BMTC - 131
CALCULUS

Indira Gandhi National Open University
School of Sciences

Block
1
ESSENTIAL PRELIMINARY CONCEPTS

Course Design Committee*

Prof. Rashmi Bhardwaj
G.G.S. Indraprastha University, Delhi

Dr. Sunita Gupta
University of Delhi

Prof. Amber Habib
Shiv Nadar University
Gautam Buddha Nagar

Prof. S. A. Katre
University of Pune

Prof. V. Krishna Kumar
NISER, Bhubaneswar

Dr. Amit Kulshreshtha
IISER, Mohali

Dr. Aparna Mehra
I.I.T., Delhi

Prof. Rahul Roy
Indian Statistical Institute, Delhi

Prof. Meena Sahai
University of Lucknow

Dr. Sachi Srivastava
University of Delhi

Faculty members
School of Sciences, IGNOU

Prof. M. S. Nathawat (Director)
Dr. Deepika
Mr. Pawan Kumar
Prof. Poornima Mital
Prof. Parvin Sinclair
Prof. Sujatha Varma
Dr. S. Venkataraman

* The course design is based on the recommendations of the Programme Expert Committee and the UGC-CBCS template.

Block Preparation Team

Prof. Amber Habib (*Editor*)
Shiv Nadar University
Gautam Buddha Nagar

Prof. Parvin Sinclair
School of Sciences
IGNOU

Course Coordinators: Prof. Parvin Sinclair and Dr. Deepika

Acknowledgement: To Sh. S. S. Chauhan for the word processing and the preparation of the CRC of this block.

November, 2018

ISBN-978-93-88498-41-8

Further information on the Indira Gandhi National Open University courses, may be obtained from the University's office at Maidan Garhi, New Delhi-110 068 and IGNOU website www.ignou.ac.in.

Printed and published on behalf of the Indira Gandhi National Open University, New Delhi by Prof. M. S. Nathawat, Director, School of Sciences.

Printed by : Akashdeep Printers, 20-Ansari Road, Daryaganj, New Delhi-110002

COURSE INTRODUCTION

This is the first mathematics course you will be studying in the Bachelor's Degree Programme. The aim of this course is to develop an understanding of basic mathematical concepts and techniques that you will require for studying other mathematics courses of the programme, as well as any further study and work you undertake in mathematics.

Calculus is divided into two broad areas, **differential calculus** and **integral calculus**. Broadly speaking, differential calculus is the study of change and integral calculus is about adding up the parts. Differential calculus helps you to find, for example, the effect of changing conditions on a system being investigated, and hence to gain control over the system. The process of this mathematical investigation uses the powerful technique of modelling the phenomena concerned. The models usually involve differential equations. Differential calculus is useful in formulating models and integral calculus is used to solve the differential equations associated with the model. Apart from well known applications in physics, mathematical models based on calculus are used for the study of population ecology, cybernetics, management practices, economics and medicine.

In this course we shall focus on integral calculus after we discuss differential calculus. However, it was integral calculus that developed first historically. This has its origins in the need for measuring lands for the purpose of revenue collection. It is said that the Egyptian river Nile changed its course often, and the lands near it, with differently curved boundaries, were required to be measured again and again for levying taxes. This led to the development of mensuration in the Egyptian civilisation. We can regard mensuration as the forerunner of integral calculus. Indeed, one meaning of the word 'quadrature', used in integration, is computing the area.

The modern development of calculus began with the work of the famous 17^{th} century mathematicians, Newton and Leibnitz, in developing differential calculus. One of the early successes of calculus was the prediction of the period of Halley's comet. As you can see, calculus provides a powerful tool for the study of not only such natural phenomenon, but also artificial entities like the stock market. Over the centuries, many European mathematicians like Euler, Lagrange, the Bernoullis, Gauss, Cauchy and Riemann contributed to the development of this subject.

Now, a few words about how this course unfolds. In the first block of this course, you will be introduced to two basic building blocks of mathematics, namely, sets and functions. In the process you would recall a lot of related mathematics you studied in school, including two important coordinate systems for representing and studying two-dimensional spaces. Next, you will get more than a glimpse of the world of complex numbers, $\mathbb{C}$. Finally, you will study ways of solving certain polynomial equations over $\mathbb{R}$.

In the second block, you will begin by studying the properties of real numbers that you will need again and again. You will also study the concept of limits and continuity, which play a central role in calculus and, more generally, in mathematics.

In the third block, you will begin your study of differential calculus with the definition of a derivative and its basic properties. You will study several formulae for the derivatives of some functions which are used often, like polynomial functions and trigonometric functions. This block ends with a discussion on higher order derivatives and the Leibnitz rule for finding higher order derivatives.

In the fourth block, you will find some applications of differentiation. You shall study how derivatives can help to get information about various geometrical properties of curves. This block ends with a discussion on the tracing of different types of curves.

In the fifth, and last, block, your focus will be on integral calculus. You will study concepts of 'integral of a function' and 'primitive of a function'. You will also study the integrals of common functions like polynomial functions and trigonometric functions. You will get an opportunity to apply some techniques of integration like the substitution method, integration by parts and reduction formulae. The block ends with some applications of integral calculus for measuring lengths, areas and volumes.

Now a word about our notation. Each block has units and each unit is divided into sections, which may be further divided into sub-sections. These sections/sub-sections are numbered sequentially, as are the exercises and important equations in a unit. Since the material in the different units is heavily interlinked, there will be a lot of cross-referencing. For this we will be using **the notation Sec. x.y to mean Section y of Unit x.**

Throughout this course the emphasis will be on techniques rather than on theory. So you may not find many proofs here. (You will be able to find the proofs of many of the theorems you study and apply here in our third semester course, Real Analysis.)

Another compulsory component of this course is **its assignment**, which covers the whole course. Your **academic counsellor at the study centre** will evaluate it, and return it to you with detailed comments. Thus, the assignment is meant to be a teaching as well as an assessment aid. Further, you will not be allowed to take the exam of this course till you submit your assignment response at your study centre. So please submit it well in time.

The course material that we have sent you is self-sufficient. If you have a problem in understanding any portion, please ask your academic counsellor for help. Also, if you feel like studying any topic in greater depth, you may consult.

A word of friendly advice here is that to learn the various techniques presented in this course, you will need to put in a lot of practice in solving problems given in the material. You should attempt to solve all the exercises in the block as you go along, before you look up the solution. As a part of the tutorials, we have added miscellaneous examples and exercises at the end of each block. You should also attempt these exercises. In addition, you may also like to look up some other books in the library of your study centre, and try to solve some exercises from these books.
Some useful books and websites are the following:

1. Essential Calculus, by James Stewart, Cengage Publication.
2. Calculus, by Anton, Bivens and Davis, Wiley Publications.
3. https://brilliant.org/courses/calculus-done-right
4. https://www.mathisfun.com/calculus
5. www.math.mit.edu/Ndjk/calculus_beginners

We have also prepared a **video programme**, which will be available at your study centre, called "**Limits**", based on the material in Block 2.

Wishing you a happy learning experience,

The Course Team

BLOCK 1 ESSENTIAL PRELIMINARY CONCEPTS

With this block, you are stepping into your study of undergraduate mathematics. You would have studied mathematics upto Class XII, wherein you would have studied some of the concepts that will be covered here. These are included because they are essential for further study, and recalling them will help you.

To begin with, in Unit 1, you will look at a basic essential concept developed in the 20^{th} century, namely, that of a set. Here you will get an opportunity to relook operations on sets and their properties.

Following this, in Unit 2, you will have occasion to recall what a function is, operations on functions as well as various kinds of functions.

In Unit 3, you will be considering two coordinate systems for two dimensions. The Cartesian system would be familiar to you, though the polar coordinate system may seem new. Linked with this is Unit 4, which focuses on introducing you to complex numbers and their properties. The link between the two units will become clear as you study them.

Finally, in Unit 5 you will study various polynomials and how to find their roots. In particular, we will be discussing polynomials upto degree 4.

At the end of this block you will find **a set of miscellaneous exercises** related to the concepts covered in this block. Please do study them, and try each exercise yourself. This will help you engage with the concepts concerned and understand them better.

A word about some signs used in the unit! Throughout each unit, you will find theorems, examples and exercises. To signify the end of the proof of a theorem, we use the sign ■ . To show the end of an example, we use $***$. Further, equations that need to be referred to are numbered sequentially within a unit, as are the exercises and figures. E1, E2, etc. denote the exercises and Fig. 1, Fig. 2, etc. denote the figures.

NOTATIONS AND SYMBOLS (used in Block 1)

$\in$ ($\notin$)	belongs to (does not belong to)
$\subseteq$ ($\subsetneq$)	is contained in (is properly contained in)
$\not\subseteq$ ($\not\supseteq$)	is not contained in (does not contain)
$\supseteq$ ($\supsetneq$)	contains (properly contains)
$A \cup B$	the union of the sets A and B
$A \cap B$	the intersection of the sets A and B
$A \setminus B$	the complement of the set B in the set A
$A \times B$	The Cartesian product of the sets A and B
A^c	complement of the set A
$\mathbb{N}$	the set of natural numbers
$\mathbb{Z}(\mathbb{Z}^*)$	the set of integers (non-zero integers)
$\mathbb{Q}(\mathbb{Q}^*)$	the set of rational numbers (non-zero rational numbers)
$\mathbb{R}(\mathbb{R}^*)$	the set of real numbers (non-zero real numbers)
$\mathbb{C}(\mathbb{C}^*)$	the set of complex numbers (non-zero complex numbers)
$\emptyset$	empty set
$\Rightarrow$ ($\Leftrightarrow$)	implies (implies and is implied by)
iff	if and only if
$\therefore$	therefore
w.r.t.	with respect to
s.t.	such that
$<$ ($\leq$)	is less than (is less than or equal to)
$>$ ($\geq$)	is greater than (is greater than or equal to)
$\exists$	there exists
$\forall$	for all
$\sum_{i=1}^{n} a_i \left(\prod_{i=1}^{n} a_i\right)$	$a_1 + a_2 + \cdots + a_n$ $(a_1 a_2 \ldots a_n)$
$f: X \to Y$	f is a function from the set X to the set Y
$\{x \mid x \text{ satisfies } P\}$	the set of all x such that x satisfies the property P
$\wp(X)$	the power set of the set X
$\lvert x \rvert$	modulus of the real, or complex number, x
$\operatorname{Re} z$	real part of the complex number z
$\operatorname{Im} z$	imaginary part of the complex number z
$\operatorname{Arg} z$	the principal argument of the complex number z
$\bar{z}$	the complex conjugate of the complex number z
$\deg f$	degree of the polynomial f
$\mathbb{R}^n$ ($\mathbb{C}^n$)	$\mathbb{R} \times \mathbb{R} \times \ldots \times \mathbb{R}$ (n times) ($\mathbb{C} \times \mathbb{C} \times \ldots \times \mathbb{C}$ (n times))

Greek Alphabets

α	Alpha	K	Kappa	σ, Σ	Sigma (Capital sigma)
β	Beta	λ	Lambda		
γ	Gamma	μ	Mu		
δ	Delta	ν	Nu	τ	Tau
ε	Epsilon	ξ	Xi	υ	Upsilon
ζ	Zeta	o	Omicron	ϕ	Phi
η	Eta	π, Π	Pi (capital pi)	χ	Chi
θ	Theta			ψ	Psi
ι	Iota	ρ	Rho	ω	Omega

UNIT 1

SETS AND OPERATIONS ON THEM

Structure **Page Nos.**

1.1 INTRODUCTION

Welcome to the world of algebra! We start our formal discussion with a basic entity necessary for doing any algebra. So, consider the collection of words that are defined in a given dictionary. A word either belongs to this collection or not, depending on whether it is listed in the dictionary or not. This collection is an example of a set, as you will see in Sec. 1.2. When you start studying any part of mathematics, you will be working with one or more sets. This is why we want to spend some time in discussing some basic concepts and properties concerning sets. These objects were first defined by the German mathematician Cantor. To start with, you will see various examples of sets and different ways of describing sets.

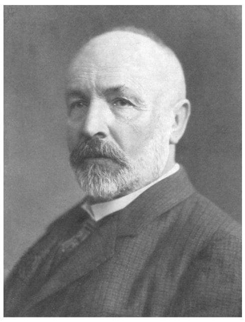


Fig. 1: Georg Cantor (1845-1918)

Then, in Sec. 1.3, we will discuss 'parts' of a set, which also form sets. In Sec. 1.4, we will consider Venn diagrams, a pictorial representation of inter-relationships between sets.

You must be familiar with the basic operations on real numbers – addition, subtraction, multiplication and division. When we apply any of these operations, we combine two real numbers at a time to obtain another real number. For instance, if r and s are two real numbers, then $r-s$ is also a real number. In a similar way, we can obtain new sets by applying certain operations on two given sets at a time. In Sec. 1.5, we shall discuss four operation on pairs of sets, namely, the complement of a set A in a set B, the union and intersection of A and B, and the Cartesian product of A and B. The operations of union, intersection and Cartesian product can be naturally applied to any number of sets, too, as you will see.

Finally, in Sec. 1.6, we shall familiarise you with certain laws relating to these operations.

As mentioned earlier, a knowledge of the material covered in this unit is necessary for studying any mathematics course. So please study this unit carefully.

And now we will list the objectives of this unit. After studying the unit, **please read this list again** and make sure that you have achieved the objectives. One way of ensuring this is to **try each exercise** given in the unit **as you get to it**. Do not go further, till you have done the exercise to your satisfaction.

Objectives

After studying this unit, you should be able to:

- identify a set;
- describe sets by the listing method or the property method;
- give examples of finite and infinite sets;
- represent relationships between sets by Venn diagrams;
- explain, and apply, the operations of complementation, intersection, union and Cartesian product on pairs of sets;
- prove, and apply, basic results pertaining to these operations;
- state, prove and apply, the distributive laws pertaining to these operations;
- state, prove and apply, the De Morgan's laws.

1.2 SETS

You may have come across various collections of objects, like a stamp collection, a coin collection, a gathering of people interested in 'kabaddi', and so on. In mathematics some of these collections are considered 'sets', and some are not. Let us see why.

Consider the coin collection of the National Museum, Delhi. Given any object, you can firstly see whether it is a coin or not. Next, if it is a coin, you can easily find out whether it is part of the Museum's collection or not. Also, whatever conclusion you reach will be the same conclusion reached by any object (person) in any part of the world. So, there is a universal agreement about whether an object belongs to this collection or not. This certainty comes because the collection is 'well-defined'.

So, a **well-defined collection** is one for which given an object, it should be quite clear to anybody, anywhere, whether the object belongs to the collection or not, regardless of who is deciding this. So, for example, the collection of all female pilots who have worked in Air India from 2000 on is well-defined because any object (person) is/was either a female pilot working in Air India in this period or not, and accordingly she/he does or does not belong to the collection. On the other hand, the collection of all intelligent human beings is not well-defined. Why? Well, a particular human being may seem intelligent to one person and not to another. So, there is no objective criterion for agreeing on who is intelligent and who is not.

This leads us to the following definition.

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

Let us look at some more examples of sets, which you may have already come across. You will be using them a great deal in this course, and in the other mathematics courses you take.

i) The set of natural numbers, denoted by $\mathbb{N}$.

ii) The set of integers, denoted by $\mathbb{Z}$.

iii) The set of rational numbers, denoted by $\mathbb{Q}$.

iv) The set of real numbers, denoted by $\mathbb{R}$.

In Unit 4 you will be studying another set of numbers, namely, the set of complex numbers, denoted by $\mathbb{C}$.

Doing the following exercises will help you to assess if you have understood the concept of a set.

E1) Which of the collections mentioned below are sets?

i) The collection of all people interested in 'kabaddi'.

ii) The collection of all those people who have been to Mars.

iii) The collection of prime numbers.

iv) The Asiatic Society Library collection.

v) The collection of all rectangles that are not squares.

vi) The collection of all funny movies.

A **prime number** is a natural number other than one, whose only factors are one and itself.

E2) Suppose you are given a stamp album. Is the collection of stamps in that album a set? Give reasons for your answer.

Now consider the objects in a set, for example, the set of all female pilots in Air India. Any such female pilot is a 'member' of the set, and a person who is/was not a female pilot with Air India is not a 'member' of the set, according to the definition below.

Definition: An object that belongs to a set is called an **element**, or **member**, of that set.
For example, 2 is an element of the set of natural numbers, $\mathbb{N}$, and (-2) is not an element of $\mathbb{N}$.

People normally use capital letters A, B, C, etc. , to denote sets. The lower case letters a, b, c, x, y, etc., are usually used to denote elements of sets. We symbolically write the statement 'a is an element of the set A' as $\mathbf{a \in A}$.
If a is not an element of A, that is, if a does not belong to A, we write this fact as $\mathbf{a \notin A}$.
So, for example, if A is the set of prime numbers, then $5 \in A$ and $9 \notin A$.

The symbol '$\in$' stands for 'belongs to'. It was suggested by the Italian mathematician Peano (1858-1932).

Now, you know that the square of a real number is always non-negative. So, will the collection of all real numbers whose square is negative be a set? Is there any such number? There isn't. Therefore, the collection does not have any element. However, ***it is well-defined*** because it is clear to anybody anywhere that this collection has no element. So, this collection is a set, but with no member, and it has a special name.

Definition: A set which has no element is called the **empty set** (or the **void set**, or the **null set**). It is usually denoted by $\emptyset$.

Try the following exercise now.

E3) Which of the following statements are true? Change those that are false to obtain correct statements.

i) $0.2 \in \mathbb{N}$

ii) $\sqrt{9} \notin \mathbb{Z}$

iii) $\sqrt{2} \in \mathbb{R}$

iv) $\sqrt{2} \in \mathbb{Q}$

v) The set of all squares that are not rectangles is $\emptyset$

vi) Any circle is a member of the set in E1 (v) above.

Recall that $\sqrt{a}$ denotes the non-negative square root of the non-negative real number a.

A set which has at least one element is called a **non-empty** set. We usually describe a non-empty set in two ways – the **listing method** and the **property method**.

Listing Method: In this method, we list all the members of the set within curly brackets. For instance, the set of all natural numbers that are factors of 10 is $\{1,2,5,10\}$. Using this notation, some people also denote the empty set by $\{\ \}$.

Remember that if you write an opening bracket '{' to start a set, you must write its corresponding closing bracket '}' after the elements of the set are listed.

But what if the set has too many elements to be able to write them all down? In this case we list some of the elements of the set, enough to see/show some pattern which its elements follow. For example, the set $\mathbb{N}$ of natural numbers can be described as $\mathbb{N} = \{1,2,3,...\}$, where you can see the pattern, namely, the next element is obtained by adding 1 to the previous one.
Similarly, the set of all even numbers strictly lying between 10 and 100 is $\{12,\ 14,\ 16\ ,\ldots,\ 98\}$. (Note that 'strictly' lying between two numbers a and b means that a and b are not included.)

'…' is the convention for showing that the elements continue accordingly.

This method of representing sets is called the **listing method** (or **tabular method**, or **roster method**).

Property Method: In the second method of describing a set, we describe its elements by a property common to all of them. As an example, consider the set S of all the stars in the sky. Here the property common to all of them is that each is a star. So, we can write this as
$S = \{s$, where s is a star in the sky$\}$.

In formal mathematics we write this as $S = \{s \mid s$ is a star in the sky$\}$.

The vertical bar after s denotes 'such that'. (Some authors use ':' instead of '|' for 'such that'.)

Now, consider the set T of all natural numbers which are multiples of 5. This set T can be written in the form
$T = \{x \mid x \in \mathbb{N}$ and x is a multiple of $5\}$.

This states that T is **the set of all x such that** x is a natural number and x is a multiple of 5. We can also write this in a slightly shorter form as
$T = \{x \in \mathbb{N} \mid x$ is a multiple of $5\}$.

This method of describing the set is called the **property method** (or **set-builder method**).

In some cases, we can use either method to represent the set under consideration. For instance, the set T above, can be described by the roster method as $T = \{5, 10, 15, \ldots\}$.

Again, the set E, of all natural numbers less than 10, can be described as

$E = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$ (by the listing method)

$$\left.\begin{array}{l} E = \{x \mid x \text{ is a natural number less than } 10\} \\ E = \{x \in \mathbb{N} \mid x < 10\} \\ E = \{x \mid x \in \mathbb{Z} \text{ and } 0 < x < 10\} \end{array}\right\} \text{(by the property method).}$$

Sometimes, however, it is difficult to represent a set by both methods. For example, take the set of stars in the sky. How would you represent it by the listing method?

Again, consider the collection of three elements {Kochi, $5, \pi$}. Firstly, is this a set? Since this is a well-defined collection of three objects, it is a set. So, how would you represent this set by the property method? The three objects do not appear to have any common property except that they belong to the given set. So the property method won't work here.

The property method can only be used for sets whose elements have a common property.

Now, consider the set E, represented above by two methods. Instinctively, you can see that both these sets are the same. How do we say this more formally?

Definition: Two sets S and T are called **equal**, denoted by $\mathbf{S=T}$, iff every element of S is an element of T and every element of T is an element of S.

'iff' is short for 'if and only if'.

So, if $A = \{1, 2, 3, 3\}$, $B = \{1, 2, 3\}$ and $C = \{2, 1, 3\}$, are $A = B$ and $B = C$? The answer to both questions is 'yes', since the elements are the same in all 3 sets.

This leads us to the following remarks.

Remark 1: While listing the elements in a set, we do not gain anything by repeating them. Therefore, the convention is that we **do not repeat the elements in a set**.

Remark 2: Changing the order in which the elements are listed in a set does not alter the set.

Remark 3: There can be several properties that define the same set. For example, two ways of describing the set $\{2\}$ by the property method are $\{x \mid 3x - 1 = 5\}$ or $\{x \mid x$ is an even prime number$\}$.

Why don't you do the following exercises now?

E4) Describe the following sets by the listing method, if possible

i) $\{x \in \mathbb{Z} \mid x$ is the largest negative integer$\}$

ii) $\{x \in \mathbb{Z} \mid 3x - 5 \le 19\}$

iii) the set of all the letters in the English alphabet.

E5) Describe the following sets by the property method, if possible.

i) $\{1, 4, 9, 16, ...\}$

ii) {Indira Gandhi National Open University, $\pi, 100^{100}$}

iii) $\{..., -6, -4, -2, 0, 2, 4, 6, ...\}$

iv) $\emptyset$

v) {September, April, June, November}

E6) Check whether or not $\{x \in \mathbb{N} \mid x^2 = 9\} = \{x \in \mathbb{R} \mid x^2 = 9\}$.

Let us now consider one way of classifying sets, depending on their 'size', in a sense.

Consider the set S, of students of IGNOU enrolled in the course 'Calculus'. How many elements do you think this set has? It may have anywhere up to 10,000 elements. So it is a set with finitely many members. Similarly, $\{x \in \mathbb{R} \mid 2x^2 - 5x + 1 = 0\}$ has two elements, and hence is finite.

Now consider the set $\mathbb{Z}$. Does this have finitely many members? For any positive integer n, you can always find an integer $n+1$ (or for a negative integer n, you can find $n-1$). This means that there is no end to the number of integers we can find. So, it is not possible to count all the elements of this set, and hence it does not have finitely many elements.

These examples lead us to the following definitions.

Definitions: i) A non-empty set with finitely many elements is called a **finite set**. A set which is not finite is called an **infinite set**.

ii) The number of elements in a finite set A is called the **cardinality of the set** A, and is denoted by $|A|$, or **card**$(\mathbf{A})$**.** (We shall talk about the cardinality of infinite sets later.)

iii) A set with only one element is called a **singleton**.

Regarding cardinality, what follows is an important point.

Remark 4: By convention, **we treat the empty set as a finite set, with cardinality zero**.

An example of a finite set is the set of stars in the sky. Do you agree? Well, if you try and count them, they seem to be far too many to count. However, according to present day astronomers they can be counted, and their number is approximately 100 billion.

Yet another example of a finite set is $\{\emptyset, -\sqrt{2}, \text{IGNOU}\}$, with cardinality 3. On the other hand, $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$ are all infinite sets.

Remark 5: Note that the set $\{0.3\}$, for example, is a singleton, and is not the same as the number 0.3. For any object x, **there is a difference between $\mathbf{x}$ and $\{\mathbf{x}\}$. In fact, $\mathbf{x} \in \{\mathbf{x}\}$.**

Why don't you try an exercise now?

E7) Which of the following sets are infinite? Give reasons for your answers.
 i) The number of spoons of water in a given drum of water.
 ii) $\{x+3 \mid x \in \mathbb{Q}\}$
 iii) $\{x \in \mathbb{R} \mid x \notin \mathbb{Q}\}$
 iv)
 v) $\{\mathbb{N}\}$
 vi) The set of points on the circumference of a circle.
 vii) $\emptyset$.

E8) Which of the following statements are true? Give reasons for your answer.
 i) $\{\mathbb{R}\} = \mathbb{R}$
 ii) $\{\emptyset\}$ is a singleton.

Now, you have been working with the sets $\mathbb{N}$ and $\mathbb{Z}$. You would have noted that every element of $\mathbb{N}$ is an element of $\mathbb{Z}$, but not vice versa. Is there a similar relationship between other sets? Let us see.

1.3 SUBSETS

Consider two sets A and B, where
A = the set of all students of IGNOU, and
B = the set of all female students of IGNOU.

Now, every female student of IGNOU is certainly a student of IGNOU. So, each element of B is also an element of A. In such a situation, we say that B **is contained in** A.
Of course, IGNOU also has some male students! So, there is at least one element x in A such that x does not belong to B. Mathematically, we write this as
$\exists\, x \in A$ s.t. $x \notin B.$
This is read as 'there exists x belonging to A such that x does not belong to B'.

'$\exists$' denotes 'there exists'.
's.t.' is short for 'such that'.

So, every element of B is an element of A, but A has some more elements too. In this situation we say that B is **properly contained** in A.
So, for example, would you agree that $\mathbb{N}$ is properly contained in $\mathbb{Z}$?
In general, we have the following definitions.

Definitions: i) A set A is a **subset** of a set B if every element of A belongs to B, and we denote this fact by $\mathbf{A \subseteq B}$. In this situation, we also say that $\mathbf{A}$ **is contained in** $\mathbf{B}$, or that $\mathbf{B}$ **contains** $\mathbf{A}$, denoted by $\mathbf{B \supseteq A}$. Here, we also say that B is a **superset** of A.

ii) If $A \subseteq B$ and $\exists\ y \in B$ such that $y \notin A$, then we say that A is a **proper subset** of B (or A is **properly contained** in B). We denote this by $\mathbf{A \subset B}$, or $\mathbf{A \subsetneq B}$.

iii) If X and Y are two sets such that X has an element x which does not belong to Y, then we say that **$\mathbf{X}$ is not contained in $\mathbf{Y}$**. We denote this fact by $\mathbf{X \not\subseteq Y}$. So, $X \not\subseteq Y$ implies that $\exists\ x \in X$ s.t. $x \notin Y$.

Let us look at a few examples of what we have just defined.

Example 1: Consider the set $A = \{1, 2, 3\}$. Is $A \subseteq A$? Is $A \subsetneq A$? Give reasons for your answer.

Solution: Since every element of A is certainly in A, we find that $A \subseteq A$. Also, there cannot be any element of A that is not in A. Therefore, A cannot be a proper subset of itself.

***

This example leads us to the following remark.

Remark 6: For any set A, you can show that $A \subseteq A$ by using the same reasoning. Thus, **any set is a subset of itself**. Also, **no set is a proper subset of itself**. (Why?)

Now consider the following example.

There are infinitely many pairs of sets A and B such that $A \not\subseteq B$ and $B \not\subseteq A$.

Example 2: Give an example of two sets, neither of which is a subset of the other.

Solution: Consider the sets $A = \{1, -1\}$ and $B = \{0, 1, 2\}$.
Is $A \subseteq B$? We find $(-1) \in A$ such that $(-1) \notin B$. $\therefore A \not\subseteq B$.
Also, $B \not\subseteq A$. (Why?)

***

Note that given any two sets A and B, one and only one of the following possibilities is true.

i) $A \subseteq B$, or

ii) $A \not\subseteq B$.

$\emptyset \subseteq A$, for any set A.

Using this fact, we can show that the **empty set $\emptyset$ is a subset of every set.**

Now, let us go back for a moment to the point before Remark 1, where we defined equality of sets. Let us see what equality means in terms of subsets. Consider the sets
A = the set of even natural numbers less than 10, and
$B = \{2, 4, 6, 8\}$.

'$\Leftrightarrow$' denotes 'implies and is implied by', or 'is equivalent to'.

Every member of A is a member of B, and vice versa. That is, $A \subseteq B$ and $B \subseteq A$. But, by our definition, we also note that $A = B$.
So, we find that $A = B$ is equivalent to $(A \subseteq B \text{ and } B \subseteq A)$ taken together.
We write this as
$A = B \Leftrightarrow (A \subseteq B$ and $B \subseteq A)$. In fact, this is true for any two sets.

Remark 7: For any two sets A and B, $A = B \Leftrightarrow A \subseteq B$ and $B \subseteq A$.

Try the following exercises now. While doing them, remember that **to show that $\mathbf{A \subseteq B}$**, for any two sets A and B, you must show that if $a \in A$, then $a \in B$, i.e., $a \in A \Rightarrow a \in B$.

Also, **to show that $\mathbf{A \not\subseteq B}$**, you must show that there is at least one element in A that is not in B, i.e., $\exists\, x \in A$ such that $x \notin B$.

'$\Rightarrow$' denotes 'implies'.

The set of all subsets of a set A is called the **power set** of A.

E9) Write down all the subsets of $\{1,2,3\}$. How many of these contain
i) no element, ii) one element, iii) two elements,
iv) three elements, v) more than three elements.

E10) Show that if $A \subseteq B$ and $B \subseteq C$, then $A \subseteq C$. (This shows that '$\subseteq$' is a **transitive relation**, as you will see in Unit 2.)

E11) Give a superset of the set {IGNOU, 0.7, Mahatma Gandhi}.

So far you have seen two methods of describing sets. There is yet another way of representing sets and the relationships between them. This is what we discuss in the next section.

1.4 VENN DIAGRAMS

It is often easier to understand a situation if we can represent it pictorially. To ease our understanding of many situations involving sets and their relationships, we represent them by simple diagrams, called **Venn diagrams**. An English logician, John Venn, invented them. To be able to draw a Venn diagram, you would need to know what a universal set is.

Fig. 2: John Venn (1834-1923)

Consider a situation involving two or more sets, for example, the set D of female film directors, and the set S of female scientists. Then, we first look for a convenient superset of all the sets under discussion. For example, here we can take this to be the set of all women. We call this superset a **universal set**, and denote it by $\mathbf{U}$. So here our universal set U is the set of all women. This is because U contains D as well as S. We could also have taken U to be the set of all humans, which would be a larger superset.

Consider another situation involving the sets of integers and rational numbers. Here we could take the set of real numbers as our universal set. We could also take $\mathbb{Q}$ as our universal set, since it contains both $\mathbb{Z}$ and $\mathbb{Q}$.

A universal set is not unique.

Remark 8: As you have seen, we have several possibilities from which we pick one as a universal set in a given situation. We usually take our universal set to be just large enough to contain all the sets under consideration.

Now, let us see how to draw a Venn diagram, using an example. To clarify what we have just said, consider the following example.

Example 3: Draw a Venn diagram to represent the sets $A=\{1,2,3\}, B=\{3,4,5\}, C=\{6,7\}$, with $U=\{1,2,\ldots,10\}$.

Solution: Let us denote A by a circle, B by an ellipse and C as another closed region. The points $8, 9$ and 10 don't lie in any of A, B or C, but they are in U. Note that 3 belongs to both A and B. Therefore, it lies in the circle as well as the ellipse. So the circle and ellipse have a common region. Also note that A and C do not have any elements in common. Therefore, the regions representing them do not cut each other. For the same reason the regions representing B and C do not cut each other. We represent the universal set U by a rectangle enclosing all these regions and points, as in Fig. 3.

Note that, in Fig. 3, we could have represented B and C in the shape of circles also, or any other shape.

