(0 \le s \le 1, o \le t \le 1)$ shows the induced map $q: Y^I \to Y^I$ given by $q(f) = f \ o \ p, \ f \in Y^I$ is homotopic to the identity map on Y^1. Since $q \mid L$ (Y; a, b) maps L (L; a, b) into itself, the assertion follows for the product operation.

Proposition 3.2.3: (i) The mapping $(f \cdot g) \to f * g$ of $L(Y; a, b) \times L(Y; b, c) \to L(Y; a, c)$ is continuous.

(ii) If $e \in L(Y; b)$ is the constant path at b, the map $f \to f * e$ of L (Y; a, b) to itself is homotopic to the identity map, and so also is the map $g \to e * g$ of L (Y; b, c) to itself.

Proof: (i) For (f, g) $\in$ L (Y; a, b) × L (Y; b, c) f, $g \in$ L (Y; a, c) f, g: $\mathbf{I} \to Y$ i.e., $f * g \in Y^1$.

Hence consider the map on L (Y; a, b) × L (Y; b, c) × I as follows $(f, g, t) \to f * g(t) \in Y$. If we show the latter is continuous, by Theorem 3.1.4, the continuity of the map $(f, g) \to f * g$ follows.

Now $$f * g(t) = \begin{cases} f[2t] & \text{if } 0 \le t \le 1/2 \\ g[2t - 1] & \text{if } 1/2 \le t \le 1 \end{cases}$$

Hence $(f, g, t) \to f \times g\ (t)$ decomposes into $(f, g, t) \to (f, t) \to (f, 2t) \to f(2t)$ if $0 \le t \le 1/2$ and $(f, g, t) \to (g, t) \to (g, 2t - 1) \to g(2t - 1)$ if $1/2 \le t \le 1$.

In each case, the first map is a projection and hence continuous; the second maps are clearly continuous, and the last maps, being evaluation maps, are also continuous. Since the two maps coincide on [L (Y; a, b) × L (Y; b, c)] × 1/2 the result follows.

1. Let p be the mapping of $I \to I$: $p[t] = \min\{1, 2t\}$. Now $f * e: I \to Y$ is given by

$$f * e(t) = \begin{cases} f(2t) & \text{if } 0 \le t \le 1/2 \\ b & \text{if } 1/2 \le t \le 1 \end{cases}$$

Consider

$$f \ o \ p(t) = \begin{cases} f(2t) & \text{if } 0 \le t \le 1/2 \\ b & \text{if } 1/2 \le t \le 1 \end{cases}$$

Since $f(1) = b$, $f * e = f \ o \ p$. Hence $f \to f * e$ is the map $f \to f \ o \ p$ and the result follows from Theorem 3.2.2.

The second part is also proved similarly.

Proposition 3.2.4: (i) The maps $(f, g, h) \to (f * g) * h$ and $(f, g, h) \to f * (g * h)$ of $L(Y; a, b) \to L(Y; b, c) \times L(Y; c, d) \to L(Y; a, d)$ are homotopic.

(ii) The map $f \to f * f^{-1}$ of $L(Y; a, b) \to L(Y; a)$ is null homotopic, and so also is the map $f \to f^{-1} * f$.

Proof: (i) Let $R(f, g, h) = f * (g * h)$. Let $p: I \to I$ be as follows

$$p(t) = \begin{cases} 2t & \text{if } 0 \le t \le 1/4 \\ t + 1/4 & \text{if } 1/4 \le t \le 1/2 \\ \frac{t+1}{2} & \text{if } 1/2 \le t \le 1 \end{cases}$$

Now $f * (g * h)(t) = \begin{cases} f(2t) & \text{if } 0 \le t \le 1/2 \\ g(4t-2) & \text{if } 1/2 \le t \le 3/4 \\ h(4t-3) & \text{if } 3/4 \le t \le 1 \end{cases}$

and $[(f * g) * h](t) = \begin{cases} f(4t) & \text{if } 0 \le t \le 1/2 \\ g(4t-1) & \text{if } 1/2 \le t \le 1/2 \\ h(2t-1) & \text{if } 1/2 \le t \le 1 \end{cases}$

Now easy to see that $[(f * g) * h] = [f * (g * h)] \ o \ p$ for all f, g, h … (1) Let q be the induced map on $Y^1 \to Y^1$ defined by $q(\phi) = \phi \ o \ p$ for $\phi \in Y$.

By Theorem 3.2.2, $q \cong$ identity on Y^1. Hence $q \ o \ r \cong R$. From (1) this simply implies $(f * g) * h \cong f * (g * h)$.

(ii) Let $p_o \mathbf{I} \to \mathbf{I}$ be the map

$$p_o(t) = \begin{cases} 2t & \text{if } 0 \le t \le 1/2 \\ 2(1-t) & \text{if } 1/2 \le t \le 1 \end{cases}$$

Define $q_o : Y^{\mathbf{I}} \to Y^{\mathbf{I}}$ by $q_o(g) = g \ o \ p_o$, $g \in Y^{\mathbf{I}}$.

Now $g \ o \ p_o(t) = \begin{cases} g(2t) & \text{if } 0 \le t \le 1/2 \\ g(2(1-t)) & \text{if } 1/2 \le t \le 1 \end{cases}$

If g^{-1} be the map $g^{-1}(t) = g(1-t)$ for $0 \le t \le 1$

Then $g \ o \ g^{-1}(t) = \begin{cases} g(2t) & \text{if } 0 \le t \le 1/2 \\ g(2t-1) & \text{if } 1/2 \le t \le 1 \end{cases}$

$$= \begin{cases} g(2t) & \text{if } 0 \le t \le 1/2 \\ g(1-2t+1) & \text{if } 1/2 \le t \le 1 \end{cases}$$

$$= g \ o \ p_o(t).$$

Define $F : \mathbf{I} \times \mathbf{I} \to \mathbf{I}$ as follows

$$F(t, s) = \begin{cases} 2t(1-s) & \text{if } 0 \le t \le 1/2 \\ 2(1-t)(1-s) & \text{if } 1/2 \le t \le 1 \end{cases}$$

Then $F : p_o \cong 0$. By Theorem 3.1.5, $p_o^+ : Y^{\mathbf{I}} \to Y^{\mathbf{I}}$ defined by $p_o^+(g) = g \ o \ p$ is null homotopic. Since $F(0, s) = 0 = F(1, s)$ for each s, it follows that $F_r^+ : Y^{\mathbf{I}} \to Y^{\mathbf{I}}$, defined by $F_s^+(g) = g \ o \ F(., s);\ g \in Y^1,\ s \in \mathbf{I},$ maps $L(Y; a, b)$ into $L(Y; a)$ for each s.

Hence the map $f \to f * f^{-1} : L(Y; a, b) \to L(Y; a)$ is null homotopic. The other part also follows similarly.

The path space $L(Y; a, b)$ is, in general, not path-connected. We call two paths $f, g \in L(Y; a, b)$ equivalent [written: $f \cong g$] if they belong to the same path component. Stated directly in terms of the maps f, g: we can write

Proposition 3.2.5: $f \sim g$ if and only if $f \cong g$ rel. (0, 1); that is, if and only if f can be deformed into g without ever moving the end points.

Proof: Clearly $f \sim g \Rightarrow$ there is a path joining f and g in $L(Y; a, b)$. Let $\phi : \mathbf{I} \to L(Y; a, b)$ be such a path. Define $F : \mathbf{I} \times \mathbf{I} \to Y$ as follows

$$F(s, t) = \phi(t)(s).$$

Then $F(s, 0) = \phi(0)(s) = f(s),\ s \in I$ and $F(s, 1) = \phi(1)(s)$

$\cong g(s),\ s \in \mathbf{I}$. Let $(s_o, t_o) \in \mathbf{I} \times \mathbf{I}$. Then $F(s_o, t_o) = \phi(t_o)(s_o)$.

Let U be an open *neighbourhood* of $\phi(t_o)(s_o)$ in Y.

There exists compact *neighbourhood* V of s in $\mathbf{I}$ such that $\phi(t_o)(V) \subset U$. Consider the open *neighbourhood* of $\phi(t_o)$ in $L(Y; a, b)$ given by $(V, U) \cap L(Y; a, b)$. By continuity of ϕ there is an open *neighbourhood* w of t * in $\mathbf{I}$ such that $\phi(t) \in (V, U) \cap L(Y; a, b)$ for all $t \in W$. Consider the open *neighbourhood* of (s_o, t_o) in $I \times I$ given by $v^o \times W$ ($v^o =$ interior of V in $\mathbf{I}$). Then $(s, t) \in V^o \times W \Rightarrow F(s, t) = \phi(t)(s) \in U$. Hence, F is continuous. F is clearly a homotopy of f and g

rel. $\{0, 1\}$. Conversely, if $F: f \cong g$ rel. $\{0, 1\}$ then define $\phi: \mathbf{I} \to L(Y; a, b)$ as follows: $\phi(t) = F(., t)$. Then ϕ is a path joining *f* and *g*.

To show $\phi(t) = F(., t)$ is a path joining of *f* and *g* in *L* (*Y*; *a*, *b*).

First, $\phi(0) = F(., 0) = f$ and $\phi(1) = F(., 1) = g$ and $\phi(t) = F(., t): \mathbf{I} \to Y$, so $\phi(t) \in Y^1$. Again $\phi(t)(0) = F(0, t) = a$ and $\phi(t)(1) = F(1, t) = b$.

Consequently $\phi(t) \in L(Y; a, b)$. Since $F: \mathbf{I} \times \mathbf{I} \to Y$ is a continuous map $\phi(t) = F(., t)$ is a map of $\mathbf{I} \to Y^1$ and by Theorem 3.1.4 is continuous. As a result ϕ is a path in *L* (*Y*; *a*, *b*) joining *f* and *g*.

Let $f \in L(Y; a, b)$ by a fixed path. For any given $y \in Y$, the path *f* induces a 'transition map' $f_R: L(Y; y, a) \to L(Y; y, b)$ by defining $f_R(g) = g * f$. Similarly, *f* induces a transition map $f_L: L(Y; b, y) \to L(Y; a, y)$ by $f_L(g) = f * g$. According to Proposition 3.2.3 transition maps are continuous.

Remarks: Before we proceed to prove the next theorem let us note the following: We have seen that $(f, g, h) \in L(Y; a, b) \times L(Y; b, c) \times L(Y; c, d) \to (f * g) * h$ and $f * (g * h)$ of *L* (*Y*; *a*, *b*) are homotopic. This is achieved by constructing a parameter transformation $p: \mathbf{I} \to \mathbf{I}$ such that $p \cong 1_1$ rel. (0, 1) and $(f * g) * h = [f * (g * h)] \, o \, p$.

Hence $(f * g) * h \cong f * (g * h) \, o \, 1_1$ rel. $\{0, 1\}$ i.e., $(f * g) * h \cong f * (g * h)$ by Proposition 3.2.4, this implies $(f * g) * h \sim f * (g * h)$. Again while proving the fact that $f \to f * f^{-1}$ of *L* (*Y*; *a*, *b*) to *L* (*Y*; *a*) is homotopic to the map $f \to e$ of *L* (*Y*; *a*, *b*) to *L* (*Y*; *a*) where *e* is the constant path at *a*, we observed that $f \to f * f$ is the map $f \, o \, p_o$ on *L* (*Y*; *a*, *b*) where $p_o: \mathbf{I} \to \mathbf{I}$ is homotopic to the map 0 (i.e., $0(t) = o$ for $t \in \mathbf{I}$) rel. $\{0, 1\}$. Hence $f \, o \, p_o \cong f \, o \, 0$ rel. $\{0, 1\}$, i.e., $f * f^{-1} \cong e$. By proposition 3*J* we then conclude $f * f^{-1} \sim e$. Similarly $f^{-1} * f \sim e$. Similarly we can show that $f * e$ and $e * f$ and if $f \sim f_1$ and $g \sim g_1$ then $f * g \sim f_1 * g_1$.

Proposition 3.2.6: (*i*) Each f_R (resp. f_L) is a homotopy equivalence with homotopy inverse $(f^{-1})_R$ [resp. $(f^{-1})_L$].

(*ii*) $f_R \cong h_R$ [resp. $f_L \cong h_L$] if and only if $f \sim h$.

Proof: (*i*) We prove this for the maps f_R; that for the f_L is similar.

(*ii*) Let $\phi: \mathbf{I} \to L(Y; a, b)$ be a path joining *f* to *h*; by proposition 3*H* (1) the map $(g, t) \to g * [\phi(t)]$ is continuous and shows that $f_R \cong h_R$.

Conversely, assume $f_R \cong h_R$

Let $F: L(Y; y, a) \times \mathbf{I} \to L(Y; y, b)$ be a homotopy between f_R and h_R. Then $f_R(g) = F(g, 0)$ and $h_R(g) = F(g, 1)$. Then $F(g,.): \mathbf{I} \to L(Y; y, b)$ is a path joining $g * f = f_R(g)$ to $h_R(g) = g * h$. Hence $g * f \sim g * h$. By the remarks above,

$$g^{-1} * (g * f) \sim g^{-1} * (g * h).$$

Again by the same remark

$$g^{-1} * (g * f) \sim (g^{-1} * g) * f \sim e * f \sim f.$$

Similar reasoning leads to the fact $g^{-1} * (g * h) \sim h$

$f \sim h$.

(*a*) Note that $(f^{-1})_R \, o \, f_R$ and $(f * f^{-1})R$ are from $L\ (Y;\ y,\ a) \to L\ (Y;\ y,\ a)$ given by $(f^{-1})_R \, o \, f_R(g) = (g * f) * f^{-1}$ and $(f * f^{-1})_R(g) = g * (f * f^{-1})$ respectively. Clearly $(f^{-1})_R \, o \, f_R$ is the map $g \to (g * f) * f^{-1}$ and $(f * f^{-1})$ is the map $g \to g * (f * f^{-1})$ and they are known to be homotopic.

Hence $(f^{-1})_R \, o \, f_R \cong (f * f^{-1})$. Again $f * f^{-1} \sim e$ where *e* is the constant loop at *a*.

So by part (b) $(f * f^{-1})_R \cong e_R$. Now $e_R(g) g * e$, i.e., e_R is the map $g \to g * e$ which is known to be homotopic to the identity map on *L* (*Y*; *y*, *a*), say, 1 consequently $1 \cong e_R \cong (f * f^{-1})_R \cong (f^{-1})_R \ o \ f_R$.

Similarly $f_R \ o \ (f^{-1})_R$ is homotopic to the identity map on *L* (*Y*; *y*, *b*). This completes the proof.

In what follows, we are concerned with loop spaces (*L*; *a*) where $a \in Y$. Each path $f \in L(Y; a, b)$ induces a transition map $f^+ : L(Y; a) \to L(Y; b)$ given by $f^+(g) = f^{-1} * (g * f)$: this is continuous. Again $(f^{-1})_L \ o \ f_R(g) = (f^{-1})_L(g * f) = f^+(g)$. Hence $(f^{-1})_L \ o \ f_R = f^+$. By Theorem 3.2.4, f^+ is a homotopy equivalence with homotopy inverse $(f^{-1})^+$.

Similarly $f^+ \cong h^+$ if and only if $f \sim h$. Hence we have proved the following theorem.

Theorem 3.2.7: (*i*) Each f^+ is a homotopy equivalence with homotopy inverse $(f^{-1})^+$.

(*ii*) $f^+ \cong h^+$ if and only if $f \sim h$.

3.3 H-STRUCTURES

We begin with a definition.

Definition: An H-structure [‘Hopf’, or’ Homotopy’ structure] is a couple [*Y*, *m*] consisting of a space *Y* and *a* continuous function *m*: $Y \times Y \to Y$ which has the following property: There exists a point $e \in Y$ such that maps $y \to m(y, e)$ and $y \to m$ (*e*, *y*) are both homotopic to the identity map of *Y*. The map *m* is called the composition law of the H-structure (*Y*, *m*), and *Y* is the carrier of the H-structure, *m* (*a*, *b*) is written as *a*.*b*. If there is no scope of confusion, we say simply that ‘*Y* is an H-structure’.

Examples:

1. Every topological group is an H-structure.
2. Given any *Y*, the loop space *L* [*Y*; y_o], with the composition $m(f, g) = f * g$ is an H-structure. *M* is called *the natural composition law* in *L* (*Y*; *Y*).
3. Let *Y* be a contractible space, and $m: Y \times Y \to Y$ any continuous map. Then (*Y*, *m*) is an H-structure, since any two maps of *Y* in itself are homotopic.

   Observe (i) that the composition law is not required to be associative, and (ii) that a given space may carry many H-structures.
4. Let *Y* be any discrete space. Select any $e \in Y$ and define *m* in any way compatible with *m* (*y*, *e*) = *m* (*e*, *y*) = *y*; we then have (*y*, *m*) as an H-structure.

Remarks: It is clear that H-structures form a class of structures that contains both the loop spaces and the topological groups. The point e in the definition is not unique in general. This is the content of the next proposition.

Proposition 3.3.1: Let $P = \{e$: both the maps $y \to m(y, e)$ and $y \to m(e, y)$ are homotopic to $1_Y\}$. Then P is a path component of Y, called the principal component of the H-structure (Y, m).

Proof: Choose $e_0 \in P$ and let C be the path-component of e. To prove $C \subset P$:

Let $e \in C$: choose a path f joining e to e_o; consider the map F: $Y \times I \to Y$ defined by $F(y, t) = m(y, f(t))$. F is clearly continuous and $F(y, 0)$ is the map $y \to m[y, e]$ and $F(y, 1)$ is the map $y \to m(y, e_o)$. Hence $y \to m(y, e)$ is homotopic to $y \to m(y, e_o)$ and therefore also homotopic to 1_Y. Similarly $y \to m(y, e)$ is homotopic to 1_y; and therefore $e \in P$.

To prove $P \subset C$: let $e \in P$. By definition of P, $m(., e) \cong 1_y$. Again $e_o \in P$. Hence $m(e_o, .) \cong 1_y$. Let F: $m(e_o, .) \cong 1_y$ and $G: m(., e) \cong 1_y$. Then $F(., 0) = m(e_o, .)$ and $F(., 1) = 1_y$ and $G(., 0) = m(., e)$ and $G(., 1) = 1_y$. Clearly $F(e, .)$ is a path. In Y joining $m(e_o, e)$ and e and $G(e_0, .)$ is a path in Y joining $m(e_o, e)$ and e_o. Hence, there is a path in Y joining e_o and e i.e., $e \in C$, since C is the path component containing e_o.

Corollary: Let (Y, m) be an H-structure. Then its principal component p, with composition law $m(p \times P)$ is an H-structure, called the induced H-structure in P.

Proof: This is immediate from the previous theorem, since the composition of two elements belonging to P is again contained in P.

Let Comp Y denote the set of all path components of Y endowed with the discrete topology. For $y \in Y$, let $[y]$ denote the path component containing y.

Proposition 3.3.2: Let $[Y, m]$ be an H-structure. Then Comp Y, with the composition law $([x],[y]) \to [m(x, y)]$ is an H-structure, called the induced H-structure in Comp Y.

Proof: The composition law is a well defined function:

Let $[x] = [x_1]$ and $[y] = [y_1]$ to show $[m(x, y)] = [m(x_1, y_1)]$, it suffices to show that there is a path in Y joining $[m(x, y)]$ and $[m(x_1, y_1)]$.

Since $m(., y)$ is a continuous map of Y to Y, $m(., y)$ maps a path into a path. Since $[x] = [x_1]$, there is a path between x and x_1 in Y, say f: $\mathbf{I} \to Y$. Then $m(., y)$ is a path joining $m(x, y)$ and $m(x_1, y)$. Arguing with $m(x_1, .)$ we can show that there is a path joining $m(x_1, y)$ to $m(x_1, y_1)$.

Since $m(e,.) \cong 1_Y$. Let $F: m(e, .) \cong 1_Y$.

Consider the path $F(y, .)$: $\mathbf{I} \to Y$. $F(y, 0) = m(e, y)$ and $f(y, 1) = y$. Similarly, there is a path joining y and $m(y, e)$. Hence, $[m(e, y)] = [y] = [m(y, e)]$. Hence, $[e].[y] = [m(e, y)] = [y] = [m(y, e)]$. Clearly Comp Y is an H-structure.

Theorem 3.3.3: Let (Y, m) be an H-structure. Then, for each locally compact T_2 space X, the space Y^X with composition law $(f, g) \to m \circ [f \times g]$ is an H-structure, called the induced H-structure in Y^X.

[*Note*: $f \times g$: $X \to Y \times Y$ is defined by $(f \times g)(x) = (f(x), g(x))$]

Proof: We first show that the composition map $\phi: Y^X \times Y^X \to Y^X$ is continuous where $\phi(f, g) = m \; o \; (f \times g)$.

The associated map $\phi: Y^X \times Y^X \times X \to Y$ is given by

$$\phi(f, g, x) = m \circ (f \times g)(x) = m(f(x), g(x)).$$

If ϕ is continuous, ϕ is continuous by Theorem 3.1.2. Observe that is the map

$$(f, g, x) \to (f(x), g(x)) \to m(f(x), g(x)).$$

The last map is evidently continuous; $(f, g, x) \to (f(x))$ and $(f, g, x) \to (g(x))$ are both evaluation maps and since X is locally compact Housdorff, by Theorem 3.1.4, both are continuous. Thus ϕ is continuous, as required. Now let $\ddot{e}$ be the constant map $\ddot{e}(x) = e, x \in X$;

Let $m_e(y) = m(y, e)$; then $\phi(f, \ddot{e}) = m \; o \; (f \times \ddot{e}) = m_e \; o \; f$.

Let $m_e^+, 1_Y^+$ be defined as follows:

$$m_e^+(f) = m_e \; o \; f$$

$$; \; f \in Y^X$$

$$1_Y^+(f) = 1_y \; o \; f$$

Since $m_e \cong l_Y$, by Theorem 3.2(2) on page 320 of Dugundji's 'Topology', $m_e^+ \cong 1_y^+$ But m_e^+ is the map $f \to m_e \; o \; f = \phi(f, \ddot{e})$ and 1_Y^+ is the identity map on Y^X. Similarly, $g \to \ddot{e}.g.$ is homotopic to the identity map, and the proof is complete.

3.4 H-HOMOMORPHISMS

Let X, Y be two H-structures.

Definition: A continuous map $f: X \to Y$ is called an *H-homomorphisms* whenever the two maps $(x, x') \to f(x) \,.\, f(x')$ and $(x, x') \to f(x, x')$ of $X \times X$ into Y are homotopic.

Definition: An H-homomorphism $f: X \to Y$ an *H-isomorphism* if there exists an H-homomorphism $g: Y \to X$ such that both $g \; o \; f \cong 1_X$ and $f \; o \; g \cong 1_Y$; in this event the H-structures are called *H-isomorphic*. Note that an H-isomorphism is necessarily a homotopy equivalence, the converse need not be true.

Result 3.4.1: Let X, Y be H-structures. Let $f: X \to Y$ be an H-homomorphism [H-isomorphism] and $g: X \to Y$ be such that $f \cong g$. Then g is also H-homomorphism [H-isomorphism] [i.e., the concepts of H-homomorphism [H-isomorphism] are homotopy class invariants].

Proof: Given that $(x, x') \to f(x) \cdot f(x')$ and $(x, x') \to f(x, x')$ are homotopic maps from $X \times X \to Y$. Consider the map $(x, x') \to g(x) \cdot g(x')$ of $X \times X \to Y$. We shall show that it is homotopic to the map $(x, x') \to f(x) \cdot f(x')$ of $X \times X \to Y$. Similarly, we shall show that the maps $(x, x') \to f(x, x')$ and $(x, x') \to g(x, x')$ are also homotopic.

By the transitivity of homotopy property the required result then follows.

Consider the maps [1] $(x, x') \to f(x) \cdot f(x')$ and [2] $(x, x') \to g(x) \cdot g(x')$ $X \times X \to Y \times Y$.

Let $F: f \cong g$. Define $H: X \times X \times I \to Y \times Y$. as follows $H(x, x', t) = (F(x, t), F(x', t)$

Then, clearly H is a continuous function. In fact H is a homotopy between the maps [1] and [2]. Note that the maps

$(x, x') \to f(x) \cdot f(x') = (x, x') \to (f(x) \cdot f(x')) \to f(x) \cdot f(x')$ and

$(x, x') \to g(x) \cdot g(x') = (x, x') \to (g(x) \cdot g(x') \to g(x) \cdot g(x')$.

Hence these two maps are also homotopic. Again note that

$(x, x') \to f(x \cdot x') = (x, x') \to x: x' \to f(x \cdot x')$ and

$(x, x') \to g(x \cdot x') = (x, x') \to x: x' \to g(x \cdot x')$.

Since *f* and *g* are homotopic, clearly the above two maps are also homotopic. Hence, the required result follows.

Let X and Y be discrete spaces with H-structures. An H-homomorphism $f: X \to Y$ satisfies $f(x \cdot x') = f(x) \cdot f(x')$. This follows from the following general result and the fact that $(x \cdot x') \to f(x) \cdot f(x')$ and $(x \cdot x') \to f(x \cdot x')$ are homotopic.

Result 3.4.2: Let X and Y be discrete topological spaces $f,\ g: X \to Y$ are two continuous maps. Then $f \cong g$ if and only $f = g$.

Proof: We shall prove that $f \cong g \Rightarrow f = g$. Let $F: X \times \mathbf{I} \to Y$ be a homotopy between *f* and *g*, then for each $x \in X$, $F(X, .): \mathbf{I} \to Y$ is a path in Y and, Y being discrete, should be a constant path.

In particular, $F(X, 0) = F(X, 1)$ i.e., $f(x) = g(x)$. Consequently $f = g$.

Remark: Thus when X and Y are both discrete, an H-homomorphism is evidently a homomorphism in the usual sense of algebra, and the prefix 'H-' will be omitted.

3.5 HOPF SPACE

We begin with some definitions.

Definition: A topological space X is called an *H-space* or *HOPF space* if X has an H-structure (X, m).

Definition: Let (X, m) be an H-space. Call $e \in X$ a *homotopy unit* if

[*i*] $m(e; e) = e$

[*ii*] The maps $x \to m(x, e)$ and $x \to m(e, x)$ are both homotopic to 1_X relative to $\{e\}$.

Examples: 1. From example 1 of 3.3 we know a topological group is an H-Space. The identity element, it is easy to see, is a homotopy unit.

2. For any space Y, the loop space $L(Y; y_o)$ is an H-space. [Example 2 of 3.3]. Here also the constant loop $C(t) = y_o$ for all $t \in I$, is a homotopy unit.
3. From the above two examples we should not conclude that each H-space has a homotopy unit. Consider the set of real numbers $\mathbf{R}$ with usual topology, $\mathbf{R}$, being contractible, is an H-space [Example 3 of 3.3]. Let $m: \mathbf{R} \times \mathbf{R} \to \mathbf{R}$ be defined by $m(x, y) = x + 1$. Then $(\mathbf{R}, m)$ is an H-space but clearly it has no homotopy unit.

Theorem 3.5.1: If (x, m) is an H-space with e as a homotopy unit then π_1 (X, e) is an abelian group.

Proof: Let $a,\ b \in \pi_1(X, e)$. To prove $a.b = b.a$. choose loops $f \in a$ and $g \in b$. Define $h: I \to X$ by $h(t) = m(f(t), g(t) = f(t)\, g(t)$ for $t \in I$. Then $h(i) = f(i)\, g(i) = m(x_o, x_o) = x_o$ $(i = 0, 1)$. Hence h is a loop based at x_o and represents an element $C \in \pi_1(X, e)$. We shall show that $c = a \cdot b$, $c = b \cdot a$. Let us consider the identity element $\lambda \in \pi_1(X, e)$. It is represented by the constant loop $k: I \to X$ with $k(t) = e$ for all $t \in 1$. Since $a \cdot e = a$, f is equivalent to the loop $f': I \to X$ defined by

$$f'(t) = \begin{cases} f(2t) & \text{if } 0 \le t \le 1/2 \\ e & \text{if } 1/2 \le t \le 1 \end{cases} \quad \text{i.e., } f' = f * k.$$

Similarly, g is equivalent to the loop $g': \mathbf{I} \to X$ defined by

$$g'(t) = \begin{cases} e & \text{if } 0 \le t \le 1/2 \\ g(2t-1) & \text{if } 1/2 \le t \le 1 \end{cases} \quad \text{i.e., } g' = k * g.$$

Let $H: f \cong f'$ rel. (0, 1) and $K: g \cong g'$ rel. (0, 1) be the relevant homotopies.

Define $\phi(t\,s) = m(H(t\,s), K(t\,s))\ t, s \in \mathbf{I}$

$\phi(t\,0) = m(H(t\,0), K(t\,0)) = h(t), t \in \mathbf{I}$ and

$\phi(0\,s) = m(H(0\,s), K(0\,s)) = m(e, e) = e = \phi(1, s), s \in \mathbf{I}$.

Let $p: X \times \mathbf{I} \to X$ be a homotopy between 1_X and $x \to m$ (x, e) and $g: X \times \mathbf{I} \to X$ be a homotopy between 1_X and $x \to m$ (e, x), both relative to $\{e\}$.

$$\phi(t, 1) = \begin{cases} m(f(2t), e) & \text{if } 0 \le t \le 1/2 \\ m(e, g(2t-1)) & \text{if } 1/2 \le t \le 1 \end{cases}$$

and $\quad \phi: \phi(., 0) \cong \phi(., 1)$ rel. (0, 1).

Define $\psi; I \times I \to X$ as follows:

$$\psi(t, s) = \begin{cases} p(f(2t), s) & \text{if } 0 \le t \le 1/2 \\ q(g(2t-1), s & \text{if } 1/2 \le t \le 1 \end{cases}$$

Then Ψ is a homotopy between $\psi(., 0)$ and $\psi(., 1)$ and $\psi(., 1) = f * g$ and $\psi(., 1) = \phi(., 1)$.

So we see that $\psi(., 0) \cong \psi(., 1) = \phi(., 1)\ \phi(., 0) = h$

But $\psi(., 0)$ represents $a.b.$ and h represents c.

Hence, $a \cdot b = c$. Similarly, we can show that $b \cdot a = c$. Hence, we obtain $a \cdot b = b \cdot a$.

Corollary: If X is a topological group with e as the identity element, then $\pi_1(X, e)$ is an abelian group.

CHAPTER 4

Higher Homotopy Groups

4.1 THE n-DIMENSIONAL HOMOTOPY GROUP

Let X be a topological space and $x_o \in X$ be given.

Definition: For every integer $n > 1$, we define the *n-dimensional homotopy group* $\pi_1(X, x_o)$ of the space X at the base point x_o recurrently by the formula

$$\pi_n(X, x_o) = \pi_{n-1}(Y, e)$$

where $Y = L(X, x_0)$ and $e(t) = x_0$ for $t \in \mathbf{I}$

Among the most important properties of homotopy groups is the following:

Theorem 4.1.1: $\pi_n(X, x_o)$ is an abelian group for every $n > 1$.

Proof: By our recurrent definition, suffices to prove the theorem for $n = 2$.

Now $\pi_2(X, x_o) = \pi_1(Y, e)$, by definition. We know the loop space $(Y, e) = (L(X, x_0), e)$ is an H-structure with e as a homotopy unit. Hence, by Theorem 3.5.1 $\pi_1(Y, e)$ is an abelian group.

Let us prove the following characterization of a simply connected space.

Theorem 4.1.2: Let X be a path-connected space. The following are equivalent.

(1) X is simply connected;

(2) Every continuous map of the unit circle S^1 into X extends to a continuous map of the closed unit disc E^2 into X.

(3) If f and g are paths in X with the same initial points and the same terminal points then $f \cong G$ rel. (0, 1).

Proof: Let g be any map of $\mathbf{S}^1$ into X. We shall prove $(1) \Leftrightarrow g$ is null homotopic, then by theorem it follows that $(1) \Leftrightarrow (2)$. To prove $(1) \Rightarrow$ say such g is null homotopic, Let $\lambda : \mathbf{I} \to S^1$ be the continuous map $\lambda(t) = \exp(2\pi \mathrm{it}), t \in \mathbf{I}$ and p_o be the point $(1, 0) \in S^1$. Then $\lambda(0) = p_o = \lambda(1)$ and is a homomorphism of (0, 1) onto $S^1 - \{p_o\}$.

Define $f : \mathbf{I} \to X$ by $f(t) = g \; o \; \lambda(t), t \in \mathbf{I}$.

Then, f is a path in X with $f(0) = g(p_o) = f(1)$, i.e., f is a loop at $f(0) = g(p_o) = x_o$ (say). Let $c(x_o)$ denote the constant loop at x_o. Since X is simply connected, $f \cong c(x_o)$, $c(x_o)$ induces a constant map on $\mathbf{S}^1$ as follows $d(z) = x_o$ for all $z \in S^1$. We show that $g \cong d$, let $F : f \cong c(x_o)$ rel. {0, 1}. Define $G : S' \times \mathbf{I} \to X$ as follows:

$$G(z, s) = \begin{cases} F(0, s) & \text{if } z = p_o \\ f(\lambda^{-1}(z), s) & \text{if } z \in s' - \{p_o\}. \end{cases}$$

G is a continuous map into X and G (., 0) = g (.) and G (., 1) = d (.) = x_o.

Hence, $G: g \cong d$. To prove any g is null homotopic $\Rightarrow$ (1), Let $f: \mathbf{I} \to X$ by a loop. At some point $x_o \leftarrow X$. To show f is homotopic to the constant loop at x_o. Define $g: S^1 \to X$ by

$$g(z) = \begin{cases} f(\lambda^{-1}(z)) & z \in S' - \{p_o\} \\ X_o & z = p_o. \end{cases}$$

g is a map of S^1 onto X. By assumption $g \cong d$ where d is a constant map of S^1 to X. Let d $(z) = x_1$ (say), $z \in S$. Consider the constant map e $(z) = x_o$, z S. Since X is a path-connected, there is a path joining x_o to x_1. This path gives a homotopy of the constant maps d and e on S^1. Hence, $g \cong e$. Consequently $g \; o \; \lambda = e \; o \; \lambda$; infact $g \; o \; \lambda \cong e \; o \; \lambda$ rel. {0, 1}.

But $g \; o \; \lambda = f$ and $e \; o \; \lambda =$ constant loop at x_o. This is the required result.

(1) $\Leftrightarrow$ (3) The fact (3) $\Rightarrow$ (1) is trivial. Hence, we prove (1) $\Rightarrow$ (3): Let f, g: $\mathbf{I} \to X$ be two paths in X with f (0) = g (0) and f (1) = g (1). To prove $f \cong g$ rel. {0, 1}, consider the loop $f * g^{-1}$ at f (0). If c_0 denotes the constant loop at f (0) then (1) $\Rightarrow$ $f * g^{-1} \cong c_o$ rel. {0, 1}. Using the remark under 3J1 we can achieve the following.

$(f * g^{-1}) * g \cong f * (g^{-1} * g) \cong f * c_1$. (where c_1 is the constant loop at f (1) = g (1) f all relative to {0, 1}. Again from $f * g^{-1} \cong c_0$ relative {0, 1}.

We get $(f * g^{-1}) * g \cong c_o * g$ relative {0, 1}. But $c_o * g \cong g$ relative {0, 1}. Hence $f \cong g$ relative {0, 1}. Hence the proof.

Let us prove the following important result about simple connectivity.

Theorem 4.1.3: Let X be a topological space such that $X = U_1 \cup U_2$ where U_1, U_2 are simply connected, $U_1 \cap U_2$ is non empty and path-connected. Then X is simply connected.

Proof: It is easy to see that X is path-connected. Hence, we have to prove that $\pi_1(x)$ is trivial. Let $x_o \in U_1 \cap U_2$. Let $f: \mathbf{I} \to X$ be a loop based at x_o. Clearly $I = f^{-1}(U_1) \cup f^{-1}(U_2)$. Since I is a constant metric space, this covering has a Lebesgue number, say $\varepsilon > 0$, we next assert that it is possible to divide I, the unit interval, into subintervals $[0, t_1]$, $[t_1, t_2]$ …, $[t_{n-1}, t_n]$, where

$$0 = t_0 < t_1 < t_2 < ... < t_{n-1} < t_n = 1,$$

Such that the following conditions hold:

[a] $f([t_i, t_{i+1}]) \subset U_1$ or $f([t_i, t_{i+1}]) \subset U_2$ for $0 \le i \le n$.

[b] $f([t_{i-1}, t_i])$ and $f([t_i, t_{i+1}])$ are not both contained in the *same* open set U_j, j = 1 or 2.

This assertion is proved as follows: Divide the interval $\mathbf{I}$ in any way whatsoever into subintervals of length less than ε.

Then condition (*a*) will hold; however condition (*b*) may not hold. If two successive subintervals are mapped by *f* into the same *Uj*, then amalgamate these two subintervals into a single subinterval by omitting the common end-point. Continue this process of amalgamation until condition (*b*) holds.

Note that for any two closed intervals [*a*, *b*] and [*c*, *d*] with $a < b$ and $c < d$, there is a *unique order* preserving linear homeomorphism *h* such that $h(a) = c$ and $h(b) = d$. We are not interested in the explicit expression for *h* but we shall denote it by the following notation:

$$[a, b] \to [c, d]$$

Now, for each *i*, $1 \le i \le n$, $[0, 1] \to [t_{i-1}, t_i] \xrightarrow{f} X$ is a path in *X*, denote by f_i [say]. Now, there exists a partition $\{0 < a_1 < a_2 < ... < a_{n-1} < a_n = 1\}$ of [0, 1], not depends on f_i *s* such that

$$f_1 * f_2 * ... * f_n = \begin{cases} [0, \alpha_1] \to [0, 1] \xrightarrow{f_1} [0, \alpha_1] \\ [\alpha_1, \alpha_2] \to [0, 1] \xrightarrow{f_2} [\alpha_1, \alpha_2] \\ ... \\ [\alpha_{v-1}, i] \to [0, i] \xrightarrow{f_n} [\alpha_{v-1}, \alpha_n] \end{cases}$$

If $p : \mathbf{I} \to \mathbf{I}$ be the continuous map defined by

$$p = \begin{cases} [0, \alpha_1] \to [0, t_1] \text{ on } [0, \alpha_1] \\ [\alpha_1, \alpha_2] \to [t_1, t_2] \text{ on } [\alpha_1, \alpha_2] \\ ... \\ [\alpha_{v-1}, \alpha_n] \to [t_{n-1}, 1] \text{ on } [\alpha_{v-1}, \alpha_n] \end{cases}$$

then $f_1 * f_2 * ... * f_n = f \ o \ p$. Since $p(0) = 0$ and $p(1) = 1$, by Theorem 3.2.2, $f \cong f \ o \ p = f_1 * f_2 * ... * f_n$, relative {0, 1}. Now, each f_i is a path in U_1 or U_2. Because of condition (b), it is clear that $f(t_i) \in U_1 \cap U_2$, $o \le i \le n$. Since $U_1 \cap U_2$ is path connected, let g_i be a path in U_1 or U_2 with $g_i(0) = f(t_{i-1})$ and $g_i(1) f(t_i)$, $2 \le i \le n$. Then $f' = f_1 * g_2 * g_3 * ... * g_n$ is a loop at x_o. Since image of $f_1 = f([0, 1])$ is connected in U_j (j = 0 or 1), f_1 is a path in U_j. By construction each g_i (2 *i n*) is a path in *U*, *U* and hence, in particular, in *U*. Consequently, f' is loop at x_0 lying inside U_j. But U_j is simply connected, thus f' *is homotopic to the constant loop at* x_o. Again, f_i and g_i are paths in either U_1 or U_2 with same initial and same terminal points $(2 \le i \le n)$. Both U_1 and U_2 being simply connected, by Theorem 4.1.2, $f_i \cong g_i$ relative (0, 1); $i \ge 2$. Hence $f' = f_1 * g_2 * g_3 * ... * g_n \cong f_1 * f_2 * f_3 * ... * f_n \cong f$ all relative to {0, 1}. Consequently *f* is homotopic to the constant loop at x_o relative {0, 1}. As a result, $\pi_1(X, x_o)$ is trivial. *X* is hence simply connected.