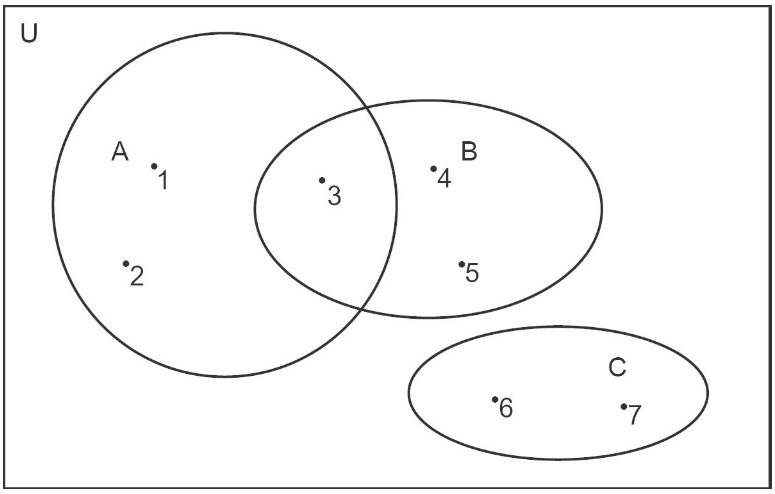


Fig. 3: A Venn diagram

***

Remark 9: Note that the relative areas of the regions in the Venn diagram do not actually represent the relative cardinalities of the sets concerned. For instance, the regions representing A and B in Fig. 3 do not have the same area though A and B have the same number of elements.

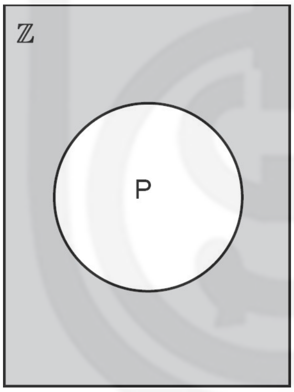


Fig. 4

So, in general, **how would we draw a Venn diagram?** Suppose we are discussing various sets $A, B, C, \ldots$. They may be finite or infinite. We choose a universal set U. So, $A \subseteq U, B \subseteq U, C \subseteq U$, and so on. We show this situation in a Venn diagram as follows:

The interior of a rectangle represents U. The subsets A, B, C, etc., are represented by the interiors of closed regions lying completely within the rectangle showing U. These regions may be in the form of a circle, ellipse, **or any other shape**. Further, we assume that **any enclosed area in a Venn diagram represents a non-empty set**.

Now, what will a Venn diagram corresponding to the situation $P = \{2, 3, 5, 7, 11, \ldots\} \subseteq \mathbb{Z}$ look like? Well, we can just take $\mathbb{Z}$ to be our universal set. Then the Venn diagram in Fig. 4 is one such diagram. If we take another set U that properly contains $\mathbb{Z}$ as our universal set, say $\mathbb{Q}$, then we get the Venn diagram in Fig. 5. So both Venn diagrams represent the relationship between the set of primes and $\mathbb{Z}$.

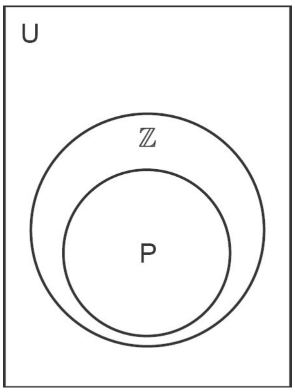


Fig. 5

Try these exercises now.

E12) How would you represent the following situation in a Venn diagram? *The set of all rectangles, the set of all squares and the set of all parallelograms.*

E13) Draw a Venn diagram to show three sets A, B and C, where A and C have some common elements, B is infinite and C is finite.

So, what is the purpose of a Venn diagram? Well, consider Fig. 6. Just one look and you know the broad situation – there are 4 sets A, B, C, D, of which

A has no element in common with the others; B and C have common elements; C and D have common elements; B and D have no common elements. This single visual gives us so much information.
That is the utility of these diagrams.

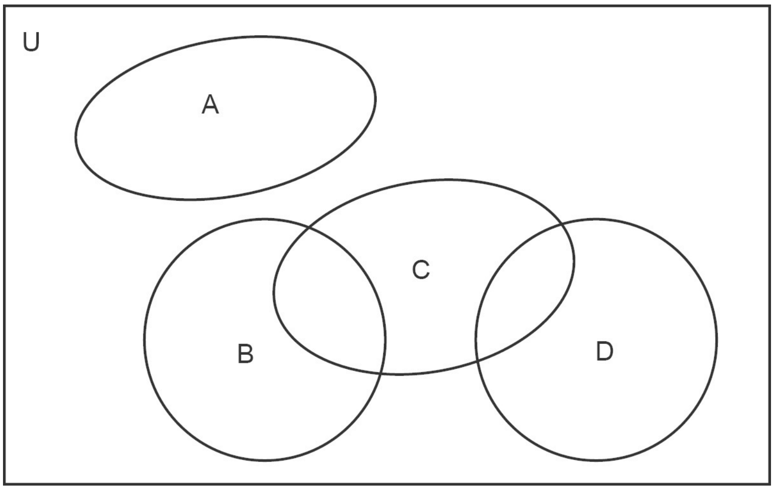


Fig. 6

Now that you are familiar with Venn diagrams, let us consider various ways in which we can create new sets from two or more sets given to us.

1.5 OPERATIONS ON SETS

Consider two sets $A = \{-1, 1\}$ and $B = \left\{1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \cdots\right\}$. We can obtain other sets from them in several ways, for example, by considering the elements in A that are not in B, or by taking the elements common to both A and B, etc. These are examples of operations on these sets. In this section, we shall discuss four operations on sets that you will be using very often.

1.5.1 Complementation

Consider the sets A and B above. There are elements of B that do not belong to A, like $\frac{1}{2}, \frac{1}{3}$, etc. These elements form a set, namely, $\left\{\frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \cdots\right\}$. Similarly, there are elements in A that are not in B, which form a set, namely, $\{-1\}$. This way of obtaining a third set from two given sets is defined below.

Definition: Let A and B be two sets. The **complement of $\mathbf{A}$ in $\mathbf{B}$**, denoted by $\mathbf{B \setminus A}$, and read as '**$\mathbf{B}$ complement $\mathbf{A}$**', is the set of elements in B which are not elements of A, that is, $\{x \in B \mid x \notin A\}$.
Similarly, $\mathbf{A \setminus B = \{x \in A \mid x \notin B\}}$.
In the special case when B is the universal set U, $B \setminus A$ is $U \setminus A$. This set is called the **complement of the set $\mathbf{A}$**, and is denoted by $\mathbf{A'}$ or $\mathbf{A^c}$.

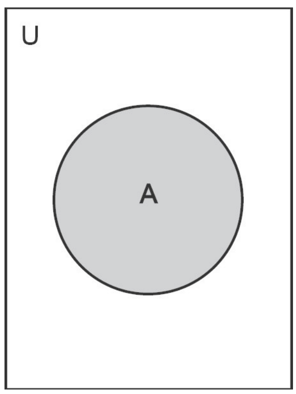


Fig. 7

If you look at Fig. 7, the unshaded part in the Venn diagram represents $A^c = U \setminus A$. In fact, this diagram also shows us that $\mathbf{x \in A^c}$ **if and only if**

$\mathbf{x \notin A}$. Similarly, the unshaded area in Fig. 4 represents the set $P = \{2, 3, 5, 7, 11, \ldots\}$, and the shaded area represents $\mathbb{Z} \setminus P$.

Remark 10: Note that if $\mathbf{A \subseteq B}$**, then,** $\mathbf{A \setminus B = \emptyset}$, since there is no element in A that is not an element of B.

Try these exercises now.

E14) i) Represent the following sets in one Venn diagram:
The set P of all prime numbers, the set $\mathbb{Z}$ and the set $\mathbb{Q} \setminus \mathbb{Z}$.

ii) Is the set $\mathbb{Z} \setminus P$ finite or infinite? Is the set $P \setminus \mathbb{Z}$ finite? Give reasons for your answers.

E15) Let A be any set. Give the sets $A \setminus A, \emptyset \setminus A, A \setminus \emptyset$ and $(A^c)^c$.

$\mathbf{B \setminus A \subseteq B}$

E16) Give an example of sets A and B such that

i) $A \setminus B \not\subseteq B$.

ii) $A \setminus B \subseteq B$.

Let us now consider another operation on sets, namely, the intersection of two or more sets.

1.5.2 Intersection

Let us consider $\mathbb{N}$ and $\{0, 1\}$. Are there any elements common to both these sets? For instance, 1 belongs to both. In fact, this is the only element common to both. Thus, $\{1\}$ is the intersection of the sets $\mathbb{N}$ and $\{0, 1\}$, according to the definition below.

Definition: Let A and B be two subsets of a universal set U. The **intersection of $\mathbf{A}$ and $\mathbf{B}$** will be the set of elements of U that are common to both A and B. This is denoted by $\mathbf{A \cap B}$, and read as 'A intersection B'. Thus, $\mathbf{A \cap B = \{x \in U \mid x \in A}$ **and** $\mathbf{x \in B\}}$.

To clarify this operation further, consider the following example.

Example 4: Let S be the set of prime numbers and T be the set $\{x \in \mathbb{Z} \mid x$ divides $10\}$. What is $S \cap T$?

Solution: We take $\mathbb{Z}$ to be our universal set. Then
$S \cap T =$ set of those integers that are prime numbers as well as divisors of 10
$= \{2, 5\}$.

***

Example 5: Let A be the set of all human beings living in Bihar, B be the set of all women and C be the set of all Indian cities. Describe $A \cap B$ and $A \cap C$.

Solution: We can take U to be the set of all human beings and all the cities of India. Then $A \cap B$ is the set of all women living in Bihar, and $A \cap C$ is the empty set.

***

You should be able to do the following exercise now.

E17) Obtain the sets $\mathbb{Z}\cap\mathbb{Q}, \mathbb{Q}\cap\mathbb{Z}, \mathbb{Z}\cap\mathbb{Z}$ and $\mathbb{Z}\cap\emptyset$. In each case, clearly state what U is.

While solving E17 you may have noticed certain facts about the operation of intersection. We explicitly list them in the following theorem. You will study the proofs of some of them, and then prove the rest yourself.

Theorem 1: For any two sets A and B in a universal set U,

i) $A\cap B\subseteq A$

ii) $A\cap B\subseteq B$

iii) $A\subseteq B \Rightarrow A\cap B = A$

iv) $A\cap A = A$

v) $A\cap\emptyset = \emptyset = B\cap\emptyset$

vi) $A\cap U = A, B\cap U = B$

vii) $A\cap B = B\cap A$ (i.e., the operation of intersection of sets is **commutative**.)

viii) $A\setminus B = A\cap B^c$

ix) If C is a subset of U such that $C\subseteq A$ and $C\subseteq B$, then $C\subseteq A\cap B$.

Proof: We will prove (i) and (ii), and leave you to prove the rest (see E18). So, to prove these facts, we need to show that every element of $A\cap B$ is an element of A as well as of B. For this, let $x\in A\cap B$. Then, by definition $x\in A$ and $x\in B$.
This is true for any element x of $A\cap B$.
Thus, $A\cap B\subseteq A$ and $A\cap B\subseteq B$. So we have proved both (i) and (ii). ■

Remark 11: The operation of intersection is meaningless unless we are clear about our universal set. However, we usually tend not to write the universal set explicitly, and assume it as understood.

Also note that if, for example, in Example 4 we had taken the universal set to be $\mathbb{Q}$, we would still have got $S\cap T = \{2, 5\}$. Thus, the choice of the universal set does not change the set $S\cap T$. In fact, the intersection of two sets just involves the elements in these two sets, and hence is independent of the choice of the universal set.

Now, using (i) and (ii) of Theorem 1, try to do the following exercises.

E18) Prove (iii) – (ix) of Theorem 1.

E19) Does Theorem 1 (ix) remain true if we replace '$\subseteq$' by '$\subsetneq$' everywhere? Give reasons for your answer.

Now, consider the set $\mathbb{Q}^-$ of negative rational numbers and the set $\mathbb{Q}^+$ of positive rational numbers. Then $\mathbb{Q}^-\cap\mathbb{Q}^+ = \emptyset$, that is, they have no elements in common. This pair of sets is an example of what we will now define.

Definition: Let A and B be two sets such that $A \cap B = \emptyset$. Then A and B are called **mutually disjoint** (or **disjoint**).

For example, the sets A and A^c in Fig. 7 are disjoint. In fact, for any set A, $\mathbf{A \cap A^c = \emptyset}$. (Why?)

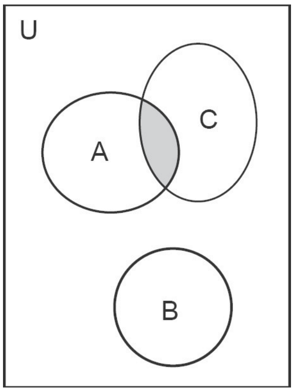


Fig. 8

Now let us represent the intersection of sets by means of Venn diagrams. Consider Fig. 8. The shaded region in Fig. 8 represents the set $A \cap C$, which is non-empty, as you can see. Also note that the regions representing A and B do not overlap, that is, $A \cap B = \emptyset$, that is, A and B are disjoint. From this diagram, we can also see that neither is $A \subseteq C$, nor is $C \subseteq A$. Further, both $C \setminus A$ and $A \setminus C$ are non-empty sets. See how much information a Venn diagram can convey!

Now, go back to Fig. 4 for a moment. What situation does it represent? It shows two sets A and B, with $A \subseteq B$. So, the shaded area shows that $A \cap B = A$.

So far we have considered the intersection of two sets. Now let us consider the intersection of more than 2 sets, through an example first.

Example 6: Let A, B and C be the sets of multiples of $3, 6$ and 10 in $\mathbb{N}$, respectively. Obtain $A \cap B \cap C$. Also draw a Venn diagram to represent all these sets.

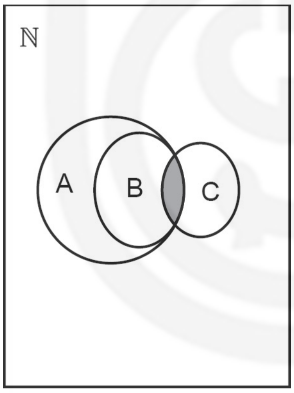


Fig. 9: The shaded portion represents $\mathbf{A \cap B \cap C}$.

Solution: Here $A = \{3x \mid x \in \mathbb{N}\}, B = \{6x \mid x \in \mathbb{N}\}, C = \{10x \mid x \in \mathbb{N}\}$.

Let us take $U = \mathbb{N}$.

$A \cap B \cap C$ will consist of all those natural numbers that belong to all three sets, A, B and C. Thus,

$A \cap B \cap C = \{x \in \mathbb{N} \mid 3, 6 \text{ and } 10 \text{ divide } x\}$

$= \{x \in \mathbb{N} \mid x = 3a,\ x = 6b,\ x = 10c \text{ for some } a, b, c \in \mathbb{N}\}$

$= \{x \in \mathbb{N} \mid 30 \text{ divides } x\}$ (since $30 \mid x$ iff '$3, 6$ and 10 divide x')

$= \{30y \mid y \in \mathbb{N}\}$. (Did you notice that 30 is the lcm of $3, 6, 10$?)

Now consider the Venn diagram in Fig. 9, where we show A, B, C and $A \cap B \cap C$ in $\mathbb{N}$. Note here that $B \subset A$ and $A \setminus B$ does not intersect C.

***

This example, and your understanding of intersection, will have helped you develop a definition of the intersection of 3 or more sets. Does this match the definition we now give?

$\forall$ denotes 'for all', or 'for every'.

Definition: The **intersection of $\mathbf{n}$** sets $A_1, A_2, \ldots, A_n$ in a universal set U is defined to be the set $\{x \in U \mid x \in A_i \ \forall\ i = 1, \ldots, n\}$. This is denoted by

$A_1 \cap A_2 \cap \ldots \cap A_n$, or $\bigcap\limits_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{A_i}$.

Let us look at another example involving the intersection of 3 sets.

Example 7: Let $S = \left\{ \frac{p}{q} \in \mathbb{Q} \,\middle|\, \frac{p}{q} = \frac{1}{2} \right\}$ and $T = \left\{ \frac{1}{n} \,\middle|\, n \in \mathbb{N} \right\}$. Find $S \cap T \cap \mathbb{Q}$.

Further, find two sets A and B, different from S or T, such that $A \cap B = S \cap T$.

Solution: Note that $S \subseteq \mathbb{Q}, T \subseteq \mathbb{Q}$. Therefore, $S \cap T \cap \mathbb{Q} = \{x \in \mathbb{Q} \mid x \in S \text{ and } x \in T\} = S \cap T$, since every element of S or of T is already an element of $\mathbb{Q}$.

Now take $x \in S \cap T$. As $x \in S$, $x = \frac{p}{q}$, for some p, q with $2p = q$, that is,

$x = \frac{p}{2p}, p \neq 0$.

As $x \in T$, $x = \frac{1}{n}$ for some $n \in \mathbb{N}$.

So, $\frac{p}{2p} = \frac{1}{n}$ shows that $n = 2$, that is, $x = \frac{1}{2}$.

Therefore, $S \cap T \cap \mathbb{Q} = S \cap T = \left\{\frac{1}{2}\right\}$.

Now for the second part of the problem. We can find many such sets A and B. For instance, take $A = \left\{x \in \mathbb{R} \;\middle|\; x \leq \frac{1}{2}\right\}$ and $B = \left\{x \in \mathbb{Q} \;\middle|\; x \geq \frac{1}{2}\right\}$.

Check that $A \cap B = S \cap T$.

***

Try the following exercises now.

E20) Let $A = \{1, 2, 3, 4\}, B = \{3, 4, 5, 6\}$ and $C = \{1, 4, 7, 8\}$.
Determine $A \cap B \cap C$. Represent all these sets in a Venn diagram.
Also verify that

i) $A \cap B \cap C = (A \cap B) \cap C$,

ii) $A \cap B \cap C = A \cap (B \cap C)$,

iii) $A \cap B \cap C = A \cap C \cap B$.

E21) If $A = \{6n \mid n \in \mathbb{N}\}, B = \{4n \mid n \in \mathbb{N}\}$ and C is the set of prime numbers, then find $A \cap B \cap C$.

What you have shown in E20 is not only true for those sets; it is true for any three sets A, B and C. (i) and (ii) say that $\cap$ is an **associative** operation. When we combine this with (vii) of Theorem 1, we see that **we can obtain the intersection of any n sets by intersecting any two of these sets at a time**.

For example, if A, B, C, D are sets, then

$$A \cap B \cap C \cap D = (A \cap B) \cap (C \cap D)$$
$$= (B \cap C) \cap (A \cap D), \text{ etc.}$$

Thus, we choose to combine those sets, whose intersection helps us to find the overall intersection more quickly.

Let us now look at another operation on sets.

1.5.3 Union

Let us again come back to the sets $\mathbb{N}$ and $\{0, 1\}$. You have seen that

$\mathbb{N} \setminus \{0, 1\} = \{2, 3, 4, \ldots\}$ and $\mathbb{N} \cap \{0, 1\} = \{1\}$. Now, consider the set consisting of all the elements in $\mathbb{N}$ along with all the elements in $\{0, 1\}$. What will this set be? It will be $\{0, 1, 2, \ldots\}$, i.e., the set of whole numbers $\mathbf{W}$. Note that 1 is repeated when we take the elements of both the sets, but we do not repeat an element in listing a set, as noted in Sec. 1.2.

Again, consider the sets $A = \{x \in \mathbb{R} \mid x \le 10\}$ and $B = \{x \in \mathbb{R} \mid x \ge 10\}$. Take the set consisting of all the elements in A and all those in B. This will be $\mathbb{R}$ because any real number will be either less than or equal to 10 or greater than or equal to 10, and 10 will belong to both A and B.

These examples lead us to define the operation we are undertaking in them.

Definition: Let A and B be two sets in a universal set U. The set of all those elements of U which belong either to A, or to B, or to both A and B, is called the **union of A and B**. This is symbolically written as $\mathbf{A \cup B}$, and is read as 'A union B'.

Thus, $A \cup B = \{x \in U \mid x \in A \text{ or } x \in B \text{ or } x \in A \cap B\}$.

However, you know that $A \cap B \subseteq A$. So, $x \in A \cap B \Rightarrow x \in A$. Therefore, we remove the repetition, which is $A \cap B$, and write

$\mathbf{A \cup B = \{x \mid x \in A \text{ or } x \in B\}}$.

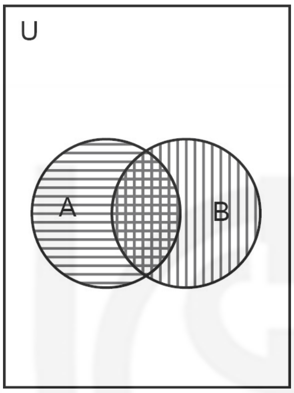


Fig. 10: The lined region is $\mathbf{A \cup B}$, of which the double lined region is $\mathbf{A \cap B}$.

So, for example, at the beginning of this sub-section you have seen that $\mathbb{N} \cup \{0, 1\} = \mathbf{W}$ and $\{x \in \mathbb{R} \mid x \le 10\} \cup \{x \in \mathbb{R} \mid x \ge 10\} = \mathbb{R}$.

Before going further we make a remark.

Remark 12: Since $A \cup B$ contains all the elements of A as well as of B, it follows that

$\mathbf{A \subseteq A \cup B, B \subseteq A \cup B}$.

In fact, $A \cap B \subseteq A \subseteq A \cup B$.

We can show this fact in a Venn diagram, as in Fig. 10.

Now let us look at another example.

Example 8: Find $\mathbb{N} \cup \mathbb{Z}$. Also show this in a Venn diagram .

Solution: $\mathbb{N} = \{1, 2, 3, \ldots\}$ and $\mathbb{Z} = \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$.

We want to find $\mathbb{N} \cup \mathbb{Z} = \{x \mid x \in \mathbb{N} \text{ or } x \in \mathbb{Z}\}$.

Since $\mathbb{N} \subseteq \mathbb{Z}$, $x \in \mathbb{N} \Rightarrow x \in \mathbb{Z}$. This tells us that $\mathbb{N} \cup \mathbb{Z} = \{x \mid x \in \mathbb{Z}\} = \mathbb{Z}$.

The Venn diagram is given in Fig. 11, in which $\mathbb{Z}$ is taken as the universal set.

***

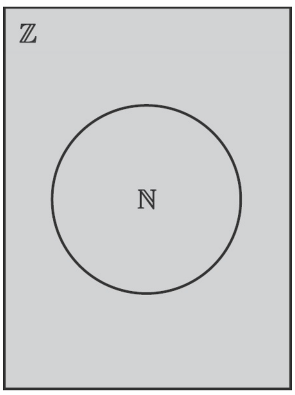


Fig. 11: The shaded region is $\mathbb{N} \cup \mathbb{Z}$, i.e., $\mathbb{Z}$.

Example 8 is a particular case of the general fact that we now state and prove.

Theorem 2: For any two sets A and B, $\mathbf{A \subseteq B \Leftrightarrow A \cup B = B}$.

Proof: Here we have to prove two statements.

One is that if $A \subseteq B$, then $A \cup B = B$.

The other is **the converse**, namely, if $A \cup B = B$, then $A \subseteq B$.

First, let us assume $A \subseteq B$, and let us prove that $A \cup B = B$.

For this, take any $x \in A \cup B$.

If $x \in A$, then $x \in B$, since $A \subseteq B$.

If $x \notin A$, then too $x \in B$, since $x \in A \cup B$.

So, every element of $A \cup B$ is an element of B, that is, $A \cup B \subseteq B$.
Also, we know that $B \subseteq A \cup B$. Therefore, using Remark 7, we have proved that $A \cup B = B$.
Conversely, assume that $A \cup B = B$. We also know that $A \subseteq A \cup B$.
Hence, $A \subseteq B$, which is what we wanted to prove. ■

You can use this theorem while solving the following exercises.

E22) For any three sets A, B and C, in a universal set U, show that

i) $A \cup A = A$.

ii) $A \cup B = B \cup A$, that is, the **operation of union is commutative**.

iii) $A \cup \emptyset = A$.

iv) If $A \subseteq C$ and $B \subseteq C$, then $A \cup B \subseteq C$.

v) $A \cup A^c = U$.

You can also see the video at https://www.youtube.com/watch?v=uR70knMr2Hg

E23) Let U be the real line $\mathbb{R}$, $A = \{x \in \mathbb{R} \mid 0 \le x \le 1\}$ and $B = \{x \in \mathbb{R} \mid 1 \le x \le 3\}$. Determine $A \cup B$.
Also find two distinct subsets S and T of $\mathbb{R}$, different from A and B, such that $S \cup T = A \cup B$. Justify your choice of S and T.

E24) What can you say about the number of elements in A and B if $A \cup B = \emptyset$?

Just as in the case of '$\cap$', we can define the union of 3 or more sets, in a very natural way.

Definition: The **union of n sets** $A_1, A_2, \ldots, A_n$ in a universal set U is the set $\{x \in U \mid x \in A_i$ for **some** i such that $1 \le i \le n\}$. This is denoted by $A_1 \cup A_2 \cup \ldots \cup A_n$, or $\bigcup_{i=1}^{n} A_i$.

Let's consider an example.

Example 9: Find $A \cup B \cup C$, where $A = \{1, 2, 3\}$, $B = \{2, 3, 4, 5\}$, $C = \{1\}$.
Also, check whether $A \cup B = (A \setminus B) \cup (A \cap B) \cup (B \setminus A)$ or not.

Solution: You can check that $A \cup B \cup C = \{1, 2, 3, 4, 5\}$.
Also, $A \cup B = \{1, 2, 3, 4, 5\}$, $A \setminus B = \{1\}$, $B \setminus A = \{4, 5\}$, $A \cap B = \{2, 3\}$.
Therefore,

$$\begin{aligned}(A \setminus B) \cup (A \cap B) \cup (B \setminus A) &= \{1, 2, 3, 4, 5\}. \\ &= A \cup B.\end{aligned}$$

***

In the example above, you can also see that $A \cup B \cup C$ is the same as $(A \cup B) \cup C$ as well as $A \cup (B \cup C)$.

What we have noted in Example 9 are particular cases of the general facts that we ask you to prove in the following exercises.

E25) For any three subsets A, B, C of a set U, show that
$$A \cup B \cup C = (A \cup B) \cup C = A \cup (B \cup C).$$

The operation of union of sets is associative.

E26) For any two sets A and B, show that
$$A \cup B = (A \setminus B) \cup (A \cap B) \cup (B \setminus A).$$
(We depict this situation in the Venn diagram in Fig. 12.)

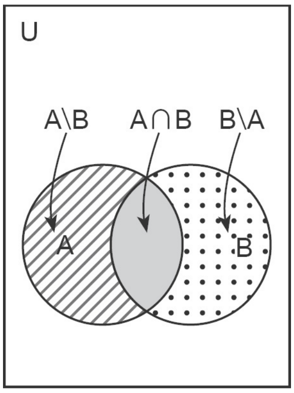


Fig. 12: $A \cup B$ is the whole filled-in region.

What you see from E22 (ii) and E25 is that **we can obtain the union of any number of sets by taking the union of any two of these sets at a time**. For example, if A, B, C, D are four sets, then

$$A \cup B \cup C \cup D = [(A \cup B) \cup C] \cup D$$
$$= A \cup [(B \cup C) \cup D]$$
$$= A \cup [D \cup (B \cup C)]$$
$$= (A \cup D) \cup (B \cup C).$$

Let us now discuss the fourth operation on sets that we had planned to at the beginning of this section.

1.5.4 Cartesian Product

Fig. 13: René Descartes

An interesting set that can be formed from two given sets is their Cartesian product, named after the French philosopher and mathematician, René Descartes (1596-1650). He also invented the Cartesian coordinate system that is used for plotting points in the XY-plane, and which we shall discuss in Unit 3. In fact, defining this operation helped Descartes to create the coordinate system, and hence, to study and understand geometry by using algebra.

Let us start by considering A, the set of first names of all the students of Class 4 of a certain Government school, and B the set of their birth dates. Then form all pairs of elements (a, b) where a is in A and b is in B. This collection is well-defined. Note that (Sarla, 17) can be an element of this set but (17, Sarla) cannot be, since Sarla is not a date and 17 is not a name. So, the **order of writing the pair** is important. We define this way of obtaining a new set below.

Definitions: i) Let A and B be two sets. The pair (a, b), in which the first element is from A and the second element is from B, is called an **ordered pair**.

ii) The **Cartesian product $A \times B$**, of the sets A and B, is the set of all ordered pairs (a, b), where $a \in A$, $b \in B$.
That is, $\mathbf{A \times B = \{(a, b) \mid a \in A \text{ and } b \in B\}}$.

If $a \neq b$, (a, b) and (b, a) are different ordered pairs.

iii) Two ordered pairs (a, b) and (c, d) are said to be **equal** (or the **same**) if $a = c$ and $b = d$.

For example, if $A = \{1, 2, 3\}, B = \{4, 6\}$, then
$A \times B = \{(1, 4), (1, 6), (2, 4), (2, 6), (3, 4), (3, 6)\}$, and
$B \times A = \{(4, 1), (4, 2), (4, 3), (6, 1), (6, 2), (6, 3)\}$.
You can see that $(1, 4) \in A \times B$, but $(1, 4) \notin B \times A$.

Therefore, $A \times B \neq B \times A$. (While studying Unit 3, you will learn to visualize $A \times B$.)

Now what does $A \times B$ look like if either $A = \emptyset$ or $B = \emptyset$? In this case $A \times B$ cannot have any elements. Thus, $A \times B = \emptyset$.

$A \times B = \emptyset$ iff $A = \emptyset$ or $B = \emptyset$.

Try these exercises now.

E27) If $A = \{2, 5\}$ and $B = \{2, 3\}$, find $A \times B, B \times A, A \times A, (A \times B) \cap (B \times A)$, $(A \times B) \cup (B \times A)$ and $A \setminus (B \times A)$.

E28) If $A \times B = \{(7, 2), (7, 3), (7, 4), (2, 2), (2, 3), (2,4)\}$, determine A and B.

E29) If $A \subseteq C, B \subseteq D$, then show that $A \times B \subseteq C \times D$.

E30) Find 3 distinct elements of $\mathbb{N} \times (\mathbb{Q} \setminus \mathbb{N})$.

E31) Give an example of a proper non-empty subset S of $\mathbb{R} \times \mathbb{R}$. Also give an element of $(\mathbb{R} \times \mathbb{R}) \setminus S$.

In the examples and exercises above, have you found any relationship between the cardinality of $A \times B$ and the cardinalities of A and B? Look again and consider the remark below.

Remark 13: i) **If $|A| = n, |B| = m$, then $|A \times B| = nm$.**

ii) **If A and B are non-empty sets and either one of them is infinite, then $A \times B$ will be infinite.**

Now that we have defined the Cartesian product of two sets, let us extend the definition to any number of sets.

Definition: Let $A_1, A_2, \ldots, A_n$ be n sets. Then their **Cartesian product** is the set $\mathbf{A_1 \times A_2 \times \ldots \times A_n = \{(x_1, x_2, \ldots, x_n) \mid x_i \in A_i \forall\, i = 1, 2, \ldots, n\}}$.
The element $(x_1, x_2, \ldots, x_n)$ is called an **n-tuple**.

For example, if $\mathbb{R}$ is the set of real numbers, then
$\mathbb{R} \times \mathbb{R} = \{(a_1, a_2) \mid a_1 \in \mathbb{R}, a_2 \in \mathbb{R}\}$,
$\mathbb{R} \times \mathbb{R} \times \mathbb{R} = \{(a_1, a_2, a_3) \mid a_i \in \mathbb{R} \forall\, i = 1, 2, 3\}$, and so on.
It is customary to write $\mathbb{R}^2$ for $\mathbb{R} \times \mathbb{R}$ and $\mathbb{R}^n$ for $\mathbb{R} \times \ldots \times \mathbb{R}$ (n times). Also, a 2-tuple is usually called an **ordered pair**, and a 3-tuple is usually called an **ordered triple**.