An Application: Suppose $\mathbf{S}^n$ stands for the closed *n*-sphere of radius 1 in $\mathbf{R}^{n+1}$; $n \ge 1$.
i.e., $\mathbf{S}^n = \{(x_1, x_2, ..., x_{n+1}) \in R^{n+1};\ ^{n+1} x_i^2 = i\}$.

Before we prove Lemma 1, we prove the following lemma first.

Lemma 1: Let $0 < \alpha < \beta$. For $n \ge 1$, $A = \{(x_1, x_2, ..., x_n) \in R^n; \alpha < \sum^n y^2 i < \beta\}$ is path-connected.

Proof: First observe that each S^n is path-connected: Let

$\mathbf{S}_+^n = \{(x_1, x_2, ..., x_{n+1}) \in \mathbf{S}^n;\ x_{n+1} \geq 0\}$ and

$\mathbf{S}_-^n = \{(x_1, x_2, ..., x_{n+1}) \in \mathbf{S}^n;\ x_{n+1} \geq 0\}$. Under the projection to the space $\mathbf{R}^n$, each $\mathbf{S}_+^n$ and $\mathbf{S}_-^n$ is homeomorphic to the closed n-dimensional ball $E^n = \{(x_1, x_2, ..., x_n) \in \mathbf{R}^n;\ y_i^2 \geq 1\}$. Since E^n is path-connected, $\mathbf{S}_+^n$ and $\mathbf{S}_-^n$ are path-connected. Again $\mathbf{S}_+^n \cap \mathbf{S}_-^n \neq \phi$. Hence $S^n = S_+^n \cup S_-^n$ is path connected. Fix n. Consider $A = \{(x_1, x_2, ..., x_n) \in \mathbf{R}^n;\ \alpha < \Sigma x_i^2 < \beta\}$. Take $(a_1, a_2, ..., a_n), (b_1, b_2, ..., b_n) \in A$

Let $c = \Sigma a_i^2,\ d = \Sigma b_i^2$

***Case 1*:** $\alpha < c < d < \beta$

Let $A_1 = \{(x_1, x_2, ..., x_n) \in \mathbf{R}^n; \Sigma x_i^2 = c\}$,

$A_2 = \{(x_1, x_2, ..., x_n) \in \mathbf{R}^n; \Sigma x_i^2 = d\}$

and $A_3 = \{(x_1, x_2, ..., x_n) \in \mathbf{R}^n; \sqrt{c} \leq x_1 \leq \sqrt{d}\}$

Clearly A_1 and A_2 are homeomorphic to S^{n+1}, hence they are path-connected by our first observation. A_3 is homeomorphic to the closed interval $[\sqrt{c}, \sqrt{d}]$ and hence, path-connected. Note that $(a_1, a_2, \ldots, a_n)$ and $(\sqrt{c}, 0, 0, ..., 0)$ belong to A_1, $(b_1, b_2, \ldots, b_n)$ and $(\sqrt{d}, 0, 0, ..., 0)$ belong to A_2 and $(\sqrt{c}, 0, 0, ..., 0)$ and $(\sqrt{d}, 0, 0, ..., 0)$ belong to A_3. Since all these three sets are path-connected, we can find paths $f, g,\ h: \mathbf{I} \to \mathbf{R}^n$ such that

$f(0) = (a_1, a_2, ..., a_n),\ g(0) = (\sqrt{d}, 0, 0, ..., 0)$ and $f(I) \subset A_1$,

$f(1) = (\sqrt{c}, 0, ..., 0),\ g(1) = (b_1, b_2, ..., b_n),\ h(0) = (\sqrt{c}, 0, ..., 0)$,

$g(\mathbf{I}) \subset A_2$ and $g(\mathbf{I}) \subset A_3$.

$h(1) = (\sqrt{d}, 0, ..., 0)$

Since $f(1) = h(0)$ and $h(1) = g(0)$ we can compose these paths to get a path $p = f * h * g: \mathbf{I} \to \mathbf{R}$ such that $p(0) = (a_1, a_2, \ldots, a_n)$ and $p(1) = (b_1, b_2, \ldots, b_n)$. Note that $A_1 \cup A_2 \cup A_3 \subset A$. Hence p is a path in A joining $(a_1, a_2, \ldots, a_n)$ and $(b_1, b_2, \ldots, b_n)$.

***Case 2*:** $\alpha < c = d < \beta$. Since $c = d$, $(a_1, a_2, \ldots, a_n)$ and $(b_1, b_2, \ldots, b_n) \in A_1$ and A_1 is path-connected.

Lemma 1: For $n \geq 2$, $\mathbf{S}^n$ can be written as a union of two open sets U_1 and U_2 of $\mathbf{S}^n$ such that $U_1 \cap U_2$ is path-connected and each of U_1 and U_2 is simply connected.

Proof: Let $n \geq 2$ be fixed. Consider the two open subsets of $\mathbf{S}^n$ given below:

$U_1 = \{(x_1, x_2, ..., x_{n+1}) \in \mathbf{S}^n;\ x_{n+1} > -1/4\}$

$U_2 = \{(x_1, x_2, ..., x_{n+1}) \in \mathbf{S}^n;\ x_{n+1} < 1/4\}$

Clearly $U_1 \cap U_2 = \mathbf{S}^n$. We shall show that U_1 and U_2 have the required properties. Look at the map $h: \mathbf{S}^n \to \mathbf{R}^n$ defined by $h\{(x_1, x_2, ..., x_{n+1}) = (y_1, y_2, ..., y_n)$

where $y_i = \dfrac{x_i}{1 + x_{n+1}}$ where $1 \leq i \leq n$.

Clearly h is a homeomorphism of U_1 onto the range.

$D = \{(y_1, y_2, ..., x_n) \in \mathbf{R}^n;\ ^n\ y_i^2 < 5/3\}$. D is an open ball of $\mathbf{R}^n$ of radius 5/3. Hence D is simply connected. As a result U_1 is simply connected.

Similar arguments show that U_2 is simply connected.

Now $U_1 \cap U_2 = \{(x_1, x_2, ..., x_{n+1}) \in \mathbf{S}^n; -1/4 < x_{n+1} < 1/4\}$. We see that

$h(U_1 \cap U_2) = \{(y_1, y_2, ..., x_n) \in \mathbf{R}^n; 3/5 <\ ^n\ y_i^2 < 5/3\}$ and this set is path connected by dint of Lemma 2. Since h is a homeomorphism, $U_1 \cap U_2$ is also path-connected.

The following theorem followa from Lemma 1 and Theorem 4.1.3.

Theorem 4.1.4: The n-sphere $\mathbf{S}^n$, $n \geq 2$, is simply connected.

Examples: Any convex subset X of $\mathbf{R}^n$ is contractible to a point. To prove this, choose an arbitrary point $x_0 \in X$, and then define $F: X \times \mathbf{I} \to X$ by the expression $F(x, t) = (1 - t)x + tx_0$ (i.e., $F(x, t)$ is the point on the line segment joining x and x_o which divides it in the ratio $(1 - t): t$. Then $F(x, 0) = x$, and $F(x, 1) = x_o$., as required. More generally, we call a subset X of $\mathbf{R}^n$ *starlike with respect to the point* $x_0 \in X$ provided the line segment joining x and x_o lies in X for any $x \in X$. Then, the same proof suffices to show that, if X is starlike with respect to x_o, it is contractible to the point x_o – a contradiction, because S^1 is not even simply connected. Consequently, $\mathbf{S}^n$ is not contractible.

4.2 HOMOTOPY INVARIANCE OF THE FUNDAMENTAL GROUP

From Theorem 2.1.1 we know that two pointed spaces having the same homotopy type as pointed spaces have isomorphic fundamental groups. We would like to establish a similar result for two path connected spaces which have the same homotopy type as spaces (no base point condition). We need some preliminary results.

Lemma: Let $F: \mathbf{I} \times \mathbf{I} \to X$ and let a_o, a_1, b_o, and b_1 be the paths in X defined by $a_i(t) = F(i, t)$ and $b_i(t) = F(t, i)$; $t \in \mathbf{I}$, $i = 0, 1$. Then $(a_0 * b_1) * (a_1^{-1} * b_0^{-1})$ is a loop in X at $F(0, 0)$ which represents the trivial element of $\pi_1(X, x_0)$ where $x_0 = F(0, 0)$.

Proof: Let a'_o, a'_1, b'_o, and b_i be the paths in $\mathbf{I} \times \mathbf{I}$ defined by $a'_1(t) = (i, t)$ and $b'_1(t) = (t, i)$; $t \in \mathbf{I}$, i =0, 1. Then $(a'_o * b'_1) * (a'^{-1}_1 * b'^{-1}_o)$ is a loop at (0, 0). F maps this loop into the loop $(a_o * b_1) * (a_1^{-1} * b_o^{-1})$ at x_o. Since $\mathbf{I} \times \mathbf{I}$, being a convex subset of $\mathbf{R}^2$, is contractible, hence $(a_o * b_1) * (a_1^{-1} * b_o^1) \cong c(0, 0)$ where $c(0, 0)$ is the constant loop at (0, 0). Therefore,

$$(a_o * b_1) * (a_1^{-1} * b_o^{-1}) = F\ o\ ((a'_o * b'_1) * (a'^{-1}_1 * b'^{-1}_o) \cong F\ o\ c(0, 0) = d(x_0)$$

where $d(x_o)$ is the constant loop at $x_o = F(0, 0)$.

Let f, g be continuous maps: $X \to Y$ and let $F: X \times \mathbf{I} \to Y$ be a homotopy between f and g, i.e., $F(x, 0) = f(x)$, $F(x, 1) = g(x)$. Choose a base point $x_o \in X$. Then, f and g induce homomorphisms

$$f*: \pi_1(X, x_o) \to (Y, f(x_0)),$$

$$g*: \pi_1(X, x_o) \to (Y, f(x_o)).$$

Let α be the path from $f(x_o)$ to $g(x_o)$ given by $\alpha(t) = F(x_o, t), t \in \mathbf{I}$. α defines an isomorphism $u: \pi_1(Y, f(x_o)) \to \pi_1(Y, g(x_0))$ by the formula

$$u(\lambda) = [\alpha]^{-1} * \lambda * [\alpha],\ \lambda \in \pi_1(Y,\ f(x_0)).$$

Theorem 4.2.1: Under the above hypothesis, the following diagram is commutative.

$$\begin{array}{ccc} \pi_1(X,\ x) & \xrightarrow{f*} & \pi_1(Y,\ f(x)) \\ & \searrow^{g*} & \downarrow u \\ & & \pi_1(Y,\ g(x)) \end{array}$$

Proof: Let $h: I \to X$ be a loop at x_o. Consider the map $G: \mathbf{I} \times \mathbf{I} \to Y$ defined by

$G(t, a) = F(h(t), s)$.

Then, we have $G(t, 0) = F(h(t), 0) = f(h(t))$, $G(t, 1) = F(h(t), 1) = g(h(t))$.

$G(0, s) = F(h(0), s) = F(x_o, s) = F(h(1), s) = G(1, s)$. i.e.,

$$G(t, 0) = f\ o\ h(t),\ G(t, 1) = g\ o\ h(t) \text{ and } G(0, s) = G(1, s) = \alpha(s).$$

Apply the Lemma above to conclude that $(\alpha * (g\ o\ h)) * (a^{-1} * (f\ o\ h)^{-1})$ is a loop at $f(x_0)$ and it represents the trivial element of $\pi_1(Y,\ f(x_0))$.

This implies $[\alpha] * g * ([h]) * [\alpha]^{-1} = f * ([h])$ or equivalently

$$g * ([h]) = [\alpha]^{-1} * f * ([h]) * [\alpha] = u(f * [h])$$

$$g* = u\ o\ f*$$

Theorem 4.2.2: If $f: X \to Y$ is a homotopy equivalence, then $f: \pi_1(X,\ x) \to \pi_1(Y,\ f(x))$ is an isomorphism for any $x \in X$.

Proof: Because $gf = g\ o\ f \cong 1_X$, we obtain the following diagram (which is commutative by Theorem 4.2.1):

$$\begin{array}{ccc} \pi_1(X,\ x) & \xrightarrow{f*} & \pi_1(Y,\ f(x)) \\ & \searrow^{g*} & \downarrow g* \\ & & \pi_1(Y,\ gf(x)) \end{array}$$

Here u is an isomorphism induced by a certain path from x to $gf(x)$. Hence, we conclude f is a monomorphism, and g is an epimorphism. Applying the same argument to the homotopy $fg = f\ o\ g \cong 1_y$ we obtain the following commutitative diagram:

$$\begin{array}{ccc} \pi_1(Y,\ f(x)) & \overset{v}{\searrow} & \\ g* \downarrow & & \\ \pi_1(X, gf(x)) & \xrightarrow{f*} & \pi_1(Y,\ fgf(x)) \end{array}$$

Therefore, we conclude g is a monomorphism. Because $f^* g^*$ is both an epimorphism and a monomorphism, it is an isomorphism. Because $g * f* = v$ and both $g*$ and v are isomorphisms, $f*$ has to be an isomorphism.

Corollary 1: If X and Y are path connected and are of the same homotopy type, then their fundamental groups are isomorphic.

Corollary 2: If X and Y are path connected and homeomorphic spaces, then their fundamental groups are isomorphic.

Remark: The above two corollaries provide a very powerful method in algebraic topology of proving that certain spaces are not of the same homotopy type (and hence are not homeomorphic).

Theorem 4.2.3: $\mathbf{S}^1$ is not a retract of $E^2 = \{(x_1, x_2) \in \mathbf{R}^2 : x_1^2 + x_2^2 \le 1\}$.

Proof: If possible let $\mathbf{S}^1$ be a retract of E^2. Let $i : \mathbf{S}^1 \to E$ be the inclusion map. Then by our assumption, there is a continuous map (i.e., retraction) $r: E^2 \to \mathbf{S}^1$ such that $r.i. = 1_s'$. Then, we have

$$\pi_1(\mathbf{S}^1, (1, 0)) \xrightarrow{j*} \pi_1(E^2(1, 0)) \xrightarrow{r*} \pi_1(\mathbf{S}^1, (1, 0)) \text{ and}$$

$$r * i * = (ri) * = (1_s^{1-}) * = \text{identity on } (\mathbf{S}^1, (1, 0)) = \mathbf{Z}.$$

This is a contradiction, because $\pi_1(E^2, (1, 0)) = \{0\}$.

One of the best known theorems of topology is the following fixed-point theorem of L.E.J. Brouwer.

Let $E^n = \{x \in \mathbf{R}^n : \| x \| \le 1\}$

Theorem 4.5.4: Any continuous map *f* of *E* into itself has at least one fixed point, i.e., a point *x* such that $f(x) = x$.

Proof: We shall prove this theorem for $n \le 2$.

First we prove that, for any integer $n > 0$, the existence of a continuous map $f : E^n \to E^n$. which has no fixed point, implies that the $(n - 1)$–sphere $\mathbf{S}^{n-1}$ is a retract of E^n. This is done by the following geometric construction. For any point $x \in E^n$, let $r(x)$ denote the point of intersection of $\mathbf{S}^{n-1}$ and the ray starting at the point $f(x)$ and going through the point x . r can be shown to be a continuous function of E^n into $\mathbf{S}^{n-1}$. If $x \in \mathbf{S}^{n-1}$. It is clear that $r(x) = x$. Therefore r is the desired retraction. For $n = 1$, $S = 1$, t is disconnected, E connected and hence S cannot be a retract of E.

For $n = 2$, the result follows from Theorem 4.2.3.

Hence the theorem cannot be proved by methods of homotopy theory. In later chapters we shall present the proof for the general case. Figure shows the situation for the case $n = 2$.

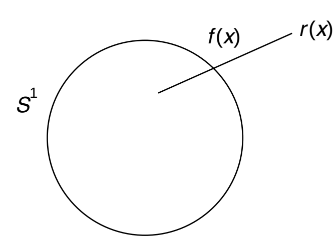


Fig. 4.1

CHAPTER 5

Surfaces, Manifolds and CW Complexes

5.1 SURFACES

Since classification of topological spaces requires the evaluation of homotopy and cohomology groups, we shall first discuss some special topological spaces of pathological interest.

Let $\mathbf{R}^n$ denote the n-dimensional Euclidean space with the usual Euclidean metric. We shall now denote $B^n = \{X \in \mathbf{R}^n ; \|X\| \geq 1\}$ = The closed unit n-ball

$\mathbf{S}^{n-1} = \{X \in \mathbf{R}^n ; \|X\| = i\}$ = The (n – 1)-sphere in $\mathbf{R}^n$

Thus, $\mathbf{S}^1 = \{(x_1, x_2), x_1^{\ 2} + x_2^{\ 2} = 1\}$ = The unit circle in $\mathbf{R}^2$

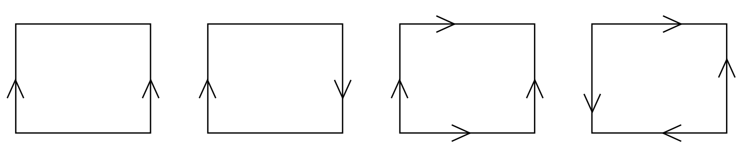
and $\mathbf{S}^2 = \{(x_1, x_2, x_3), x_1^{\ 2} + x_2^{\ 2} + x_3^{\ 2} = 1\}$ = The unit sphere in $\mathbf{R}^3$

$\mathbf{I} = \{X \in \mathbf{R} ; 0 \leq x \leq i\}$ = The unit interval in $\mathbf{R}$

$\mathbf{I}^2 = \mathbf{I} \times \mathbf{I} = \{(x, y), 0 \leq x \leq 1, 0 \leq y \leq 1\}$ = The unit square in $\mathbf{R}^2$

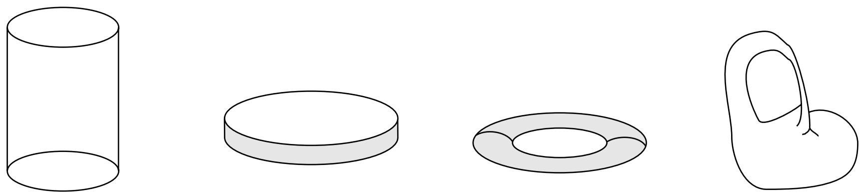
$\mathbf{S}^1 \times \mathbf{S}^1 = \{x, y\}; x \in \mathbf{S}^1, y \in \mathbf{S}^1\}$ = The torus having the surface of a doughnut.

Observe that a torus can be obtained by topological identification of the edges of a square as indicated below. Surfaces like Cylinder, Mobius Band, K leins bottle can also be obtained by topological identification as follows.



Cylinder Mobius band Torus Klein's bottle

Note the same direction i.e., arrowhead implies that the two edges are to be glued similarly but opposite arrowheads implies that the corresponding edges are to be glued after a twist of 180°. The surfaces are pictured below.



The rigorous mathematical method of defining the surfaces like Cylinder, Mobius band, Torus and Kleins bottle is the topological identification. Below we demonstrate the method for Cylinder and Mobius band only. The others follow similarly.

Let $X = \mathbf{I} \times \mathbf{I}$. Define an equivalence relation ~ on X by $(x, y) \sim (x', y')$ if $f\ (x, y) = (x', y')$ or $\{x, x'\} = \{0, 1\}$ and $y = y'$. Then $X/\sim$ with the quotient topology is in fact homeomorphic to the cylinder.

For mobius band, we define the equivalence relation ~ on X by $(x, y) \sim (x'. y')$ if $f\ (x, y) = (x', y')$ or $\{x, x'\} = \{0, 1\}$ and $y = 1 \sim y'$.

Then $X/\sim$ is homeomorphic to the Mobius band.

One should note $\mathbf{S}^1$ can also be obtained from I by a similar identification. The torus is thus $S^1 \times S^1$ equiped with the product topology.

Projective Spaces

The real projective space $\mathbf{P}^n$ is defined by identification as follows:

Define ~ on $\mathbf{S}^n$ as $x \sim y$ if $f\,x = -y$ i.e., x and $-x$ of S are identified by ~.

Then $\mathbf{S}^n/\sim$ is called the *Real Projective Space* and is denoted by $\mathbf{P}^n$ or more specifically $\mathbf{RP}^n$. The space $\mathbf{P}^2$ is usually called the *Real Projective Plane* and hence is obtained from S^2 by topological identification.

It is easy to observe that $\mathbf{P}^n$ just as $\mathbf{S}^{n-1}$ is a closed subspace of $\mathbf{S}^n$.

One can define the complex projective space similarly, infact, starting with the complex n-sphere $\mathbf{S}^{2n+1} \subset C^{n+1}$ one defines.

$$(z_1, z_2, \ldots, z_{n+1}, 0, ..) \sim (\lambda\, z_1, \ldots, \lambda\, z_{n+1}, 0, ...)$$

for each complex number λ with $|\lambda| = 1$. The resulting quotient space is the complex projective n-space and is denoted by $\mathbf{CP}^n$ to distinguish from $\mathbf{RP}^n$.

Connected Sum

Intuitively the connected sum of two surfaces is obtained by removing small open discs from both the surfaces and then glueing along their boundaries.

Thus the double torus is the connected sum of two torus. The formal definition is given below:

Definition: Let X and Y be two topological spaces, then the connected sum W of X and Y, written as $X \times Y$ is a topological space W defined as follows:

Let $f : U \to V$ be a homeomorphism where $U \subset X$ and $V \subset Y$. Then define an equivalence ~ on $X\ U\ Y$ as $Z_1 \sim Z_2$ if $Z_1 = Z_2$ or $z_1 = f\ (z_2)$ or $z_2 = f\ (z_1)$.

Then, $W = X \cup Y$ is the connected sum of the spaces X and Y. Note f does the connection of the two spaces. In particular, this connection can be effected by taking only two points x_o in X and y_o in Y and then defining ~ on $X \cup Y$ as $z_1 \sim z_2$ iff $z_1 = z_2$ or $z_1 = x_o$ and $z_2 = y_o$ or $z_1 = y_o$ and $z_2 = x_o$. Note the definition can be extended to finitely many spaces.

One of the classic theorem on surfaces is the following.

Classification Theorem of Surfaces

Every compact surface is either homeomorphic to a sphere or to a connected sum of a sphere with some tori or to a connected sum of a sphere with some projective planes.

5.2 MANIFOLD

Mainfolds constitute a general class of topological spaces with nice behaviour and hence form an important component of the study of topological spaces from algebraic stand point.

Definition: An *n-dimensional manifold* or simply *n-manifold* is a Hausdorff space in which each point has an open neighbourhood homeomorphic to the open *n*-dimensional disc $D^n = \{X \in \mathbf{R}^n : \|X\| < 1\}$.

Clearly the circle S^1 is a *I*-manifold, S^2 is a 2-manifold and similarly S^n is an *n*-manifold. The following observations are interesting:

Observations 1: Every non-empty open subset of an *n*-manifold is an *n*-manifold.

Observations 2: If S is an *m*-manifold and T is an *n*-manifold, then $S \times T$ is an $(m + n)$ manifold.

Observations 3: If X is a *G*-space where G is a finit group acting freely on X, then X/G is an *n*-manifold if X is so.

Observations 4: Every compact *n*-manifold is homeomorphic to a subspace of some Euclidean space.

Definition: A *surface* is a compact connected 2-manifold. Thus the mobius band, torus, double torus, sphere, Klein's bottle, real projective plane are surfaces.

Definition: A surface is said to be *orientable* if it does not contain a mobius strip within it. A surface is called *non-orientable* if it is not orientable.

Thus a torus is orientable but a Klein's bottle is non-orientable. Similarly a sphere and a double torus are orientable surfaces.

Definition: An *orientable surface* is said to be of genus m if it is expressible as the connected sum of $\mathbf{S}^1$ and m replicas of a torus.

A non- orientable surface is said to be of genus m if it is expressible as the connected sum of $\mathbf{S}^1$ and m replices of the real projective plane $\mathbf{P}^2$.

Definition: An *n-manifold M* with boundary ∂M is a Hausdorff space M in which each point has an open neighbourhood homeomorphic to either $\mathbf{R}^n$ or the upper half space of $\mathbf{R}^n$, i.e., $\{x \in \mathbf{R}^n, x_n \geq 0\}$.

Here $\partial M = \{x \in M;\ x$ has neighbourhoods homeomorphic to the upper half space but not to $\mathbf{R}^n\}$.

Therefore, a surface with boundary is a compact connected 2 manifold with boundary.

It is easy to observe that the boundary of an *n*-manifold is an $(n - 1)$ - manifold.

As for example; the space B^n is an *n*-manifold with boundary $\mathbf{S}^{n-1}$.

The Mobius strip is a surface with boundary. As regards surfaces the following result is interesting.

Theorem 5.2.1: Two surfaces are homeomorphic if and only if their fundamental groups are isomorphic. For proof, consult Spanier [25].

Definition: A surface is imply connected if it is path-connected and its fundamental group is trivial.

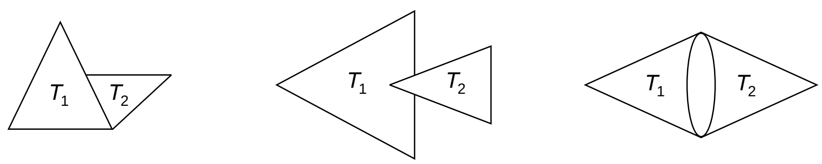
One simple characterization of simply connected surfaces is the following:

Theorem 5.2.2: A surface is simply connected if and only if it is homeomorphic to $\mathbf{S}^2$.

Proof: Easy.

Definition: A triangulation of a compact surface S consists of a finite family of closed subsets $\{T_1, T_2, T_n\}$ which cover S and a family of homeomorphisms $\phi_1 : T_i \to S$. Intuitively, triangulation of surface means division of the surface into several triangles which make up the entire surface.

So the following triangulations are not allowable

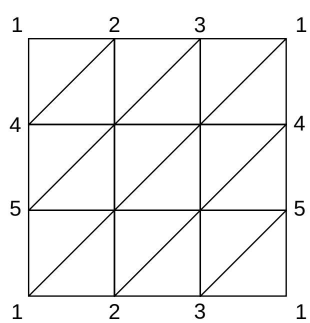


For a triangulations to be allowable, the following conditions must be fulfilled.

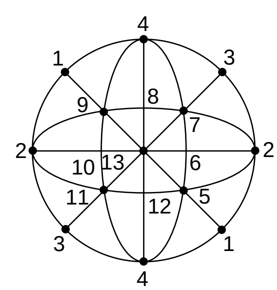
(i) Each side is an edge of exactly two triangles.

(ii) Let v be a vertex of a triangulation. Then, one can arrange the family of all triangles with the common vertex v in cyclic order such that, $T_o, T_1, T_2, \ldots T_{n-1}\ T_n = T_o$ such that T_i and T_{i+1} have an edge in common for $0 \le i \le n-1$.

Below we show some triangulations.



Torus



Projective Plane

Remarks: The notion of triangulation can be extended to arbitrary topological spaces as follows:

A *triangulation* (K, φ) of a topological space X consists of a simplicial complex K and a hoeomorphism $\phi : |K| \to X$. If X has a triangulation i.e., if X is triangulable, X is called a *polyhedron*. In other words, every polyhedron is triangulable. This is not difficult to prove.

Rado's Theorem: A surface is triangulable if and only if it has a countable basis for its topology.

Whether or not the n-manifolds ($n > 2$) are triangulable is still an open problem.

However for classfication, Euler devised a characteristic for triangulable surfaces.

Definition: If S is a *triangulable surface* with v vertices, e edges and t triangles, then $v - e + t$ is called the *Euler's characterstic* of the surface S and is denoted by $\chi(S)$.

Thus $\chi(S) = v - e + t$.

One thing should be observed that the Euler's characteristic of a surface is independent of the mode of tringulation and as such it is well-defined.

One can easily see χ (Torus) = 0

χ (Sphere) = 2

χ (Projective Plane) = 1.

The following two results are classic in this area.

Theorem 5.2.3: If S_1 and S_2 are compact surfaces, then $\chi(S_1 \# S_2) = \chi(S_1) + \chi(S_1) - 2$.

Theorem 5.2.4: Two compact surfaces are homeomorphic if and only if their Euler's characteristics are equal and surfaces are both orientable or both non-orientable.

Covering Space: Covering spaces form a class of topological spaces that have peculiar properties. We just recall the definition here.

Definition: A *covering space* of a topological space X is a pair $(\bar{X}, p)$ where X is a topological space and $p: \bar{X} \to X$ is a continuous map such that each point x of X has a path-connected open neighbourhood U for which each path-component of p^{-1}(U) is mapped topologically onto U by p.

Any open neighbourhood U that satisfies the condition stated above is called an *elementary neighbourhood* and p is called the *projection.* As for example, (**R**, p) is a covering space of $\mathbf{S}^1$ where, $p: \mathbf{R} \to \mathbf{S}^1$ is defined as $p(x) = e^{ix}$. Similarly, if $f: Y \to X$ is an homeomorphism, then (Y, f) is a covering space of X. It is easy to observe that every space is a covering space of itself.

An important result regarding covering spaces is the following:

Covering Space Theorem: If $p: B \to X$ is a covering projection, $b_o \in B$, $x_o = p(b_o)$ then $p*: \pi_n(B, b_o) \cong \pi_n(X, x_o)$, $n \geq 2$ and p maps $\pi_1(B, b_o)$ isomorphically into $\pi_1(X, x_o)$.

Proof: See Spainer [25].

5.3 CW COMPLEXES

The computation of homology groups for spaces is not as easy as it appears.

For surfaces like torus even, it is too long and tedius.

A technique was developed by J.H.C. Whitehead in this regard and the related concept, viz., CW Complex generalized to a great extent the spaces that are usually dealt in this connection.

Definition: A *CW Complex* is a Hausdorff space X which is the disjoint union of homeomorphs e * of the cell B^n a finite union of subsets of X each of which is a homeomorphic image of a cell of dimension less than n.

A CW Complex is finite if it is a union of a finitely many homeomorphs of B^n.

It is easy to observe that a finite CW Complex is compact.

One can see easily that most of the commonly known manifolds are CW Complexes.

As for example, the sphere, the torus, the Klein's bottle and the Projective Spaces are CW Complexes.

An interesting result is that any topological space is approximable by CW Complexes which is unique upto homotopy.

5.4 FIBRE BUNDLES

A covering space is in a sense the product of its base space and a discrete space. Its direct generalization is the concept of fibre bundle and for a fibre bundle, the total space is locally the product of its base space and its fibre.

Definition: A *fibre bundle* is a quadruple (E, B, F, p) where E is the total space, B is the base space, F is a fibre and $p: E \to B$ is a bundle projection such that there exists an open covering $\{U_\alpha\}$ of B and for each U_α there exists a homeomorphism $\phi_\alpha : U_\alpha \times F \to p^{-1}(U_\alpha)$ such that the composite.

$$U_\alpha \times F \longrightarrow p^{-1}(U_\alpha) \xrightarrow[p]{} U_\alpha$$

is the projection to the first factor.

The set $p^{-1}(b)$ is called the fibre at $b \in B$

We note

(1) F is homeomorphic to $p^{-1}(b)$

(2) The bundle projection $p: E \to B$ and the projection $B \times F \to B$ are locally equivalent.

CHAPTER 6

Simplicial Homology Theory

The exquisite world of algebraic topology came into existence out of our attempts to solve topological problems by the use of algebraic tools and this revealed to us the nice interplay between algebra and topology which causes each to reinforce interpretations of the other there by breaking down the often artificial subdivision of mathematics into different branches and emphasizing the essential unity of all mathematics. The homology theory is the main branch of algebraic topology and plays the main role in the classification problems of topological spaces. There are various approaches to the study of this theory such as geometric approach, abstract approach and axiomatic approach. Since geometric approach appeals easily to our intuition, we shall start with geometric approach, the relevant development being called simplicial homology theory. The abstraction of this having two-way development will be called *singular homology theory* in one direction and *Cech Homology Theory* in the other. The singular theory is more appropriate to homotopy theory while the Cech theory is more appropriate to homotopy theory while the Cech theory is suitable for the study of manifolds and fibre bundles. Thus the simplicial theory is in a sense the intersection of these two theories. There is another approach due to Eilenburg and Steenrod popularly known as axiomatic approach. Since such an approach always evolves historically as a successor to the intuitive approach, ours is no exception to it. But the limitations imposed on us to write such a book in so few pages will hardly give us any opportunity to go into the details of proofs.

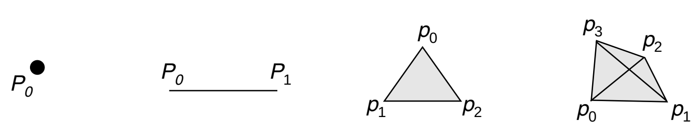
6.1 SIMPLEX AND SIMPLICIAL COMPLEX

We introduce the notation of simplex in a finite dimensional Euclidean space. It should however be noted that this notation of simplex can be generalized to infinite dimensional spaces also and infact to general topological ventor spaces. Intuitively, simplices are the building blocks of this simplicial homology theory.

Definition: A set of points $p_o, p_1, \ldots p_n$, in $\mathbf{R}^m$ $(n \geq m)$ is said to be *affinely independent* if the vectors $p_1 - p_o, p_2 - p_o, \ldots, p_n - p_o$ are linearly independent.

An n-simplex Δ_n is the set of points obtained by taking all convex combinations of a set of $n + 1$ affinely independent points in $\mathbf{R}^m$ $(m \geq n)$ called the vertices of the simplex Δ_α.

Thus, a point may be taken as *o*-simplex, a closed segment is a 1-simplex, a closed triangle is a 2-simplex, a closed tetrahedron is a 3-simplex.



0-simplex 1-simplex 2-simplex 3-simplex

Observation 1: Every point p of an n-simplex $(p_o, p_1 ; - ; p_n)$ can be written as

$$p = \lambda \ p_o + \lambda_1 \ p + ... + \lambda_n \ p_n$$

where $\lambda_i \geq 0$ for all i and $\sum \lambda_i = 1$

The ordered real numbers $(\lambda_0, \lambda_1, ..., \lambda_n)$, is referred to as the *barycentric co-ordinates* of the point p.

Observation 2: Every point p admits of a unique expression in *barycentric* coordinates.

Remark: Sometimes a simplex Δ_n may be expressed in terms of its vertices as $(p_o, p_1, ..., p_n)$ and in that case the simplex is said to be spanned by its vertices.

Observation 3: The boundary of an n-simplex is composed of $(n - 1)$–simplices. For example, the boundary of the above 2-simplex is composed of three I-simplices namely (p_o, p_1), (p_1, p_2) and (p_2, p_o) the boundary of the above 3-simplex is made of four 2-simplices namely

$$(p_o \ p_1 \ p_2), (p_o \ p_2 \ p_3), (p_o \ p_1 \ p_3) \text{ and } (p_1 \ p_2 \ p_3)$$

By we shall denote the boundary of Δ_n by Δ_n.

Thus $\Delta_1 = (p_o) \cup (p_i) = (p_o \ p_1) \cup (p_1 \ p_2) \cup (p_2 \ p_o)$.

Definition: By an *m-face* of a simplex Δ_n we mean a subset Δ_m spanned by any set of m vertices of where $m < n$ i.e., the set of points obtained by taking all convex combinations of any m-vertices of Δ_n.

Observe that a 2-simplex Δ_2 has its boundary composed three faces, each of which is a 1-simplex and a 3-simplex has four faces, each of which is a 2-simplex. Thus, an n-simplex will have $(n + 1)$ faces each of which is an $(n - 1)$ simplex.

Definition: A *finite geometric simplicial complex* or simply complex in $\mathbf{R}^m$ is a finite collection K of simplices Δ_p of $\mathbf{R}^m$ subject to the conditions.

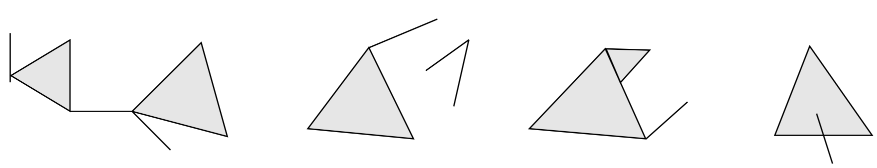
(*a*) If $\Delta_\rho \in K$ and Δ_q is a face of Δ_p written as $\Delta_q < \Delta_p$, then $\Delta_q \in K$.

(*b*) If $\Delta_\rho \cap \Delta_q \neq \phi$. $\Delta_\rho \in K, \Delta_q \in K,$ then $\Delta_\rho \cap \Delta_q \in K$.

Thus, if a complex K has a simplex as its object, then K contains all its faces also.

The dimension of a complex will be defined as the least upper bound of the dimensions of its simplices.

The following examples illustrate the situation



a complex   a complex   not a complex   not a complex

6.2 TRIANGULATION

Definition: The union of all simplices of a complex K will be called the underlying *polyhedron* of K and will be denoted by $|K|$. Thus $|K|$ is the set of all points of the complex K the topologised by the Enclidean metric.

The complex K is said to be a *dissection* or *triangulation* of the polyhedron K.

Observation 1: If $\Delta \in K,$ then $\Delta \subset |K|$.

Observation 2: The polyhedorn $|K|$ of a complex K is the set of points that constitute the simplices of the complex K and it has the relative Euclidean topology. Thus from above it follows every polyhedron is compact.

Definition: A subset of a complex K which itself is a complex is called a *subcomplex* of K.

The subcomplex of a given complex K consisting of all simplices of dimension $\leq r$ is called the r-skeleton of the complex K and is denoted as $K^{(r)}$.

Thus, the O-skeleton $K^{(c)}$ consists of all vertices of K.

Observation 1: The union and the intersection of subcomplexes are subcomplexes.

Observation 2: If K_1 and K_2 are two subcomplexes of K, then $|K_1 \cup K_2| = |K_1| \cup |K_2|$ and $|K_1 \cap K_2| = |K_1| \cap |K_2|$

Definition: If a subset of K_o of a polyheydron $|K|$ is such that there exists a subcomplex whose underlying polyhedron is $|K_0|$, then it is called a *subpolyhedron*.