Try the following exercise now.

E32) Which of the following belong to the Cartesian product $\mathbb{Q} \times \mathbb{Z} \times \mathbb{N}$? Give reasons for your answers.

i) $(3, 0)$ ii) $\left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)$ iii) $(1, 1, 1)$

iv) $\left(\frac{1}{2}, -5, \sqrt{2}\right)$ v) $\{-2, 2, 3\}$.

In this section we have discussed four operations on sets. You have also noted a relation them between some of these operations, in E26 for example. Let us now see if there are any other relationships satisfied by these operations that help us in carrying them out more efficiently.

1.6 LAWS ON OPERATIONS

You must be familiar with the law of distributivity that connects the operations of multiplication and addition of real numbers, namely,
$a \times (b+c) = a \times b + a \times c \; \forall \; a, b, c \in \mathbb{R}$.
Here we say that multiplication distributes over addition. We have similar laws governing the operations on sets that you have studied so far. You shall study two kinds of such laws in this section.

1.6.1 Distributive Laws

The name of these laws may give you an idea of what to expect here – may be some process like that in arithmetic? Let us look at an example.

Example 10: Check whether or not $\mathbb{N} \cap (\mathbb{Q} \cup \mathbb{R}) = (\mathbb{N} \cap \mathbb{Q}) \cup (\mathbb{N} \cap \mathbb{R})$. Also, if you interchange $\cap$ with $\cup$ in the equation above, will you still get an equation?

Solution: To check whether or not $\mathbb{N} \cap (\mathbb{Q} \cup \mathbb{R}) = (\mathbb{N} \cap \mathbb{Q}) \cup (\mathbb{N} \cap \mathbb{R})$, note that $\mathbb{Q} \cup \mathbb{R} = \mathbb{R}$, since $\mathbb{Q} \subseteq \mathbb{R}$.
Therefore, $\mathbb{N} \cap (\mathbb{Q} \cup \mathbb{R}) = \mathbb{N} \cap \mathbb{R} = \mathbb{N}$, since $\mathbb{N} \subseteq \mathbb{R}$. … (1)
Also $\mathbb{N} \cap \mathbb{Q} = \mathbb{N}$ and $\mathbb{N} \cap \mathbb{R} = \mathbb{N}$.
Therefore, $(\mathbb{N} \cap \mathbb{Q}) \cup (\mathbb{N} \cap \mathbb{R}) = \mathbb{N} \cup \mathbb{N} = \mathbb{N}$. … (2)
Thus, from (1) and (2), we see that $\mathbb{N} \cap (\mathbb{Q} \cup \mathbb{R}) = (\mathbb{N} \cap \mathbb{Q}) \cup (\mathbb{N} \cap \mathbb{R})$ … (3)
So, in this case $\cap$ distributes over $\cup$.

LHS stand for 'left hand side' of an equation. Similarly, RHS stands for 'right hand side' of the equation.

Now, if we interchange $\cap$ with $\cup$ in (3) above, we get $\mathbb{N} \cup (\mathbb{Q} \cap \mathbb{R}) = (\mathbb{N} \cup \mathbb{Q}) \cap (\mathbb{N} \cup \mathbb{R})$. Note that the LHS and RHS are both equal to $\mathbb{Q}$. Hence, this is a correct equation too, i.e., in this case $\cup$ distributes over $\cap$.

***

What this example shows us **is a particular** case of the following theorem.

Theorem 3 (Laws of Distributivity): Let A, B and C be three subsets of a universal set U. Then

i) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$, that is, **intersection distributes over union**;

ii) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$, that is, **union distributes over intersection**.

Proof: We will prove (i) and ask you to prove (ii) (see E33).

i) We know that two sets are equal if and only if each is a subset of the other. So, we will show that

$A \cap (B \cup C) \subseteq (A \cap B) \cup (A \cap C)$ and … (4)

$(A \cap B) \cup (A \cap C) \subseteq A \cap (B \cup C)$ … (5)

Now, $x \in A \cap (B \cup C)$

$\Rightarrow$ $x \in A$ and $x \in B \cup C$

$\Rightarrow$ $x \in A$ and ($x \in B$ or $x \in C$)

$\Rightarrow$ ($x \in A$ and $x \in B$) or ($x \in A$ and $x \in C$)

$\Rightarrow$ $x \in A \cap B$ or $x \in A \cap C$

$\Rightarrow$ $x \in (A \cap B) \cup (A \cap C)$

So, we have proved the first inclusion, (4).

To prove (5), let $x \in (A \cap B) \cup (A \cap C)$

$\Rightarrow$ $x \in A \cap B$ or $x \in A \cap C$

$\Rightarrow$ ($x \in A$ and $x \in B$) or ($x \in A$ and $x \in C$)

$\Rightarrow$ $x \in A$ and ($x \in B$ or $x \in C$)

$\Rightarrow$ $x \in A$ and $x \in B \cup C$

$\Rightarrow$ $x \in A \cap (B \cup C)$

So we have proved (5), and hence (i) of Theorem 3. ■

Did you notice that our argument for proving (5) is just the reverse of our argument for proving (4)? In fact, we could have combined the proofs of (4) and (5), using the **two-way** implication, as follows:

$x \in A \cap (B \cup C)$

$\Leftrightarrow$ $x \in A$ and $x \in B \cup C$

$\Leftrightarrow$ $x \in A$ and ($x \in B$ or $x \in C$)

$\Leftrightarrow$ ($x \in A$ and $x \in B$) or ($x \in A$ and $x \in C$)

$\Leftrightarrow$ $x \in A \cap B$ or $x \in A \cap C$

$\Leftrightarrow$ $x \in (A \cap B) \cup (A \cap C)$

This proves Theorem 3 (i).

In Fig. 14 we have given a Venn diagrammatic representation of Theorem 3(i).

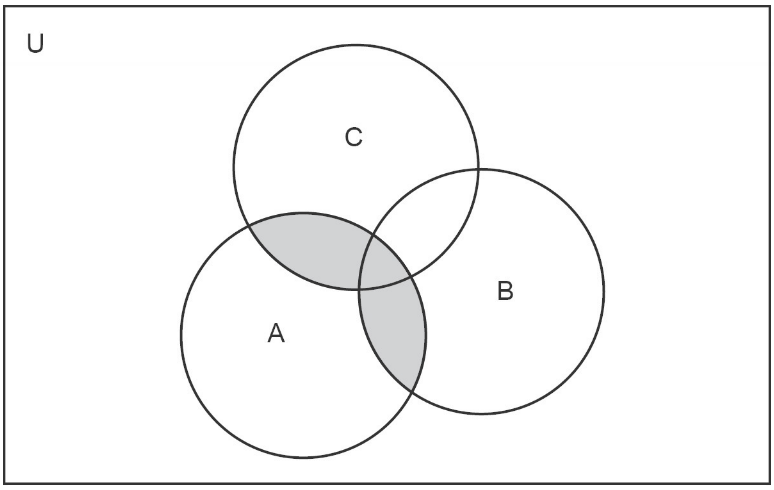


Fig.14: $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

Now try to solve the following exercise, using the two-way implication, $\Leftrightarrow$.

E33) Prove (ii) of Theorem 3. Also represent this statement in a Venn diagram.

Let us now consider some properties relating the Cartesian product of sets with the other operations on sets.

Theorem 4: For any 3 sets A, B, C, prove that $(A \cup B) \times C = (A \times C) \cup (B \times C)$, that is, **the Cartesian product of sets distributes over the union of sets**.

Proof: Firstly, since $A \subseteq A \cup B,\ A \times C \subseteq (A \cup B) \times C$.
Similarly, $B \times C \subseteq (A \cup B) \times C$.
Therefore, $(A \times C) \cup (B \times C) \subseteq (A \cup B) \times C$.
So, we need to prove that $(A \cup B) \times C \subseteq (A \times C) \cup (B \times C)$. … (6)
For this, consider any ordered pair $(x, y) \in (A \cup B) \times C$.
$\Rightarrow \ x \in A \cup B,\ y \in C$.
$\Rightarrow \ x \in A$ or $x \in B,\ y \in C$.
$\Rightarrow \ (x, y) \in A \times C$ or $(x, y) \in B \times C$.
$\Rightarrow \ (x, y) \in (A \times C) \cup (B \times C)$
Hence, (6) is proved, and the theorem is proved. ■

Why not try an exercise now?

E34) Prove that $(A \cap B) \times C = (A \times C) \cap (B \times C)$, that is, **the Cartesian product distributes over intersection**.

You may wonder what use these properties are. Let us consider an example.

Example 11: Let A be the set of irrational numbers and $B = \{x \in \mathbb{R} \mid x < 3\}$
Find $(A \cap B) \cup (\mathbb{Q} \cap B)$.

Solution: Now, try to find $A \cap B$ as well as $\mathbb{Q} \cap B$. Then find their intersection. This is one way of solving the problem, but a rather difficult one. On the other hand, note that $A \cup \mathbb{Q} = \mathbb{R}$. So, applying the distributive laws, we find that $(A \cap B) \cup (\mathbb{Q} \cap B) = (A \cup \mathbb{Q}) \cap B = \mathbb{R} \cap B = B$.
So, distributivity has simplified our calculations!

***

Let us now consider some laws involving the operation of complementation.

Fig. 15: De Morgan

1.6.2 De Morgan's Laws

We will now discuss two laws that relate the operation of finding the complement of a set to that of the intersection or union of sets. These are known as De Morgan's laws, after the British mathematician Augustus De Morgan (1806-1871). Let us first consider an example.

Example 12: Let $A = \{-2, 2\},\ B = \{1, 3, 5, 7, \ldots\},\ U = \mathbb{Z}$. Check whether or not $(A \cup B)^c = A^c \cap B^c$.

Solution: Here $(A \cup B)^c = \mathbb{Z} \setminus \{-2, 1, 2, 3, 5, 7, \ldots\}$

$= \{\ldots, -5, -4, -3, -1, 0, 4, 6, 8, \ldots\}$.

Also, $A^c \cap B^c = (\mathbb{Z} \setminus \{-2, 2\}) \cap (\mathbb{Z} \setminus \{1, 3, 5, \ldots\})$

$$= \{\ldots, -4, -3, -1, 0, 1, 3, 4, 5, \ldots\} \cap \{\ldots, -3, -2, -1, 0, 2, 4, 6, \ldots\}$$

$$= \{\ldots, -5, -4, -3, -1, 0, 4, 6, 8, \ldots\}.$$

So, $(A \cup B)^c = A^c \cap B^c$.

***

This interesting way in which the operation of complementation 'interchanges' $\cup$ and $\cap$, is true for all sets, not just those in the example above. Let us state these laws, and prove them now.

Theorem 5 (De Morgan's Laws): For any two sets A and B in a universal set U,

i) $(A \cap B)^c = A^c \cup B^c$,

ii) $(A \cup B)^c = A^c \cap B^c$.

Proof: As in the case of Theorem 3, we will prove (i), and ask you to prove (ii) (see E35). So, let us take $x \in (A \cap B)^c$. Now,

$x \in (A \cap B)^c = U \setminus (A \cap B)$

$\Leftrightarrow$ $x \notin A \cap B$, (that is, x does not belong to both A and B.)

$\Leftrightarrow$ $x \notin A$ or $x \notin B$ (because if $x \in A$ and $x \in B$, then $x \in A \cap B$)

$\Leftrightarrow$ $x \in A^c$ or $x \in B^c$

$\Leftrightarrow$ $x \in A^c \cup B^c$

So $(A \cap B)^c = A^c \cup B^c$. ■

The De Morgan's laws can be quite useful for using the operations efficiently. Consider the following example.

Example 13: Consider the sets $\mathbb{Z}$, $A = \{x \in \mathbb{Q} \mid x \geq \sqrt{2}\}$, $U = \mathbb{Q}$. Find $A^c \cup \mathbb{Z}^c$.

Solution: Now, $A^c = \{x \in \mathbb{Q} \mid x < \sqrt{2}\}$ and $\mathbb{Z}^c = \mathbb{Q} \setminus \mathbb{Z}$.

Can we write $A^c \cup \mathbb{Z}^c$ in a manner that its elements are clear to us? It doesn't seem so.

However, by De Morgan's laws, $A^c \cup \mathbb{Z}^c = (A \cap \mathbb{Z})^c$.

Now, $A \cap \mathbb{Z} = \{x \in \mathbb{Z} \mid x \geq \sqrt{2}\} = \{2, 3, \ldots\}$

So $(A \cap \mathbb{Z})^c = \mathbb{Q} \setminus (A \cap \mathbb{Z}) = \mathbb{Q} \setminus \{2, 3, 4, \ldots\}$

So $A^c \cup \mathbb{Z}^c = \mathbb{Q} \setminus \{2, 3, \ldots\}$, a clear way of looking at the required set.

***

Now try the following exercises.

E35) Prove (ii) of Theorem 5.

E36) Verify De Morgan's laws for A and B, where $A = \{1, 2\}$, $B = \{2, 3, 4\}$. (For convenience, you can take $U = \{1, 2, 3, 4\}$, i.e., $U = A \cup B$. Of course, the laws will continue to hold true with any other U.)

E37) For any three sets A, B, C in a universal set U, prove that $(A \setminus B) \times C = (A \times C) \setminus (B \times C)$. Also write down what the set U could be.

E38) Is the statement, 'For any three sets $A, B, C, (A \setminus B) \times C = A \setminus (B \times C)$' true? Give reasons for your answer.

By now you would be familiar with sets, several operations on sets, and laws relating these operations. Let us take an overview of what we have covered in this unit.

1.7 SUMMARY

In our discussion on sets we have brought out the following points:

1. A set is a well-defined collection of objects.

2. The listing method and property method for representing sets. Some sets can be represented by both methods, and some by only one of them.

3. The concepts of 'subset', 'proper subset', 'superset', 'universal set', 'finite set' and 'infinite set'.

4. The sets A and B **are equal** iff A is a subset of B and B is a subset of A, that is, $A = B$ iff $A \subseteq B$ and $B \subseteq A$.

5. A is **not a subset** of B if $\exists$ $a \in A$ s.t. $a \notin B$.

6. The pictorial representation of sets and their relationships by Venn diagrams, and the utility of this type of representation.

7. The definition, and examples, of the complement of a set A in a set B, denoted by $B \setminus A$.
   $B \setminus A = \{x \in B \mid x \notin A\}$.
   When B is a universal set, $B \setminus A$ is called the complement of A, and is denoted by A' or A^c.

8. The definition, and examples, of the intersection of two sets A and B in a universal set U, denoted by $A \cap B$.
   $A \cap B = \{x \in U \mid x \in A$ **and** $x \in B\}$.
   This definition extends to more than two sets, in a natural manner, as
   $A_1 \cap A_2 \cap \ldots \cap A_n = \{x \in U \mid x \in A_i \; \forall \; i = 1, \ldots, n\}$.

9. Several properties of the operation of intersection of sets.

10. The definition, and examples, of the union of two subsets A and B of a universal set U, denoted by $A \cup B$.
    $A \cup B = \{x \in U \mid x \in A$ **or** $x \in B\}$.
    This definition extends to more than two sets, in a natural way, as
    $A_1 \cup A_2 \cup \ldots \cup A_n = \{x \in U \mid x \in A_i$ for some $i = 1, \ldots, n\}$.

11. Several properties of the operation of union of sets.

12. The definition, and examples, of the Cartesian product of the sets A and B, denoted by $A \times B$.
 $A \times B = \{(x, y) \mid x \in A, y \in B\}$.
 This definition extends to the Cartesian product of n sets, as follows:
 $A_1 \times A_2 \times \ldots \times A_n = \{(x_1, x_2, \ldots, x_n) \mid x_i \in A_i \ \forall\, i = 1, \ldots, n\}$.

13. Several properties of the Cartesian product of sets.

14. The statement, and proof, of the **distributive laws:** For any three sets A, B, C,
 $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$;
 $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.

15. The statement, and proof, of **De Morgan's laws:** For any two sets A and B,
 $(A \cup B)^c = A^c \cap B^c$;
 $(A \cap B)^c = A^c \cup B^c$.

Now, we suggest that you **go back to the objectives given in Sec. 1.1,** and see if you have achieved them. One way of checking this is to solve all the exercises in the unit. If you would like to know what our solutions are, we have given them in the next section. But please do not look at them until you have tried to solve all the exercises on your own.

1.8 SOLUTIONS/ANSWERS

E1) (ii) – (v) are sets.
 (i) is not a set because one person may think Asha, for example, is interested in the game, while another may think she is not. So there are no clear-cut criteria for assessing interest, and hence the collection is not well-defined.
 Similarly, (vi) is not a set.

E2) A given stamp collection is a set because given any object it is clear to anybody firstly, whether it is a stamp or not; and secondly, if it is a stamp, then whether it belongs to the collection or not.

E3) (iii), (v) are true. There could be several ways of making changes to the rest. For instance, altering $\in$ to $\notin$, or picking an appropriate number or set to make the statement true. For example, (i) can be changed to $0.2 \notin \mathbb{N}$, or $0.2 \in \mathbb{Q}$.

E4) i) $\{-1\}$
 ii) This is the set of all integers less than or equal to 8, i.e., $\{\ldots, -3, -2, \ldots, 6, 7, 8\}$.
 iii) $\{a, b, c, \ldots, x, y, z\}$

E5) i) $\{x^2 \mid x \in \mathbb{N}\}$
 ii) This cannot be described by the property method since the three elements have nothing in common.
 iii) $\{x \mid x$ is an even integer$\}$
 iv) We can have several representations (see Remark 3). For example,

$\emptyset = \{x \in \mathbb{N} \mid x$ is both odd as well as even}, or

$\emptyset = \{x \in \mathbb{N} \mid x < 0\}$.

v) $\{x \mid x$ is a month with only 30 days}.

E6) The set on the LHS is $\{3\}$. The set on the RHS is $\{-3, 3\}$. Therefore, they are not equal.

E7) i) The number of spoons of water may be 1000 or 1 million, but it is still some number. Thus, the set is finite.

ii) If x and y are different rational numbers, then $x+3$ and $y+3$ are different rational numbers also. So, for each $x \in \mathbb{Q}$, there is an element of the given set, and all these elements are as many as the number of elements in $\mathbb{Q}$. Since $\mathbb{Q}$ has infinitely many elements, this set will be infinite too.

iii) This is the set of irrational elements, and is infinite.

iv) $\{\mathbb{N}\}$ is a singleton, having only one element, $\mathbb{N}$. Thus, it is finite.

v) Between any two points on a circle, there is always another point on the circle. Therefore, this set is infinite.

vi) By convention, this set is finite.

If a finite set A has n elements, then its power set has 2^n elements.

E8) i) False, since on the RHS is an infinite set, and on the LHS is a singleton.

ii) True, since the only element in the set is $\emptyset$.

E9) $\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1,3\}, \{2, 3\}, \{1, 2, 3\}$

As you can see, the answers to (i) – (iv) are $1, 3, 3, 1$, respectively.

As the cardinality of the given set is 3, no subset can have more than 3 elements. Thus, the answer to (v) is 'none'.

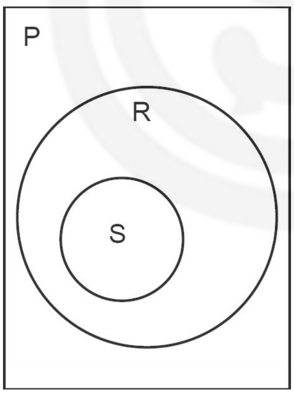


Fig. 16

E10) Let $x \in A$.

Then $A \subseteq B \Rightarrow x \in B$.

Now, $x \in B$ and $B \subseteq C \Rightarrow x \in C$.

$\therefore \quad x \in A \Rightarrow x \in C$

Since this is true for any $x \in A$, $A \subseteq C$.

E11) There are infinitely many possibilities here. One could be the given set itself! Another could be $\{3, 0.7,$ IGNOU, Mahatma Gandhi}.

E12) S, R, P are the sets of squares, rectangles and parallelograms, respectively. Since $S \subseteq R \subseteq P$, we have taken $U = P$ in the Venn diagram in Fig. 16.

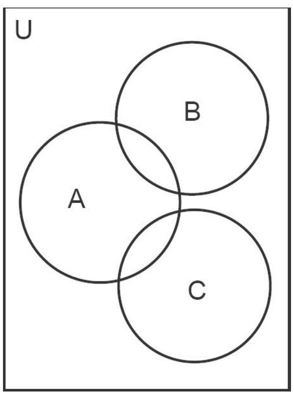


Fig. 17

E13) One possible diagram is given in Fig. 17. Note that just because B is infinite and C is finite does not mean that the regions depicting them need be of different areas.

E14) i) Note that $P \subseteq \mathbb{Z} \subseteq \mathbb{Q}$. We take $\mathbb{Q}$ as U (see Fig. 18).

ii) $\mathbb{Z} \setminus P = \{x \in \mathbb{Z} \mid x$ is not a prime number}.

This certainly contains the set of all negative numbers as a subset, which is infinite. Hence, $\mathbb{Z} \setminus P$ is infinite.

$P \setminus \mathbb{Z} = \{p \mid p$ is a prime number and $p \notin \mathbb{Z}\}$

$= \emptyset$, since every prime number is an integer.

So, $P \setminus \mathbb{Z}$ is finite.

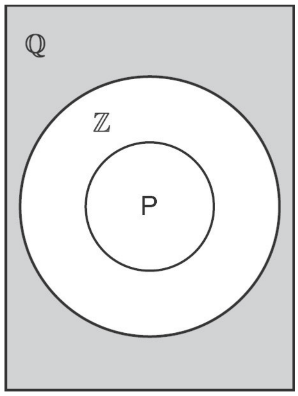


Fig. 18: The shaded area is $\mathbb{Q} \setminus \mathbb{Z}$.

E15) $A \setminus A = \emptyset$, $\emptyset \setminus A = \emptyset$, $A \setminus \emptyset = A$,
$(A^c)^c = A$, since $x \in A \Leftrightarrow x \notin A^c \Leftrightarrow x \in (A^c)^c$.

E16) i) $\mathbb{Q} \setminus \mathbb{N} \not\subseteq \mathbb{N}$, for example. There are many such examples.

ii) This can only be true if $A \setminus B = \emptyset$, because if $\exists x \in A \setminus B$, then $x \notin B$. But $A \setminus B \subseteq B \Rightarrow x \in B$. So we reach a contradiction, unless $A \setminus B = \emptyset$. So any two sets A and B for which $A \setminus B = \emptyset$ will be a valid example.

E17) $\mathbb{Z}, \mathbb{Z}, \mathbb{Z}, \emptyset$, respectively.
In all the cases U could be $\mathbb{Q}$. Of course, there are several other choices for U too.

E18) iii) From (i) we know that $A \cap B \subseteq A$. We need to prove that in this case $A \subseteq A \cap B$.
For this, let $x \in A$. Then, since $A \subseteq B$, $x \in B$. Thus, $x \in A \cap B$.
$\therefore \quad A \subseteq A \cap B$.
Since $A \subseteq A \cap B \subseteq A$, we see that $A = A \cap B$.

iv) Applying (iii) with $A = B$, we get $A \cap A = A$.

v) $A \cap \emptyset \subseteq \emptyset$, applying (ii).
Also $\emptyset \subseteq A \cap \emptyset$, since $\emptyset$ is a subset of every set.
$\therefore \quad A \cap \emptyset = \emptyset$.

vi) Since $A \subseteq U$, by (iii) above we get the result.

vii) $A \cap B \subseteq B$ and $A \cap B \subseteq A$. $\therefore \quad A \cap B \subseteq B \cap A$.
Similarly, $B \cap A \subseteq A \cap B$.
$\therefore \quad A \cap B = B \cap A$.

viii) $x \in A \setminus B \Leftrightarrow x \in A$ and $x \notin B$
$\Leftrightarrow \quad x \in A$ and $x \in B^c$
$\Leftrightarrow \quad x \in A \cap B^c$.
$\therefore \quad A \setminus B = A \cap B^c$.
(Note that we have used the **two-way implication**, $\Leftrightarrow$, at each stage, and hence, simultaneously shown that $A \setminus B \subseteq A \cap B^c$ and $A \cap B^c \subseteq A \setminus B$.)

ix) Let $x \in C$. Then $C \subseteq A \Rightarrow x \in A$. Similarly, $x \in B$. Therefore, $x \in A \cap B$. Hence, $C \subseteq A \cap B$.

E19) No. For example, if $A = \{1, 2, 3\}$, $B = \{1, 2, 4\}$, $C = \{1, 2\}$, then $C \subsetneqq A$ and $C \subsetneqq B$, but C is not **properly** contained in $A \cap B$; in fact, it is exactly **equal** to $A \cap B$.

E20) $A \cap B \cap C = \{4\}$

$(A \cap B) \cap C = \{3,4\} \cap C = \{4\}$

$A \cap (B \cap C) = A \cap \{4\} = \{4\}$

$A \cap C \cap B = (A \cap C) \cap B = \{1, 4\} \cap B = \{4\}$

E21) $A \cap B \cap C = (A \cap B) \cap C = \{12\, n \mid n \in \mathbb{N}\} \cap C = \emptyset$, since C is the set of primes and no element of $A \cap B$ is a prime.

E22) i) Since $A \subseteq A$, by Theorem 2, $A \cup A = A$.

ii) $x \in A \cup B \Leftrightarrow x \in A$ or $x \in B$

$\Leftrightarrow$ $x \in B$ or $x \in A$.

$\Leftrightarrow$ $x \in B \cup A$.

$\therefore$ $A \cup B = B \cup A$.

iii) $\emptyset \subseteq A \Rightarrow A \cup \emptyset = A$, by Theorem 2.

iv) Let $x \in A \cup B$. Then $x \in A$ or $x \in B$. In either case $x \in C$, since $A \subseteq C$ and $B \subseteq C$. Thus,

$x \in A \cup B \Rightarrow x \in C$.

$\therefore$ $A \cup B \subseteq C$.

v) By (iv) above $A \cup A^c \subseteq U$.

Now, let us show that $U \subseteq A \cup A^c$.

Let $x \in U$. Then, either $x \in A$ or $x \notin A$, that is, either $x \in A$ or $x \in U \setminus A = A^c$.

So, $x \in A \cup A^c$.

Therefore, $U \subseteq A \cup A^c$.

Hence, the equality is proved.

E23) $A \cup B = \{x \in \mathbb{R} \mid 0 \le x \le 3\}$.

Take $S = \{x \in \mathbb{Q} \mid 0 \le x \le 3\}$ and $T = \{x \in \mathbb{R} \setminus \mathbb{Q} \mid 0 \le x \le 3\}$.

Then $S \cup T \subseteq A \cup B$.

Also, for any $x \in A \cup B$, x is either rational or irrational. Accordingly, $x \in S$ or $x \in T$.

Hence, $A \cup B \subseteq S \cup T$.

Thus, $S \cup T = A \cup B$.

There can be several other pairs S and T that satisfy this requirement. Look for at least one more pair.

E24) Since $A \subseteq A \cup B = \emptyset$, we see that $A \subseteq \emptyset$. Also, $\emptyset \subseteq A$ always.

$\therefore$ $A = \emptyset$. Similarly, $B = \emptyset$. Thus, $|A| = |B| = 0$.

E25) Firstly, let us show that $(A \cup B) \cup C = A \cup B \cup C$.

Since $A \cup B \subseteq A \cup B \cup C$ and $C \subseteq A \cup B \cup C$,

by E22, $(A \cup B) \cup C \subseteq A \cup B \cup C$.

Conversely, let $x \in A \cup B \cup C$. Then $x \in A$ or $x \in B$ or $x \in C$.

If $x \in A$, then $x \in A \cup B \subseteq (A \cup B) \cup C$.

Similarly, if $x \in B$ or $x \in C$ then $x \in (A \cup B) \cup C$.

Thus, $A \cup B \cup C \subseteq (A \cup B) \cup C$.

Therefore, $A \cup B \cup C = (A \cup B) \cup C$.

You can, similarly, show that $A \cup B \cup C = A \cup (B \cup C)$.

E26) $A \setminus B \subseteq A \subseteq A \cup B,\ B \setminus A \subseteq B \subseteq A \cup B,\ A \cap B \subseteq A \cup B.$

$\therefore\ (A \setminus B) \cup (A \cap B) \cup (B \setminus A) \subseteq A \cup B.$

Conversely, let $x \in A \cup B$. Then $x \in A$ or $x \in B$.
Now, there are only three possibilities for x:

i) $x \in A$ but $x \notin B$, that is, $x \in A \setminus B$, or

ii) $x \in A$ and $x \in B$, that is, $x \in A \cap B$, or

iii) $x \in B$ but $x \notin A$, that is, $x \in B \setminus A$.

Thus, $A \cup B \subseteq (A \setminus B) \cup (A \cap B) \cup (B \setminus A)$.

So we have proved the result.

E27) $A \times B = \{(2, 2), (2, 3), (5, 2), (5, 3)\}$

$B \times A = \{(2, 2), (3, 2), (2, 5), (3, 5)\}$

$A \times A = \{(2, 2), (2, 5), (5, 2), (5, 5)\}$.

$(A \times B) \cap (B \times A) = \{(2, 2)\}$

$(A \times B) \cup (B \times A) = \{(2, 2), (2, 3), (5, 2), (5, 3), (3, 2), (2, 5), (3, 5)\}$.

$A \setminus (B \times A) = \{2, 5\}$.

E28) $A =$ set of all the first elements in each pair $= \{7, 2\}$.

$B =$ set of all the second elements in each pair $= \{2, 3, 4\}$.

E29) Let $(x, y) \in A \times B$. Since $x \in A \subseteq C$, $x \in C$. Similarly, $y \in D$.
So, $(x, y) \in C \times D$. Hence, $A \times B \subseteq C \times D$.

E30) For instance, $(1, 0), \left(1, \frac{1}{2}\right), (1, 37.525252\ldots)$.

There are infinitely many elements you can pick from.

E31) There can be several examples. One is $S = \{(0, 0)\}$. Then
$(\mathbb{R} \times \mathbb{R}) \setminus S = \{(x, y) \mid x, y \in \mathbb{R}$ and $x, y \neq 0\}$.

E32) (i) is not, since it is only an ordered pair, and not a triple.

(ii) is not, since $\frac{1}{2} \notin \mathbb{N} \cup \mathbb{Z}$.

(iii) is, since $1 \in \mathbb{Q} \cap \mathbb{Z} \cap \mathbb{N}$.

(iv) is not, since $\sqrt{2} \notin \mathbb{N}$

(v) is not, since this is not an ordered triple; it is a set of three elements.

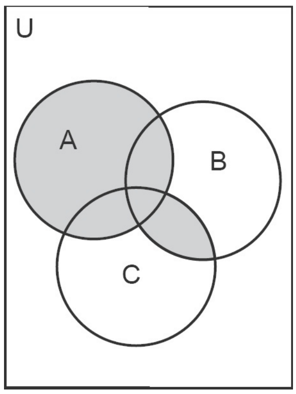


Fig.19: $\mathbf{A \cup (B \cap C) = (A \cup B) \cap (A \cup C)}$.

E33) $x \in A \cup (B \cap C)$

$\Leftrightarrow x \in A$ or $x \in B \cap C$

$\Leftrightarrow x \in A$ or $(x \in B$ and $x \in C)$

$\Leftrightarrow (x \in A$ or $x \in B)$ and $(x \in A$ or $x \in C)$

$\Leftrightarrow x \in A \cup B$ and $x \in A \cup C$

$\Leftrightarrow x \in (A \cup B) \cap (A \cup C)$.

Hence, Theorem 3 (ii) is proved. (A visual representation of this is given in Fig. 19.)

E34) $(x, y) \in (A \cap B) \times C$

$\Leftrightarrow x \in A \cap B$ and $y \in C$

$\Leftrightarrow (x \in A$ and $x \in B)$ and $y \in C$

$\Leftrightarrow (x, y) \in A \times C$ and $(x, y) \in B \times C$

$\Leftrightarrow (x, y) \in (A \times C) \cap (B \times C)$

Hence the equality.