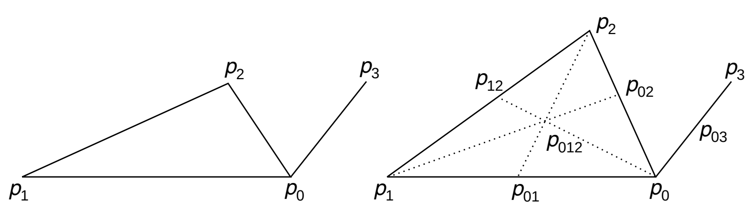
Let $(p_o \; p_1 \; ... \; p_n)$ be an n-simplex. Then we can define a barycenter (or center of gravity) of this simplex as a distinct point $p_{o\,1\,2\,...\,n}$ defined as

$$p_{o\,1\,2\,...\,n} = \frac{1}{n+1} \sum_{i=o}^{n} p_i$$

6.3 BARYCENTRIC SUBDIVISION

Definition: A subdivision of a complex into simplices whose vertices are finite subsets of the vertices and the barycentres such that the subscripts of p's satisfy a total order relation will be called a *barrycentric subdivision*.

The following will illustrate the situation better. Consider the complex defined below by the vertices p_1, p_2 and p_3.



Let p_{01}, p_{12}, p_{02}, p_{03} denote the barycentres of $(p_0 \; p_1)$, $(p_1 \; p_2)$, $(p_0 \; p_2)$ and $(p_0 \; p_3)$ respectively and let $p_{o\,1\,2}$ denote the barycentre of $(p_0 \; p_1 \; p_2)$.

The simplices obtained by dissecting the given complex will be $(p_0 \; p_{01} \; p_{012})$, $(p_1 \; p_{01} \; p_{012})$, $(p_1 \; p_{12} \; p_{012})$, $(p_2 \; p_{12} \; p_{012})$, $(p_2 \; p_{02} \; p_{012})$, $(p_0 \; p_{02} \; p_{012})$, $(p_0 \; p_{03})$ and $(p_3 \; p_{03})$. Observe in each simplex the sub-scripts satisfy an ordering by inclusion. Such a sub-division will be called the *first barycentric subdivision*, of the given complex K and it will be denoted by K'. One can make another set of dissection in the same way to obtain the second barycentric subdivision K''. If the process is continued, one will be able to get the *n*th barycentric subdivision K^n. Thus $K^n = (K^{n-1})'$. One should observe that the more we continue the dissection process, the less will be the diameters of the simplices obtained thereby. In fact the following is true.

Theorem 6.1.1: For any simplicial complex *K* and any $\varepsilon > 0$, there exists a natural number *n* such that the simplices of the barycentric subdivision K^n have diameters $< \varepsilon$.

One can immediately derive the following corollary:

Corollary: Every polyhedron has for every $\varepsilon > o$ a triangulation with simplices of diameters $< \varepsilon$.

Definition: For a vertex *p* of a complex *K*, the star of *p* written as $St_k\ [p] = p < \Delta \in K$ where $p < D$ means *p* is a vertex of a simplex *D* i.e., it is the union of all simplices of *K* which have *p* as a vertex.

6.4 SIMPLICIAL MAP

Definition: A map $\phi : |K| \to |L|$ from the polyhedron of a complex *K* to the polyhedron of another complex *L* is said to be *a simplicial map* if for every simplex $(p_o\ p_1 \ldots p_n) \in K$, the points $\phi\ (p_0), \ldots .\ \phi\ (p_n)$ are the vertices of a certain simplex in *L*.

One should note that a simplicial map carries a simplex to another simplex not necessarly of the same dimension.

6.5 SIMPLICIAL APPROXIMATION

Definition: Let $f : |K| \to |L|$ be a continuous map from a polyhedron |*K*| to a polyhedron |*L*|. Then a simplicial approximation of *f* is defined as a triple (K, L, ϕ) where *K* and *L* are the associated complexes, $|K|$ and $|L|$ the triangulations of *K* and *L*, and ϕ is a simplicial map such that for every $St_k\ (P) \subset f^{-1}\ (st_1(\phi\ (p)))$.

One of the important achievements in the direction of triangulating of a space is the following theorem and its corollary.

Theorem 6.5.1: For every polyhedron *L*, there exists an ε_α such that if *f* and *g* are continuous mappings of an arbitrary polyhedron *K* into *L* and $|\ f - g\ | < \varepsilon$, then *f* and *g* have a common simplicial approximation $(K^{(n)}, L, \phi)$ for some $n \geq 1$.

Proof: See Dold [6].

Corollary: Every continuous mapping of a polyhedron into a polyhedron has a simplicial approximation.

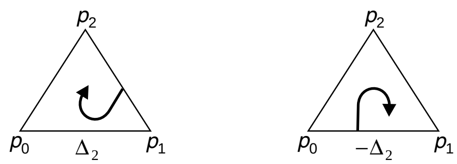
Orientation of a simplex means an ordering of its vertices.

Definition: Two orderings of the vertices are said to determine the same *orientation* of the simplex if and only if an even permutation transforms one ordering into the other; if the permutation is odd, the orientations are said to be opposite.

Thus, there are basically two orientations of a simplex. If Δ_n has the orientation $(p_0\ p_1 \ldots p_n)$, then the same simplex with opposite orientation will be denoted as $-\Delta_n$.

For example, if $\Delta_2 = (p_0\ p_1\ p_2)$, then $(p_1\ p_2\ p_0) = (p_2\ p_0\ p_1) = \Delta_2$ but

$$(p_0\ p_2\ p_1) = (p_2\ p_1\ p_0) = (p_1\ p_0\ p_2) = -\Delta_2.$$



Let $\Delta_n = (p_0\ p_1\ ...\ p_n)$. An oriented $(n - 1)$ face opposite p_i written by $(p_0\ p_1\ ...\ \hat{p}_i\ ...\ p_n)$ is said to be *coherently oriented* with $(-1)^i$, i.e., $(p_0\ p_1\ ..\ \hat{p}_i\ ..\ p_n)$ is coherently oriented with Δ_n if i is even and *noncoherently oriented* if i is odd. Thus $(p_0\ p_1)$ has coherent orientation with the simplex $(p_0\ p_1\ p_2)$ while $(p_0\ p_2)$ is non-coherently oriented with $(p_0\ p_1\ p_2)$.

$(p_0\, p_1)$ Coherently oriented $(p_0\, p_2)$ Non-coherently oriented

A complex K is said to be *oriented* if each of its simplices is endowed with some orientation. In fact an orientation of a complex K is a function which assigns to each simplex of K one of the oriented simplices determined by it and hence an oriented complex is written as (K, α).

Remark: To start with it may appear that orientation is a must for defining homology group of a simplicial complex, but it is a proven fact that the homology group of a complex is independent of its orientation.

6.6 HOMOLOGY GROUP

If G is an additive abelian group and A is a subset of G, then the smallest subgroup containing A is called the *group generated by A* and the set A is called the *set of generators* of this group. Every element of this group admits of an expression of the form $m_1\ a_1 + m_2\ a_2 + ... + m_k\ a_k$ where $... \alpha_k \in A$, m_1, m_2 ... are integers, i.e., sum of integral multiples of finitely many elements of A.

The generated group is said to be *free* if such an expression for every element is unique.

We shall use this notion now to construct a sequence groups, called *groups of chains*.

Definition: Let (K, α) be an oriented complex. The free group generated by all 0-simplices of K is called the *group of 0-dimensional chains* and is denoted as $C_o(K, \alpha)$ or simply $C_o(K)$. Similarly, the free group generated by all 1-simplices of K is the group of 1-dimensional chains and is denoted as $C_1(K)$. One defines $C_n(K)$ as the group of n-dimensional chains.

Thus, we have a sequence of the groups of chains

$$C_o(K), C_1(K), C_2(K), ..., C_n(K), C_{n+1}(K), ...$$

One should observe for a finite dimensional complex K, the above sequence is finite.

We next define a sequence of homomorphisms for the above sequence of additive abelian groups.

Let $\Delta_n = (p_o \; p_1 \; ... \; p_n)$ be an oriented *n*-simplex belonging to the complex *K*.

Let us define map $\partial_n : \Delta_n \to C_{n-1}(K)$ as follows

$$\partial_n (\Delta_n) = \Sigma (-1)^i \; (p_o \; p_1 \; .. \; p_i \; ... \; p_n).$$

Then since Δ_n is a generator of $C_n(K)$ and ∂_n is to be a homomorphism on $C_n(K)$, we can extend this map to $C_n(K)$ as follows:

For any element $C_n = m_i \; \Delta_n^1 + m_2 \; \Delta_n^2 + ... + m_k \; \Delta_n^k \in C_n(K)$,

define $\partial_n (C_n) = m_1 \partial_n (\Delta_n^1) + m_2 \partial_n (\Delta_n^2) + ... + m_k \; \partial_n (\Delta_n^k)$

Such homomorphisms $\partial_n \; (n \geq 1)$ are called *boundary operators*. One can easily verify now that $\partial_{n-1} \; \partial_n = 0$.

The proof follows by first taking a generator Δ_n and proving $\partial_{n-1} \; \partial_n \; (\Delta_n) = 0$ and then taking any chain C instead of a generator.

We thus get a sequence of groups of chains with associated boundary operators as follows:

$$C_o \; (K) \xleftarrow{\partial_1} C_1 \; (K) \xleftarrow{\partial_2} C_2 \; (K) \xleftarrow{\partial_3} ... \; \xleftarrow{\partial_n} C_n \; (K) \xleftarrow{\partial n \; + 1} ...$$

Let us now consider the group $C_n(K)$ of *n*-dimensional chains and the two adjoining groups as follows:

$$C_{n-1} \; (K) \xleftarrow{\partial \, n} C_n \; (K) \xleftarrow{\partial \, \mathrm{n+1}} C_{n+1} \; (K)$$

Definition: The *kernel* of the homomorphism ∂_n which is a subgroup of $C_n(K)$ is called the group of *n-dimensional cycles* of the complex (K, α) and is denoted as $Z_n(K, \alpha)$ or simply $Z_n(K)$.

The image of the homomorphism ∂_{n+1} which is a (normal) subgroup of $C_n(K)$ is called the group of *n-dimensional boundaries* of the complex (K, α) and is denoted as $B_n(K, \alpha)$ or simply $B_n(K)$.

Note that the fact $\partial_{n-1} \; \partial_n = 0$ proves that $B_n(K)$ is a (normal) subgroup of $Z_n(K)$ and hence one can define the quotient group $Z_n(K)/B_n(K)$.

The *nth homology group* $H_n(K)$ is defined to be the quotient group $Z_n(K)/B_n(K)$. Thus, $H_n \; (K) = Z_n \; (K)/B_n \; (K), \; n \geq 1$.

By taking ∂_0 as a homomorphism mapping from $C_o(K)$ to the empty set ϕ one can define $H_o \; (K)$ also.

Two *n*-dimensional cycles z_1 and z_2 are said to be homologous if $z_1 - z_2 \in B_n(K)$.

It is not difficult to prove that if α and β are two distinct orientations of the same complex, the corresponding homology groups of all dimensions are isomorphic. Thus one can define homology group of an unoriented complex as the homology group of the said complex given any orientation to it.

Some Results

The following results are of computational importance. Their proofs are easy.

Let *G* denote the Coefficient Group, e.g., **Z**. Let $\cong$ denote isomorphism.

Theorem 6.6.1: $H_0(K) \cong G$ if and only if K is connected.

Theorem 6.6.2: $H_p(K) \cong \sum_{r=1}^{n} H_p(K_r)$ where *K* is the union of *n* disjoint connected components $K_1, ... K_n$.

Example 1: Let K be a connected 1-dimensional complex or linear graph, with α_0 vertices and α_1 edges.

Then $H_o(K) \cong G, H_1(K)$ is free abelian with p_1 generators where $p_1 = \alpha_1 - \alpha_o + 1$ i.e., $H_1(K) \cong G \oplus G \oplus \ldots \oplus G$ (p_1 many)

Example 2: If $\mathbf{S}^2$ denoted a *s*-sphere, then $H_o(\mathbf{S}^2) = G, H_1(\mathbf{S}^2) = 0$, and $H_2(\mathbf{S}^2)$ is free abelian with one generator.

Example 3: Since the torus is a topological product of two copies of $\mathbf{S}^1$ $H_o(K) \cong G, H_1(K) \cong G \oplus G,$ $H_2(K) \cong G,$ where K is the torus.

Example 4: If K is a sphere with p handles, then $H_o(K) \cong G, H_1(K) \cong Z, H_2(K)$ is free abelian with $2p$ generators.

Example 5: If K is the n-dimensional Projective space $\mathbf{P}^n$. Then $H_o(K) = Z,$ $H_r(K) = 0$ if r is odd and $H_r(K)$ is free abelian with two generators if r is even and not equal to zero.

6.7 HUREWICZ THEOREM

Next let $\pi_n\ (X, A, x_o)$ denote the nth relative homotopy group of X with x_o as the base point. Then the following theorem due to Hurewicz establishes a relation between homotopy groups and homology groups.

Hurewicz Theorem: Let the subspace A of X be arcwise connected and let X and A be simply connected. Let $\pi_i(X, A, x_o) = 0$ for $2 \leq i < n$. Then, $\pi_n\ (X, A, x_o) \cong H_n(X, A)$.

Proof: See Dold [8]

Thus, the first non-zero homology group and the first non-zero homotopy group have the same dimension and they are isomorphic.

An immediate consequence of the above theorem is the direct sum theorem:

Direct Sum Theorem: Let A be a retract of X and $f: X \to A$ be a retraction.

Let $i: A \to X$ and $j: X \to (X, A)$ be inclusion maps and let $x \in A$.

Then, $\pi_n(X) \cong \pi_n(A) \oplus \pi_n(X, A)$.

6.8 CO-CHAIN, CO-CYCLE, CO-BOUNDARY AND CO-HOMOLOGY

The concepts of the homology group and its related notions respond reasonably well to our intuition and geometric ideas and are essentially contained in the pioneering work of Poincare done towards the end of the nineteenth century. The concept of the dual of homology groups came much later through the hands of Alexander, Cech, Kolmogoroff and Whitney around 1935. Much of the algebra of homology and cohomology smacks of the theory of vector spaces. Just as the dual of a vector space means the space of linear functionals over the vector space, the cohomology group of a complex is also a group (in fact a ring) of functions rather homomorphisms from the group of chains to the coefficient group. For simplicity we take $\mathbf{Z}$ as the coefficient group here. (We can take any commutative ring as the coefficient group).

Let (K, α) be an oriented complex. Let $C^n(K)$ denote the free group generated by all homomorphisms $C^n : C^n(K) \to \mathbf{Z}$. Then $C^n(K)$ is called the group of *n-dimensional cochains* of the complex *K*. Thus an *n*-cochain is a homomorphism defined by its value on each *n*-simplex. Given an $(n-1)$ – cochain e^{n-1}, we define an *n*-cochain $\partial^{n-1} C^{n-1}$ by the rule $\partial^{n-1} C^{n-1}(\Delta_n) = C^{n-1}(\partial_n \Delta_n)$ where ∂_n is the boundary operator defined on $C_n(K)$. Thus ∂^{n-1} is the adjoint of ∂_n and raises dimension by 1. We now see that ∂^{n-1} is a homomorphism since

$$\begin{aligned}(\delta^{n-1} C_1^{n-1} + \delta^{n-1} C_2^{n-1})(\Delta_n) &= \delta^{n-1} C_1^{n-1}(\Delta_n) + \delta^{n-1} C_2^{n-1}(\Delta_n) \\ &= C_1^{n-1}(\partial_n \Delta_n) + C_2^{n-1}(\partial_n \Delta_n) \\ &= (C_1^{n-1} + C_2^{n-1})(\partial_n \Delta_n) \\ &= \delta^{n-1}(C_1^{n-1} + C_2^{n-1})(\Delta_n)\end{aligned}$$

Now, using the fact $\partial_{n-1} \partial_n = 0$ one can easily prove that $\delta^n \delta^{n-1} = 0$ i.e., $\delta^2 = 0$.

This homomorphion $\delta^n : C^n(K) \to C^{n-1}(K)$ is called the *co-boundary operator*. We thus have a sequence of groups of co-chains and associated co-boundary as follows:

$$C^o(K) \xrightarrow{\delta_0} C^1(K) \xrightarrow{\delta_1} C^2(K) \xrightarrow{\delta_2} \dots \xrightarrow{\delta_{n-1}} C^n(K) \xrightarrow{\delta_n} C^{n+1}(K) \xrightarrow{\delta_{n+1}} \dots$$

Let us now restrict our definition to the nth group of co-chains $C^n(K)$, and its two adjoining groups.

Definition: The kernel of the homomorphism δ^n which is a sub-group of $C^n(K)$ is called the *group of n-dimensional co-cycles* of the complex K and is denoted as $Z^n(K)$.

The image of the homomorphiom δ^{n-1} which is a (normal) sub-group of $C^n(K)$ is called the *group of n-dimensional co-boundaries* of the complex *K* and is denoted as $B^n(K)$.

The fact that $\delta^n \delta^{n-1} = 0$ proves that $B^n(K)$ is a (normal) sub-group of $Z^n(K)$ and hence one can define the quotient group $Z^n(K)/B^n(K)$.

The *nth cohomology group* $H^n(K)$ is defined to be the quotient group $Z^n(K)/B^n(K)$. Thus $H^n(K) = Z^n(K)/B^n(K)$.

6.9 CUP PRODUCT

The cohomology theory has an added advantage over homology theory because of an algebraic nature which is missing in homology theory. It is defacto the ring structure of the cohomology group *H* (*K*), the direct sum of $H^n(K)$ achieved through defining a multiplication type operation *U* called the *Cup product*.

Before defining the cup product we recall that coefficient group **Z** is infact a ring and hence multiplication is defined there. Here, the cup product is defined for two cochains, one *m*-cochain and the other an *n*-cochain and the result of the cup product gives an *m* + *n* cochain. Let $\Delta_{m+n} = (p_o \ p_1 \dots p_{m+n})$.

Define F_m and F_n as linear maps: $F_m : \Delta_{m+n} \to \Delta_m$ and $F_n = \Delta_{m+n} \to \Delta_n$, where F_m maps the vertices of Δ_m to the vertices $p_0, p_1, \dots p_m$ of Δ_{m+n} and *G* maps the vertices of Δ_n to the vertices P_0, P_1, ... P_{m+n} of Δ_{m+n}. Note $\Delta_m < \Delta_{m+n}$, $\Delta_n < \Delta_{m+n}$. Let $C^m \in C^m(K)$ and $C^n \in C^n(K)$.

Then define a cochain $C^m \cup C^m(K)$ as $(C^m \cup C^n)(\Delta_{m+n}) = C^m(F_m \Delta_{m+n}).C^n(F_n \Delta_{m+n})$ i.e., $(C^m \cup C^n)(p_o, p_1, \dots p_{m+n}) = C^m(p_0 \ p_1 \dots p_m).C^n(p_{m+1} \ p_{m+2} \dots p_{m+n})$.

The cochain $C^m \cup C^n$ being defined for any simplex can be lifted to any cochain of $C^{m+n}(K)$.

It is now routine to verify that this cup product of cochains of $C(K)$, i.e., $\sum C^n(K)$ is associative and distributive over addition and since **Z** has an identity element, $C(K)$ has also an identity element and thus $C(K)$ turns out to be a ring of cochains. Also the set $Z(K) = \sum Z^n(K)$, the direct sum of cocycles is a subring of $C(K)$ and the set $B(K) = \sum B^n(K)$, the direct sum of co-boundaries, is a two-sided ideal of $Z(K)$. Thus the quotient ring $Z(K)/B(K)$ is defined and is called the *cohomology ring* of K.

Remarks 1: It is an important result to note that if two topological spaces are homeomorphic, then their cohomology rings are isomorphic. Thus the ring structure is a topological invariant.