E35) $x \in (A \cup B)^c \Leftrightarrow x \notin A \cup B$

$\Leftrightarrow \quad x \notin A$ and $x \notin B$

$\Leftrightarrow \quad x \in A^c$ and $x \in B^c$

$\Leftrightarrow \quad x \in A^c \cap B^c$.

So, $(A \cup B)^c = A^c \cap B^c$.

E36) $A \cup B = U, \therefore (A \cup B)^c = \emptyset$

Also $A^c = \{3, 4\}, B^c = \{1\}$

$\therefore \quad A^c \cap B^c = \emptyset$

$\therefore \quad (A \cup B)^c = A^c \cap B^c$

Further, $A \cap B = \{2\}$. $\therefore (A \cap B)^c = \{1, 3, 4\}$

$\therefore \quad A^c \cup B^c = (A \cap B)^c$.

E37) Since $A \setminus B \subseteq A, (A \setminus B) \times C \subseteq A \times C$ … (7)

If $(x, y) \in (A \setminus B) \times C$, then $x \in A \setminus B$, i.e., $x \notin B$.

$\Rightarrow \quad (x, y) \notin B \times C$ … (8)

(7) and (8) tell us that $(A \setminus B) \times C \subseteq (A \times C) \setminus (B \times C)$ … (9)

Conversely, let $(x, y) \in (A \times C) \setminus (B \times C)$.

Then $(x, y) \in A \times C$ and $(x, y) \notin B \times C$.

So, $x \in A, y \in C$ and $x \notin B$.

Therefore, $x \in A \setminus B, y \in C$.

Thus, $(x, y) \in (A \setminus B) \times C$

This proves that $(A \times C) \setminus (B \times C) \subseteq (A \setminus B) \times C$ … (10)

(9) and (10) together prove the equality.

One possibility for U is $(A \times C) \cup B$.

E38) Any element of $(A \setminus B) \times C$ is an ordered pair (a, c), where $a \in A, a \notin B$ and $c \in C$. On the other hand, any element of $A \setminus (B \times C)$ is an element of A and not a pair. Therefore, you should expect the statement to be false. You can show it by taking, for example,

$A = \{1, 2\}, B = \{1, 3\}, C = \mathbb{N}$.

Then $(A \setminus B) \times C = \{2\} \times \mathbb{N}$, while $A \setminus (B \times C) = \{1, 2\}$.

So $(A \setminus B) \times C \neq A \setminus (B \times C)$.

UNIT 2

FUNCTIONS

Structure **Page Nos.**

2.1 INTRODUCTION

Now that you are familiar with sets and operations on them, we shall focus on one of these operations in this unit. You will see how the Cartesian product is used in different ways, under different names.

We start the unit with a discussion in Sec. 2.2 on relations, which are just subsets of $A \times B$, with A and B being sets. We also study relations with certain properties that make them reflective, symmetric, transitive or equivalence relations.

In Sec. 2.3 we go on to focus on certain relations which are called functions. This is a concept you may have worked with while studying mathematics in school. Apart from looking at functions generally, we will consider functions with certain special properties that make them injective, surjective or bijective.

In the next section, namely, Sec. 2.4, we shall look at a way of joining two functions called 'the composition' of those functions. You will see that the composition of two functions may not always be defined. You will also study the conditions under which their composition can be defined. In this section you will also see why a bijective function is invertible.

Finally, in Sec. 2.5, we look at a particular kind of function from $S \times S \to S$, where S is a set. This is essentially an operation on pairs of elements of S. This is why it is called a **binary** operation on S. You will see that you have actually been working with binary operations from primary school on. In fact, you will find these operations through and through in the mathematics that you do henceforth too.

Objectives

After studying this unit, you should be able to:

- explain what a relation from a set A to a set B is;
- give examples of reflexive, symmetric or transitive relations;
- define, and give examples of, a function from a set A to a set B;
- explain when a function is an injection, surjection or bijection;
- obtain the composition of two functions under appropriate conditions;
- define, and give examples of, binary operations on sets.

2.2 RELATIONS

You are already familiar with the concept of a relationship between people. For example, a parent-child relationship exists between two people A and B if and only if A is a parent of B or B is a parent of A. Similarly, we can find relationships between integers, for example, two integers have a relationship if one is a factor of the other. However, there is a difference between 'relationship' and 'relation' in English usage. Similarly, there is a difference in these words in mathematical usage. For example, if R is the relation 'is a factor of' on $\mathbb{N}$, then 1 is a factor of 5, but 5 is not a factor of 1, and we write this as $1R5$, but not $5R1$. So, we see that R relates a pair of elements, **and the order matters**. Here, if we treat R as a subset of $\mathbb{N}\times\mathbb{N}$, then $R=\{(x,y)\in\mathbb{N}\times\mathbb{N} \mid xRy\}$, and $(1, 5)\in R$ but $(5, 1)\notin R$.

In mathematics, a relation R on a set S is a particular kind of relationship between the elements of S. If $a\in S$ is related to $b\in S$ by means of this relation, we write $\mathbf{aRb}$, or $(a, b)\in R$, and this is exactly how we define a relation on a set.

Definition: A **relation** R on a set S is a subset of $S\times S$.

For example, if R is the relation 'is greater than' on $\mathbb{Q}$, then $3R2$ (because $3>2$), and $\frac{1}{n}R\frac{1}{n+1}\ \forall n\in\mathbb{N}$. Thus, here

$R\subseteq\mathbb{Q}\times\mathbb{Q}$, $(3, 2)\in R$, $\left(\frac{1}{n}, \frac{1}{n+1}\right)\in R$ for any $n\in\mathbb{N}$, and $(0, 2)\notin R$.

Remark 1: Since a relation on S is a subset of $S\times S$, two relations R_1 and R_2 on S will be distinct if the sets R_1 and R_2 are different, i.e., $R_1\neq R_2$. For instance, consider $S=\{1, 2,3\}$, $R_1=\{(1, 1), (1, 2)\}$ and $R_2=\{(2, 1), (1,2)\}$. Then R_1 and R_2 are both relations on S, being subsets of $S\times S$. But $R_1\neq R_2$.

Try the following exercises now, which deal with relations on a set.

E1) Let $\mathbb{N}$ be the set of all natural numbers and R the relation $\{(a, a^2) \mid a\in\mathbb{N}\}$. State whether the following statements are true or false. Also give reasons for your answers.

i) $2R3$, ii) $3R9$,

iii) $9R3$, iv) $(\sqrt{3}, 3) \in R$.

E2) Give two distinct relations on the set of courses of IGNOU.

We now look at some particular kinds of relations, which you will be using very often in other mathematics courses.

Definition: A relation R defined on a set S is said to be

i) **reflexive** if $aRa\ \forall\ a \in S$, i.e., $(a, a) \in R\ \forall\ a \in S$.

ii) **symmetric** if $aRb \Rightarrow bRa\ \forall a, b \in S$, i.e., $(a, b) \in R \Rightarrow (b, a) \in R$.

iii) **transitive** if aRb and $bRc \Rightarrow aRc\ \forall\ a, b, c \in S$, i.e., if $(a, b) \in R$ and $(b, c) \in R$, then $(a, c) \in R$.

To help you get used to these concepts, consider the following examples.

Example 1: Consider the relation R on the set H of those human beings who were alive in 2016, given by 'aRb iff a and b had the same weight on Jan. 1st, 2016, for any two human beings a and b in H'. Is R reflexive, symmetric or transitive? Justify your answers.

Solution: Any human being certainly has the same weight as herself. So hRh for every $h \in H$. Thus, R is reflexive.
If $a, b \in H$ such that a and b have the same weight, then b and a certainly have the same weight. So $aRb \Rightarrow bRa\ \forall\ a, b \in H$. Thus, R is symmetric.
Finally, for $a, b, c \in H$ if a and b have the same weight, and b and c have the same weight, then a and c have the same weight too.
So, aRb and $bRc \Rightarrow aRc$. Thus, R is transitive.

***

Example 2: Consider the relation R on $\mathbb{Z}$ given by 'aRb if and only if $a \gneq b$'. Determine whether or not R is reflexive, symmetric or transitive.

'$a \gneq b$' denotes 'a is strictly greater than b'.

Solution: Since $a \gneq a$ is not true, aRa is not true. Hence, R is not reflexive.
If $a \gneq b$, then certainly $b \gneq a$ is not true. That is, aRb does not imply bRa.
Hence, R is not symmetric.
Since $a \gneq b$ and $b \gneq c$ implies $a \gneq c$, we find that aRb, bRc implies aRc.
Thus, R is transitive.

***

Example 3: Let S be a non-empty set. Let $\wp(S)$ denote **the power set** of S, that is, the set of all subsets of S, i.e., $\wp(S) = \{A \mid A \subseteq S\}$.
Define the relation R on $\wp(S)$ by $R = \{(A, B) \mid A \subseteq B\}$.
Check whether or not R is reflexive, symmetric or transitive.

Solution: Since $A \subseteq A\ \forall \in \wp(S), (A, A) \in R\ \forall\ A \in \wp(S)$. Thus, R is reflexive.
Let $(A, B) \in R$. Then $A \subseteq B$. However, B need not be contained in A. For example, $(\emptyset, S) \in R$ but $(S, \emptyset) \notin R$. Thus, R is not symmetric.
If $A \subseteq B$ and $B \subseteq C$, then $A \subseteq C\ \forall\ A, B, C \in \wp(S)$, that is, $(A, B) \in R$ and $(B, C) \in R \Rightarrow (A, C) \in R$ (see E10 of Unit 1). Thus, R is transitive.

***

Remark 2: In all the examples in this section so far, when we have mentioned relations from X to X or X to Y, X and Y have been non-empty. However, a relation can be defined on the empty set too. This is called the **empty,** or **null**, **relation**.

You may like to try the following exercises now.

E3) Consider the relation 'aRb iff $a = b$' on $\mathbb{R}$. Check whether R is reflexive, symmetric or transitive.

E4) The relation $R \subseteq \mathbb{N}\times\mathbb{N}$ is defined by $(a, b) \in R$ iff 5 divides $(a-b)$ in $\mathbb{N}$. Is R
i) reflexive?
ii) symmetric?
iii) transitive?
Give reasons for your answers.

E5) Give examples to show why the relation in E1 is not reflexive, symmetric or transitive.

E6) Check whether or not $R = \{(A, B) \mid A \not\subset B,\ A, B \in \wp(S)\}$ is transitive, where S has at least two elements.

The relationship in E3 is reflexive, symmetric and transitive. This is an example of what we now define.

Definition: A relation R on a set S is called an **equivalence relation** if R is reflexive, symmetric and transitive.
You will study, and use, such relations quite a bit in the other mathematics courses.

Let us now generalize 'relation on a set' to a relation from one set to another. You have seen that a relation on a set X is a subset of $X\times X$. So, you may expect the generalized form to have the following definition.

Definition: Let X and Y be two sets. A **relation from X to Y** is a subset of $X\times Y$.
For example, let H be the set of humans beings, $W \subseteq H$ be the set of women drivers in India and Y the set of driving licences valid in India on Oct. 2nd, 2018. Then $R = \{(w, y) \mid w \in W, y \in Y\}$ is a relation from H to Y. Note that $R = W\times Y$ is a subset of $H\times Y$.

Another example is the subset R of $H\times L$, where H is the set of humans and L is the set of languages in the world in 2018, where $(h, \ell) \in R$ iff h uses the language ℓ. So, for example, $(x, I) \in R$, where I is the Indian sign language and x is a user of this language.

Let us now look at a particular type of such relations, in which the choice of the second element of the pair takes great importance.

2.3 FUNCTIONS

Consider the relation $R = H\times L$ that we gave at the end of Sec.2.2 above.

Here, we also say xRy if $(x, y) \in R$. Note that x is/was a human being and y is a language extant in 2018.

Now, let us take a subset R' of R, where '$xR'y$ if and only if $x \in H$ and y is the first language of x'. Now since each person has one and only one first language, given x there is one and only one $y \in L$ such that $(x, y) \in R'$. So $R' \subseteq H \times L$.

Note that

i) For each $x \in H$, there is **at least one** $y \in L$ such that $(x, y) \in R'$.

ii) For each $x \in H$, there is **only one** $y \in L$ such that $(x, y) \in R'$.

Such a relation R' is an example of a 'function', as you will soon see.

Let us consider another example. Let H be the set of human beings and W be the set of women. Let R_M be the relation from H to W such that for each $h \in H, hR_M w$, where w is h's birth mother. Then we see that

i) For each $h \in H \; \exists \; w \in W$ such that $hR_M w$.

ii) For each $h \in H$ there is only one w such that $hR_M w$.

iii) There could be several elements of H, say $h_1, h_2, \ldots, h_n$, with the same birth mother m, i.e., $h_1 R_M m, h_2 R_M m, \ldots, h_n R_M m$.

The properties of R_M given above tell us that R_M is a 'function'.

So, what is a 'function'?

This definition of function was given by Dirichlet in 1837, and has been used since then.

Definitions: i) A **function** f from a non-empty set A to a non-empty set B is a relation from A to B which associates with every element of A one and only one element of B. This is written as $\mathbf{f : A \to B}$, or $\mathbf{A \xrightarrow{f} B}$.

ii) If f associates with $a \in A$, the element $b \in B$, we write $\mathbf{f(a) = b}$, and b is called the **value of the function f at a**.

iii) A is called the **domain** of f, and B is called the **co-domain** of f.

iv) The set $f(A) = \{f(a) \mid a \in A\}$ is called the **range** of f. As you can see, the range of f is a subset of the co-domain of f, i.e., $f(A) \subseteq B$.

Note that if $f : A \to B$, then

i) **for each** element of A, we associate some element of B.

ii) for each element of A, we associate **only one** element of B. So, if $a \in A$ and $f(a) = b_1$ as well as $f(a) = b_2$, then $b_1 = b_2$, i.e., if (a, b_1) and (a, b_2) are elements of f, then $b_1 = b_2$.

iii) two or more elements of A can be associated with the same element of B, i.e., there can be $a_1, a_2 \in A$ with $a_1 \neq a_2$ and $(a_1, b) = (a_2, b)$.

Remark 3: Some other terms commonly used for 'function' are 'map', 'mapping', 'transformation', 'operator'.

Consider an example. Let $A = \{1, 2, 3\}$ and $B = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$. Define $f : A \to B$ by $f(1) = 1, f(2) = 4, f(3) = 9$. Then f is a function with domain A, co-domain B and range $\{1, 4, 9\}$.

In this case, note that $f(x) = x^2$ for each $x \in A$. We can write this as
$f : A \to B : f(x) = x^2$ or $f = \{(x, x^2) \mid x \in A\}$.
If we define $g : A \to B$ by $g(1) = 1, g(2) = 1, g(3) = 4$, then g is also a function, whose domain and co-domain are the same as those of f, namely, A and B, respectively. But the range of g is $\{1, 4\}$.

Remark 4: The example above tells us that several different functions can be defined with the same domain and co-domain.

Remark 5: When a relation from A to B does not satisfy the requirements for being a function, then we can say that this is not a **well-defined function**.
For example, if $A = \{1, 2, 3, \ldots, 10\}$ and $B = \{1, 5, 7\}$, $\alpha : A \to B$ with $\alpha(1) = 1, \alpha(1) = 5$ is not well-defined. (Why?)

Try some exercises now.

E7) Let X be the set of residents of Kochi and Y be the set of all 10-digit numbers. Define a mapping from X to Y, clearly giving its domain and range.

E8) i) Consider $f : \mathbb{N} \to \mathbb{Q} : f(n) = \dfrac{n}{n+1}$. Show that f is a function.

ii) Define a function g whose domain and co-domain are the same as that of f above, but there is at least one x in their domain for which $f(x) \neq g(x)$.

E9) Define a relation from $\mathbb{N}$ to $\mathbb{Q}$ which is not a function. Justify your choice of relation.

E8 leads us to the notion defined below.

Definition: Two functions **f and g are said to be equal if**

i) Domain $f =$ Domain g, and

ii) $f(x) = g(x)\ \forall x \in$ Domain f (which implies that Range $f =$ Range g).

So, for example, the function $f : \mathbb{R} \to \mathbb{R} : f(x) = x - 1$ and $g : \mathbb{R} \to \mathbb{R} : g(0) = 7, g(x) = x - 1, \forall x \neq 0, x \in \mathbb{R}$ are not equal, since $f(0) \neq g(0)$.

Try a related exercise now.

E10) 'If f and g are two functions with domain A and range B, then $f = g$.' Is this statement true or false? Give reasons for your answer.

Now let us look at functions with special properties.

'f is one-one' is briefly shown as 'f is $1-1$'.

Definition: A function $f : A \to B$ is called a **one-one** (or **injective**) function if f relates different elements of A to different elements of B, i.e., if $a_1, a_2 \in A$ and $a_1 \neq a_2$, then $f(a_1) \neq f(a_2)$.

In other words, **$\mathbf{f}$ is $\mathbf{1-1}$ if $\mathbf{f(a_1)=f(a_2) \Rightarrow a_1 = a_2}$.**
If f is injective, it is also called an **injection.**

For example, consider the function
$f : \{1, 2, 3\} \to \{1, 2, \ldots, 10\} : f(1) = 1, f(2) = 4, f(3) = 9$. You will find it takes distinct elements of the domain to distinct elements of the co-domain. So f is $1-1$.

Now, let us consider another example of sets and functions.
Let $A = \{1, 2, 3\}, B = \{p, q\}$. Let $f : A \to B$ be defined by $f(1) = q, f(2) = p, f(3) = p$. Then f is a function, with range of $f = B =$ co-domain of f. (f is pictorially represented in Fig. 1). This is an example of an onto function, as you shall see.

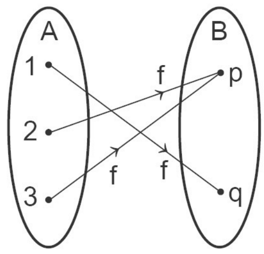


Fig. 1: $\mathbf{f : A \to B}$

Definition: A function $f : A \to B$ is called an **onto** (or **surjective**) function if the range of f is the same as its co-domain B. In other words, f is onto if $\mathbf{f(A) = B}$. This means that **for any $\mathbf{b \in B}$, there is an $\mathbf{a \in A}$ such that $\mathbf{f(a) = b}$.**
If f is surjective, it is called a **surjection**.
You will come across this kind of function very often in your mathematics courses.

The word '*surjection*' is derived from the French word '*sur*', meaning 'on'.

Let us consider another example of a surjective function. Consider two non-empty sets A and B. We define the function $\pi_1 : A \times B \to A : \pi_1((a, b)) = a$. π_1 is called the **projection** of $A \times B$ on A. You can see that the range of π_1 is the whole of A, since for any $a \in A, \pi_1(a, b) = a$, whatever b may be. Therefore, π_1 is onto. Note that if B has more than one element, then π_1 is not $1-1$. (Why?)
Similarly, $\pi_2 : A \times B \to B : \pi_2((a, b)) = b$, the projection of $A \times B$ on B, is a surjective function.
Since π_1 (or π_2) is an onto function, we also say π_1 (respectively π_2) is a function from $A \times B$ **onto** A (respectively B).

And now, we define a function that has a combination of both the properties you have just studied.

Definition: If a function $f : A \to B$ is both one-one and onto, it is called a **bijective function**, or a **bijection**.

Let us consider some examples. The first one is of a function that you will use again and again.

Example 4: Let A be any non-empty set. The function $I_A : A \to A : I_A(a) = a$ is called the **identity function** on A. Show that I_A is bijective.

Solution: For any $a \in A, I_A(a) = a$. Thus, the range of I_A is the whole of A. That is, I_A is onto.
I_A is also $1-1$ because if $a_1, a_2 \in A$ such that $a_1 \neq a_2$, then $I_A(a_1) \neq I_A(a_2)$.
Thus, I_A is bijective.

***

$f : A \to B : f(a) = c \; \forall \, a \in A$ is called a **constant function.**

Example 5: Define a function $f : \{1, 2, 3\} \to \{1, 2, 3\} : f(1) = f(2) = f(3) = 1$. This function is an example of a constant function. Check whether f is bijective or not.

Solution: The domain of f is $\{1, 2, 3\}$ and range is the singleton $\{1\} \neq \{1, 2, 3\}$. So f is not surjective.

Next, $f(1) = f(2)$ and $1 \neq 2$, so that f is not injective either.

***

Try the following exercises now.

E11) Let $f : \mathbb{N} \to \mathbb{N}$ be defined by $f(n) = n + 5$. Write f as a subset of the Cartesian product of its domain and range. Prove that f is one-one but not onto.

E12) Let $f : \mathbb{Z} \to \mathbb{Z} : f(n) = n + 5$. Prove that f is both one-one and onto.

E13) What must X be like for the constant function $f : X \to \{c\}$ to be injective? Is f surjective? Give reasons for your answers.

Before going further let us briefly look at finiteness. In Unit 1 we have discussed finite and infinite sets. Let use see what this means mathematically.

Example 6: Show that **a set A is finite if and only if** there is a bijection between A and $\{1, 2, \ldots, n\}$, for some $n \in \mathbb{N}$.

Solution: Let A be a finite set with $|A| = m$. Let $A = \{a_1, a_2, \ldots, a_m\}$.

Define $f : A \to \{1, 2, \ldots, m\} : f(a_i) = i$.

You can check that f is injective as well as surjective, and hence, bijective.

Conversely, let g be a bijective function from A to $\{1, 2, \ldots, n\}$, for some $n \in \mathbb{N}$. Since g is onto, its range is the same as its co-domain, i.e., $\{g(a) \mid a \in A\} = \{1, 2, \ldots, n\}$.

So $|g(A)| = n$.

Since g is $1-1$, $|A| = |g(A)| = n$, so A is finite.

***

Remark 6: Usually, the condition in Example 6 is treated as the definition of finiteness.

Now let us look at a very important way of producing new functions from given ones.

2.4 COMPOSITION OF FUNCTIONS

Let us start with taking the two functions $f = \{(x, x^2) \mid x \in \mathbb{N}\}$ and $g = \{(z, -z) \mid z \in \mathbb{Z}\}$. Here the range of f is a subset of $\mathbb{Z}$, which is the domain of g. Let us define h by 'combining' f and g as follows:

Domain $h =$ Domain f, Co-domain $h =$ Co-domain g and

Range $h \subseteq$ Range g.

For $x \in \mathbb{N}$, take $h(x) = g(f(x)) = g(x^2) = -x^2$ (see Fig. 2).

So $h : \mathbb{N} \to \mathbb{Z} : h(x) = g(f(x))$.

More generally, if $f : A \to B$ and $g : B \to C$ are functions, there is a natural way of combining g and f to yield a new function $h : A \to C$, as below:

For each $x \in A$, $h(x)$ is defined by the formula $h(x) = g(f(x))$.

Note that $f(x) \in B$. Therefore, $g(f(x))$ is defined as an element of C.
This function h is called the **composition of g and f** and is written as $\mathbf{g \circ f}$.
The domain of $g \circ f$ is A and its co-domain is C, i.e., $g \circ f : A \to C$ (see Fig. 2).

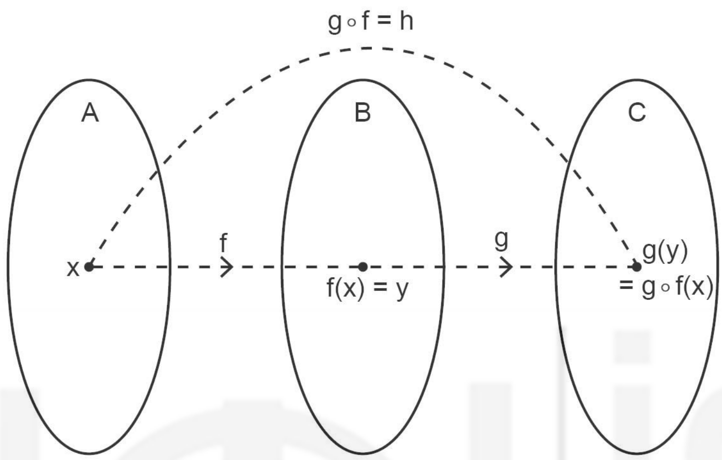


Fig. 2: Composition of f and g

Let us consider some examples of this.

Example 7: Let $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = x^2$ and $g(x) = x + 1$. Is $g \circ f$ or $f \circ g$ defined? If yes, what is $g \circ f$, and what is $f \circ g$?

Solution: We observe that the range of f is a subset of $\mathbb{R}$, the domain of g. Therefore, $g \circ f$ is defined.

By definition, $\forall\ x \in \mathbb{R}, g \circ f(x) = g(f(x)) = f(x) + 1 = x^2 + 1$.
Now, let us see if $f \circ g$ is defined. Again, since the range of g is a subet of the domain of f, $f \circ g$ is defined. So,
$\forall\ x \in \mathbb{R}, f \circ g(x) = f(g(x)) = [g(x)]^2 = (x+1)^2$.
So $f \circ g$ and $g \circ f$ are both defined. But $g \circ f \neq f \circ g$. (For example, $g \circ f(1) \neq f \circ g(1)$.)

***

Example 8: Let $A = \{1, 2, 3\}, B = \{p, q, r\}$ and $C = \{x, y\}$. Let $f : A \to B$ be defined by $f(1) = p, f(2) = p, f(3) = r$. Let $g : B \to C$ be defined by $g(p) = x, g(q) = y, g(r) = y$. Determine if $f \circ g$ and $g \circ f$ can be defined.

Solution: Here Domain $f = A$, Range $f = \{p, r\}$, Domain $g = B$, Range $g = C$. For $f \circ g$ to be defined, it is necessary that the range of g should be a subset of the domain of f. As C is not a subset of $A, f \circ g$ cannot be defined.

Since the range of f is a subset of the domain of g, we see that $g \circ f$ is defined. Also $g \circ f : A \to C$ is such that

$g \circ f(1) = g(f(1)) = g(p) = x$,

$g \circ f(2) = g(f(2)) = g(p) = x$,

$g \circ f(3) = g(f(3)) = g(r) = y$.

Therefore, $g \circ f$ is surjective. Note that g is also surjective.

***

Remark 7: Note that for $g \circ f$ to be defined all we need is that Range $f \subseteq$ Domain g. Thus, Co-domain f need not be the same as Domain g.

So, let us formally define the operation of composition now.

Definition: Let $f : A \to B$ and $g : C \to D$ be two functions such that Range $f \subseteq C$. Then the **composition of g and f** is the function $\mathbf{g \circ f} : A \to D : g \circ f(a) = g(f(a))$.

Now for some exercises on the composition of functions.

E14) In each of the following parts, both f and g are functions from $\mathbb{R}$ to $\mathbb{R}$. Define $f \circ g$ and $g \circ f$.

i) $f(x) = 5x,\ g(x) = x + 5$,

ii) $f(x) = 5x,\ g(x) = x/5$.

E15) Give an example, with justification, of functions f and g such that neither $f \circ g$ nor $g \circ f$ are defined.

E16) Let $f : A \to B$ and $g : C \to D$, with $B \subseteq C$, be two functions. Which of the following statements are true? Give reasons for your answers.

i) Range $g \supseteq$ Range $g \circ f$.

ii) Range $g =$ Range $g \circ f$.

iii) If f is injective, so is $g \circ f$.

iv) If g is surjective, so is $g \circ f$.

While doing E16, a question may have risen in your mind about whether there is any other kind of relationship between the properties of $g \circ f$ and those of f and g. For instance, what can we expect if $g \circ f$ is onto, or is $1-1$? In this context, consider the following theorem.

Theorem 1: Let $f : A \to B$ and $g : C \to D$ be two functions such that $g \circ f$ is defined. Then

i) if $g \circ f$ is injective, so is f.

ii) if $g \circ f$ is surjective, so is g.

Proof: Note that $g \circ f : A \to D : g \circ f(x) = g(f(x))$.

Let us prove (i) first. We are given that $g \circ f$ is $1-1$. To show that f is

injective, let $a_1, a_2 \in A$ such that $f(a_1) = f(a_2)$.
Then $g(f(a_1)) = g(f(a_2))$, as g is well-defined.
So, $g \circ f(a_1) = g \circ f(a_2)$.
This implies that $a_1 = a_2$, since $g \circ f$ is injective.
So, we have shown that $f(a_1) = f(a_2) \Rightarrow a_1 = a_2$, i.e., f is injective.

Now let us prove (ii). Here we are given that $g \circ f$ is onto. To show that g is onto, let $d \in D$. Since $g \circ f$ is surjective, there is $a \in A$ such that $g \circ f(a) = d$. This means that $g(f(a)) = d$. Thus, given any $d \in D, \exists f(a) \in C$ with $g(f(a)) = d$, which shows that g is surjective. ■

In the theorem above, did you notice that when $g \circ f$ is injective we have not said anything about the injectivity of g? Do you expect g to be injective? For example, if $f : \mathbb{N} \to \mathbb{R} : f(n) = \sqrt{n}$ and $g : \mathbb{R} \to \mathbb{R} : g(x) = x^2$, then $g \circ f : \mathbb{Z} \to \mathbb{R} : g \circ f(n) = n$. So $g \circ f$ is $1-1$, but g is not. (Why?)

We now come to a theorem which shows us that the identity function behaves like the number $1 \in \mathbb{R}$ does for multiplication. That is, if we take the composition of any function f with a 'suitable' identity function (see Example 4), we get the same function f.

Theorem 2: Let A and B be sets.

i) For any function $f : A \to A$, $f \circ I_A = I_A \circ f = f$.

ii) For any function $g : B \to A$, $I_A \circ g = g$ and $g \circ I_B = g$.

Proof: We shall prove (i) here, and leave the proof of (ii) to you (see E17).

(i) Since both f and I_A are defined from A to A, both the compositions $f \circ I_A$ and $I_A \circ f$ are defined. Moreover,
$\forall\, x \in A,\ f \circ I_A(x) = f(I_A(x)) = f(x)$, so $f \circ I_A = f$.
Also, $\forall\, x \in A,\ I_A \circ f(x) = I_A(f(x)) = f(x)$, so $I_A \circ f = f$. ■

To complete the proof of Theorem 2, try the next set of exercises.

E17) Prove Theorem 2(ii).

E18) Show that if f and g are two functions such that $f \circ g$ is onto, then g need not be onto.

While doing E17, did you note why I_A was used in the first equality and I_B in the second equality? This is what we meant when we said 'suitable' identity function earlier. We need to pick the identity function of the set that allows the composition concerned to be defined.

Now, in the case of the set of non-zero real numbers, $\mathbb{R}^*$, you know that given $x \in \mathbb{R}^*$ $\exists\, y \in \mathbb{R}^*$ such that $xy = 1$. This number y is called the inverse of x. Similarly, we can define an inverse function for some functions. Here, instead of multiplication of numbers, we shall consider the composition of functions.

For example, consider $f : \mathbb{R} \to \mathbb{R} : f(x) = x + 3$. If we define

$g:\mathbb{R}\to\mathbb{R}:g(x)=x-3$, then both $f\circ g$ and $g\circ f$ are defined. Further, $f\circ g(x)=f(g(x))=g(x)+3=(x-3)+3=x\ \forall\ x\in\mathbb{R}$.

Hence, $f\circ g=I_{\mathbb{R}}$. You can also verify that $g\circ f=I_{\mathbb{R}}$.

In this case we call g the inverse of f, as you will see in the definition that follows.

Definition: Let $f:A\to B$ be a given function. If there exists a function $g:B\to A$ such that $f\circ g=I_B$ and $g\circ f=I_A$, then g is called the **inverse** of f, and we write $\mathbf{g=f^{-1}}$.

So, in the example before the definition, $g=f^{-1}$ and $f=g^{-1}$. Note that in this example, f adds 3 to x and g does the opposite – it subtracts 3 from x. So, essentially f and g nullify each other's actions on x. Thus, the key to finding the inverse of a given function is: **to get $\mathbf{f^{-1}}$ try to get back $\mathbf{x}$ from $\mathbf{f(x)}$**.