Remarks 2: The ring structure of the cohomology classes provides considerable freedom and flexibility in the computation of the products.

Example 1: If $\mathbf{P}^n$ denotes the real projective space then

$$H^o(\mathbf{P}^n) = \mathbf{Z} \text{ and } H^{2q}(\mathbf{P}^n) = Z_2, \quad 0 < 2q < n$$

$$H^{2q-1}(\mathbf{P}^n) = 0, \quad 0 < 2q-1 < n$$

$$H^n(\mathbf{P}^n) = \mathbf{Z} \text{ if } n \text{ is odd.}$$

Example 2: If **K** denotes the Klein's bottle, then $H^2(\mathbf{K})$ is non-trivial, though $H^2(\mathbf{K}) = 0$

Remark: There are surfaces for which the homology and cohomology groups are equal.

Example 3: For $n > 0$, $H^k(S^n, G) = G$ for $k = 0$ and $k = n$

Example 4: If **T** denotes the torus, then $H^1(\mathbf{Y}) = \mathbf{Z}_2$ and $H^2(\mathbf{T}) = \mathbf{Z}$.

CHAPTER 7

Singular Homology Theory

7.1 SINGULAR HOMOLOGY GROUP

One of the greatest successes of the combinatorial topology has been the extension of Homology Theory to general topological spaces. In what discussed above it is clear that Homology groups can be defined for a special kind of space, namely, compact polyhedron and the complexes obtained there were finite althrough. Singular Homology theory extends the notion of Homology groups for general topological spaces by associating with each space a chain complex. A continuous map induces homology homomorphisms in an obvious way and as a consequence it follows that homotopic maps induce the same homomorphisms. There is a natural homomorphism also from homology groups to singular homology groups. In the following lines we give a sketch of notions relevant in singular homology theory.

Let X be a topological space and G be an additive abelian group.

Definition: A *singular p-simplex* on a topological space X is a continuous map σ of the standard Euclidean p-simplex Δ_p into X.

Note that a singular p-simplex is a map and not a set.

Definition: A *singular n-chain* on a topological space X over an additive abelian group G is an element of the free (abelian) group generated by all singular n-simplices and is denoted as $C_n(K, G)$.

Let the symbol $(x_0, x_1, ..., x_n)$ denote the linear map mapping the vertices $p_o, p_1, ... p_n$ of Δ_n to $x_0, x_1, ..., x_n$ respectively of X.

Then the boundary operator d_n acting on the singular n-simplex $(x_o, x_1, ..., x_n)$ is defined as $d_n(x_o, x_1, ..., x_n) = \sum_{i=0}^{n} (-1)^i (x_o, x_1, ..., x_i ..., x_n)$ where the circumflex x_i denotes the omission of the marked vertex and hence carries the usual meaning as before.

Now, if $\alpha = \sum g_i \sigma_i$ be a n-chain on a topological space X, where σ_i are singular n-simplices and g; are elements of the coefficient group, then the boundary of d_α is defined as $\sum g_i \; d_n \; \sigma_i$.

Thus, we can define the *boundary operator* ∂ on the free abelian group $S_n(X)$ generated by all singular n-chains as follows:

$$\partial_n(\alpha) = \partial_n(\sum g_i \; \sigma_i) = g_i \sum \partial_n(\sigma_i)$$

It is easy to prove that $\partial_{n-1} \partial_n = 0$ i.e., $\partial^2 = 0$. Noting that $S_n(X)$ is analogous to the simplicial complex $C_n(K)$, we see that the graded complex $S(X) = \{S_n(X)\}$.

Written in a sequence, it looks like

$$\cdots \xrightarrow{\partial_{n+2}} S_{n+1}(X) \xrightarrow{\partial_{n+1}} S_n(X) \xrightarrow{\partial_n} S_{n-1}(X) \xrightarrow{\partial_{n-1}}$$

The image of ∂_{n+1} denoted by $B_n(X)$, is then called the *subgroup of singular n-boundaries* of $S_n(X)$ and the kernel of ∂_n, denoted by $Z_n(X)$, is called the *subgroup of singular n-cycles* of $S_n(X)$.

The result $\partial_{n-1}\,\partial_n = 0$ then proves that $Z_n(X)$ is a normal subgroup of $B_n(X)$. Hence, we can define the quotient group $B_n(X)/Z_n(X)$, called the *singular nth homology group* of the topological space *X*. This quotient group is denoted by $H_n(X)$ or $H_n(X, G)$ where *G* is the coefficient group.

Thus, $H_n(X)$ consists of equivalence classes of singular *n*-cycles under the usual equivalence relation $C \sim C'$ if and only if $C - C' \in Z_n(X)$ where $C, C' \in B_n(X)$. Such cycles are called *homologous cycles*.

From above the following result is obvious.

Theorem 7.1.1: If *X* is a one point space, then

$$H_n(X) \approx \begin{cases} G & \text{if } n = 0 \\ 0 & \text{if } n \neq 0 \end{cases}$$

Proof: Easy:

Another simple but interesting result is

Theorem 7.1.2: If *X* is path-connected, then $H_o(G) \approx G$

Proof: Routine and can be seen in Dold or Spanier

To see now whether homeomorphisms of topological spaces lead to any algebraic relation between the corresponding homology groups, we observe the following:

Let $f : X \to Y$ be a continuous map from a topological space *X* to another topological space *Y*.

Define a map $f*$ from $S_n(X)$ to $S_n(Y)$ as follows: $f*(\Sigma\, m_i\, \sigma_i) = \Sigma\, m_i\; f\;(\sigma_i)$.

This proves $f*(Z_n(X)) \subset Z_n(Y)$ and $f*(B_n(X)) \subset B_n(Y)$.

Thus, *f* is a homomorphism of groups from $H_n(X)$ to $H_n(Y)$ defined by

$$f*(\Sigma\, m_i\, \sigma_i) = \Sigma\, m_i\; f\;(\sigma_i).$$

This homomorphism is known as the *induced homomorphism*.

The following theorem is of great importance.

Theorem 7.1.3: Two homeomorphic topological spaces have isomorphic homology groups but not conversely.

Proof: Easy to observe that if $f : X \to Y$ is a homeomorphism, then the induced homomorphism is actually an isomorphism.

Thus in a sense homology is a functor from topology to algebra, more precisely, from the category of topological spaces and continuous functions to the category of groups and homomorphisms.

In fact, the definition of homology groups can be extended in a natural way to topological pairs and hence the homology can be thought of as a functor from the category of topological pairs and continuous functions to the category of groups and homomorphisms.

In this connection one should note that if two continuous maps *f* and *g* from *X* to *Y* are homotopic then the induced maps are equal. This gives us the following version of Hurewicz theorem for singular homology theory.

Hurewicz Theorem: For a path connected space *X*,

$$\pi\ (X, x_0) \approx H_1\ (X) \text{ where } x_o \in X.$$

Proof: See Dold or Spanier.

From above it is thus clear that the calculation of fundamental groups of path-connected spaces can be done with the homology groups.

From the above result one immediately derives.

Theorem 7.1.4: Two surfaces S_1 and S_2 are homeomorphic if and only if $H_1(S_1) \approx H_1(S_2)$.

7.2 MAYER VIETORIS SEQUENCE

Just as Van Kampen's theorem in homotopy theory gives a useful way of computting fundamental groups of topological spaces in particular surfaces, the following theorem gives a convenient method of computing homology groups of many spaces.

Mayer Vietoris Theorem: If $X = U_1 \cup U_2$ is a topological space with two open subsets U_1 and U_2 then there are homomorphisms $\phi_n : H_n\ (X) \to H_{n-1}\ (U_1 \cap U_2)$ such that the following sequence is exact.

$$\ldots H_{n+1}(X) \xrightarrow{\phi n+1} H_n(U_1 \cap U_2) \xrightarrow{i} H_n(U_i) \oplus H_n(U_2) \xrightarrow{j} H_n(X) \xrightarrow{\phi n} H_{n-1} (U_1 \cup U_2) \to$$

If further, $Y = V_1 \cup V_2$, V_1 and V_2 open subsets of *Y* and $f : X \to Y$ is continuous with $f(U_i) \subset V_i$, then $(f \mid U_1 \cap U_2) * \phi = \phi\ f*$

Proof: See Dold [8].

The exact sequence referred in the above theorem is called the *Mayer-Vietoris sequence* and plays a significant role in the computational problems. The homomorphisms $\{\phi_n\}$ are called connecting homomorphisms.

As for example, using Mayer-Vietoris sequence one can easily find out the homology group of the Real Projective Plane.

The following result follows easily from above.

Theorem 7.2.1: $H_p(\mathbf{S}^n) \approx \begin{cases} \mathbf{Z} & \text{if } p = 0, n \\ 0 & \text{if otherwise.} \end{cases}$

Proof: Use Mayer Vietoris sequence inductively with $U_1 = \{x \in \mathbf{S}^n ; x_n < -\tfrac{1}{2}\}$ and

$$U_1 = \{x \in \mathbf{S}^n ; x_n > \tfrac{1}{2}\}$$

7.3 SINGULAR COHOMOLOGY

Just as the simplicial cohomology group is the direct sum of a family of graded groups, each of which is a free group consisting of all homomorphisms defined on the simplicial homology groups with values in a field *G*, usually, the set of real numbers or complex numbers, so is the singular cohomology group of a topological space.

Thus, if $H_n(X)$ denotes the singular n^{th} homology group of X, then $H(X) = \oplus H_n(X)$ is the direct sum of the graded homology groups $\{H_n(X)\}$. One can now define in a way analogous to that for simplicial cohomology theory, cohomology groups $H^n(X)$ for each $H_n(X)$.

To be little more precise, let $S_n(X)$ denote the singular n^{th} homology complex and $S_n(X)$ denote the free group generated by all homomorphisms from $S(X)$ to the field G, usually taken as C. Then $S^n(X)$ is called the *singular n^{th} co-complex*.

If $\delta^n : S^n(X) \to S^{n+1}(X)$ is defined as $\delta^n C^n(\sigma_n) = C^{n+1}(\partial_n \sigma_n)$ where σ_n is a singular n-simplex, $C^n \in S^n(X)$ and ∂_n is the boundary operator on $S_n(X)$, then δ^n can be extended over $S^n(X)$. This map δ^n is called the *singular coboundary operator*.

From the fact $\delta_{n-1}\,\delta_n = 0$ follows that $\delta^n\,\delta^{n-1} = 0$

Thus, one gets graded groups $\{S^n(X)\}$ indexed by N usually represented as

$$\ldots S^{n-1}(X) \xrightarrow{\delta_{n-1}} S^n(X) \xrightarrow{\delta_n} S^{n+1}(X) \longrightarrow \ldots$$

By virtue of $\delta^2 = 0$ it is easy to prove now that the image δ^{n-1} of usually denoted by $B^n(X)$ and called the *group of singular n-coboundaries* contains the kernel of δ^n, usually denoted by $Z^n(X)$ and called the group of singular n-cocycles, as a normal subgroup and hence $B^n(X)/Z^n(X)$ is a group. The family of all such cohomology groups make up a graded cohomology group $H(X) = \{H^n(X)\}$.

One can define as in the case of simplicial cohomology theory, the cup product on the direct sum of the graded cohomology group, denoted by $H(X)$. This structure available on the cohomology group $H(X)$ and not on the homology group makes cohomology theory functionally more efficient in the classification of topological spaces.

7.4 AXIOMATIZATION OF HOMOLOGY THEORY

The great success of singular homology theory as a generalization of simplicial homology theory in the classification problem prompted many to devise very many ways of generalizing the classical simplicial theory. Čech homology theory is one such generalization which achieved also remarkable success but surely not as elegantly as the singular theory. Further that all these theories coincide on a large class of spaces e.g., CW complexes perhaps influenced Eilenberg and Steenrod to study homology theory through an axiomatic definition of homology theory.

Definition: A *homology theory* (H, ∂) consists of (a) a covariant functor H from the category of topological pairs to the category of graded abelian groups and homomorphisms of degree 0, i.e., $H(X, A) = \{H_n(X, A)$, (b) a natural transformation ∂ of degree–1 from the functor H on (X, A) to the functor H on (A, ϕ), i.e., $\partial(X, A) = \{\partial n(X, A); H_n(X, A) \to H_{n-1}(A)\}$ such that the following axioms are satisfied.

Homotopy Axiom: If $f_0, f_1 : (X, A) \to (Y, B)$ are homotopic, then

$$H(f_0) = H(f_1) : H(X, A) \to H(Y, B),$$

Exactness Axiom: For any pair (X, A) with inclusion map $I : A \subset X$ and $j : X \subset (X, A)$, there is an exact sequence

$$\ldots \xrightarrow{\partial n+1} H_n(A) \xrightarrow{H n(\mathrm{i})} H_n(X) \xrightarrow{H n(\mathrm{j})} H_n(X, A) \xrightarrow{\partial n} H_{n-1}(A) \xrightarrow{H n-1(\mathrm{i})}$$

Excision Axiom: For any pair (X, A) if A is an open subset of X such that $U \subset A$, then the excision map $j: (X - U, A - U) \subset (X, A)$ induces an isomorphism $H_j: H(X - U, A - A) \approx H(X, A)$.

Dimension Axiom: For a one-point space X, the following must be true:

$$H_n(X) \approx \begin{cases} \mathbf{Z} & \text{if } n = 0 \\ 0 & \text{if } n \neq 0 \end{cases}$$

Singular cohomology theory with coefficients in G satisfies all the above conditions and hence is an example of a cohomology theory in the axiomatic approach also.

It is not difficult to prove using Mayer Vietories sequence that singular homology theory satisfies all the requirements of Eilenberg Steenrod and hence is a homology theory in the axiomatic approach.

The corresponding cohomology theory has been defined axiomatically by Eilenberg and Steenrod as follows:

Definition: A *cohomology theory* (H^*, ∂^*) with coefficients in G consists of (a) a contravariant functor H^* from the category of topological pairs the category of graded R-modules and (b) a natural transformation $\partial^*: H^*(A) \to H^*(X, A)$ of degree 1 such that the following axioms are satisfied.

Homotopy Axiom: If $f_o, f_1(X, A) \to (Y, B)$ are homotopic, then

$$H^*(f_o) = H(f_1): H^*(Y, B) \to H^*(X, A),$$

Exactness Axiom: For any pair (X, A) with the inclusion maps $i: A \subset X$ and $j: X \subset (X, A)$, there exists an exact sequence.

$$\ldots \xrightarrow{\partial *} H^n(X, A) \xrightarrow{H^n(j)} H^n(X) \xrightarrow{H^n(i)} H^n(A) \xrightarrow{\partial *} H^{n+1}(X, A) \longrightarrow \ldots$$

Excision Axiom: For any pair (X, A) if U is open in X such that $U \subset A^o$, then the excision map $j: (X - U, A - U) \subset (X, A)$ induces an isomorphism.

$$H^*(j): H^*(X, A) \approx (X - U, A - U)$$

Dimension Theorem: For a one-point space X, the following must be true:

$$H^n(X) \approx \begin{cases} \mathbf{Z} & \text{if } n = 0 \\ 0 & \text{if } n \neq 0 \end{cases}$$

Singular cohomology theory with coefficients in G satisfies all the above conditions and hence is an example of a cohomology theory in the axiomatic approach.

7.5 DUALITY THEOREMS

The observation that the isomorphism of the homology groups implies the isomorphism of the cohomology groups leads to the question whether there is any relation between the homology groups and the cohomology groups. We shall see that there do not exist any such connection for the special types of very general topological spaces like manifolds and CW complexes.

The following is one such result stated without proof:

Poincare Duality Theorem: If X is an orientable compact triangulable homology n-manifold, then for all positive integer $p \leq n$, $H^p(X, G) \approx H_{n-p}(X, G)$ where G is an arbitrary coefficient group.

The corresponding result for relative manifold is known as Lefschetz Duality Theorem.

Lefschetz Duality Theorem: If (X, A) is an orientable relative n-manifold, then for any positive integer $p \le n$, we have

$$H^p(X, A; G) \approx H_{n-p}(|X| - |A|; G)$$

and $$H_p(X, A; G) \approx H^{n-p}(|X| - |A|; G) \text{ for all } G.$$

7.6 ČECH THEORY

Simplicial homology theory defined homology groups for complexes which are made of simplices and the definition could be easily extended to the underlying polyhedrons which are essentially subspaces of the Euclidean space $\mathbf{R}^n$ with relativized topology. Singular homology theory extended the idea to arbitrary topological spaces, thus enlarging the scope of classifying general topological spaces by means of homology groups. As mentioned earlier this is not the only generalization of the simplicial theory. In fact there are several generalizations, the most notable among them being the Čech theory. This theory turns out to be very satisfactory in the case of cohomology. So in what follows we discuss Čech's approach to cohomology theory.

We begin with the recapitulation of some basic notions.

Definition: Let $\boldsymbol{A}$ be a collection of subsets of a topological space X. The *nerve* of $\boldsymbol{A}$ denoted by $N(\boldsymbol{A})$ is then defined as the abstract simplicial complex made of simplices whose vertices are elements of $\boldsymbol{A}$ and the simplices are finite subcollections $\{A_1, A_2, ..., A_n\}$ of such that $A_1 \cap A_2 \cap ... \cap A_n \neq \phi$.

Note that the definition is usually given for a cover of X. So, if $\boldsymbol{B}$ is a refinement of $\boldsymbol{A}$, the vertex map $g: \boldsymbol{B} \to \boldsymbol{A}$ defined by choosing $g(\boldsymbol{B})$ to be an element of $\boldsymbol{A}$ that contains $\boldsymbol{B}$ induces a simplicial map.

$$\bar{g}: N(\boldsymbol{B}) \to N(\boldsymbol{A}).$$

Now the Čech cohomology theory is defined as follows:

Let Δ be a directed set consisting of all open covers of a topological space X where the order relation < is defined as $\boldsymbol{A} < \boldsymbol{B}$ if $\boldsymbol{B}$ is a refinement of $\boldsymbol{A}$ for $\boldsymbol{A}, \boldsymbol{B} \in \Delta$. We then construct a direct system of abelian groups and homomorphisms as follows:

Assign to each $\boldsymbol{A} \in \Delta$ the group $H^n(N(\boldsymbol{A}); G)$ so that if $\boldsymbol{A} < \boldsymbol{B}$, then the homomorphism $f_{AB}: H^n(N(\boldsymbol{A}); G) \to H^n(N(\boldsymbol{B}); G)$ induced by refinement, satisfies the conditions:

(*i*) $f_{AA}: H^n(N(\boldsymbol{A}); G) \to H^n(N(\boldsymbol{B}); G)$ is the identity,

(*ii*) if $\boldsymbol{A} < \boldsymbol{B} < \boldsymbol{C}$, then $f_{BC} \circ f_{AB} = f_{AC}$.

The direct limit $\lim\limits_{A \in \Delta} (N(A); G)$ is then called the *nth Čech cohomology group* of X.

One can similarly define the reduced Čech cohomology group. By this abstract generalization how much is gained is a natural question. The following theorem answers partly this significant question.

Theorem 7.6.1: For a simplicial complex K, the Čech cohomology groups and the simplicial homology groups are isomorphic.

But this result should not convince one to believe that this situation prevails almost everywhere in homology theory. The following example suggests that the singular cohomology groups and the Čech cohomology group may be different also for some topological spaces.

Example: If X is the closed topologist's sine curve, then

$$H^1(X) = 0 \text{ for singular cohomology}$$

But $$H^1(X) = \boldsymbol{Z} \text{ for Čech cohomology}$$

We conclude this discussion by referring to a duality theorem establishing a connection between Čech cohomology and singular homology groups.

Alexander Pontryagin Duality Theorem

If $\phi \neq A \subset \mathbf{S}^n$ is closed, then

$$\underset{\text{(Cech cohomology)}}{H^p(A)} \approx \underset{\text{(Singular cohomology)}}{H_{n-p-1}(\mathbf{S}^n - A)}$$

Note that the result fails if Čech cohomology is replaced by singular cohomology. The above result evinces the versatality of homology theory in general.

CHAPTER 8

Manifold Analysis

The first impulse to generalize differential and integral calculus came from the attempts to define the related concepts over a general topological space than on the real line or the complex plane. But for achieving a reasonable success the objective was brought down to a manifold defined and studied in the following lines.

8.1 SOME DEFINITIONS

We begin with the notion of a manifold.

Definition: A *real n-manifold* is a topological space M every point of which has a neighbourhood homeomorphic to an open subset of $\mathbf{R}^n$.

To achieve greater success we shall often endow M with one or two additional structures, viz.,

(*a*) M is Hausdorff

(*b*) M has a countable basis, i.e., M is second countable.

(*c*) M is paracompact.

We shall generally assume (*a*) but occasionally (*b*) or (*c*) with specific mention.

Sometimes an *n*-manifold will be referred to as a manifold only or topological manifold or abstract manifold.

A *complex n-manifold M* is a topological space which has a covering by neighbourhoods each homeomorphic to an open subset of $\mathbf{C}^n$.

Here also we shall generally assume M to be Hausdorff and paracompact.

A real manifold with boundary is defined as follows:

Definition: A real n-manifold with boundary is a topological space each point of which has a neighbourhood homeomorphic to an open subset of $\mathbf{R}^n$ or to the set $\{x \in \mathbf{R}^n;\, x_i \geq 0 \text{ for some } i\}$.

Example 1: The real line $\mathbf{R}$ is a real manifold.

Example 2: The Euclidean space $\mathbf{R}^n$ is an n-manifold.

Example 3: The unitary space $\mathbf{C}^n$ is a complex n-manifold.

Example 4: A closed interval $[a, b]$ is a 1-manifold with boundary.

Example 5: An open cube such as $(0, 1) \times (0, 1) \times \ldots \times (0,1)$ (*n* times) is an n-manifold but the closed cube $[0, 1]^n$ is an n-manifold with boundary.

Example 6: The open disc $\{z \in \mathbf{C}; |z| < 1\}$ is a complex 1-manifold and the product of *n* such discs is a complex n-manifold.

Example 7: The mobius band is a real 2-manifold with boundary.

Example 8: The torus is a real 2-manifold.

Example 9: The 1-sphere $\mathbf{S}^1$ (circle) is a real 1-manifold and the 2-sphere $\mathbf{S}^2$ is a real 2-manifold.

Example 10: The projective space $\mathbf{P}^n$ is a real n-manifold.

Example 11: A paraboloid, an ellipsoid, a hyperboloid are examples of real 2-manifolds. A hyperquadric in $\mathbf{R}^{n+1}$ given by $x^2 + x^2 + ... + x_n^2 = a^2$ defines a real n-manifold.

Example 12: Any open subset of $\mathbf{R}^n$ is a real n-manifold.

In fact the following is true:

Proposition 8.1.1: Every open subset of an n-manifold is an n-manifold.

Proof: Obvious.

Proposition 8.1.2: The product of finitely many manifolds is also a manifold.

Proof: Easy.

Definition: Let M be an n-manifold and let U be an open subset of M. Then U is homeomorphic to an open subset of $\mathbf{R}^n$ under some map ϕ. The map ϕ is called a *coordinate map*, the function $\phi_{io}\phi = x_i$ are called *coordinate functions* and the pair (U, ϕ) is called a *coordinate system.*

A coordinate system (U, ϕ) is called a *cubic coordinate system* if $\phi(U)$ is an open cube about the origin in $\mathbf{R}^n$. If $m \in U$ and $\phi(m) = 0$, then the coordinate system is said to be *centered at m*.

The number n is called the *dimension* of the manifold M if $\{U_\alpha\}$ covers M. If the corresponding coordinate maps are denoted by ϕ_α, the family $\{\phi_\alpha\}$ is called the *coordinate neighbourhood* of M.

Definition: An n-manifold M is said to *differentiable of class* C^k if

(i) M is a Hausdorff space,

(ii) For any two neighbourhoods U and V of a point m, the corresponding local coordinates of m in $\phi(U)$ and $\phi(V)$ are connected by a homeomorphism which is differentiable (analytic in the case of complex manifold), i.e., $\phi_{Uo}\phi_V^{-1}$ is C^k (analytic).

An n-manifold is called *differentiable* if it is C^k for all $k \geq 0$, i.e., $\phi_{Uo}\phi_V^{-1}$ is C^∞ for every pair of open sets U and V of M.

It is usually worthwhile taking M to be second countable for many reasons to be clear in course of time.

Example 1: The general linear group $GL(n, \mathbf{R})$ of all $n \times n$ non-singular matrices with real entries is a differentiable manifold. This becomes obvious if we identify the points of $\mathbf{R}^{n2}$ with the $n \times n$ matrices of $GL(n, \mathbf{R})$. Then the determinant of a matrix becomes a continuous function from $GL(n, \mathbf{R})$. i.e., $\mathbf{R}^{n2}$ to $\mathbf{R}$.

For functions of several real or complex variables we shall generally take manifolds which are themselves subsets of $\mathbf{R}^n$ or $\mathbf{C}^n$. In such situations one observes that f is $\mathbf{C}^k$ if $\partial^i/\partial x_I$ exists and is continuous on an open subset of the manifold. Thus f is C^o if f is continuous.

The following results are almost obvious.

Proposition 8.1.3: Every open subset of a differentiable manifold is a differentiable manifold.

Proposition 8.1.4: If M_1 and M_2 are two differentiable manifolds of dimensions respectively n_1 and n_2, then the topological product $M_1 \times M_2$ is a differentiable manifold of dimension $n_1 + n_2$.

8.2 GERMS OF A FUNCTION

We define this concept in a round about way. Let M be a complex manifold.

Definition: A function $f: M \to \mathbf{C}$ is said to have the *same germ* as $g: M \to \mathbf{C}$ if there exists a neighbourhood U of $m \in M$ such that $f(x) = g(x)$ for every $x \in U$.

Now, we can define an equivalence relation for all complex valued functions on M as follows: $f \sim g$ if f has the same germ as g. The equivalence classes of ~ are called the *germs* of M. The germ corresponding to f will be denoted by **f**.

Note a germ **f** has a well-defined value at $m \in M$ given by $f(m)$.

Let F_m denote the set of all germs at a particular point m of M. Then the following is true.

Proposition 8.2.1: F_m is a vector space.

Proof: In fact the natural addition and scalar multiplication are given by

$$(\mathbf{f} + \mathbf{g})(m) = f(m) + g(m)$$

$$\alpha\mathbf{f}(m) = \alpha f(m)$$

The verification is a routine check.

In fact some thing more is true. F_m is an algebra, if we define the product of two germs as follows:

$$(\mathbf{f} \cdot \mathbf{g})(m) = f(m)\, g(m)$$

It is easy to note the set of all germs vanishing at m is a two-sided ideal of F_m.

8.3 SHEAFS

The sort of relationships between holomorphic functions and their respective germs lead generally to the study of sheafs.

Definition: A *sheaf* of abelian groups over a topological space M is a topological space S together with a mapping $\pi: S \to M$ such that the following conditions are fulfilled:

(*i*) π is a local homeomorphism,

(*ii*) For each $x \in M$, the set $\pi^{-1}(x)$, called the *stalk over x*, has the structure of an abelian group.

(*iii*) The group operations are continuous in the topology of S.

The map π is usually referred to as the projection map. A sheaf S over a topological M together with the projection π will be denoted by (S, π, M). A stalk of S over x will be sometimes denoted by S_x.

A topological space T is called a *subsheaf* of the sheaf (S, π, M) if

(*i*) T is open in S

(*ii*) $\pi(T) = M$

(*iii*) For each point $x \in M$, the stalk T_x is a subgroup of S_x.

A *section* of the sheaf S over U is a continuous mapping $f: U \to S$ such that $\pi_o f = \pi/U$.

Remark: The above definition of sheaf defined for abelian groups can be extended to sheafs of rings or modules or algebras in a natural way.

The following result is immediate from the definition.

Proposition 8.3.1: The set $\Gamma(U, S)$ of all sections over U is an abelian group.

The proof of the above is obvious if the operations are defined as

$$(f+g)(x) = f(x) + g(x)$$

$$(-f)(x) = -f(x) \text{ for } f, g \in \Gamma(U, S)$$

and the zero of the group is the zero section which assigns the zero of the stalk $\pi^{-1}(x)$ to every $x \in U$.

It is easy to observe that if V is a subset of U, there is a homomorphism $\rho_{VU} : \Gamma(U, S) \to \Gamma(V, S)$ defined by the restriction.

These observations lead to the following notions.

Defintion: A *presheaf* of abelian groups over M consists of

(*i*) a basis of open sets of M,

(*ii*) an abelian group S_U assigned to each open set U of the basis.

(*iii*) a homomorphism $\rho_{VU} : S_U \to S_V$ associated to each inclusion $V \subset U$ such that $\rho WV_o \rho VU = \rho WU$ whenever $W \subset V \subset U$.

There is a natural construction which associates to every presheaf over M a sheaf S over the same space M. For this suppose a presheaf is given. For every point $x \in M$, consider the family U_x of those open sets U of the basis of M such that U contains the point x. Then U_x is a partially ordered family under the inclusion relation. Let the direct limit group of S_U be denoted by S_x, i.e., $S_x = \lim_{U \in Ux} S_U$.

It now follows by straightforward verification that $S = \cup S_x$ is a sheaf of abelian groups with the projection $\pi : S \to M$ defined by $\pi(S_x) = x$.

In fact $S_x = \pi^{-1}(x), x \in M$ are the stalks which are abelian groups. The topology of S is defined as follows:

To any element $f \in S_U$ associate the point set

$$\rho(f) = \bigcup_{x \in U} \rho_{xU}(f) \subset S$$

where $\rho_{xU}(f)$ is the equivalence class in $S_x^* = \bigcup_{x \in Ux} S_U$, the parent set of the direct limit.

[Here two elements $f_U \in S_U \subset S_x^*$ and $f_V \in S_V \subset S_x^*$ are equivalent if there is a set $W \in U_x$ such that $W \subset U \cap V$ and $\rho_{WU}(f_U) = \rho_{WV}(g_V)$]. The set $\rho(f)$ is a basis of a topology under which the projection mapping $\pi : S \to M$ is a local homeomorphism. Further the group operation is continuous as $\rho(f) - \rho(g) = \rho(f - g)$. Thus S is a sheaf.

We now give some simple examples of sheafs.

Example 1: Let M be an open domain (i.e., a connected open set) in $\mathbf{C}^n$. To each open set U of M we associate a ring θ_U of holomorphic functions in U. If $U \subset V$ are two open sets and $f \in \theta_V$, then the restriction of f to U is off course an element of $\rho_{VU}(f)$ and when $U \subset V \subset W$, it is clear that

$\rho_{UV}\rho_{VW} = \rho_{UW}$. Thus, the collection of rings θ_U together with the restriction mappings ρ_{UV} forms a presheaf of holomorphic functions over *D*. The sheaf associated with this sheaf is known as the sheaf of germs of holomorphic functions over *M* and is usually denoted by θ.

Example 2: Let *M* be an open domain of $\mathbf{C}^n$. To each open subset *U* of *M* associate the ring C_U of continuous functions over *U*. The ring together with the natural restriction mappings defines a presheaf over *M* and the sheaf obtained as the direct limit of the presheafs is the sheaf *C* of germs of continuous functions on *M*. If in particular one takes C_U the ring of constant functions, the corresponding sheaf is known as the constant sheaf.

Example 3: If *M* is an open domain in $\mathbf{C}^n$ and $U \subset M$ be open, then the sheafs associated with the ring $C_U{}^{(r)}$ of r-times differentiable functions of the underlying 2n real variables and also the ring $C_U{}^{(\infty)}$ of infinitely differentiable functions as in example 2 are called the sheafs of r-differentiable functions of the real coordinates on *M* and the sheaf of infinitely differentiable functions on *M* respectively.

Example 4: One can in an exactly the same manner introduce the sheaf a_{pq} of germs of complex valued $\mathbf{C}^\infty$-forms of the type (*p*, *q*). In particular, we will write $a = a_{oo}$ to denote the sheaf of germs of complex valued C^∞ functions.

Example 5: The sheaf C_{pq} of germs of complex valued C^∞-forms of the type (*p*, *q*) which are closed under $\bar{\partial}$ plays a fundamental role in manifold analysis. We write $\theta = C_{OO}$, the sheaf of germs of holomorphic functions.

Example 6: The θ^* of germs of holomorphic functions which vanish nowhere has group operation defined on each stalk by the multiplication of germs of holomorphic functions.

Definition: Let *M* be a complex manifold. Let A_{pq} denote the module of all complex valued C^∞-forms of the type (*p*, *q*) over the ring of complex valued C^∞ functions. Then, a form $a \in A_{pq}$ is called $\bar{\partial}$-closed if $\bar{\partial}a = 0.$ Let C_{pq} denote the space of $\bar{\partial}$-closed forms of the type (*p*, *q*). Then the quotient group $D_{pq}(M) = C_{pq}/\bar{\partial}A_{pq-1}$ is called a *Dolbeault group* of *M*.

If we start with a real manifold *M* and if A_r denotes the space of all real valued C^∞ forms of degree *r* and C_r be the subspace of the forms of A_r which are annihilated by *d*, i.e., $da = 0$ for $a \in A_r$, then the quotient group C_r/dA_{r-1} is called a *de Rham group* of *M* and is denoted by $R_r(M)$.

The following is a fundamental theorem in manifold analysis.

Dolbeault Grothendieck Lemma: Let *D* be the polydisc $|z_i| < r_i, 1 \le i \le m \text{ in } C^m$ and D' be the smaller polydisc $|z_i| < r_i' < r_i$. Let *a* be a form of type $(p, q), q \ge 1$ in *D* such that $\bar{\partial}a = 0$. Then, there exists a form β of type (*p*, *q* –1) in *D* such that $\bar{\partial}\beta = a$ in D'.

Definition: Let $\pi : S \to M \text{ and } \tau : S \to M$ be two sheaves of abelian groups over the same space *M*. A *sheaf mapping* $\phi: S \to T$ is a continuous mapping such that $\pi = \tau_o \phi$, i.e., a mapping which preserves the stalks $\phi(\pi^{-1}(x)) \subset \tau^{-1}(x)$. The mapping ϕ is called a *sheaf homomorphism* if its restriction to every stalk is a homomorphism of groups.

If $\psi: N \to M$ is a third sheaf over *M*, the sequence of sheaves

$$0 \to S^{i} \to T^{\phi} \to N \to 0$$

connected by homomorphisms is called an *exact sequence* if at each stage the kernel of one homomorphism is identical with the image of the preceding homomorphism. Such a situation we usually describe by saying that S is a subsheaf of T and N is the quotient sheaf of T by S.

It follows from the Dolbeault Grothendieck lemma that the sequence

$$0 \to C_{pq} \xrightarrow{i} a_{pq} \xrightarrow{\bar{\partial}} C_{p\,q+1} \to 0$$

is exact. Here i is the inclusion homomorphism and $\bar{\partial}$ is the homomorphism on sheaves induced by the $\bar{\partial}$-operator. The Dolbeault Grothendieck lemma says that $\bar{\partial}$-operator is onto and the exactness of the sequence at the other stages is obvious.

8.4 COHOMOLOGY WITH COEFFICIENT SHEAF

Let M be a paracompact Hausdorff space. Let $U = \{U_i\}$ be a locally finite open covering of M. The nerve $N(U)$ of the covering U is a simplicial complex whose vertices are the members U_i of the covering such that $U_{io}, U_{i1}, ..., U_{iq}$ span a q-dimensional simplex if and only if the inetersection $U_{io} \cap U_{i1} \cap U_{i2} \cap ... \cap U_{iq} \neq \phi$. Let $\varphi: S \to M$ be a sheaf of abelian groups over M.

A *q-chain* of $N(U)$ with coefficients in the sheaf S is a function f which associates to each q-simplex $\sigma = (U_{io}, U_{i1}, ..., U_{iq}) \in N(U)$ a section $f(\sigma) \in \Gamma(U_{io} \cap U_{i1} \cap U_{i2} \cap ... \cap U_{iq}, S)$.

A *q-coboundary operator*

$$\delta^q : C^q\ (N(U), S) \to C^{q+1}\ (N(U), S)$$

is defined as follows:

For $f \in C^q\ (N(U), S)$ and $\sigma = (U_{io}, U_{i1}, ..., U_{iq})$, define $\delta^q f \in C^{q+1}\ (N(U), S)$ as

$$(\delta^q f)(\sigma) = \sum_{i=o}^{q+1} (-1)^i\ \rho_\sigma f(U_o, ..., U_{i-1}, U_{i+1}, ..., U_{q+1})$$

where ρ_σ denotes the restriction of the sections to the open set $U_o \cap U_1 \cap U_2 \cap ... \cap U_{q+1}$.

It is straightforward to verify that

$$\delta^{q+1}\delta^q = 0,\ q \geq 0, \text{ i.e., } \delta^2 = 0.$$

The kernel of δ^q is called the *group of q-cocycles* and is denoted by $Z^q(N(U), S)$ and the image of δ^{q+1} is called the *group of q-coboundaries* and is denoted by $B^q(N(U), S)$.

As a consequence of $\delta^2 = 0$ one gets

$$B^q(N(U), S) \subset Z^q(N(U), S)$$

i.e., every q-coboundary is a q-cocycle. Hence, one defines the quotient group

$$H^q(N(U), S) = Z^q(N(U), S)/B^q(N(U), S),\ B^o = 0.$$

This group is called the *qth cohomology group* of the nerve $N(U)$ with the coefficient sheaf S.

The zeroth cohomology group has the simple interpretation.

$$H^o(N(U), S) = \Gamma(M, S).$$

By a standard process initiated by Čech, one can now pass from the cohomology group $H^q(N(U), S)$ relative to all the locally finite open coverings U of M to the cohomology group $H^q(M, S)$, $q \geq 0$ of the space M itself.

Let $\pi: S \to M$ be a sheaf of abelian groups over M and let $U = \{U_i\}$ be a locally finite open covering of M. A partition of unity of the sheaf S subordinate to the covering U is a collection of sheaf homomorphisms $\eta_i: S \to S$ with the properties:

(i) η_i is the zero map in the open neighbourhood of $M - U_i$,

(ii) $\sum \eta_i = 1$, the identity mapping of S.

A sheaf S of abelian groups is called *fine* if it admits of a partition of unity subordinate to any locally finite open covering.

An example of a fine sheaf is a_{pq}.

Examples of sheaves which are not fine are

(i) the constant sheaf,

(ii) the sheaf C_{pq}.

Fine sheaves play a catalytic role in cohomology theory of sheaves because of the following theorem:

Theorem 8.4.1: If S is a fine sheave, then $H^q(N(U), S) = 0$, for $q \geq 0$.