For another example, let $f:\mathbb{R}\to\mathbb{R}$ be defined by $f(x)=3x+5$. How can we get back x from $3x+5$? One way is to "first subtract 5 and then divide by 3". So, we try $g:\mathbb{R}\to\mathbb{R}:g(r)=\frac{r-5}{3}$. And we find

$$g\circ f(x)=g(f(x))=\frac{f(x)-5}{3}=\frac{(3x+5)-5}{3}=x\ \forall\ x\in\mathbb{R}.$$

Also, $f\circ g(x)=3(g(x))+5=3\left[\frac{(x-5)}{3}\right]+5=x\ \forall\ x\in\mathbb{R}$.

Let's see if you've understood the process of obtaining the inverse of a function.

E19) What is the inverse of

i) $f:\mathbb{R}\to\mathbb{R}:f(x)=\frac{x}{3}$?

ii) $g:\mathbb{Q}\to\mathbb{Q}:g(r)=2r+\frac{5}{7}$?

The discussion above may have triggered the question: Do all functions have an inverse? To answer this, consider the following example.

Example 9: Let $f:\mathbb{R}\to\mathbb{R}$ be the constant function given by $f(x)=1\ \forall x\in\mathbb{R}$. Does the inverse of f exist?

Solution: If f has an inverse $g:\mathbb{R}\to\mathbb{R}$, we have $f\circ g=I_{\mathbb{R}}$, i.e., $\forall\ x\in\mathbb{R}, f\circ g(x)=x$. Now take $x=5$, for instance. We should have $f\circ g(5)=5$, i.e., $f(g(5))=5$. But $f(g(5))=1$ since $f(x)=1\ \forall\ x$.

So we reach a contradiction. Therefore, f has no inverse.

***

In view of this example, you may ask for the conditions under which f will have an inverse. The answer is given by the following theorem.

Theorem 3: A function $f:A\to B$ has an inverse if and only if f is a bijection.

Proof: Here we have to prove two statements:

i) If f is bijective, then f has an inverse.

ii) If f has an inverse, then f is bijective.

So, firstly, suppose f is bijective. We shall define a function $g: B \to A$ and prove that $g = f^{-1}$.

Let $b \in B$. Since f is onto, $B = f(A) = \{f(a) \mid a \in A\}$. So there is some $a \in A$ such that $f(a) = b$. Since f is one-one, there is only one such $a \in A$. We take this unique element a of A as $g(b)$. That is, given $b \in B$, we define $g(b) = a$, where a is that element of A for which $f(a) = b$.

Note that, since f is onto, we are simply defining $g: B \to A$ by $g(f(a)) = a$. This automatically ensures that $g \circ f = I_A$.

We still need to show that $f \circ g = I_B$. For this, let $b \in B$ and $g(b) = a$. Then $f(a) = b$, by definition of g. Therefore, $f \circ g(b) = f(g(b)) = f(a) = b$. Hence, $f \circ g = I_B$.

So, $f \circ g = I_B$ and $g \circ f = I_A$. This proves that $g = f^{-1}$.

Now let us prove (ii). Suppose f has an inverse, and $g = f^{-1}$. We must prove that f is bijective, that is, f is one-one and onto.

Now $f \circ g = I_B$ and $g \circ f = I_A$. From Example 4, you know that I_A and I_B are both $1-1$ and onto, so $f \circ g$ and $g \circ f$ are both $1-1$ and onto.

Now, from Theorem 1, since $g \circ f$ is $1-1$, so is f. Also, since $f \circ g$ is onto, so is f.

Thus, f is one-one and onto.

Hence, the theorem is proved. ■

Thus, by applying the theorem above to the function f in Example 9, we would immediately know that f^{-1} does not exist, since f is not injective.

Try the following exercise now.

E20) Consider the functions f_1, f_2, f_3 from $\mathbb{R}$ to $\mathbb{R}$, defined as below. For each, determine whether it has an inverse and, when the inverse exists, find it.

i) $f_1(x) = x^2 \ \forall \ x \in \mathbb{R}$;

ii) $f_2(x) = 0 \ \forall \ x \in \mathbb{R}$;

iii) $f_3(x) = 11x + 7 \ \forall \ x \in \mathbb{R}$.

In this section we have looked at an operation on functions. Earlier, you studied operations on sets, like union and intersection. Let us look at all these operations in a more general setting.

2.5 BINARY OPERATIONS

You are familiar with the operations of addition and multiplication on the set of real numbers. Addition is a function which associates with $(a, b) \in \mathbb{R} \times \mathbb{R}$ the element $a + b$ of $\mathbb{R}$. So, it is a function from $\mathbb{R} \times \mathbb{R}$ to $\mathbb{R}$. In other words,

$+ : \mathbb{R}\times\mathbb{R} \to \mathbb{R} : +(a, b) = a + b$. Can you see that multiplication is also a function from $\mathbb{R}\times\mathbb{R}$ to $\mathbb{R}$? These functions can be performed on any pair of elements of $\mathbb{R}$. They are examples of binary operations, which we now define.

Definition: A **binary operation on a non-empty set S** is a function from $S\times S$ to S.

Thus, a binary operation on S takes a pair of elements of S and associates a unique element in S to them. The word 'binary' means 'involving pairs'. It is customary to denote a binary operation by a symbol such as $+, \circ, *$, etc.
As mentioned earlier, $+$ and $\times$ are binary operations on $\mathbb{R}$.
Another example is $* : \mathbb{R}\times\mathbb{R} \to \mathbb{R} : a * b = \frac{a+b}{2}$.
As yet another example, take a set X, and take its power set $\wp(X)$. Then $\cap : \wp(X)\times\wp(X) \to \wp(X) : \cap(A, B) = A \cap B$ is a binary operation.

Some binary operations can have special properties which we now define.

Definition: A binary operation $*$ on a set S is said to be
i) **closed on a subset** T of S if $t_1 * t_2 \in T \ \forall\ t_1, t_2 \in T$.
ii) **commutative if** $a * b = b * a \ \forall\ a, b \in S$.
iii) **associative if** $(a * b) * c = a * (b * c) \ \forall\ a, b, c \in S$.

For example, the operations of addition and multiplication on $\mathbb{R}$ are commutative as well as associative. But, subtraction is neither commutative nor associative on $\mathbb{R}$. Why? Is $a - b = b - a$ for $a, b \in \mathbb{R}$? Or is $(a - b) - c = a - (b - c) \forall a, b, c \in \mathbb{R}$? No, as you know. Also, subtraction is not closed on $\mathbb{N} \subseteq \mathbb{R}$, because, for example, $1 \in \mathbb{N}, 2 \in \mathbb{N}$ but $1 - 2 \notin \mathbb{N}$.

Try an exercise now.

E21) Let X be a set and $\wp(X)$ its power set.

i) Show that complementation and union are binary operations on $\wp(X)$. Which of these operations are commutative, and which are associative? Give reasons for your answers.

ii) Is the Cartesian product a binary operation on $\wp(X)$? Give reasons for your answer.

Let us now look at a property connecting two binary operations. In Sec. 1.6, you have seen some examples of distributivity. You would recall that $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ for subsets A, B, C of a set U. This shows that the operation of union distributes over the operation of intersection. Let us define this concept more generally.

Definition: If $\circ$ and $*$ are two binary operations on a set S, we say that $*$ **is distributive** over $\circ$ (or $*$ **distributes over** $\circ$**)** if
$a * (b \circ c) = (a * b) \circ (a * c)$, and
$(b \circ c) * a = (b * a) \circ (c * a) \ \forall\ a, b, c \in S$.

Consider an example.

Example 10: Let $a * b = \frac{a+b}{2} \forall\, a, b \in \mathbb{R}$. Prove that the operation of multiplication in $\mathbb{R}$ is distributive over $*$.

Solution: We have to see whether $a(b*c) = ab*ac$ and $(b*c)a = ba*ca$.

Now $a(b*c) = a\frac{(b+c)}{2} = \frac{ab+ac}{2} = ab*ac$.

Also $(b*c)a = \frac{(b+c)}{2}a = \frac{ba+ca}{2} = ba*ca$.

Hence, multiplication is distributive over $*$.

***

Try an exercise now.

E22) For the following binary operations defined on $\mathbb{R}$, determine whether they are commutative or associative. Are any of them closed on $\mathbb{N}$?

i) $x \oplus y = x + y - 5$

ii) $x * y = 2(x+y)$

iii) $x \Delta y = \frac{x-y}{2}$

Also check if $\oplus$ distributes over Δ in $\mathbb{R}$.

We end our discussion on functions here. Of course, you will be extending your learning from this unit while studying every mathematics course. For now, let us quickly summarise what we have discussed in this unit.

2.6 SUMMARY

In this unit, you have studied the following points.

1) The definition, and examples, of a relation from a set S to a set T.

2) A relation R on a set S is
   i) reflexive if $aRa\ \forall\, a \in S$, i.e., $\{(a, a) \mid a \in S\} \subseteq R$;
   ii) symmetric if $aRb \Rightarrow bRa \forall\, a, b \in S$, i.e., $(a, b) \in R \Rightarrow (b, a) \in R\ \forall\, a, b \in S$;
   iii) transitive if $aRb, bRc \Rightarrow aRc\ \forall\, a, b, c \in S$, i.e., $(a, b) \in R, (b, c) \in R \Rightarrow (a, c) \in R\ \forall\, a, b, c \in S$.
   iv) an equivalence relation if it is reflexive, symmetric and transitive.

3) The definition, and examples, of a function with domain S, co-domain T and range R.

4) Two functions f and g are equal iff Domain f = Domain g and $f(x) = g(x) \forall\, x \in$ Domain f.

5) A function $f : A \to B$ is
   i) injective (or $1-1$) if $a_1, a_2 \in A$ with $a_1 \neq a_2 \Rightarrow f(a_1) \neq f(a_2)$;
   ii) surjective (or onto) if for each $b \in B\ \exists\, a \in A$ such that $f(a) = b$;
   iii) bijective (or 1-to-1) if f is injective and surjective.

6) The composition of functions $f: A \to B$ and $g: C \to D$, where the range of g is a subset of A, is $f \circ g: C \to B: f \circ g(c) = f(g(c))$.

7) When a function is invertible, and then defining its inverse explicitly.

8) The definition, and examples, of a binary operation on a set S.

9) When a binary operation on S is closed on a subset of S, or is commutative or associative.

10) When one binary operation on a set S distributes over another binary operation on S.

2.7 SOLUTIONS/ANSWERS

E1) $(2, 3) \notin R$ since $3 \neq 2^2$. Similarly, $(9, 3) \notin R$. $(\sqrt{3}, 3) \notin R$ since $\sqrt{3} \notin \mathbb{N}$. Thus, (ii) is the only true one since $3R3^2$, with $3 \in \mathbb{N}$.

E2) If S is the set of courses of IGNOU, then consider
$R_1 = \{(C_1, C_2) \mid C_1, C_2$ are first level courses of IGNOU$\}$,
$R_2 = \{(M, N) \mid M$ is a course of the BA prog. and N is a course of the B.Sc. programme$\}$.
Then R_1 and R_2 are subsets of $S \times S$ which are not equal.

E3) R is reflexive because $a = a \ \forall \ a \in \mathbb{R}$.
R is symmetric because $a = b \Rightarrow b = a \ \forall \ a, b \in \mathbb{R}$.
R is transitive because $a = b, b = c \Rightarrow a = c \ \forall \ a, b, c \in \mathbb{R}$.

E4) i) R is not reflexive because for any $a \in \mathbb{N}, a - a = 0$. So, there is no natural number n for which $5n = a - a$.

ii) If $5 \mid (a-b)$, then $\exists\, n \in \mathbb{N}$ such that $5n = a - b$. If $\exists\, m \in \mathbb{N}$ such that $5m = b - a = -5n$, then $m = -n$, which is a contradiction since n and $-n$ both cannot be in $\mathbb{N}$. So R is not symmetric.

iii) If $5 \mid (a-b)$ and $5 \mid (b-c)$ in $\mathbb{N}$, then check that $5 \mid (a-c)$ in $\mathbb{N}$. So, R is transitive.

E5) There can be several examples. We give the following:
$(2, 2) \notin R$.
$(2, 4) \in R$, but $(4, 2) \notin R$.
$(2, 4) \in R, (4, 16) \in R$, but $(2, 16) \notin R$.

E6) Consider the set $S = \{a, b\}$. Then $\{a\}R\{b\}$ and $\{b\}R\{a\}$, but $\{a\}R\{a\}$ is not true. Therefore, R is not transitive.

E7) If we assume that every resident of Kochi has one and only one mobile phone number, then we can define $f: X \to Y: f(x) =$ mobile no. of x.
Another way could be if we list all the elements of X from 1, (i.e., $00\ldots01$ in the 10-digit format) onwards, calling it his/her resident number, then we can define
$g: X \to Y: g(x) =$ resident number of x.

The domains of f and g are the set of residents of Kochi.
Range $f = \{$ Mob. No. of $x \mid x \in X\}$,
Range $g = \{$ Resident no. of $x \mid x \in X\}$.

E8) i) For each $n \in \mathbb{N}$, there is $\frac{n}{n+1} \in \mathbb{Q}$.

Also, if $n, m \in \mathbb{N}$ such that $n = m$, then $\frac{n}{n+1} = \frac{m}{m+1}$.

So, for each $n \in \mathbb{N}$, there is a **unique** element $q \in \mathbb{Q}$, such that $f(n) = q$. Here $q = \frac{n}{n+1}$.

Hence f is well-defined.

ii) There can be several such functions from $\mathbb{N}$ to $\mathbb{Q}$. For example, take $g(n) = n$, or $h(n) = \frac{n+1}{n}$. You can check that $f(n) \neq g(n)$ and $f(n) \neq h(n)\ \forall n \in \mathbb{N}$.

E9) For example, take the relation R from $\mathbb{N}$ to $\mathbb{Q}$, $R = \{(1, 1), (1, -1), (2, 2), (2, -2), \ldots\}$. Since an element of $\mathbb{N}$ is not uniquely mapped under R to an element of $\mathbb{Q}$ (e.g., $(2, 2)$ and $(2, -2)$ are both in R), R is not a function.

E10) This is false. For instance, take $f : \mathbb{R} \to \mathbb{R} : f(x) = x$ and $g : \mathbb{R} \to \mathbb{R} : g(x) = x + 1$. Then the domains of f and g are $\mathbb{R}$, which is also their ranges. But $f(x) \neq g(x)\ \forall x \in \mathbb{R}$.

E11) Domain $f = \mathbb{N}$, Range $f = \{6, 7, \ldots\}$, Co-domain $f = \mathbb{N}$. So $f = \{(n, n+5) \mid n \in \mathbb{N}\} \subseteq \mathbb{N} \times$ Range f.
f is not surjective since Range $f \neq$ Co-domain f.
Now, to check whether f is injective, suppose $f(m) = f(n)$, where $m, n \in \mathbb{N}$.
Then $m + 5 = n + 5$, which implies $m = n$. Thus, f is $1-1$.

E12) In this case, for any $z \in \mathbb{Z}, \exists\, z - 5 \in \mathbb{Z}$ such that $f(z-5) = z$. So, Range $f = \mathbb{Z}$. Therefore, f is onto.
As in E11, f is $1-1$. Hence f is bijective.

E13) Firstly, by definition, $X \neq \emptyset$.
Next, if X has more than one element, say, x_1, x_2, then $x_1 \neq x_2$, but $f(x_1) = c = f(x_2)$. This would mean that f is not injective. So, for f to be injective, X must be a singleton.
f is surjective because Range $f =$ Co-domain $f = \{c\}$.

E14) i) $f \circ g(x) = f(g(x)) = f(x+5) = 5(x+5)\ \forall\ x \in \mathbb{R}$.
$g \circ f(x) = g(f(x)) = g(5x) = 5x + 5\ \forall\ x \in \mathbb{R}$.

ii) $f \circ g(x) = x, g \circ f(x) = x$.

Note that $f \circ g = g \circ f$ in (ii) but not in (i).

E15) There are several examples. One is the following:

Define $f:\mathbb{N}\to\mathbb{Z}:f(n)=-n$ and $g:\mathbb{N}\to\mathbb{Q}:g(n)=\frac{1}{n}$.

Then Range $f=\{-1,-2,-3,\ldots\}\not\subset\mathbb{N}=$ Domain g.

Range $g=\left\{1,\frac{1}{2},\frac{1}{3},\cdots\right\}\not\subset\mathbb{N}=$ Domain f.

Thus, neither $g\circ f$ nor $f\circ g$ are defined.

E16) $g\circ f:A\to D$.

i) True, because if $y\in$ Range $g\circ f$, then $\exists\, x\in A$ such that $g\circ f(x)=y$. This means $g(f(x))=y$, which shows that $y\in$ Range g.
So Range $g\circ f\subseteq$ Range g.

ii) False, because for example, take $f:\mathbb{N}\to\mathbb{Z}:f(n)=-n$ and $g:\mathbb{Z}\to\mathbb{Z}:g(z)=z^2$. Then $g\circ f:\mathbb{N}\to\mathbb{Z}:g\circ f(n)=n^2$. So, $0\in$ Range g, but $0\notin$ Range $(g\circ f)$.

iii) False. For example, take $f:\mathbb{Z}\to\mathbb{Z}:f(z)=-z$, and g as in (ii) above. Here, check that f is $1-1$, but $g\circ f$ is not.

iv) False. For example, take $f:\mathbb{N}\to\mathbb{Z}:f(n)=-n$ and $g:\mathbb{R}\to\mathbb{R}:g(x)=x+5$. Then check that g is surjective, but $g\circ f:\mathbb{N}\to\mathbb{R}:g\circ f(n)=5-n$ is not surjective.

E17) Note that $I_B:B\to B,\ g:B\to A,\ I_A:A\to A$.
So $I_A\circ g$ and $g\circ I_B$ are well-defined, and $I_A\circ g:B\to A,\ g\circ I_B:B\to A$.
Also $I_A\circ g(b)=g(b)\ \forall\ b\in B$ and $g\circ I_B(b)=g(b)\ \forall\ b\in B$. Hence the result.

E18) For example, in Example 8 $g\circ f$ is onto, but f is not. You can find several other examples.

E19) i) To get x from $\frac{x}{3}$, we need to multiply by 3. So, let us define $g:\mathbb{R}\to\mathbb{R}:g(r)=3r$.
Then $g\circ f$ and $f\circ g$ are both defined.
Further, $g\circ f(x)=g\left(\frac{x}{3}\right)=x=I_{\mathbb{R}}(x)$
and $f\circ g(x)=f(3x)=x=I_{\mathbb{R}}(x)$.

ii) To get r from $2r+\frac{5}{7}$, we need to first subtract $\frac{5}{7}$ and then divide by 2. So, let us define $h:\mathbb{Q}\to\mathbb{Q}:h(x)=\frac{1}{2}\left(x-\frac{5}{7}\right)$. Then check that $h\circ g$ and $g\circ h$ are both defined, and are equal to $I_{\mathbb{Q}}$.

E20) i) $f_1:\mathbb{R}\to\mathbb{R}:f_1(x)=x^2$ is not $1-1$, since, for example, $f_1(1)=f_1(-1)$ and $1\neq-1$. So f_1 doesn't have an inverse.

ii) f_2 is neither an injection nor a surjection. Hence f_2 doesn't have an inverse.

iii) f_3 is bijective, and hence has an inverse. In fact, we can easily see that x can be extracted from $11x+7$ by first subtracting 7 and then dividing by 11.

So if we define $g:\mathbb{R}\to\mathbb{R}$ by $g(y)=\dfrac{y-7}{11}$, then check that $f_3\circ g = g\circ f_3 = I_{\mathbb{R}}$. So f_3 has the inverse g.

E21) i) For any two subsets A, B of X, we have $A\setminus B, A\cup B$ are subsets of X. So, complementation and union are binary operations on $\wp(X)$.
Take X to be $\mathbb{N}$. Then $\{1, 2\}\setminus\{2, 3\}\neq\{2, 3\}\setminus\{1, 2\}$, for example.
So complementation is not commutative.
Similarly, you can see that
$(\{1, 2, 3, 4\}\setminus\{1, 2\})\setminus\{4, 5\}\neq\{1, 2, 3, 4\}\setminus(\{1,2\}\setminus\{4, 5\})$, so that complementation is not associative.
However, as you have seen in Sec. 1.5, union is both commutative and associative.

ii) For any two subsets A, B of a non-empty set X, $A\times B$ is not a subset of X. This is a subset of $X\times X$. Therefore, Cartesian product is not a binary operation on $\wp(X)$.

E22) i) Check that $x\oplus y = y\oplus x$ and $(x\oplus y)\oplus z = x\oplus(y\oplus z)\forall x, y, z\in\mathbb{R}$. However, for example, $1\oplus 1=-1\notin\mathbb{N}$. So $\oplus$ is not closed on $\mathbb{N}$.

ii) $*$ is commutative, but not associative since $(x*y)*z = 2[2x+2y+z]$, and $x*(y*z)=2[x+2y+2z]$.
However, for $x, y\in\mathbb{N}, x*y\in\mathbb{N}$, so that $*$ is closed on $\mathbb{N}$.

iii) $x\Delta y$ is neither commutative nor associative. Further, Δ is not closed on $\mathbb{N}$.
Now, for $x, y, z\in\mathbb{R}$, consider

$$x\oplus(y\Delta z) = x+(y\Delta z)-5 = x+\frac{y-z}{2}-5.$$

Also,

$$(x\oplus y)\Delta(x\oplus z) = (x+y-5)\Delta(x+z-5) = \frac{(x+y-5)-(x+z-5)}{2}$$

$$=\frac{y-z}{2}.$$

So, $x\oplus(y\Delta z)\neq(x\oplus y)\Delta(x\oplus z)$
Therefore, $\oplus$ does not distribute over Δ.

UNIT 3

2D COORDINATE SYSTEMS

Structure **Page Nos.**

3.1 INTRODUCTION

We will first briefly discuss a system you would have studied in school, namely, the Cartesian coordinate system. This system of representing points in a plane was introduced in the early 17^{th} century by René Descartes, the French mathematician you read about in Unit 1. It is said that René Descartes (pronounced re-nay daycart) was a sick child and was, therefore, allowed to remain in bed till quite late in the mornings. One day when he was lying in bed, he saw a spider near one corner of the ceiling. Its movement led Descartes to think about the problem of determining its position on the ceiling. He decided that it was sufficient to know the eastward and the northward distances of the spider from the corner of the ceiling! This is supposed to have sown the seed for the development of the subject known as **coordinate geometry**.

In Sec. 3.2 you will study the idea of coordinates and how Descartes' system helps in studying all of plane geometry. In particular, you will study various algebraic representations of a straight line in Sec. 3.3. In Sec. 3.4 you will work with graphs, which are a geometric way of viewing functions, using the Cartesian system. As you go through Sec. 3.3 and Sec. 3.4, you would see the close interaction between algebraic and geometric representations of various curves.

Up to this point, you would have thought that Descartes was the only person to come out with a method for positioning points in a plane. However, in Sec. 3.5

you will study another system for locating points in a plane, which is very important in your study of Calculus and Analysis. This is the polar coordinate system, known from ancient times in a limited non-formal way. This system will be very useful for studying Unit 4 also.

Objectives

After studying this unit, you should be able to:

- explain what a Cartesian coordinate system is;
- use the Cartesian system to give different algebraic representations of straight lines in a plane;
- geometrically represent functions by their graphs in the Cartesian coordinate system;
- explain the polar coordinate system, and its relationship with the Cartesian coordinate system in which the origin and pole coincide, and the x-axis and polar axis coincide;
- geometrically represent functions by their graphs in the polar coordinate system.

3.2 THE CARTESIAN COORDINATE SYSTEM

You would be familiar with the concepts of point, line, circle, etc. You would also recall that real numbers are represented on a number line like the one in Fig.1. Here O denotes the number zero, the other points on $\overrightarrow{\mathrm{OX}}$ denote the positive real numbers and the points on $\overrightarrow{\mathrm{OX}'}$ denote the negative real numbers. If we choose the point A on $\overrightarrow{\mathrm{OX}}$ such that the length of OA is 1 unit, then A denotes 1 on this line.

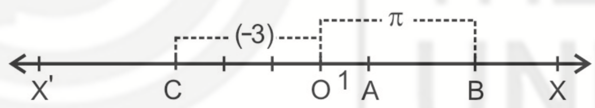


Fig. 1: The real number line

So, for example, π will be denoted by B, where the distance OB is approximately 3.1416. Similarly, -3 will be denoted by C, where the distance OC is thrice the distance OA, with C on $\overrightarrow{\mathrm{OX}'}$, that is, the 'negative' side.

Remark 1: Note that the number line need not be shown horizontal. It could also be vertical or at any other angle to the horizontal.

Now suppose you put a dot on a sheet of paper [as in Fig.2 (a)], which I can't see. Now, I ask you to describe the position of the dot on the paper in a manner that I can visualise it (remember, I can't see the paper and the dot). How would you do this? Perhaps you would say, "The dot is in the upper half of the paper", or "It is near the left edge of the paper". Do any of these statements fix the position of the dot precisely and unambiguously? No! But, if you say "The dot is 2 cm away from the left edge of the paper", it helps to give me some idea, but still does not fix the position of the dot precisely. But, if you also tell me that the dot is also at a distance of 9 cm above the bottom edge, then I can tell exactly where the dot is!

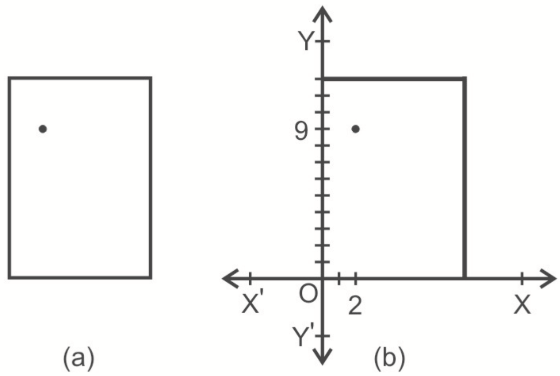


Fig. 2: Positioning the dot precisely requires knowing its distance from at least two non-parallel lines

So, what did you do? You fixed the position of the dot by specifying its distance from two lines perpendicular to each other, namely, the line along the left edge of the paper, and the line along the bottom edge of the paper [Fig. 2 (b)].

In fact, this is the basis of the Cartesian coordinate system. We can take any point O in a plane and fix it. Then draw a horizontal line $\mathrm{X'OX}$ and a vertical line $\mathrm{Y'OY}$ through O. These lines divide the whole plane of the paper into four parts, which are called **quadrants** (first, second, third and fourth), as shown in Fig. 3.

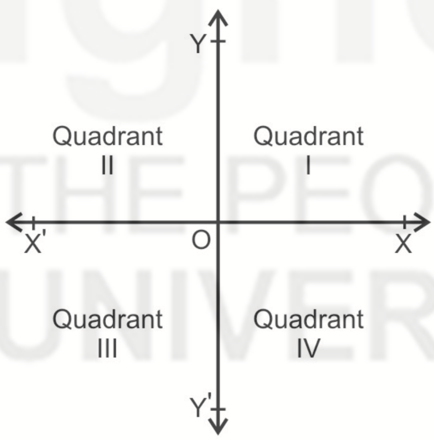


Fig. 3: The x-axis, y-axis and the four quadrants

The point O is called the **origin** and the lines $\mathrm{X'OX}$ and $\mathrm{Y'OY}$ are called the **x-axis** and the **y-axis**, respectively. Starting with the origin O, we mark off the points $1, 2, 3, \ldots$ at equal distances along $\overrightarrow{\mathrm{OX}}$ as in Fig. 1, and $-1, -2, -3, \ldots$, similarly, on $\overrightarrow{\mathrm{OX'}}$. In a similar manner, we write $1, 2, 3, \ldots$ along $\overrightarrow{\mathrm{OY}}$ and $-1, -2, -3, \ldots$ along $\overrightarrow{\mathrm{OY'}}$. $\overrightarrow{\mathrm{OX}}$ and $\overrightarrow{\mathrm{OY}}$ are the positive parts of the **axes** (plural of axis), and $\overrightarrow{\mathrm{OX'}}$ and $\overrightarrow{\mathrm{OY'}}$ are the negative parts. Now, recall how we represent any point on a plane, using this frame. Let P be any point in the plane. We have already drawn the origin and the axes. Through P we draw lines $\overrightarrow{\mathrm{PL}}$ and $\overrightarrow{\mathrm{PM}}$ perpendicular to the x-axis and y-axis, respectively (see Fig. 4), where L lies on the x-axis and M lies on the y-axis. Then PL is the distance of P from the x-axis and PM is the distance from the y-axis. Also note that $\mathrm{PM} = \mathrm{OL}$ and $\mathrm{PL} = \mathrm{OM}$.

AB denotes the length of the segment $\overline{\mathrm{AB}}$.

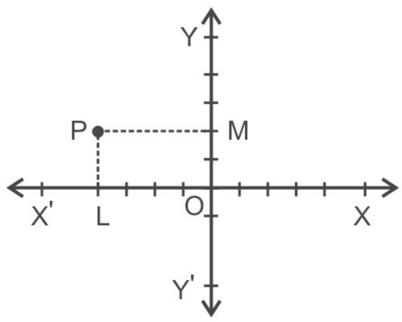


Fig. 4

The number that L represents on the x-axis is called the **x-coordinate** (or **abscissa**) of $\mathbf{P}$. The number that M represents on the y-axis is called the **y-coordinate** (or **ordinate**) **of** $\mathbf{P}$. The two taken together are known as the **coordinates of** $\mathbf{P}$, and are written as a pair, (OL, OM). Thus, if $OL = -4$ units and $OM = 2$ units, the coordinates of P are $(-4, 2)$. The expression $\mathbf{P(-4, 2)}$ indicates that P has coordinates -4 and $2, -4$ being the x-coordinate and 2 being the y-coordinate. Note that we write the x-coordinate first and then write the y-coordinate. This is a convention accepted by all users of this system. Using this convention, the point $(1, -4)$ will not be the same as the point $(-4, 1)$, as you can see in Fig. 5.

The points (x, y) and (y, x) are different unless $x = y$.

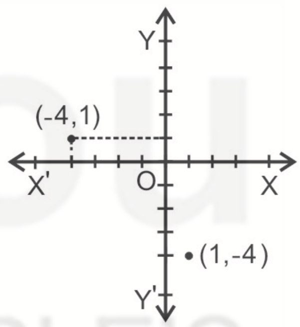


Fig. 5

This way of representing points in a plane is called the two-dimensional **Cartesian coordinate system**, after the mathematician Descartes who invented it. Here's an important related remark.

Remark 2: i) Points for which both coordinates are positive lie in Quadrant I.

ii) Points for which both coordinates are negative lie in Quadrant III.

iii) Points for which the x-coordinate is negative and y-coordinate is positive lie in Quadrant II.

iv) Points for which the x-coordinate is positive and y-coordinate is negative lie in Quadrant IV.

v) **Points on the axes do not lie in any quadrant**.

So, uptil now we have seen that if we know the position of a point P on a plane, then we can find a pair of real numbers (x, y) corresponding to P in the Cartesian coordinate system. The first number in the pair is called the x-coordinate and the second is called the y-coordinate, and the pair are called the coordinates of P.

Now, what about the other direction, i.e., given any pair of real numbers (x, y), can we find a unique point P in the plane for which x, y are its coordinates? To answer this, let us consider some pair, say, $(3, -2)$. Let A and B denote 3 along $\overrightarrow{OX}$ and -2 along $\overrightarrow{OY'}$ in Fig. 6.

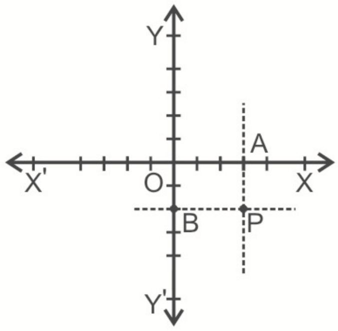


Fig. 6

Through A draw a line parallel to the y-axis, and through B draw a line parallel to the X-axis. What are the coordinates of the point P, at which they intersect? You can check that they are (OA, OB), that is, $(3, -2)$.

A Cartesian coordinate system represents the set $\mathbb{R}\times\mathbb{R}$.

In the same way, you can now see how, **given any $(x, y) \in \mathbb{R}\times\mathbb{R}$, there is a unique point P in the plane such that the coordinates of P are (x, y).**

Thus, you have seen that there is a 1-to-1 correspondence between points in a plane and a Cartesian coordinate system in which $X'OX$ and $Y'OY$ are fixed.

Try these exercises now.

E1) Plot the points with Cartesian coordinates $(-2, 2.5), (2, -2.5), (3.45, 0)$, $(0, -3.45)$, respectively.

E2) In which quadrants of the plane would (x, y) lie if:

i) $xy > 0$?

ii) $xy < 0$?

iii) $xy = 0$?

Justify your answers.

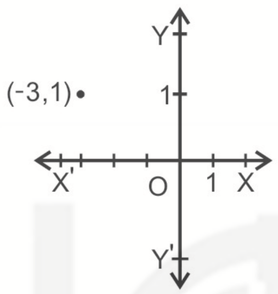


Fig. 7: The point $(-3, 1)$ in a Cartesian coordinate system in which the unit along the x-axis is different from the unit along the y-axis.

Let me end this section with an important remark.

Remark 3: In this section, one of the points you studied was how to mark off points along the x and y axes. In the examples used here, you have used the same unit along the x-axis and the y-axis. So, for example, if $OA = 1$ unit along the x-axis, and $OB = 1$ unit along the y-axis, then $OA = OB$. However, this need not be so. You can use one unit along the x-axis and another along the y-axis (see Fig. 7).

Let us now look at the way lines in a plane can be represented in the Cartesian system.

3.3 EQUATIONS OF A LINE

A two-dimensional space is a representation of a plane along with a coordinate system.

In this section we aim to discuss ways of representing straight lines in two-dimensional spaces algebraically. Since you may be familiar with the matter from school, we shall cover the ground quickly. We start with lines parallel to either of the axes.

3.3.1 Lines Parallel to the Axes

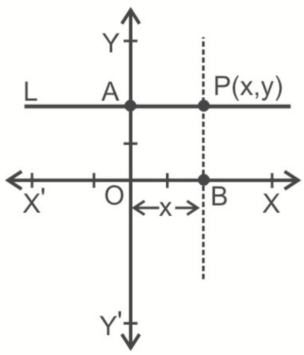


Fig. 8: The line L represents $y = 2$.

Consider a line L parallel to the x-axis, which intersects the y-axis in $A(0, 2)$. Now, take any point $P(x, y)$ on it, as in Fig. 8. You can see that the perpendicular from P onto the y-axis is along the line L, and hence intersects the y-axis in A. Thus, $y = 2$. Thus, for any point P on the given line, we find that its ordinate is 2.

Conversely, take any point Q with coordinates $(x, 2)$. This means that the perpendicular from Q onto the y-axis intersects it at the point $A(0, 2)$. So,

this perpendicular is along L. Therefore, Q will lie on L. This is regardless of the value of x.
Thus, this line L consists precisely of all those points whose coordinates (x, y) satisfy $y = 2$. **So L represents the equation $y = 2$.**

Similarly, **for any line parallel to the x-axis**, which intersects the y-axis in $(0, a), a \in \mathbb{R}$, its equation will be

$$\mathbf{y = a}. \qquad \ldots(1)$$

What do you expect the equation of a line parallel to the y-axis to be? Think about this while doing the following exercises.

E3) Find the equation of a line parallel to the y-axis, and which intersects the y-axis in
 i) $(-3, 0)$
 ii) $(b, 0)$, for some constant b.

E4) What are the equations of the coordinate axes?

If you have solved E3, you would have seen that any line parallel to the y-axis is of the form

$$\mathbf{x = b}, \text{ for some } b \in \mathbb{R}. \qquad \ldots(2)$$

Now let us obtain four forms of the equation of a line which is not parallel to either of the axes.

3.3.2 Slope-intercept Form

Consider any line not parallel to either axis. It will make an angle, say $\alpha, \alpha \neq 0, \pi/2$, with the positive direction of the x-axis, measured **in the anti-clock-wise direction**, and will cut the y-axis in $A(0, c)$, say. Then, if $P(x, y)$ is any point on this line, from Fig. 9 you can see that $\tan\alpha = \frac{y-c}{x}$.

Thus, $y = x\tan\alpha + c$,

i.e., $\mathbf{y = mx + c}$, where $m = \tan\alpha$. …(3)

m gives the slant of the line L to the horizontal. It is called the **slope** of L. c is called the **intercept** of L on the y-axis. Thus, $y = mx + c$ is called the **slope-intercept form** of the equation of a line.

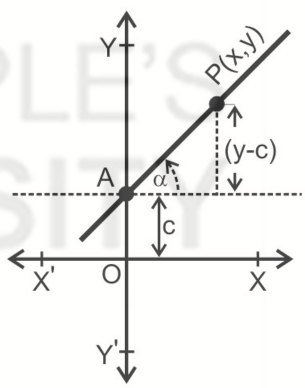


Fig. 9: L is given by $\mathbf{y = x\tan\alpha + c}$.

Let us consider an example.

Example 1: Find the equation of the line which makes the same angle with the x-axis as with the y-axis, and which intercepts the y-axis in $(0, -\pi)$.

Solution: Let the angle made by the line with the x-axis be α. Then consider Fig. 10. The line passes through $A(0, -\pi)$. Let the line intersect the x-axis in P. If we draw $\overrightarrow{AM}$ parallel to $\overrightarrow{OX}$, then $\angle MAP = \alpha$.
Also, we are given that $\angle PAY = \alpha$. Thus, $2\alpha = \pi/2$. Hence $\alpha = \pi/4$. So, the equation of the line is $y = x\tan\frac{\pi}{4} + (-\pi)$, that is,

$y = x - \pi$.

'Intercept' on an axis is the distance from the origin of the point at which the line cuts the axis concerned.

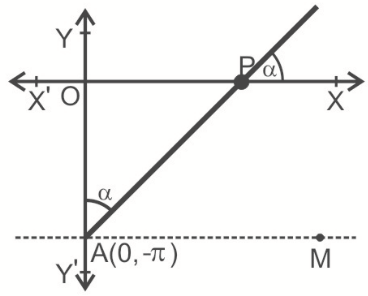


Fig. 10

***

Try some exercises now.

E5) Find the equation of the line that cuts off an intercept of 1 from the negative direction of the y-axis, and is inclined at 120° to the x-axis.

E6) What is the equation of the line passing through the origin and making an angle θ with the x-axis, where $\theta \neq 0, \frac{\pi}{2}$?

Let us now see another way of representing lines algebraically, closely related to the form (3).

3.3.3 Point-slope Form

Now, suppose we know the slope m of a line but not the intercept it makes with the y-axis. We also know that a point (x_1, y_1) lies on the line. Then, can we obtain the line's equation? Let $P(x, y)$ be any point on the line, with slope $m = \tan\alpha$, and $Q(x_1, y_1)$ be the given point. Now, from Fig. 11, we see that

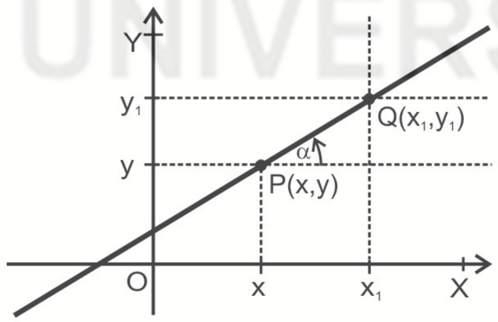


Fig. 11: $\mathbf{m = \frac{y_1 - y}{x_1 - x}}$.

$m = \tan\alpha = \frac{y_1 - y}{x_1 - x} = \frac{y - y_1}{x - x_1}$. (**Note** that $x_1 \neq x$, since the line is not parallel to the y-axis.) Thus,

$$\mathbf{y - y_1 = m(x - x_1)}, \quad \ldots(4)$$

which is the **point-slope** form of a line.

Let us consider an example.

Example 2: Find the equation of a line L which is parallel to the line

$3y = x - \sqrt{3}$, and which passes through $\left(-1, -\frac{1}{7}\right)$.

Solution: Since L is parallel to $3y = x - \sqrt{3}$, that is, $y = \frac{1}{3}x - \frac{1}{\sqrt{3}}$, its slope will be the same as the slope of this line, namely, $\frac{1}{3}$. Also, we know that L passes through $\left(-1, -\frac{1}{7}\right)$. Thus, by (4), its equation is

$$y - \left(-\frac{1}{7}\right) = \frac{1}{3}(x - (-1)),$$

$$\Leftrightarrow \quad y + \frac{1}{7} = \frac{1}{3}x + \frac{1}{3}$$

$$\Leftrightarrow \quad y = \frac{1}{3}x + \frac{4}{21}.$$

***

Try an exercise now which helps you see the close relationship between the forms (3) and (4).

E7) If L is a line with slope m, and which cuts off an intercept of length c from the y-axis, how would you use (4) to obtain its equation?

Let us consider yet another form now.

3.3.4 Two-point Form

Now, suppose we have a line which is not parallel to the axes but we don't know its slope. Can we find its equation? Yes, if we know two distinct points lying on it. Let us see how.

Let $P(x_1, y_1)$ and $Q(x_2, y_2)$ be two distinct points on a line L, which is not parallel to the axes. Let $R(x, y)$ be any point on it (see Fig. 12).

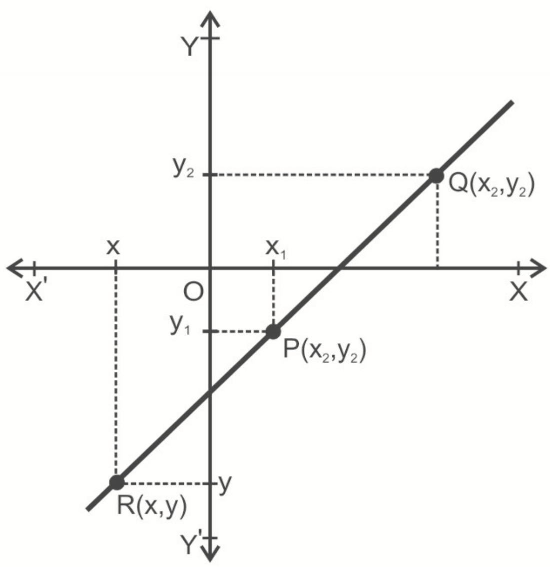


Fig. 12: $\frac{y - y_1}{y_2 - y_1} = \frac{x - x_1}{x_2 - x_1}$.

Then the slope of L, say m, will be $m = \frac{y_1 - y}{x_1 - x}$. (Note that $x_1 - x \neq 0$. (Why?))

You can see that m will also be $\frac{y_2 - y_1}{x_2 - x_1}$.

Thus, $\frac{y_1 - y}{x_1 - x} = m = \frac{y_2 - y_1}{x_2 - x_1}$

$\Leftrightarrow \quad \frac{y_1 - y}{y_2 - y_1} = \frac{x_1 - x}{x_2 - x_1}$ (Note that $y_2 \neq y_1$.)

$$\Leftrightarrow \quad \mathbf{\frac{y - y_1}{y_2 - y_1} = \frac{x - x_1}{x_2 - x_1}} \qquad \ldots(5)$$

which is known as the **two-point form** of the line.

As you can see, the slope of the line is actually known — it is $\mathbf{\frac{y_2 - y_1}{x_2 - x_1}}$.

Also, **its intercept on the y-axis, which can be obtained by finding the intersection of (5) with the y-axis**, i.e., $x = 0$, is given by

$$\frac{y - y_1}{y_2 - y_1} = \frac{-x_1}{x_2 - x_1}$$

i.e., $y = y_1 - x_1\left(\frac{y_2 - y_1}{x_2 - x_1}\right)$.

Let us consider an example.

Example 3: Find the equation of the line whose intercept on the x-axis is 2 and which passes through $\left(\frac{1}{\pi}, \pi\right)$.

Solution: We know that the line required passes through $(2, 0)$ and $\left(\frac{1}{\pi}, \pi\right)$.

Hence, using (5), its equation is $\frac{y - 0}{\pi - 0} = \frac{x - 1/\pi}{2 - 1/\pi}$

$\Leftrightarrow \quad y = \frac{\pi}{2\pi - 1}(\pi x - 1)$.

***

Try an exercise now, which actually leads you to another well known form of the equation of a line.

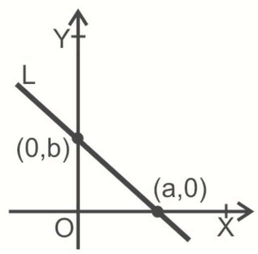


Fig. 13: L is given by $\mathbf{\frac{x}{a} + \frac{y}{b} = 1}$.

(6) is called the **intercept form** of the equation of L.

E8) i) Suppose we know that the intercept of a line on the x-axis is 2 and on the y-axis is -3. Then show that its equation is $\frac{x}{2} - \frac{y}{3} = 1$.
(**Hint:** See how you can use (5).)

ii) More generally, if a line L cuts off an intercept $a(\neq 0)$ on the x-axis and $b(\neq 0)$ on the y-axis (see Fig.13), then show that its equation is

$$\mathbf{\frac{x}{a} + \frac{y}{b} = 1} \qquad \ldots(6)$$

Now, have you noticed a characteristic that is common to all the equations of lines you have studied here? They are all linear in two variables, that is, of the form $ax+by+c=0$, where $a, b, c \in \mathbb{R}$ and at least one of a and b is non-zero. This is not a coincidence, as the following theorem tells us. (We shall not formally prove this result.)

Theorem 1: A linear equation in two variables represents a straight line in two-dimensional space. Conversely, the equation of a straight line in the plane is a linear equation in two variables.

So, for example, $2x+3y-1=0$ represents a line. What is its slope? To find this, we rewrite it as $y=-\frac{2}{3}x+\frac{1}{3}$. The slope is the coefficient of x, namely $-\frac{2}{3}$. Do you agree that its intercepts on the x and y axes are $\frac{1}{2}$ and $\frac{1}{3}$, respectively? You can check this by finding its intersection with the x and y axes respectively.

Why don't you try an exercise now?

E9) i) Find the equation of the line parallel to $y+x+1=0$ and passing through $(0, 0)$.

ii) What is the equation of the line perpendicular to the line obtained in (i) above, and passing through $(2, 1)$?

Let us now stop our discussion on lines and linear equations, and move on to more general equations. We shall discuss a concept that will help us to represent functions geometrically in the rest of the course.

3.4 GRAPH OF A FUNCTION

In Unit 2, you studied functions. There, you saw that the function f is a subset of the Cartesian product $A \times B$ with special properties, where A is the domain of f and B is the co-domain of f. In this section we shall look at the geometrical representation of f, using Cartesian coordinates, when $A \times B \subseteq \mathbb{R} \times \mathbb{R}$.

In the previous section you have already studied this for one type of function. Do you agree? Isn't a line a representation of a function from $\mathbb{R}$ to $\mathbb{R}$? For instance, take any line, say $2x=3y+7$. This is the set of points $\left\{\left(x, \frac{2x-7}{3}\right) \middle| x \in \mathbb{R}\right\} \subseteq \mathbb{R} \times \mathbb{R}$. Check that this relation satisfies all the requirements of a function. So, this can also be written as the function $f: \mathbb{R} \to \mathbb{R} : f(x) = \frac{1}{3}(2x-7)$.

So, this function is represented by the line $2x=3y+7$, i.e., $y=\frac{2}{3}x-\frac{7}{3}$, which has slope $\frac{2}{3}$ and y-intercept $-\frac{7}{3}$.

In general, the function $f:\mathbb{R}\to\mathbb{R}:f(x)=ax+b,\ a,\ b\in\mathbb{R}$, is represented by a line with slope a and y-intercept b, as in Fig. 14. This representation of f is called **the graph of f**.

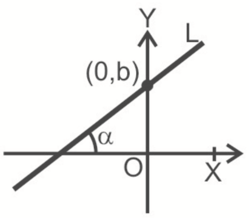


Fig. 14: L represents $f:\mathbb{R}\to\mathbb{R}:f(x)=ax+b$, where $a=\tan\alpha$.

More generally, given any function from $A\subseteq\mathbb{R}$ to $B\subseteq\mathbb{R}$, we have the following definition.

Definition: Let f be a function with domain A and co-domain B, where A and B are subsets of $\mathbb{R}$. The **graph of f** is the subset $\{(x,f(x))\mid x\in A\}$ of $\mathbb{R}\times\mathbb{R}$.

Thus, when this set is represented using the Cartesian system, we see it as the geometrical view of the function. As you have seen, if the function is a linear polynomial, its graph is a straight line. Let us consider some more examples.

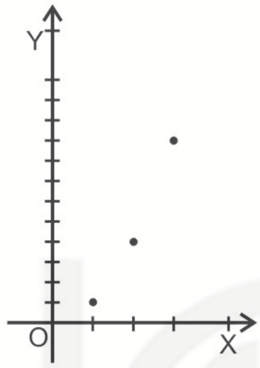


Fig. 15

Example 4: Draw the graph of the function $f:\{1,2,3\}\to\{1,2,\ldots,10\}:f(1)=1,\ f(2)=4,\ f(3)=9$.

Solution: We plot the points in the set $\{(1,1),(2,4),(3,9)\}$ in a Cartesian system, as in Fig. 15. Thus, the graph of f comprises the three points shown in Fig. 15.

***

Example 5: Draw the graph of the constant function $f:\mathbb{R}\to\mathbb{R}:f(r)=1$.

Solution: The graph of this function is the set of points representing $\{(r,1)\mid r\in\mathbb{R}\}$, that is, the line $y=1$ shown in Fig. 16.

***

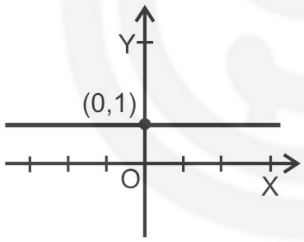


Fig. 16: $y=1$.

Example 5 leads into the following remark.

Remark 4: Look at the graph in Fig. 16. Take any value b along the y-axis, $b\neq1$. You can see that the line $y=b$ does not intersect the graph at all. But b lies in the co-domain $\mathbb{R}$ of f. What does this show? It tells us that **f is not surjective** (see Unit 2). More generally, by looking at the graph of a function, you can tell it is not onto if the line $y=b$ does not intersect the graph for even one value of b in the co-domain of the function.

Here's an example about checking surjectivity.

Example 6: Draw the graph of the function given in E8(i), Unit 2. Hence, check whether this function is onto, by inspecting its graph.

Solution: $f:\mathbb{N}\to\mathbb{Q}:f(n)=\dfrac{n}{n+1}$.

Thus, the graph is as in Fig. 17. Note that we have used different units along the x-axis and along the y-axis, for convenience.

You can see that the graph comprises infinitely many points, getting nearer and nearer to the line $y=1$, but lying below it even when n gets very very large.

Note that 1 lies in the co-domain $\mathbb{Q}$ of f, but $y=1$ does not intersect the graph of f. This tells us that f is not onto.

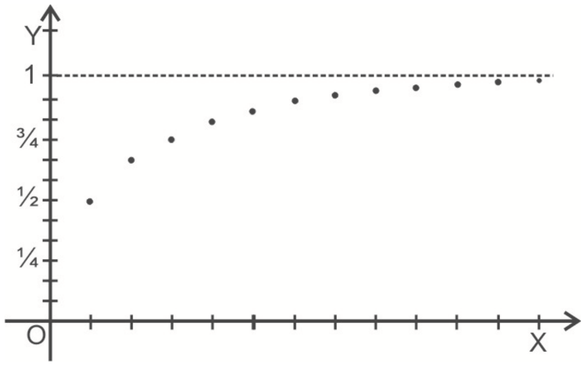


Fig. 17: The graph of $\mathbf{f : \mathbb{N} \to \mathbb{Q} : f(n) = \frac{n}{n+1}}$.

***

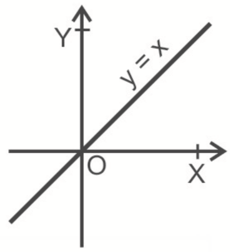


Fig. 18: The graph of $\mathbf{f : \mathbb{R} \to \mathbb{R} : f(r) = r}$.

Example 7: Draw the graph of the identity function from $\mathbb{R}$ to $\mathbb{R}$.

Solution: $f : \mathbb{R} \to \mathbb{R} : f(r) = r$.

Thus, the graph is the set $\{(r, r) \mid r \in \mathbb{R}\}$, which is drawn in Fig. 18. You can see that this is the line $y = x$.

***

Example 7 leads into an important point.

Remark 5: If you inspect the graph in Fig. 18 and take any line parallel to the x-axis, say $y = b$, it will intersect the graph in one point only, namely, (b, b). This property of the graph tells us that f is $1-1$ (see Unit 2). More generally, **if $\mathbf{f : A \to B \subseteq \mathbb{R}}$, and $\forall$ $\mathbf{b \in B}$ $\mathbf{y = b}$ intersects the graph of $\mathbf{f}$ in at most one point, then $\mathbf{f}$ is $\mathbf{1-1}$.** If the intersection is two or more points for even one $b \in B$, then f will not be $1-1$.

Example 8: Draw the graph of the function $g : \mathbb{R} \to \mathbb{R} : g(x) = x^2$. Hence, decide whether

i) g is $1-1$ or not,

ii) g is onto or not.

Solution: The graph of g is the set of points $\{(x, x^2) \mid x \in \mathbb{R}\}$. As you can see, this is the parabola $y = x^2$, shown in Fig. 19.

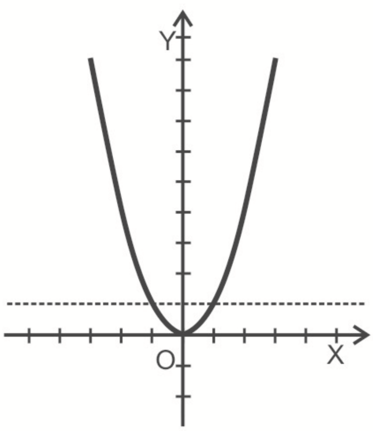


Fig. 19: The curve represents $\mathbf{\{(x, x^2) \mid x \in \mathbb{R}\}}$.

i) From the graph, you can see that, for example, the line $y = 1$ intersects the graph in two points $(1, 1)$ and $(-1, 1)$. Hence g is not $1-1$.

ii) From the graph you can see that for any $\alpha \in \mathbb{R}, \alpha < 0, y = \alpha$ does not intersect the graph. Hence g is not onto.

***

f defines the **modulus function**,

$$|x| = \begin{cases} x, & x \geq 0 \\ -x, & x < 0 \end{cases}$$

$\mathbb{R}^- = \{x \in \mathbb{R} \mid x < 0\}$

$\mathbb{R}^+ = \{x \in \mathbb{R} \mid x > 0\}$

Example 9: Draw the graph of $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x$, if $x \geq 0$ and $f(x) = -x$, if $x < 0$. Further, by inspecting the graph, decide if f is bijective or not.

Solution: To draw the graph of this function notice that it is defined differently for negative real numbers, $\mathbb{R}^-$, and for non-negative numbers, $\mathbb{R}^+ \cup \{0\}$. So, the graph is the set union $\{(x, x) \mid x \in \mathbb{R}^+\} \cup \{(0, 0)\} \cup \{(x, -x) \mid x \in \mathbb{R}^-\}$. It is shown in Fig. 20.

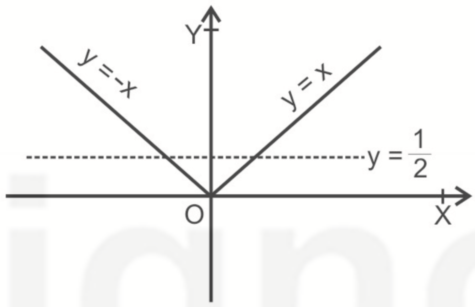


Fig. 20: Graph of $|x|$.

In Example 9 you can also show that f is not bijective by showing that it is not surjective.

You can see that on $\mathbb{R}^+$ it is the line $y = x$ and on $\mathbb{R}^-$ it is the line $y = -x$. Further, by looking at it, you can see that the line $y = \frac{1}{2}$ intersects it in two points. Hence f is not injective, and thus, not bijective.

***

Why don't you do some related exercises now?

E10) Draw the graphs of the functions given below:

i) $f : \mathbb{N} \to \mathbb{N} : f(n) = n + 5$;

ii) $g : \mathbb{Z} \to \mathbb{Z} : g(n) = n + 5$;

iii) $h : \mathbb{R} \to \mathbb{R} : h(r) = r + 5$.

What is the difference you note in the three graphs above? Include differences regarding injectivity and surjectivity.

E11) Draw the graph of $f : \mathbb{R} \to \mathbb{R} : f(x) = \begin{cases} 1, & x \geq 0 \\ -1, & x < 0 \end{cases}$.

Hence decide whether f is bijective or not.

E12) Draw the graph of the function $f : \mathbb{R} \to \mathbb{R}$ which represents the line given by $\frac{x}{\pi} + \frac{y}{\sqrt{2}} = 1$. Also check whether f is bijective or not.

So far we have been working with Cartesian coordinates. But there is another very useful coordinate system for locating points in a plane. Let's see what it is.

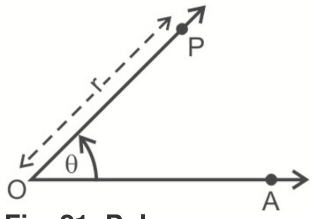


Fig. 21: Polar coordinates of P with O as pole, $\overrightarrow{OA}$ as polar axis, r as the radius of P and θ as the polar angle of P.

3.5 THE POLAR COORDINATE SYSTEM

You have come across the concepts of 'angle' and 'radius' before. Using these concepts one can determine the position of any point in a plane. Let us see how.

We first fix a point O in the plane called the **pole**. Then we fix an axis, usually a horizontal ray through O, called the polar axis, shown as $\overrightarrow{OA}$ in Fig. 21. Then we can locate any point P in the plane, if we know the distance OP, say r, and the angle AOP, say θ radians. r is called the **radial coordinate**, or **radius**, of P and θ is called the **angular coordinate**, or **polar angle**, of P. Thus, given a point P in the plane, we can represent it by a pair of coordinates (r, θ), where r is the "directed distance" of P from O and θ is $\angle AOP$, measured in radians **in the anticlockwise direction**. We use the term "directed distance" because r can be negative also. For instance, the point P in Fig. 22 can be represented by $\left(5, \frac{5\pi}{4}\right)$ or $\left(-5, \frac{\pi}{4}\right)$, because the positive distance is along the ray which makes an angle $\frac{5\pi}{4}$ with the polar axis. In general, given (r, θ), r is positive if it is measured along the ray which makes an angle θ with the polar axis, and r is negative if it is measured along the ray which makes an angle $\pi+\theta$ with the polar axis.

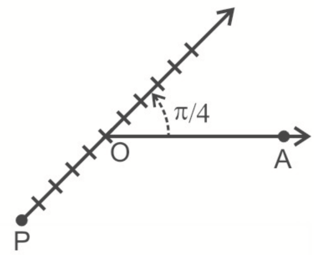


Fig. 22: P's polar coordinates are $\left(-5, \frac{\pi}{4}\right)$.

Note that by this method the pole O corresponds to $(0, \theta)$, for any angle θ.

A point has many different pairs of polar coordinates.

Thus, you have seen that for any point P, there is a pair of real numbers (r, θ) that corresponds to it. They are called the **polar coordinates** of P. These coordinates are not unique, as you have seen.

Now, if we keep θ fixed, say $\theta = \pi/3$, and let r take on all real values, we get the line $\overleftrightarrow{OP}$ (see Fig. 23), where $\angle AOP = \pi/3$.

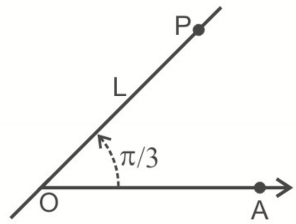


Fig. 23: The line L is given by $\theta = \pi/3$.

Similarly, keeping r fixed, say $r = 2$, and allowing θ to take all real values, the point $P(2, \theta)$ traces a circle of radius 2, with centre at the pole (see Fig. 24).

Remark 6: A negative value of θ means that the angle has magnitude $|\theta|$, but is taken in the clockwise direction. Thus, for example, the point $\left(2, -\frac{\pi}{2}\right)$ is also represented by $\left(2, \frac{3\pi}{2}\right)$.

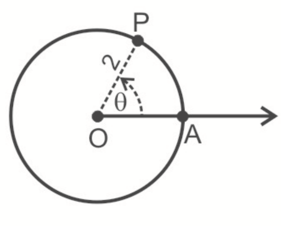


Fig. 24: The circle $r = 2$.

As you have probably guessed, the Cartesian coordinates (x, y) and polar coordinates (r, θ) are very closely related. Can you find the relationship? From Fig. 25 you would agree that the relationship is

$$x = r\cos\theta,\ y = r\sin\theta. \qquad \ldots(7)$$

This gives us

$$r = \sqrt{x^2 + y^2}, \ \theta = \tan^{-1}\frac{y}{x} \qquad \ldots(8)$$

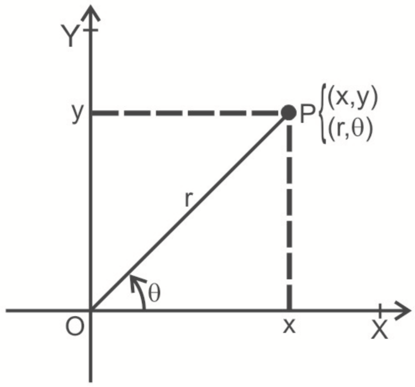


Fig. 25: Polar and Cartesian coordinates.

Note that the origin and the pole are coinciding here, as are the x-axis and the polar axis. This is usually the situation.

Remark 7: If the polar coordinates of a point are to be unique, we need to restrict θ to $0 \le \theta < 2\pi$. (The pole will then have the polar coordinates $(0, 0)$.) This restriction is consistent with the uniqueness in the Cartesian system, as you can see.

We use (7) and (8) often while dealing with equations, particularly where the relationship between two points is easily given in terms of r and θ. For example, the Cartesian equation of the circle $x^2 + y^2 = 25$, reduces to the simple polar form $r = 5$. So we may prefer to use this simpler form rather than the Cartesian one.

The equation of a curve in terms of r and θ is called its **polar equation**. You will often use this form in this course and other mathematics courses. To give you a flavour, let us consider an example.

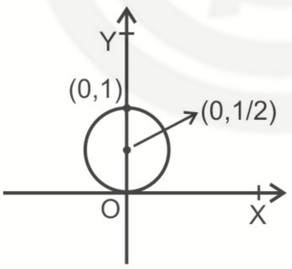


Fig. 26: The graph of $r = \sin\theta$.

Example 10: Draw the graph of the curve $r = \sin\theta$.

Solution: $r = \sin\theta$ is the curve given by $x^2 + y^2 = y$, in Cartesian form. This is the same as the equation $x^2 + \left(y - \frac{1}{2}\right)^2 = \frac{1}{4}$, which is a circle with radius $\frac{1}{2}$ and centre $\left(0, \frac{1}{2}\right)$. We have graphed this in Fig. 26.

***

Now, try the following exercises to get used to polar coordinates.

E13) Draw the graph of the curve $r\cos\theta = 1$

E14) Find the Cartesian forms of the equations

i) $r^2 = 3r\sin\theta$.

ii) $r = a(1 - \cos\theta)$, where a is a constant.

The history of the development, and use, of the polar coordinate system is very interesting. In fact, it goes back to the 2nd century BC, when the ancient Greek astronomer Hipparchus applied it for locating the stars. Nearer to our times, the 17th century German mathematician, Jacob Bernoulli, is credited with formalising this system, and using it to study curves.

The polar coordinate system, is useful for studying various properties of curves, and for drawing these curves. You will find it immediately useful in the next unit, while studying the geometric representation of a complex number.

Let us now summarise what you have studied in this unit.