Definition: A *sheaf* homomorphism $I: S \to T$ induces a homomorphism $\Gamma(U, S) \to \Gamma(U, T)$ for every open set U of M and hence a homomorphism

$$i^q : C^q(N(U), S) \to C^q(N(U), T),\ q \geq 0.$$

This leads to an induced homomorphism

$$i^q : C^q(M, S) \to C^q(M, T),\ q \geq 0$$

We now describe a homomorphism

$$\delta^{q+} : H^q(M, S) \to H^{q+1}(M, T),\ q \geq 0$$

as a result of the exact sequence

$$0 \to S \xrightarrow{i} T \xrightarrow{\phi} N \to 0$$

We take a covering U and write down as follows a diagram of cochain groups and their connecting homomorphisms.

$$\begin{array}{ccccccccc}
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 \to & & C^q(N(U), S) & \to & C^q(N(U), T) & \to & C^q(N(U), N) & \to & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 \to & & C^{q+1}(N(U), S) & \to & C^{q+1}(N(U), T) & \to & C^{q+1}(N(U), N) & \to & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 \to & & C^{q+2}(N(U), S) & \to & C^{q+2}(N(U), T) & \to & C^{q+2}(N(U), N) & \to & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & &
\end{array}$$

This diagram is commutative in the sense that the image of a cochain depends only on its final position and is independent of the paths taken. Moreover the horizontal sequences are exact. In fact, to an element of $H^q(M, N)$ we take a representative q-cocycle, i.e., an element $u \in C^q(N(U), S)$ satisfying $\delta^{q+1}(u) = 0$, there exists $v \in C^q(N(U), T)$ such that $\varphi^q v = u$. Then $\varphi_0^{q+1} \delta^{q+} v = \delta^{q+} u = 0$ and there exists $w \in C^{q+1}(N(U), S)$ satisfying $i_0^{q+2} \delta^{q+1+} w = \delta_0^{q+1+} i^{q+1} w = \delta_0^{q+1+} \delta^{q+} v = 0$ so that $\delta^{q+1+} w = 0$. By chasing the diagram further it can be shown that the element of $H^{q+1}(N(U), S)$ defined by w is independent of the various choices made. This defines the homomorphism δ^{q+}.

A fundamental fact in cohomology theory is the result:

Theorem 8.4.2: If the sequence $0 \to S \overset{i}{\to} T \overset{\varphi}{\to} N \to 0$ is exact, then the sequence of cohomology groups

$$0 \overset{\delta_{1+}}{\to} H^0(M, S) \overset{i_0}{\to} H^0(M, N) \overset{\varphi_0}{\to} H^1(M, S) \overset{\delta_{0+}}{\to} H^1(M, T) \overset{i_1}{\to} H^1(M, N) \overset{\varphi_1}{\to} H^2(M, S)$$

is exact.

We apply this result to the exact sequence

$$0 \to C_{pq} \overset{i}{\to} a_{pq} \overset{\bar{\partial}}{\to} C_{pq+1} \to 0$$

A part (section) of the induced sequence of cohomology groups will be

$$\ldots \to H^{r-1}(M, a_{pq}) \overset{\bar{\partial}}{\to} (M, C_{pq+1}) \overset{\delta q+}{\to} H^r(M, C_{pq}) \overset{i}{\to} H^r(M, a_{pq}) \to \ldots$$

which is also exact.

Since the sheaf a_{pq} is fine, we have

$$H^r(M, a_{pq}) = 0, \; r \geq 1$$

From the exactness of the above sequence the following isomorphisms follow:

$$H^r(M, C_{pq}) \cong H^{r-1}(M, C_{pq+1}) \cong \ldots \cong H^r(M, C_{pq+r-1}) \cong H^0(M, C_{pq+r}) / \bar{\partial} H^0(M, C_{pq+r-1}),$$

the last expression being the Dolbeault group $D_{pq+r}(M)$.

Taking $r = q$, $q = 0$, we get

$$D_{pq}(M) \cong H^q(M, C_{p0})$$

Definition: The sequence

$$0 \to C_{pq} \overset{i}{\to} a_{pq} \overset{\bar{\partial}}{\to} C_{pq+1} \to 0$$

can be combined into one sequence as

$$0 \to C_{p0} \overset{i}{\to} a_{p0} \overset{\bar{\partial}}{\to} C_{p1} \overset{\bar{\partial}}{\to} \ldots \overset{\bar{\partial}}{\to} a_{pq} \to \ldots$$

which is exact by Dolbeault Grothendieck lemma. The subsheaf of a_{pq} which is the image of the preceding homomorphism and the kernel of the succeeding one is precisely C_{pq}. Since a_{pq} is fine, the above sequence is called a fine resolution of the sheaf C_{po}.

A similar but simpler situation prevails in the case of a real differentiable manifold *M*.

Let $a^r b$ the sheaf of germs of C^∞ -real valued differentiable forms of degree *r* and let C^r be the sub sheaf of a^r consisting of the germs of closed r-forms.

Then the sequence

$$0 \rightarrow R \xrightarrow{1} a^0 \xrightarrow{d} a^1 \xrightarrow{d} \ldots \xrightarrow{d} a^r \xrightarrow{d} \ldots$$

where *R* is the constant sheaf of real numbers and *i* is the inclusion map, is exact.

The above sequence is a fine resolution of the sheaf *R*. From the exactness of the same follows the deRham isomorphism:

$$Rr(M) \quad \cong \quad H^r(M, R)$$

(*r*-dimensional deRham group of *M*)

CHAPTER 9

Fibre Bundles

The study of fibre bundles makes an important component of algebraic topology for many reasons. On one hand it helps classification of the topological spaces and on the other gives remarkable results in physics, differential geometry and many other areas so far as applications are concerned. In this chapter we show only a few aspects of this theory.

9.1 VECTOR BUNDLES

We begin with some definitions.

Definition: A ***k-dimensional vector bundle*** ξ over the field F is a bundle (X, p, B) such that

(i) For each $b \in B$, $p^{-1}(b)$ is a k-dimensional vector space.

(ii) Each point of B has an open neighbourhood U and an U-isomorphism $h_U : U \times F \to p^{-1}(U)$ whose restriction to $\{b\} \times F$ is a vector space isomorphism onto $p^{-1}(b)$ for each b in U.

The condition (ii) above is called the *local triviality condition* and the U-isomorphism is called a *local coordinate chart* of ξ.

The tangent bundle over the n-sphere $\mathbf{S}^n$ denoted by $\tau(\mathbf{S}^n)$ and the tangent bundle over the n-real projective space RP^n, denoted by $\tau(RP^n)$ are simplest examples of vector bundles.

Definition: The *Stiefel variety* of orthonormal k-frames in $\mathbf{R}^n$, denoted by $V_k(\mathbf{R}^n)$ is the subspace $(v_1, v_2, ..., v_k) \in (\mathbf{S}^{n-1})^k$ such that the inner product $(v_i, v_j) = \delta_{ij}$.

That $\mathbf{S}^{n-1}$ is compact implies $V_k(\mathbf{R}^n)$ is compact.

The *Grassman variety* of k-dimensional subspaces of $\mathbf{R}^n$, denoted by $G_R(\mathbf{R}^n)$ is the set of k-dimensional subspaces of $\mathbf{R}^n$ equipped with the quotient topology which makes the function $(v_1, v_2, ..., v_k) \to sp(v_1, v_2, ..., v_k)$ from $V_k(\mathbf{R}^n)$ onto $G_k(\mathbf{R}^n)$ continuous. Clearly the Grassman variety $G_k(\mathbf{R}^n)$ is a compact manifold. It is also true that

$$G_k(\mathbf{R}^n) \subset G_k(\mathbf{R}^{n+1}) \subset G_k(\mathbf{R}^{n+2}) \subset ... \subset G_k(\mathbf{R}^\infty)$$

where $G_k(\mathbf{R}^\infty) = \bigcup_{n \geq k} G_k(\mathbf{R}^n)$ has the induced topology.

It is easy to see that $V(\mathbf{R}^n) = \mathbf{S}^{n-1}$ but $G(\mathbf{R}^n) = \mathbf{RP}^{n-1}$.

The canonical k-dimensional vector bundle γ_k of the product bundle $(G_k(\mathbf{R}^n) \times \mathbf{R}^n, p, G_k(\mathbf{R}^n))$ such that the total space consists of pairs $(v, x) \in G_k(\mathbf{R}^n) \times \mathbf{R}^n$ with $x \in V$. One can define γ_k as y_k on $G_k(\mathbf{R}^\infty)$.

Definition: If $\xi \equiv (X, p, B)$ is a bundle and $A \subset B$, then the restricted bundle $\xi|_A$ is a bundle whose total space is $p^{-1}(A)$ and $p|_A$ is the projection.

Definition: If $\xi \equiv (X, p, B)$ is a bundle and $f: B_1 \to B$ be any map, then the induced bundle of ξ under f, denoted by $f^*(\xi)$ is the bundle whose total space is the subspace of all pairs $(b, x) \in B_1 \times X$ such that $f(b_1) = p(x)$ and the map $(b_1, x) \to b_1$ is the projection.

9.2 A HOMOTOPY PROPERTY OF VECTOR BUNDLES

In this section a homotopy property of a vector bundle is established which will lead to an isomorphism in the next section.

We shall need the following two lemmas.

Lemma 1: If $\xi \equiv (X, p, B)$ be a vector bundle where $B = B_1 \cup B_2$, $B_1 = A \times [a, c]$, $B_2 = A \times [c, b]$, $a < c < b$ and $\xi | B_1 \equiv (X_1, p_1, B_1)$ and $\xi|B_2 \equiv (X_2, p_2, B_2)$ are trivial bundles, then ξ itself is trivial.

Proof: Let $h_i: B_i \times F \to Xi$ be a B_i-isomorphism for $i = 1, 2$ and let $g_i = h_i|(B_1 \cap B_2) \times F$.

Then $h = g_2^{-1} g_1$ is an $A \times \{c\}$-isomorphism of the trivial bundles and h is given by $h(x, y) = (x, \eta(x)y)$ where $(x, y) \in (B_1 \cap B_2) \times F$ and $\eta: A \to GL(k, F)$ is a map. Now this h can be prolonged to a B_2-isomorphism $w: B_2 \times F \to B_2 \times F$ by defining w as $w(x, s, y) = (x, s, \eta(x)y)$ when $x \in A$, $y \in F$, $t \in [c, b]$. The bundle-isomorphisms $u_1: B \times F \to X_1$ and $u_2 w: B_2 \times F \to X_2$ are equal on the closed set $(B_1 \cap B_2) \times F$. Hence, there exists an isomorphism $u: B \times F \to X$ such that $u|B_1 \times F = u_1$ and $u|B_2 \times F = u_2 w$. This completes the proof.

Lemma 2: For every vector bundle over $B \times I$, then there exists an open covering $\{U_i\}$ of B such that $\xi|U_i \times I$ is trivial where $I = [0, 1]$.

Proof: Follows from Lemma 1.

Theorem 9.2.1: If r is a map defined by $r(b, t) = (b, 1)$ where $(b, t) \in B \times I$ and if $\xi \equiv (X, p, B \times I)$ is a vector bundle where B is paracompact, then there exists a mapping $f: X \to X$ such that $(f, r): \xi \to \xi$ is a morphism and f is an isomorphism of fibres.

Proof: Since B is paracompact, by Lemma 2 there exists a locally finite open covering $\{U_i\}$ of B such that $\xi|_{U_i \times I}$ is trivial. Let $\{\eta_i\}$ be a partition of unity subordinate to $\{U_i\}$, i.e., sup $\eta_i \subset U_i$ and 1 = max $\eta_i(b)$ for each $b \in B$.

Let $h_i: U_i \times I \times F \to p^{-1}(U_i \times I)$ be a $U_i \times I$-isomorphism which follows from triviality. Then define $(f_i, r_i): \xi \to \xi$ as $r_i(b, t) = (b, \max(\eta_i(b), t))$ and f_i is the identity outside $p^{-1}(U_i \times I)$ and $f_i(h_i(b, t, x)) = h_i(b, \max(\eta_i(b), t), x)$ for each $(b, t, x) \in U_i \times I \times F$.

Applying the well-ordering principle to I we get, for each $b \in B$, an open neighbourhood $U(b)$ such that $U_i \cap U(b) \neq \phi$ for each $i \in I(b)$ where $I(b)$ is a finite subset of I. We now define

$r = r_{i(1)} \dots r_{i(n)}$ on $U_i \times U(b)$

and $f = f_{i(1)} \dots f_{i(n)}$ on $p^{-1}(U(b) \times I)$

where $i(b) = \{i(1),\dots,i(n)\}$ and $i(1) < i(2) < \dots < i(n)$.

Since r_i and f_i are identities for $i \notin I(b)$, they are infinite combinations of maps all but finitely many of which are identities near a point and since each f_i is an isomorphism on each fibre, f is an isomorphism on each fibre.

Corollary 9.2.2: There exists an isomorphism (f, r): $\xi \,|(B \times \{0\} \to \xi \,|\, (B \times \{1\})$

Proof: Immediate from the above theorem.

Theorem 9.2.3: If f, g: $B \to B'$ are two homotopic maps, B is paracompact and ξ is a vector bundle on B', then the induced bundles $f^*(\xi)$ and $g^*(\xi)$ are B-isomorphic.

Proof: Let H: $B \times I \to B'$ be the homotopy of f and g, i.e., $H(x, 0) = f(x)$ and $H(x, 1) = g(x)$. Then, $f^*(\xi)$ is isomorphic to $H^*(\xi)|B \times \{0\}$ and $g^*(\xi)$ is isomorphic to $H^*(\xi)|B \times \{1\}$. Since, by the above corollary, $H^*(\xi)|B \times \{0\}$ and $H^*(\xi)|B \times \{1\}$ are isomorphic, $f^*(\xi)$ and $g^*(\xi)$are B-isomorphic.

Corollary 9.2.4: Every vector bundle over a contractible paracompact space is trivial.

Proof: Let f: $B \to B$ be the identity and c: $B \to B$ be a constant map. By contractibility, f is homotopic to c. Hence $f^*(\xi)$ and $c^*(\xi)$ are isomorphic by the above theorem. But $f^*(\xi)$ is isomorphic to ξ since f is the identity map and $c^*(\xi)$ is isomorphic to $(B \times F, p, B)$. Hence ξ is isomorphic to the trivial vector bundle $(B \times F, p, B)$.

9.3 A REPRESENTATION THEOREM

We begin with a definition.

Definition: A *Gauss map* of a vector bundle $\xi \equiv (X, p, B)$ in F^n is a map g: $X \to F^m$ such that g is a linear monomorphism when restricted to $p^{-1}(b)$ for each $b \in B$.

Note that not for every vector bundle a Gauss map exists. But it is easy to see that if the base space B is paracompact, then such a map exists. In fact it is not difficult to show that the existence of a Gauss map g: $X \to F^{\infty}$ is equivalent to the condition of being isomorphic to the induced bundle γ_k of the canonical vector bundle under a suitable map f. Thus the following theorem is immediate from the above observation and can be considered as a classification theorem.

Theorem 9.3.1: Every vector bundle ξ over a paracompact space B is B-isomorphic to $f^*(\gamma_k)$ for some f: $B \to G_k(F^{\infty})$.

A simple argument and the definition of Gauss map now help infer the following:

Theorem 9.3.2: If f, g: $B \to G_k(F^{\infty})$ be two maps such that $f^*(\gamma_k)$ is B-isomorphic to $g^*(\gamma_k)$, then f is homotopic to g.

We are now in a position to state and prove the main theorem.

Theorem 9.3.3: The set of isomorphism classes of k-dimensional vector bundles over a paracompact space B is in a natural bijective correspondence with the set of homotopy classes of maps of B into the Grassman manifold of k dimensional subspaces of an infinite dimensional space.

Proof: Denoting the set of homotopy classes from B into $G_k(\mathbf{R}^\infty)$ by $[B, G_k(\mathbf{R}^\infty)]$ we define a function φ_B for each paracompact space B as follows :

$$\varphi_B([f]) = \{f*(\gamma_k)\} \in \text{Vect}_k(B)$$

where $[f] \in [B, G_k(R^\infty)]$ and $\text{Vect}_k(B)$ denote the set of isomorphism classes of k-dimensional vector bundles over B and $\{\gamma_k\}$ denotes the class containing γ_k .

The well-definedness now follows from Theorem 9.2.3 above. The Theorem 9.3.2 now gives the injectivity of φ. The subjectivity of φ is a consequence of the fact that a vector bundle of isomorphism classes is isomorphic to $f*(\gamma_k)$ for some f and the corresponding $[f] \in [B, G_k(\mathbf{R}^\infty)]$.

Remark 1: The above representation theorem help classify vector bundles and thus reduces the computation of isomorphism classes of vector bundles to that of the homotopy classes of maps from a paracompact base space to the Grassman manifold $G_k(\mathbf{R}^\alpha)$.

Remark 2: A series of almost analogous results can be derived without the assumption of paracompactness of the base space.

Bibliography

1. Alexandroff, P. and H. Hopf; *Topologgie I*, Springer Verlag, Berlin, 1935.
2. Alexandroff, P.S.; *Combinatorial Topologie*, Vol I–III, Graylock Press, Rochester, N.Y., 1956–60.
3. Arnold, B.H.: *Intuitive Concepts in Elementary Topology*, Prentice Hall, Inc., Eaglewood, Cliff, N.J., 1964.
4. Atiyah, M.F. and F. Hirshbruch; *Vector Bundles and Homogeneous Spaces in Differential Geometry*, A.M.S. Proceeding Symp. P. Maths.3, 7–38, 1967.
5. Bott, R.; *Homogeneous Vector Bundles*, Ann. Math (2), 66, 201–248, 1957.
6. Chatterjee, D.; *Mathematical Analysis*, Kalyani, 93.
7. Chatterjee, D.; *Topology*, Academy, 1992.
8. Dold, A.; *Algebraic Topology*, Springer verlag, 1976.
9. Dugundji, J.; *Topology*, Prentice Hall, New Delhi, 1975.
10. Eilenberg, S. and N.E. Steenrod; *Foundations of Algebraic Topology*, Princeton Univ. Press, Princeton, N.J., 1952.
11. Greenberg, M.J.; *Lectures on Algebraic Topology*, Benjamin, New York, 1967.
12. Hall, D.W. and G.L. Spencer; *Elementary Topology*, John Wiley, N.Y., 1955.
13. Hilton, P.J.; *An Introduction to Homotopy Theory*, Cambridge Univ. Press, London, 1953.
14. Hilton, P.J. and S. Wylie; *Homology Theory, An Introduction to Algebraic Topology*, Cambridge Univ. Press, London, 1960.
15. Hu, S.T.; *Homotopy Theory*, Academic Press, N.Y., 1959.
16. Husemoller, D.; *Fibre Bundles*, McGraw-Hill, New York, 1966.
17. Kuratowski, C.; *Topology I–II*, Polska Akademic Nauk, Warsaw, 1948.
18. Lefschetz, S.; *An Introduction to Topology*, Princeton Univ. Press, N.J., 1949.
19. Lefschetz, S.; *Algebraic Topology*, A.M.S. Coll. Publ. Vol. 27, 1942.
20. Massey, W.S.; *Homology and Cohomology Theory*, Mercel Dekker, N.Y., 1978.
21. Milnor, J.; *Differential Topology*, Note, Deptt. of Mathematics, Princeton University, 1958.
22. Munkres, J.R.; *Elementary Differential Topology*, Princeton Univ. Press, N.J., 1963.
23. Patterson, E.M.; *Topology*, Oliver and Boyd, London, 1956.
24. Pontryagin, L.S.; *Foundations of combinatorial Topology*, Graylock Press, Rochester, N.Y., 1952.

25. Spanier, E.H.; *Algebraic Topology*, McGraw-Hill, New Delhi, 1966.
26. Wallace, A.H.; *An Introduction to Algebraic Topology*, Pergamon Press, N.Y., 1957.
27. Whyburn, G.T.; *Analytic Topology*, A.M.S. Coll Publ. Vol. 28, 1942.
28. Whitehead, J.H.C.; *Combinatorial Homotopy*, Bull. A.M.S. 55 (1959), pp 213–245.
29. Willard, S; *General Topology*, Addison Wesley, 1970.
30. Wilder, R.L.; *Topology of Manifolds*, A.M.S. Coll. Publ. Vol. 32, 1949.