3.6 SUMMARY

In this unit, we have briefly run through certain elementary concepts of two-dimensional analytical geometry. In particular, we have covered the following points:

1. The 2D Cartesian coordinate system represents two-dimensional space giving a one-to-one correspondence between the points in space and those in $\mathbb{R}\times\mathbb{R}$.

2. In the Cartesian system any line parallel to the x-axis is represented by $y = a$, and any line parallel to the y-axis is represented by $x = b$, for some constants a and b.

3. The equation of a line in
   i) slope-intercept form is $y = mx + c$,
   ii) point-slope form is $y - y_1 = m(x - x_1)$,
   iii) two-point form is $\frac{y - y_1}{y_2 - y_1} = \frac{x - x_1}{x_2 - x_1}$,
   iv) intercept form is $\frac{x}{a} + \frac{y}{b} = 1$.

4. **The polar coordinate system:** A point P in a plane can be represented by a pair of real numbers (r, θ), where r is the directed distance of P from the pole O, and θ is the angle that OP makes with the polar axis, measured in radians in the anticlockwise direction. r (called the radius) and θ (called the polar angle) are the **polar coordinates** of P. They are related to the Cartesian coordinates (x, y) of P by $r = \sqrt{x^2 + y^2}$ and $\theta = \tan^{-1}\frac{y}{x}$.

5. r and θ uniquely represent points in a plane only if we insist that either $0 \le \theta < 2\pi$, or $-\pi \le \theta < \pi$.

In the next unit we shall discuss numbers that are represented by the Cartesian and polar systems. But, before going to it, **please make sure that you have achieved the unit objectives listed in Sec. 3.1**. One way of checking this is to ensure that you have done all the exercises in the unit. Our solution to these exercises are given in the following section.

3.7 SOLUTIONS/ANSWERS

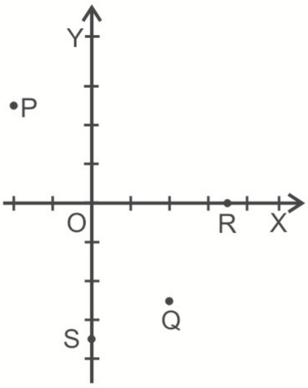


Fig. 27

E1) The points $P(-2, 2.5), Q(2, -2.5), R(3.45, 0), S(0, -3.45)$ are shown in Fig. 27.

E2) i) $xy > 0 \Rightarrow x > 0, y > 0$ or $x < 0, y < 0$.
Thus, (x, y) will lie in the Quadrants I or III.

ii) Arguing as in (i) above, (x, y) will lie in the Quadrants II or IV.

iii) $xy = 0 \Rightarrow x = 0$ or $y = 0$.
Thus, $(x, y) = (0, y)$ or $(x, y) = (x, 0)$. Accordingly, the point lies on the y-axis or the x-axis.

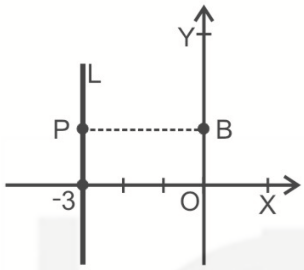


Fig. 28: L represents $x = -3$.

E3) i) Consider the line L in Fig. 28. This is parallel to the y-axis and passes through $(-3,0)$. Take any point $P(x, y)$ on L. Draw a perpendicular from P onto the y-axis, meeting it in B. Then $x = PB = -3$. Thus, any point on L is of the form $(-3, y), y \in \mathbb{R}$. Conversely, if you plot any point $(-3, r), r \in \mathbb{R}$, it will lie on L. Thus, the equation of L is $x = -3$.

ii) Arguing as in (i) above, the equation of this line will be $x = b$.

E4) The x and y-axes are given by $y = 0$ and $x = 0$, respectively.

E5) In Fig. 29 we have drawn the line. Its equation is $y = mx + c$, where $c = -1$ and $m = \tan 120^\circ = -\sqrt{3}$.
Thus, the required equation is $y = -(\sqrt{3}x + 1)$.

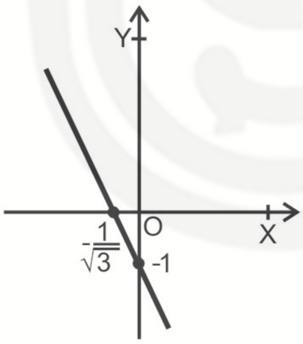


Fig. 29: $y = -(\sqrt{3}x + 1)$

E6) Here $c = 0$, since the line intersects the y-axis in $(0, 0)$. Thus, the equation is $y = x \tan\theta$.

E7) We are given that the slope of L is m, and $(0, c)$ lies on it. Thus, using (4), we get its equation as $y - c = m(x - 0)$, that is, $y = mx + c$.

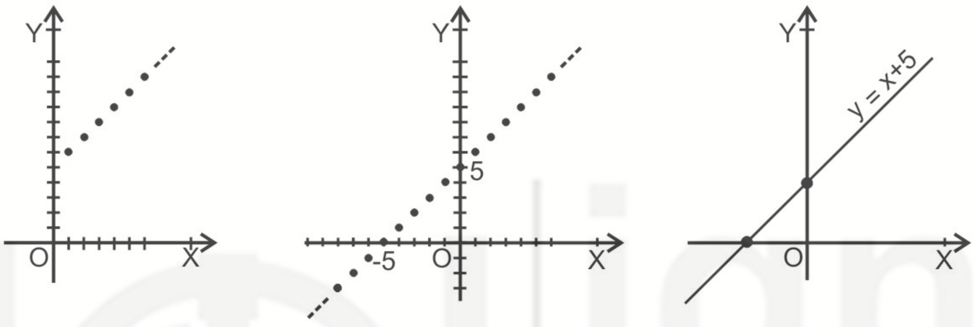
E8) i) $(2, 0)$ and $(0, -3)$ lie on the line. Thus, its two-point form is
$$\frac{y-0}{-3-0} = \frac{x-2}{0-2}, \text{ that is, } 2y = 3(x-2).$$
We can rewrite this as $\frac{x}{2} - \frac{y}{3} = 1$.

ii) $(a, 0)$ and $(0, b)$ lie on the line. Thus, its equation is
$$\frac{y-0}{b-0} = \frac{x-a}{0-a} \Leftrightarrow \frac{x}{a} + \frac{y}{b} = 1.$$

E9) i) The slope of $y + x + 1 = 0$ is -1. Thus the slope of any line parallel to the given line is -1. Thus, the line required is of the form $y + x + c = 0$, where $c \in \mathbb{R}$. Since $(0, 0)$ lies on it, $0 + 0 + c = 0$, that is, $c = 0$. Thus, the required line is $y + x = 0$.

ii) The slope of the line $y + x = 0$ is -1, that is, $\tan\left(\frac{\pi}{2} + \frac{\pi}{4}\right)$. Thus, the angle that any line perpendicular to it makes with the x-axis is $\left(\frac{\pi}{2} + \frac{\pi}{4}\right) \pm \frac{\pi}{2} = \frac{\pi}{4}, \frac{5\pi}{4}$. You should check that both these values give the same line. So, the slope of the line will be $\tan\frac{\pi}{4} = 1$.
Thus, the equation of the line required is of the form $y = x + c$, where $c \in \mathbb{R}$.
Since $(2, 1)$ lies on it, $1 = 2 + c$, so that $c = -1$.
Thus, the line required is $y = x - 1$.

E10) The graphs of f, g and h are shown in Fig. 30 (a), (b), (c), respectively.



(a) **(b)** **(c)**

Fig. 30

As you can see, in Fig. 30(a), the graph shows infinitely many points, all lying in the first quadrant. In Fig. 30(b), you have these points as well as one more on the x-axis and on the y-axis, four in the second quadrant, and infinitely many more in the third quadrant. In Fig. 30(c), you find the line $y = x + 5$. Also, all the points in the graphs of f and g actually lie on this line. Thus, the graph of f is a proper subset of the graph of g, which is a proper subset of the graph of h.

Considering injectivity, you can check that all three are injective. Considering surjectivity, (i) is not surjective because, for example, $1 \in \mathbb{N}$, but $y = 1$ does not intersect the graph. However, you can check that (ii) and (iii) are onto.

E11) Here, the graph of f comprises two parts, namely, $y = 1$ for $\mathbb{R}^+ \cup \{0\}$, and $y = -1$ for $\mathbb{R}^-$. Hence, the graph is as shown in Fig. 31. **We circle the point $(-1, 0)$ to show that it is not included**.

Looking at the graph, you can see that, for example, $0 \in \mathbb{R}$, but $y = 0$ (that is, the x-axis) does not intersect the graph. Hence, f is not $1-1$. Hence, f is not bijective.

You can also show it is not bijective by checking that it is not onto.

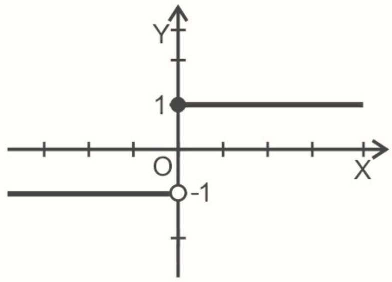


Fig. 31: The graph of $\mathbf{f}:\mathbb{R}\to\mathbb{R}:\mathbf{f(x)}=\begin{cases}\mathbf{1}, & \mathbf{x\geq 0}\\ \mathbf{-1}, & \mathbf{x<0}\end{cases}$

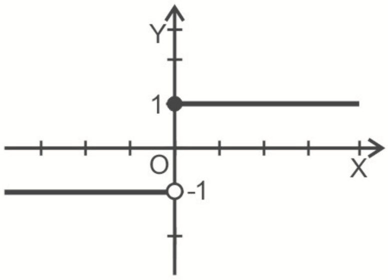


Fig. 32

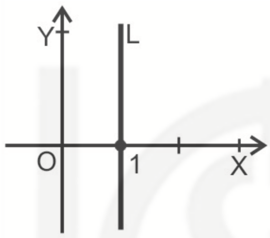


Fig. 33

E12) The given equation is linear. Hence it is represented by a line. Now, comparing the given equation with that in E8(ii), you can see that the intercepts of this line on the x and y-axes are π and $\sqrt{2}$, respectively. Thus, its graph is as shown in Fig. 32.
You can check the graph and show that it is bijective.

E13) The curve is $x=1$, if we convert to Cartesian coordinates. This is the line L in Fig. 33.

E14) i) Since $r^2=x^2+y^2$ and $y=r\sin\theta$, the equation becomes $x^2+y^2=3y$.

ii) The equation becomes

$$\sqrt{x^2+y^2}=a\left(1-\frac{x}{\sqrt{x^2+y^2}}\right), \text{ since } \cos\theta=\frac{x}{r}=\frac{x}{\sqrt{x^2+y^2}}$$

$$\Rightarrow\ x^2+y^2+a\,(x-\sqrt{x^2+y^2}\,)=0.$$

UNIT 4

COMPLEX NUMBERS

Structure **Page Nos.**

4.1 INTRODUCTION

In your studies so far you have dealt with natural numbers, integers, rational numbers and real numbers. You would also know that a shortcoming in $\mathbb{N}$ led mathematicians of several centuries ago to define negative numbers. Hence, the set $\mathbb{Z}$ was born. For similar reasons $\mathbb{Z}$ was extended to $\mathbb{Q}$ and $\mathbb{Q}$ to $\mathbb{R}$ at various stages in history. Then came a point when mathematicians found that $\mathbb{R}$ was not enough, for example, when looking for solutions of equations like $x^2+1=0$. Since $x^2+1=0$ has no solution in $\mathbb{R}$, for a long time it was accepted that this equation has no solution. The Indian mathematicians Mahavira (in 850 A.D.) and Bhaskara (in 1150 A.D.) clearly stated that the square root of a negative quantity does not exist. Then, in the 16th century the Italian mathematician Cardano tried to solve the quadratic equation $x^2-10x+40=0$. He found that $x_1=5+\sqrt{-15}$ and $x_2=5-\sqrt{-15}$ satisfied the equation. But then, what is $\sqrt{-15}$? He, and other mathematicians, tried to give expressions like this some meaning. Even while making mathematical models of real life solutions, the mathematicians of the 17th and 18th centuries were coming across more and more examples of equations which had no real roots. To overcome this shortcoming, the concept of a complex number slowly came into being. It was the famous mathematician Gauss (1777-1855) who used, and popularised, the name 'complex number' for numbers of the type $5+\sqrt{-15}$.

Fig. 1: Cardano (1501 - 1576) acknowledged the existence of imaginary numbers in his book Ars Magna, published in 1545

In the early 1800s, a geometric representation of complex numbers was developed. This representation finally made complex numbers acceptable to all mathematicians. Since then complex numbers have seeped into all branches of mathematics. In fact, they have even been found necessary for developing several areas in modern physics and engineering.

In this unit, you will have an opportunity to familiarise yourself with complex numbers in Sec. 4.2. There are several different ways of representing complex numbers, which you shall study in Sec. 4.3. Next, in Sec. 4.4, you will study the basic algebraic operations on complex numbers. Finally, in Sec. 4.5, we shall acquaint you with De Moivre's theorem, and you will see why this is considered so important.

We would like to reiterate that whichever mathematics course you study, you will need the understanding of complex numbers that we have tried to give you through this unit. So please go through it carefully, do every exercise as you come to it and ensure that you have achieved the following objectives.

Objectives

After studying this unit, you should be able to:

- define a complex number;
- describe the algebraic, geometric and polar representations of a complex number;
- add, subtract, multiply and divide two complex numbers;
- apply De Moivre's theorem to prove trigonometric identities;
- apply De Moivre's theorem for finding the nth roots of $z \in \mathbb{C}$, where $n \in \mathbb{N}$.

4.2 WHAT A COMPLEX NUMBER IS

When you consider the linear equation $2x+3=0$, you know that it has a solution, namely, $x = \frac{-3}{2}$. But, can you always find a real solution of the equation $ax+b=0$, where $a, b \in \mathbb{R}$ and $a \neq 0$? Is the required solution $x = \frac{-b}{a}$? It is, since $a\left(\frac{-b}{a}\right)+b=0$.

Now, what happens if we try to look for real solutions of any quadratic equation over $\mathbb{R}$? Consider one such equation, namely, $x^2-1=0$, that is, $x^2=1$. This equation has two real solutions, $x=1$ and $x=-1$. But, what about the equation $x^2+1=0$, that is, $x^2=-1$? Does this equation have a solution in $\mathbb{R}$? Since the square of any real number must be non-negative, there is no $x \in \mathbb{R}$ such that $x^2=-1$. But, as discussed in Sec. 4.1, several equations like $x^2=-1$ were coming up in studies undertaken by mathematicians. In fact, from about 250 A.D. onwards, mathematicians have been coming across quadratic equations, arising from real life situations, which did not have any real solutions. It was in the 16th century that the Italian mathematicians, Cardano and Bombelli, started a serious discussion on extending the number system to include square roots of negative numbers. In the next two hundred years, more and more instances were discovered in which the use of square roots of negative numbers helped in finding the solution of real-life problems.

In 1777, the Swiss mathematician Euler introduced the term "imaginary unit", which he denoted by i. He defined $i = \sqrt{-1}$. Soon after, the great mathematician Carl Friedrich Gauss introduced the term 'complex numbers' for numbers such as $3i\left(=3\sqrt{-1}\right)$, $1+i\left(=1+\sqrt{-1}\right)$ or $-2+i\sqrt{5}\left(=-2+\sqrt{-5}\right)$. Nowadays, not only are these numbers accepted, they are heavily used in every field of mathematics and its applications.
[For a brief history of complex numbers, you can also see the web link **www.math.uri.edu/~merino/spring06/mth562/ShortHistoryComplexNumbers2006.pdf**]

Fig.2: Gauss (1777-1855) is often referred to as 'the prince of mathematics'

So, what you have seen is that over the centuries the number systems were extended as the need was felt — from $\mathbb{N}$ to $\mathbb{N} \cup \{0\}$, to $\mathbb{Z}$, to $\mathbb{Q}$, to $\mathbb{R}$, and then to complex numbers.

Let us formally define a complex number now.

Definition: i) A **complex number** is an expression of the form $x + iy$, where x and y are real numbers, and $i = \sqrt{-1}$ is the imaginary unit.

ii) x is called the **real part**, and y is called the **imaginary part**, of the complex number $x + iy$. We write $x = \text{Re}(x + iy)$ and $y = \text{Im}(x + iy)$.

Caution: i) Remember that i is not a real number.

ii) $\text{Im}(x + iy)$ is the real number y, and not the number iy.

We denote the set of all complex numbers by $\mathbb{C}$.
So, $\mathbb{C} = \{x + iy \mid x, y \in \mathbb{R}\}$.
By convention, we will usually denote an element of $\mathbb{C}$ by z. So, whenever we will talk of a complex number z, we will mean $z = x + iy$ for some $x, y \in \mathbb{R}$. In fact, $\mathbf{z = Re\,z + i\,Im\,z}$.
This is usually called the **algebraic** (or **standard**, or **rectangular**) **representation of a complex number**.

There is another convention that we follow while writing complex numbers, which we give in the following remark.

Remark 1: When you go through Sec. 4.4.2, you will see that $iy = yi \; \forall \; y \in \mathbb{R}$. That is why we can write the complex number $x + iy$ as $x + yi$ also.
By convention, we write any complex number $x + iy$, for which $y \in \mathbb{Q}$, as $x + yi$. For example, we prefer to write $2 + 1i$, $2 + \frac{3}{2}i$ and $2 + \frac{5}{9}i$ instead of $2 + i1$, $2 + i\frac{3}{2}$ and $2 + i\frac{5}{9}$, respectively.
But, if $z \in \mathbb{C}$ is of the form $z = a + i\sqrt{b}$, $b \in \mathbb{R}$, then we prefer to write z in this form and not as $z = a + \sqrt{b}\;i$.
Further, **we write $\mathbf{1i}$ as $\mathbf{i}$ and $\mathbf{0i}$ as $\mathbf{0}$.**

Now that you know what a complex number is, would you agree that the following belong to $\mathbb{C}$?

$$0,\ 1,\ 5+\sqrt{-15},\ 3i,\ \sqrt{2},\ \sqrt{-2}$$

Each of them is a complex number because

$$0 = 0 + 0i$$
$$1 = 1 + 0i$$

$$5+\sqrt{-15}=5+i\sqrt{15}$$
$$3i=0+3i$$
$$\sqrt{2}=\sqrt{2}+0i$$
$$\sqrt{-2}=0+i\sqrt{2}$$

$\sqrt{-a}=i\sqrt{a}\ \forall a\geq 0$.

From these examples, you may have realised that some complex numbers can have their real part or their imaginary part equal to zero. We have names for such numbers.

Definition: Consider a complex number $z=x+iy$.
If $y=0$, we say z is **purely real**.
If $x=0$, and $y\neq 0$, we say z is **purely imaginary**.

We usually write the purely real number $x+0i$ as x only, and write the purely imaginary number $0+iy$ as iy only, as you may have noted in the examples above.

Try these exercises now.

E1) Complete the following table:

z	Re z	Im z
$\frac{1+\sqrt{-23}}{2}$		
i		
	0	0
$\frac{-1+\sqrt{3}}{5}$		

E2) Is $\mathbb{R}\subseteq\mathbb{C}$? Give reasons for your answer.

So, you have seen that, given $x+iy\in\mathbb{C}$, we associate with it the unique point $(x,\ y)\in\mathbb{R}^2$. The converse is also true. That is, given $(x,\ y)\in\mathbb{R}^2$, we can associate with it the unique complex number $x+iy$. This means that the following definition of a complex number is equivalent to our previous definition.

Definition: A complex number is an ordered pair of real numbers.
In the language of sets (see Unit 1), we can say that $\mathbb{C}=\mathbb{R}\times\mathbb{R}$.

With the help of this definition can you say when two complex numbers are equal?

Definition: Two complex numbers $(x_1,\ y_1)$ and $(x_2,\ y_2)$ are **equal** iff $x_1=x_2$ and $y_1=y_2$.
In other words, $x_1+iy_1=x_2+iy_2$ iff $x_1=x_2$ and $y_1=y_2$.
Thus**, two elements of $\mathbb{C}$ are equal if and only if their real parts are equal and their imaginary parts are equal**.
So, for example, $\frac{-1+\sqrt{-3}}{2}=\frac{-1}{2}+i\frac{\sqrt{3}}{2}$, but $\frac{-1+\sqrt{-3}}{2}\neq\frac{-1}{2}+i\frac{1}{2}$.
Here's an exercise about this, now.

E3) For which real values of k and m is

i) $\sqrt{k}+3i=\sqrt{3}+im$?

ii) $ki-m\in\mathbb{R}$?

Now, given any complex number, we can define a related complex number in a very natural way, as follows.

Definition: Let $z=x+iy\in\mathbb{C}$. The **complex conjugate of z** (or simply, **the conjugate of z**) is the complex number $\mathbf{\bar{z}=x-iy}$.
Thus, $\operatorname{Re}\bar{z}=\operatorname{Re}z$ and $\operatorname{Im}\bar{z}=-\operatorname{Im}z$.

For example, if $z=15+i$ then $\bar{z}=15-i$.
Again, if $z=0$, then $\operatorname{Re}z=\operatorname{Im}z=0$, so that $\operatorname{Re}\bar{z}=\operatorname{Im}\bar{z}=0$. Hence, $\bar{z}=0$ also.

In Section 4.4.2 you will see one important use of the complex conjugate. But, for now, here are some exercises for practice!

E4) Obtain the conjugates of $-5,\ \sqrt{-5},\ 2+3i,\ 2-3i$.

E5) For which $z\in\mathbb{C}$, will $z=\bar{z}$?

E6) For any $z\in\mathbb{C}$, show that $\bar{\bar{z}}=z$.

The conjugate of the conjugate of z is $z\,\forall\,z\in\mathbb{C}$.

In this section we have defined a complex number, giving the algebraic, or standard, method of representing complex numbers. Now let us consider other ways of representing such numbers.

4.3 DIFFERENT REPRESENTATIONS OF A COMPLEX NUMBER

You know that we can geometrically represent real numbers on the number line. In fact, there is a one-one correspondence between real numbers and points on the number line. You have also seen that $\mathbb{C}=\mathbb{R}\times\mathbb{R}$. So, using your understanding developed by studying Sec. 3.2, can you think of a way of representing complex numbers geometrically? Let's see.

4.3.1 Geometric Representation

Since a complex number z is given by a pair of real numbers, $\operatorname{Re}z$ and $\operatorname{Im}z$, your study of Unit 3 may have given you the idea that led mathematicians in the 18th and 19th century to think of representing complex numbers as points in a plane. This geometric representation was given in the early 1800s. It is called an **Argand diagram**, after the Swiss mathematician J. R. Argand, who propagated this idea. Interestingly, this idea was first described a few years earlier by the mathematician Wessel.

Let us see what an Argand diagram is. In a Cartesian coordinate system, take the axes OX and OY in the XOY plane. From Unit 3, you know that any point in the plane is determined by its Cartesian coordinates. Now we consider any complex number $x+iy$. We represent it by the point in the plane with Cartesian coordinates (x, y). Thus, the real parts of the complex numbers are plotted along the horizontal axis, and the imaginary parts are plotted along the vertical axis. This representation of complex numbers is called an Argand diagram.

The x-axis is called the **real axis** and the y-axis is called the **imaginary axis** in an Argand diagram.

For example, in Fig. 3, P represents the complex number $2+3i$, whose real part is 2 and imaginary part is 3. And what number does P' in Fig. 3 represent? P' corresponds to $2-3i$, the conjugate of P.

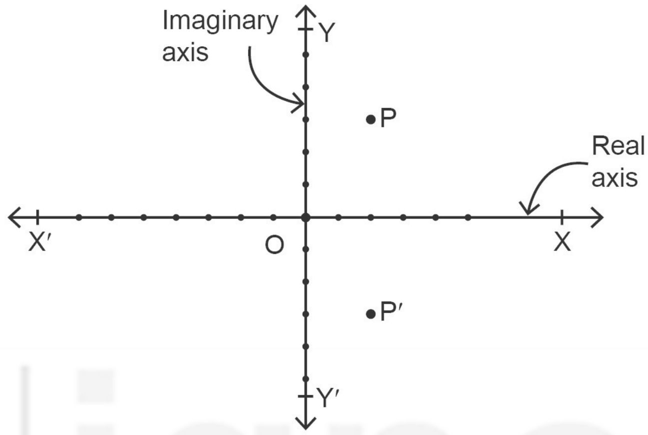


Fig. 3: An Argand diagram

From Fig. 3, you may have observed that **in an Argand diagram the point that represents $\bar{z}$ is the reflection in the real axis of the point that represents z, for any $z \in \mathbb{C}$.**

The **reflection** of a point (a, b) in the x-axis is the point $(a, -b)$, and in the y-axis is $(-a, b)$.

Now, you know that any real number is a complex number (called a purely real number). Where would the purely real numbers lie in an Argand diagram? Wouldn't they lie along the real axis? Similarly, the purely imaginary numbers lie along the imaginary axis.

Try these exercises now.

E7) a) Represent the following elements of $\mathbb{C}$ in an Argand diagram: $3, -1+i, \overline{-1+i}, i$.

b) Represent the sets $S_1 = \{2+iy \mid y \in \mathbb{R}\}$, $S_2 = \{x+3i \mid x \in \mathbb{R}\}$ and $S_3 = \{x+ix \mid x \in \mathbb{R}\}$ in an Argand diagram.

E8) Give the algebraic representation of the elements of $\mathbb{C}$ represented geometrically by the points $\left(\frac{-1}{2}, \frac{1}{3}\right)$, $(2, 0)$ and $(0, -2)$ in an Argand diagram.

Let us now consider another way of representing a complex number.

4.3.2 Polar Representation

Consider any non-zero complex number $z = x + iy$. We represent it by P in the Argand diagram in Fig. 4. We can represent this point uniquely by its Cartesian coordinates, as you know. But, you also know from Unit 3 that we can represent it by its polar coordinates (r, θ). Recall that the distance OP is r. We call it the **modulus of z**, and denote it by $|z|$. Further, θ is called **an argument** of z.

Now, let us merge the pole and polar axis with the real and imaginary axes in an Argand diagram, as in Fig. 5.

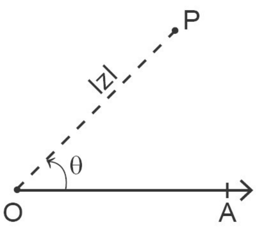


Fig. 4: The polar representation of a complex number with polar coordinates $|z|$ and θ

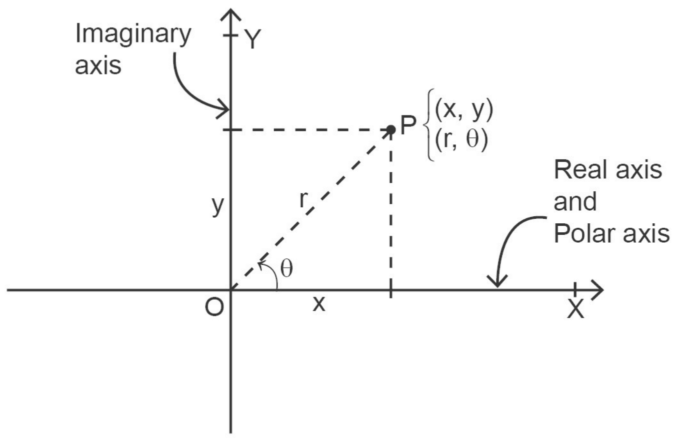


Fig. 5: The relationship between the Cartesian and the polar coordinates

Then, $|z| = \sqrt{x^2 + y^2}$.

Also, if we write $|z| = r$, then from Fig. 5 you can see that $\sin\theta = \frac{y}{r}$ and $\cos\theta = \frac{x}{r}$.

$\therefore \quad x = r\cos\theta,\ y = r\sin\theta$ …(1)

Here are some important remarks on the modulus and argument of z.

Remark 2: From Unit 3, you know that (r, θ) are unique only if we restrict θ to $0 \le \theta < 2\pi$ or to $-\pi < \theta \le \pi$. In the case of complex numbers, $|z| \ge 0\ \forall\ z \in \mathbb{C}$. Also, by convention for uniqueness the restriction on θ is $-\pi < \theta \le \pi$. We call this unique value of θ **the principal argument** of z, and denote it by $\mathbf{Arg}\ \mathbf{z}$.

Thus, for a complex number, the polar coordinates (r, θ) satisfy $r \ge 0$ and $-\pi < \theta \le \pi$.

Remark 3: i) $z \in \mathbb{C}$, but $|z| \in \mathbb{R}$.

ii) If z is real, what is $|z|$? It is just the absolute value of z (see Example 9, Unit 3).

iii) For $z = 0, |z| = 0$, and its argument is not defined.

From (1) and Remark 2, we are led to a definition.

Definition: Given any complex number, we can write it as **$z = r(\cos\theta + i\sin\theta)$, where $r = |z|$ and $\theta = \mathrm{Arg}\, z$.**
This is called the **polar form** of z.

Note that, given $z = x + iy$, we can use (1) to obtain $\mathrm{Arg}\, z = \tan^{-1}\left(\frac{y}{x}\right)$.

However, as more than one angle between $-\pi$ and π have the same tan value, for instance, $\tan^{-1}(1) = \pi/4$ and $\frac{-3\pi}{4}$, we must draw an Argand diagram to find the quadrant in which the point (x, y) lies and hence obtain the right value of $\operatorname{Arg} z$.

Let us look at an example.

Example 1: a) Obtain the modulus and principal argument of $\sqrt{3}+i, -\left(\sqrt{3}+i\right)$ and $\overline{\sqrt{3}+i}$. Hence obtain their polar forms.

b) If $|z| = 2$ and $\operatorname{Arg} z = \frac{\pi}{3}$, obtain the algebraic representation of z.

c) If $|z| = \frac{\pi}{2}$ and $\operatorname{Arg} z = \frac{\pi}{2}$, obtain the algebraic representation of z.

Solution: Let $z_1 = \sqrt{3}+i,\ z_2 = -\left(\sqrt{3}+i\right)$ and $z_3 = \overline{\sqrt{3}+i}$.

a) Now $\operatorname{Re} z_1 = \sqrt{3},\ \operatorname{Im} z_1 = 1$. Thus, $\sqrt{3}+i$ corresponds to $\left(\sqrt{3}, 1\right)$, which lies in the first quadrant. We find that

$|z_1| = \sqrt{(\operatorname{Re} z_1)^2 + (\operatorname{Im} z_1)^2} = 2$, and

$$\operatorname{Arg} z_1 = \tan^{-1}\left(\frac{\operatorname{Im} z_1}{\operatorname{Re} z_1}\right) = \tan^{-1}\left(\frac{1}{\sqrt{3}}\right) = \frac{\pi}{6} \text{ or } \frac{-5\pi}{6}.$$

Since z_1 lies in the first quadrant, $\operatorname{Arg} z_1$ must be between 0 and $\frac{\pi}{2}$.

Thus, $\operatorname{Arg} z_1 = \frac{\pi}{6}$. Hence, the polar form of $\sqrt{3}+i$ is $2\left(\cos\frac{\pi}{6} + i\sin\frac{\pi}{6}\right)$.

Now, let us consider $z_2 = -\left(\sqrt{3}+i\right) = -\sqrt{3} + (-1)i$. Here,

$|z_2| = 2$ and $\operatorname{Arg} z_2 = \tan^{-1}\left(\frac{1}{\sqrt{3}}\right)$, just as for z_1.

But $-\left(\sqrt{3}+i\right)$ lies in the 3rd quadrant. Hence, $\operatorname{Arg} z_2 = \frac{-5\pi}{6}$.

Thus the polar form of z_2 is $2\left(\cos\frac{5\pi}{6} - i\sin\frac{5\pi}{6}\right)$.

Finally, let us look at $z_3 = \overline{\sqrt{3}+i} = \sqrt{3} - i$.

You can check that $|z_3| = 2$ and

$\operatorname{Arg} z_3 = \tan^{-1}\left(-\frac{1}{\sqrt{3}}\right) = -\tan^{-1}\left(\frac{1}{\sqrt{3}}\right) = \frac{-\pi}{6}$, since z_3 lies in the 4th quadrant.

Thus, the polar form of z_3 is $2\left(\cos\frac{\pi}{6} - i\sin\frac{\pi}{6}\right)$.

b) We know that $z = |z|(\cos(\operatorname{Arg} z) + i\sin(\operatorname{Arg} z))$

$$= 2\left(\cos\frac{\pi}{3} + i\sin\frac{\pi}{3}\right) = 2\left(\frac{1}{2} + i\frac{\sqrt{3}}{2}\right)$$

$$= 1 + i\sqrt{3}.$$

c) Here $z = \frac{\pi}{2}\left[\cos\frac{\pi}{2} + i\sin\frac{\pi}{2}\right] = \frac{\pi}{2}i$

***

Try the following exercises now.

E9) Write down the polar forms of the complex numbers listed in E7 (a).

E10) Find the relationship between the moduli and principal arguments of z and $\bar{z}$, for $z \in \mathbb{C}$.

'Moduli' is the plural of 'modulus'.

E11) Show that $\{z \in \mathbb{C} \mid |z| = 1\}$ is the set of points on the circle $x^2 + y^2 = 1$ in $\mathbb{R}^2$.

You will find the polar form of a complex number very useful when you study the multiplication and division of complex numbers in the next section.

4.4 ALGEBRAIC OPERATIONS

You are familiar with the operations of addition, subtraction, multiplication and division in $\mathbb{R}$. In this section we will discuss these operations in $\mathbb{C}$. Let us first consider the first two operations.

4.4.1 Addition and Subtraction

Take any two complex numbers, say $z_1 = 3 + 2i$ and $z_2 = 4 + \frac{1}{2}i$. What do you expect $z_1 + z_2$ to be? Wouldn't you just add the real parts of both and the imaginary parts of both to get this? If so, you would be right, i.e.,

$$z_1 + z_2 = (3+4) + \left(2 + \frac{1}{2}\right)i = 7 + \frac{5}{2}i.$$

Let us define this process formally for any two complex numbers.

Definition: The **sum of two complex numbers** $z_1 (= x_1 + iy_1)$ and $z_2 (= x_2 + iy_2)$ is the complex number $z_1 + z_2 = (x_1 + x_2) + i(y_1 + y_2)$.
In terms of ordered pairs,
$(x_1, y_1) + (x_2, y_2) = (x_1 + x_2, y_1 + y_2)$.

Let us look at an example.

Example 2: Find the sum of
i) $3 + i$ and $-2 + 4i$,
ii) -5 and $5 - i$.

Solution: i) $(3 + i) + (-2 + 4i) = (3 + (-2)) + (1 + 4)i = 1 + 5i$.

ii) $(-5) + (5 - i) = (-5 + 0i) + (5 - i) = (-5 + 5) + (0 - 1)i = -i$.

***

Example 3: Show that any complex number is the sum of a purely real number and a purely imaginary number.

Solution: Take $z = x + iy = (x + 0i) + (0 + iy)$.

Here $x+0i$ is purely real and $0+iy$ is purely imaginary, hence the result.

***

In the following exercises we ask you to verify some very important properties of addition in $\mathbb{C}$.

$z+\bar{z}\in\mathbb{R}\,\forall\, z\in\mathbb{C}$

E12) i) Find the sum of $2+3i$ and $\overline{2+3i}$.

ii) Show that $z+\bar{z}=2\operatorname{Re} z$ for any $z\in\mathbb{C}$.

E13) Show that $\overline{z_1+z_2}=\overline{z_1}+\overline{z_2}\ \forall\ z_1, z_2\in\mathbb{C}$.

E14(a) tells us that addition in $\mathbb{C}$ is commutative.
E14(b) says that addition in $\mathbb{C}$ is associative.

E14) a) Show that $z_1+z_2=z_2+z_1$ for any $z_1, z_2\in\mathbb{C}$.

b) Show that $(z_1+z_2)+z_3=z_1+(z_2+z_3)$ for any $z_1, z_2, z_3\in\mathbb{C}$.

E15) Find an element $a+ib\in\mathbb{C}$ such that $z+(a+ib)=z\ \forall\ z\in\mathbb{C}$.

If you have solved these exercises, you must have realised that the addition in $\mathbb{C}$ satisfies most of the properties that addition in $\mathbb{R}$ satisfies. Also, because of what you proved in E15, we say that $0+i0(=0)$ is the **additive identity** in $\mathbb{C}$.

Now, can you define subtraction in $\mathbb{C}$? For this, let us first define $-z$, for $z\in\mathbb{C}$. You may already have come up with the following definitions, which are very natural.

Definitions: i) Given $z=x+iy\in\mathbb{C}$, $-z$ is the complex number $(-x)+i(-y)$.

ii) The **difference** z_1-z_2 **of two complex numbers** $z_1(=x_1+iy_1)$ and $z_2(=x_2+iy_2)$ is defined by
$z_1-z_2=z_1+(-z_2)=(x_1-x_2)+i(y_1-y_2)$.

So, what do you think $z-z$ is, for any $z\in\mathbb{C}$? Let's see. Take $z=x+iy$.
Then $z-z=(x-x)+i(y-y)=0$, the additive identity in $\mathbb{C}$.
This tell us that for any $z\in\mathbb{C}$, $(-z)$ is the **additive inverse** of z.

Try the following exercises now.

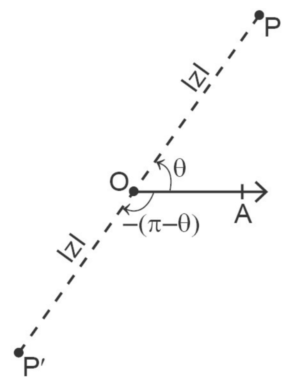


Fig. 6: z and $-z$ are represented by P and P', respectively.

E16) i) Find $(3-2i)-\overline{(3-2i)}$.

ii) Find $z-\bar{z}$, for any $z\in\mathbb{C}$.

E17) Find the relationship between

i) $|z|$ and $|-z|$,

ii) $\operatorname{Arg} z$ and $\operatorname{Arg}(-z)$, for any $z\in\mathbb{C}$ (see Fig. 6).

We will now make a brief remark on the graphical representation of the sum of complex numbers.

Remark 5: The addition of two complex numbers has an interesting geometrical representation. Consider an Argand diagram (Fig. 7) in which we represent two complex numbers (x_1, y_1) and (x_2, y_2) by the points P and Q. If we complete the parallelogram with adjacent sides OP and OQ, the fourth vertex R represents the sum $(x_1, y_1)+(x_2, y_2)$. In vector algebra you will come across a similar parallelogram law of addition.

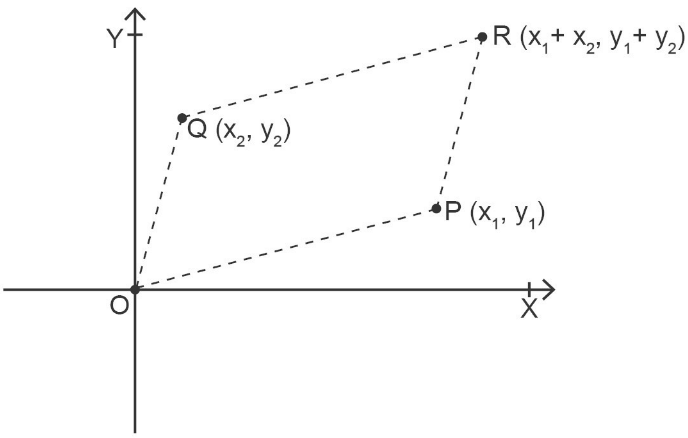


Fig. 7: The geometric representation of addition in $\mathbb{C}$.

So far you have seen how naturally we have defined addition (and subtraction) in $\mathbb{C}$ by using addition (and subtraction) in $\mathbb{R}$. Let us see if we can do the same for multiplication and division. You may find the polar form more useful than the standard form for these operations.

4.4.2 Multiplication and Division

Let us begin by considering two complex numbers in their polar forms, say $z_1 = 2 = 2(\cos 0 + i\sin 0)$ and $z_2 = -3 = 3(\cos\pi + i\sin\pi)$.
Notice that z_1 and z_2 are actually real numbers, and we know that

$$z_1z_2 = -6 = 6(\cos\pi + i\sin\pi) = (2)\ (3)\{\cos(0+\pi) + i\sin(0+\pi)\}.$$
$$= |z_1||z_2|\{\cos\ (\text{Arg } z_1 + \text{Arg } z_2) + i\sin\ (\text{Arg } z_1 + \text{Arg } z_2)\}$$

This may help you see why we define multiplication in $\mathbb{C}$ as below.

Definition: The **product** of two complex numbers $z_1 = r_1(\cos\theta_1 + i\sin\theta_1)$ and $z_2 = r_2(\cos\theta_2 + i\sin\theta_2)$ is defined to be the complex number

$$\mathbf{z_1z_2 = r_1r_2[\cos(\theta_1+\theta_2+2k\pi) + i\sin(\theta_1+\theta_2+2k\pi)]}$$
$$\mathbf{= |z_1||z_2|\{\cos(\text{Arg } z_1 + \text{Arg } z_2 + 2k\pi)}$$
$$\mathbf{+ i\sin(\text{Arg } z_1 + \text{Arg } z_2 + 2k\pi)\},}$$

where $k \in \mathbb{Z}$ is such that $-\pi < \text{Arg } z_1 + \text{Arg } z_2 + 2k\pi \le \pi$.

So, for example, if $z_1 = \sqrt{2}\left(\cos\frac{\pi}{6} + i\sin\frac{\pi}{6}\right)$ and $z_2 = \sqrt{7}(\cos\pi + i\sin\pi)$, then

$$z_1z_2 = \sqrt{2}\left(\cos\frac{\pi}{6} + i\sin\frac{\pi}{6}\right)\sqrt{7}(\cos\pi + i\sin\pi)$$
$$= \sqrt{2}\sqrt{7}\left[\cos\left(\frac{\pi}{6}+\pi\right) + i\sin\left(\frac{\pi}{6}+\pi\right)\right]$$
$$= \sqrt{2}\sqrt{7}\left(\cos\frac{7\pi}{6} + i\sin\frac{7\pi}{6}\right)$$

$$= \sqrt{2}\sqrt{7}\left[\cos\left(-\frac{5\pi}{6}\right) + i\sin\left(-\frac{5\pi}{6}\right)\right], \text{ since } -\pi < \text{Arg } z \le \pi.$$

Thus, $z_1 z_2 = |z_1||z_2|\{\cos(\text{Arg}\, z_1 + \text{Arg}\, z_2 - 2\pi) + i\sin(\text{Arg}\, z_1 + \text{Arg}\, z_2 - 2\pi)\}$.

In Fig. 8, you can see a visual representation of multiplication of two complex numbers. As you can see, this involves scaling and rotation of vectors.

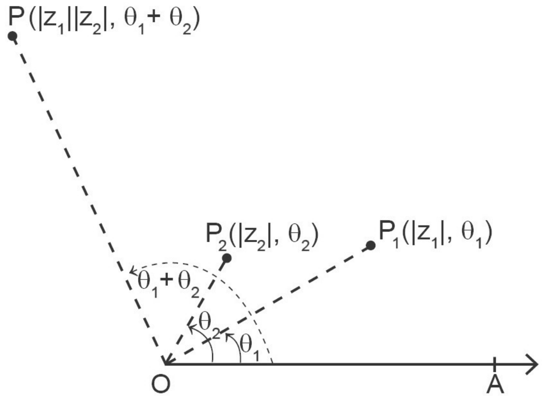


Fig. 8: P represents the product, in polar form, of the complex numbers represented by P_1 and P_2.

Let us consider another example.

Example 4: Obtain the product of $z_1 = 2(\cos 1 + i\sin 1)$ and $z_2 = \cos 3 + i\sin 3$ in polar form.

Solution: Here $|z_1| = 2$, $\text{Arg}\, z_1 = 1$, $|z_2| = 1$, $\text{Arg}\, z_2 = 3$.

Therefore, $z_1 z_2 = 2\{\cos(1+3) + i\sin(1+3)\}$
$= 2(\cos 4 + i\sin 4)$.

Note that $\text{Arg}(z_1 z_2) \ne 4$, since $4 > \pi$. We need to choose an integer k such that $-\pi < 4 + 2k\pi \le \pi$. $k = -1$ serves the purpose. Thus, $\text{Arg}(z_1 z_2) = 4 - 2\pi$.

Hence $z_1 z_2 = 2\{\cos(4 - 2\pi) + i\sin(4 - 2\pi)\}$.

***

Try some exercises now, which will help you see some interesting properties of multiplication.

E18) Find $z.1, z.0, z.i, z.\bar{z} \;\forall\; z \in \mathbb{C}$.

E19) a) Show that $z_1 z_2 = z_2 z_1 \;\forall\; z_1, z_2 \in \mathbb{C}$.

b) Show that $(z_1 z_2) z_3 = z_1 (z_2 z_3) \;\forall\; z_1, z_2, z_3 \in \mathbb{C}$.

If you have done E19(a) and (b), then you have shown that multiplication in $\mathbb{C}$ is commutative and associative, respectively.

Now, let us consider division. If I want to find $\frac{z_1}{z_2} = z_1.z_2^{-1}$, all I need to do is look for a nice way to find z_2^{-1} in polar form, and then multiply this with z_1 to

get the result. So, for obtaining z^{-1}, $z \neq 0$, consider an example. Take
$z = 3\left(\cos\frac{\pi}{4} + i\sin\frac{\pi}{4}\right)$. Then

$$z^{-1} = \frac{1}{3\left(\cos\frac{\pi}{4} + i\sin\frac{\pi}{4}\right)} = \frac{1}{3}\frac{1}{\left(\cos\frac{\pi}{4} + i\sin\frac{\pi}{4}\right)}\left(\frac{\cos\frac{\pi}{4} - i\sin\frac{\pi}{4}}{\cos\frac{\pi}{4} - i\sin\frac{\pi}{4}}\right)$$

$$= \frac{1}{3}\frac{\left(\cos\frac{\pi}{4} - i\sin\frac{\pi}{4}\right)}{\left(\cos^2\frac{\pi}{4} + \sin^2\frac{\pi}{4}\right)}$$, using $(a+ib)\,(a-ib) = a^2 - i^2b^2 = a^2 + b^2$.

$$= \frac{1}{3}\left(\cos\frac{\pi}{4} - i\sin\frac{\pi}{4}\right)$$

So $|z^{-1}| = \frac{1}{|z|}$ and $\text{Arg}(z^{-1}) = \text{Arg}(-z)$.

More generally, using the same steps we see that if $z = r(\cos\theta + i\sin\theta)$, and $z \neq 0$, then

$$z^{-1} = \frac{1}{r}(\cos\theta - i\sin\theta). \qquad \text{...(2)}$$

Now, if $z_1 = r_1(\cos\theta_1 + i\sin\theta_1)$ and $z_2 = r_2(\cos\theta_2 + i\sin\theta_2)$,

then $\frac{z_1}{z_2} = \frac{r_1(\cos\theta_1 + i\sin\theta_1)}{r_2(\cos\theta_2 + i\sin\theta_2)}$

$= \frac{r_1}{r_2}(\cos\theta_1 + i\sin\theta_1)(\cos\theta_2 - i\sin\theta_2)$, using (2)

$= \frac{r_1}{r_2}\{\cos(\theta_1 - \theta_2) + i\sin(\theta_1 - \theta_2)\}$

So, $$\frac{z_1}{z_2} = \frac{r_1}{r_2}\{\cos(\theta_1 - \theta_2 + 2k\pi) + i\sin(\theta_1 - \theta_2 + 2k\pi)\}, \qquad \text{...(3)}$$

where $k \in \mathbb{Z}$ is such that $-\pi < \theta_1 - \theta_2 + 2k\pi \leq \pi$.

Let us consider an example.

Example 5: Find $\frac{z_1}{z_2}$ in polar form, where $z_1 = 6(\cos\pi + i\sin\pi)$ and
$z_2 = \sqrt{2}\left(\cos\frac{\pi}{4} - i\sin\frac{\pi}{4}\right)$.

Solution: Here $\left|\frac{z_1}{z_2}\right| = \frac{|z_1|}{|z_2|} = \frac{6}{\sqrt{2}} = 3\sqrt{2}$, and

$\text{Arg}\left(\frac{z_1}{z_2}\right) = \text{Arg}\, z_1 - \text{Arg} z_2 + 2k\pi$, such that $-\pi < \text{Arg}\left(\frac{z_1}{z_2}\right) \leq \pi$.

Thus, $\text{Arg}\left(\frac{z_1}{z_2}\right) = \pi - \left(\frac{-\pi}{4}\right) - 2\pi = -\frac{3\pi}{4}$.

So $\frac{z_1}{z_2} = 3\sqrt{2}\left(\cos\frac{3\pi}{4} - i\sin\frac{3\pi}{4}\right)$.

***

Try the following exercise now.

E20) Find the polar forms of z_1 and z_2, where $z_1 = -6$ and $z_2 = 1+i$. Hence obtain the polar forms of $z_1 z_2$ and $\frac{z_1}{z_2}$.

Let us now consider these operations on complex numbers using the standard form. The route is slightly circuitous.

Let $z_1 = x_1 + iy_1$ and $z_2 = x_2 + iy_2$. Then, from Unit 3 you know that if the polar coordinates of z_1 and z_2 are (r_1, θ_1) and (r_2, θ_2), respectively, then $x_1 = r_1\cos\theta_1,\ y_1 = r_1\sin\theta_1,\ x_2 = r_2\cos\theta_2,\ y_2 = r_2\sin\theta_2$.

You also know that

$\cos(A+B)$
$= \cos A\cos B - \sin A\sin B,$
$\sin(A+B)$
$= \sin A\cos B + \cos A\sin B.$

$$\begin{aligned} z_1 z_2 &= r_1 r_2\{\cos(\theta_1+\theta_2) + i\sin(\theta_1+\theta_2)\} \\ &= r_1 r_2(\cos\theta_1\cos\theta_2 - \sin\theta_1\sin\theta_2) + ir_1 r_2(\sin\theta_1\cos\theta_2 + \cos\theta_1\sin\theta_2) \\ &= (x_1x_2 - y_1y_2) + i(x_1y_2 + x_2y_1). \end{aligned}$$

Accordingly, we have the following definition.

Definition: The **standard form of the product** $z_1 z_2$, of two complex numbers $z_1 = x_1 + iy_1$ and $z_2 = x_2 + iy_2$, is given by

$\mathbf{z_1 z_2 = (x_1x_2 - y_1y_2) + i\ (x_1y_2 + x_2y_1)}$.

Or, in the language of ordered pairs,

$(x_1, y_1).(x_2, y_2) = (x_1x_2 - y_1y_2,\ x_1y_2 + x_2y_1)$.

For example, $(1, 2)\ (-3, 2) = [1.(-3) - 2.2, 1.2 + (-3)2] = (-7, -4)$.

Let us check and see what i^2 is according to this definition.

$i^2 = i.i = (0+i)\ (0+i) = (0-1) + i(0-0) = -1$, which is as it should be!

While solving E18 and E19 you have noted some properties of multiplication in $\mathbb{C}$. You will discover some more properties if you try the following exercises.

E21) Obtain $(x, y)\ .\ (1, 0), (x, y).(0, 1)\ (x, y).(0, 0), (x, 0).(y, 0)$ and $(x, y).(1, 1)\ \forall\ (x, y) \in \mathbb{C}$.

E22) Show that $(x, y).\left(\frac{x}{x^2+y^2}, \frac{-y}{x^2+y^2}\right) = (1, 0)\ \forall\ (x, y) \in \mathbb{C}\setminus\{0\}$.

(Note that $x^2 + y^2 \neq 0$, since $(x, y) \neq (0, 0)$.)

If you've solved these exercises, you must have realised that

i) $\mathbf{z.1 = z}\ \forall\, \mathbf{z} \in \mathbb{C}$, that is, 1 is the **multiplicative identity** of $\mathbb{C}$;

ii) $\mathbf{z.0 = 0}\ \forall\, \mathbf{z} \in \mathbb{C}$;

iii) $i(x+iy) = -y + ix\ \ \forall x, y \in \mathbb{R}$;

iv) if $z_1, z_2 \in \mathbb{R}$, then our definition of multiplication coincides with the usual one for $\mathbb{R}$;

v) for any non-zero element $z \in \mathbb{C}\ \exists z' \in \mathbb{C}$, such that $zz' = 1$. In this case we say that z' is the **multiplicative inverse** of z. So $z' = \frac{1}{z}$.

Now let us see how to obtain the standard form of the quotient of a complex number by a non-zero complex number. We will use a process similar to the one you used for obtaining $\frac{z_1}{z_2}$ earlier (before Example 5). Consider an example.

Example 6: Obtain $\frac{2+3i}{1-i}$ in standard form.

Solution: Let us multiply and divide $\frac{2+3i}{1-i}$ by $\overline{1-i} = 1+i$. We get

$$\left(\frac{2+3i}{1-i}\right)\left(\frac{1+i}{1+i}\right) = \frac{(2+3i)\,(1+i)}{(1-i)\,(1+i)} = \frac{-1+5i}{1+1} = \frac{-1}{2} + \frac{5}{2}i.$$

So, $\frac{2+3i}{1-i} = \frac{-1}{2} + \frac{5}{2}i$.

$z\bar{z} = |z|^2 \in \mathbb{R}\ \forall z \in \mathbb{C}.$

***

If you've understood the way we have solved the example, you will have no problem in doing the following exercises.

E23) Obtain $\frac{-2+i}{\sqrt{-3}+i\sqrt{-4}}$ in standard form, and hence in polar form.

E24) For $a, b, c, d \in \mathbb{R}$ and $c^2 + d^2 \neq 0$, write $\frac{a+ib}{c+id}$ in standard form.

E25) Show that $\frac{1}{z} = \frac{1}{|z|^2}\bar{z}\ \forall\ z \in \mathbb{C}\setminus\{0\}$. Hence show that

$$\left|\frac{1}{z}\right| = \frac{1}{|z|}\ \forall\ z \in \mathbb{C}\setminus\{0\}.$$

E26) Represent $z_1, z_2, \bar{z}_2, \frac{1}{z_2}$ and $\frac{z_1}{z_2}$ in an Argand diagram, where

$z_1 = -1-2i,\ z_2 = 2-3i$.

We will use multiplication and division in the polar form a great deal in the next section. Before going to it, let us give you a rule that relates '$+$' and '$\times$' in $\mathbb{C}$. Do you know of such a law in $\mathbb{R}$? You must have used the distributive law often enough. It says that $a(b+c) = ab + ac\ \forall a, b, c \in \mathbb{R}$. The same law holds for $\mathbb{C}$. Why don't you try and show this (see E27)?

Multiplication distributes over addition in $\mathbb{C}$.

E27) i) Check that

$$(1+i)\left\{\left(\sqrt{2}-3i\right) + (5+i)\right\} = (1+i)\left(\sqrt{2}-3i\right) + (1+i)\,(5+i).$$

ii) Show that $z_1(z_2 + z_3) = z_1z_2 + z_1z_3 \; \forall \, z_1, z_2, z_3 \in \mathbb{C}$.

Now let us discuss a theorem which is very useful for complex numbers.

4.5 APPLICATIONS OF DE MOIVRE'S THEOREM

In the previous section, you studied that if
$z_1 = r_1(\cos\theta_1 + i\sin\theta_1)$ and $z_2 = r_2(\cos\theta_2 + i\sin\theta_2)$, then
$z_1z_2 = r_1r_2\{\cos(\theta_1 + \theta_2) + i\sin(\theta_1 + \theta_2)\}$.
In particular, if $z_1 = z_2$, then $r_1 = r_2, \theta_1 = \theta_2$, and hence we find that
$z_1^2 = r_1^2(\cos 2\theta_1 + i\sin 2\theta_1)$.

Fig. 9: De Moivre (1667-1754)

In fact, this is a particular case of a very nice formula, that uses the theorem below. This is **De Moivre's theorem**, named after the French mathematician Abraham De Moivre. It may amuse you to know that De Moivre never explicitly stated this result. But he seems to have known it and used it in his writings of 1730. It was the mathematician Euler who explicitly stated and proved this result in 1748. We shall not be proving the result here, but shall state it and discuss some of its consequences.

Theorem 1 (De Moivre's theorem): $(\cos\theta + i\sin\theta)^n = \cos n\theta + i\sin n\theta$, for any $n \in \mathbb{Z}$ and any angle θ.

This statement is so simple, and so beautiful. For instance, an immediate implication of this theorem is that if
$z = r(\cos\theta + i\sin\theta) \in \mathbb{C}$, then $\forall \; n \in \mathbb{Z}$,

$$z^n = r^n(\cos\theta + i\sin\theta)^n$$
$$= r^n(\cos n\theta + i\sin n\theta), \text{ using De Moivre's theorem.}$$

So, for example, $z^7 = r^7(\cos 7\theta + i\sin 7\theta)$ and
$z^{-100} = r^{-100}\{\cos(100\theta) - i\sin(100\theta)\}$.

What we have shown is that

$$\boxed{[r(\cos\theta + i\sin)]^n = r^n(\cos n\theta + i\sin n\theta) \; \forall r \geq 0, \, \theta \in \mathbb{R}, \, n \in \mathbb{Z}.}$$

This equality is true not just for a particular value of θ, or of r, or of n. It is true for all values of these variables as shown in the box. Such an equality is called an **identity**. Thus, an identity is an equality that is true for all applicable values of the variables involved. We will consider some identities now, which are proved by using De Moivre's theorem.

4.5.1 Trigonometric Identities

One of the most useful applications of Theorem 1 is in proving identities that involve trigonometric ratios like $\sin\theta, \cos\theta$, etc. Let us look at an example.

Example 7: Find a formula for $\cos 4\theta$ in terms of $\cos\theta$ and $\sin\theta$, for any $\theta \in \mathbb{R}$.

Solution: By De Moivre's theorem,

$$(\cos\theta + i\sin\theta)^4 = \cos 4\theta + i\sin 4\theta, \; \forall \; \theta \in \mathbb{R}. \quad \ldots(4)$$

We can also expand the left hand side of (4) by using the binomial expansion. Then

$$(\cos\theta + i\sin\theta)^4 = (\cos\theta)^4 + {}^4C_1(\cos\theta)^3(i\sin\theta) + {}^4C_2(\cos\theta)^2(i\sin\theta)^2$$
$$+ {}^4C_3\cos\theta(i\sin\theta)^3 + (i\sin\theta)^4$$
$$= \cos^4\theta + 4i\sin\theta\cos^3\theta - 6\sin^2\theta\cos^2\theta - 4i\sin^3\theta\cos\theta + \sin^4\theta \qquad \ldots(5)$$

${}^nC_m = \dfrac{n!}{m!(n-m)!}$
$\forall\, n,\ m \in \mathbb{N},\ n \geq m$.

Now, you know that two complex numbers are equal iff their real parts and their imaginary parts are equal.
Thus, comparing the real parts in (4) and (5), we get
$\cos 4\theta = \cos^4\theta - 6\sin^2\theta\cos^2\theta + \sin^4\theta$.

***

You can try the following exercise on similar lines.

E28) Find formulae for $\cos 3\theta$ in terms of $\cos\theta$, and for $\sin 3\theta$ in terms of $\sin\theta$.

Now, for any $m \in \mathbb{N}$ let us look at z^m, where $z \in \mathbb{C}$ such that $|z| = 1$. Then, by De Moivre's theorem
$z^m = \cos m\theta + i\sin m\theta$, and
$z^{-m} = \cos(-m)\theta + i\sin(-m)\theta = \cos m\theta - i\sin m\theta$.
Thus, $z^m + z^{-m} = 2\cos m\theta$, and $z^m - z^{-m} = 2i\sin m\theta$. …(6)
We can use these relations to express $\cos^m\theta$ and $\sin^m\theta$ in terms of $\cos m\theta$ and $\sin m\theta\ \forall\ m \in \mathbb{Z}$. Let us consider an example.

Example 8: Expand $2^{4n-2}(\cos^{4n}\theta + \sin^{4n}\theta)$ in terms of the cosines or sines of multiples of θ, where $\theta \in \mathbb{R}$.

Solution: Putting $m = 1$ in Equation (6), we get

$2\cos\theta = z + \dfrac{1}{z}$ and $2i\sin\theta = z - \dfrac{1}{z}$.

$$\therefore \quad 2^{4n}\cos^{4n}\theta = \left(z + \frac{1}{z}\right)^{4n}$$

$$= z^{4n} + 4nz^{4n-1}\frac{1}{z} + {}^{4n}C_2 z^{4n-2}\frac{1}{z^2} + \cdots + {}^{4n}C_{2n}z^{2n}\frac{1}{z^{2n}} + \cdots + 4nz\frac{1}{z^{4n-1}} + \frac{1}{z^{4n}},$$

by the binomial expansion.

$$= \left(z^{4n} + \frac{1}{z^{4n}}\right) + 4n\left(z^{4n-2} + \frac{1}{z^{4n-2}}\right) + \cdots + {}^{4n}C_{2n}. \qquad \ldots(7)$$

Also, $2^{4n}\sin^{4n}\theta = \left(z - \dfrac{1}{z}\right)^{4n}$, since $i^{4n} = (i^4)^n = 1$.

$$= \left(z^{4n} + \frac{1}{z^{4n}}\right) - 4n\left(z^{4n-2} + \frac{1}{z^{4n-2}}\right) + \cdots + {}^{4n}C_{2n} \qquad \ldots(8)$$

Thus, (7) and (8) give

$$2^{4n}(\cos^{4n}\theta + \sin^{4n}\theta) = 2\left(z^{4n} + \frac{1}{z^{4n}}\right) + 2\left({}^{4n}C_2\right)\left(z^{4n-4} + \frac{1}{z^{4n-4}}\right) + \cdots + 2\left({}^{4n}C_{2n}\right)$$

$= 2\{2\cos 4n\theta + 2\left({}^{4n}C_2\right)\cos(4n-4)\theta + \ldots\} + 2\left({}^{4n}C_{2n}\right)$, using (6).

$$\therefore\ 2^{4n-2}(\cos^{4n}\theta+\sin^{4n}\theta)=\cos 4n\theta+{}^{4n}C_2\cos(4n-4)\theta+\cdots+\frac{1}{2}{}^{4n}C_{2n}.$$

***

The procedure we have shown in Example 8 is very useful for solving differential equations involving trigonometric functions, as you will see in the 2nd semester course. It is also useful for finding the Laplace transform of such functions.

Why don't you try some related exercises now?

E29) Apply De Moivre's theorem to prove that

i) $\cos 2\theta=\cos^2\theta-\sin^2\theta$,

ii) $\sin 2\theta=2\sin\theta\cos\theta$.

E30) Expand $\cos^6\theta-\sin^6\theta$ in terms of the cosines of multiples of θ.

Let us now look at another area in which we can apply De Moivre's theorem with great success.

4.5.2 Roots of a Complex Number

Let us take any non-zero real number r. If $r>0$, then it has two square roots, $\sqrt{r}$ and $-\sqrt{r}$ in $\mathbb{R}$. If $r<0$, then $-r>0$. So, $(-r)$ has two distinct square roots, $\pm\sqrt{-r}\in\mathbb{R}$. The question is that if $z\in\mathbb{C},\ z\neq 0$, then does z also have two distinct square roots in $\mathbb{C}$? In fact, the set of complex numbers has a much stronger property, which is a major reason for its importance in mathematics. This property is:

> Given any $n\in\mathbb{N}$ and $z\in\mathbb{C},\ z\neq 0$, we can find **distinct** $z_1,\ldots,z_n\in\mathbb{C}$ such that $z_k^n=z\ \forall\ k=1,\ldots,n$.

Thus, every complex number has $\mathbf{n}$ distinct nth roots in $\mathbb{C}$, where $\mathbf{n}\in\mathbb{N}$. To find all these roots, we need De Moivre's theorem as well as the following theorem.

Theorem 2: Let x be a **positive** real number and $n\in\mathbb{N}$. Then there is one and only one **positive** real number b such that $b^n=x$.

We denote the unique positive nth root obtained in Theorem 2 by $\mathbf{x^{1/n}}$.

We shall not prove the existence of b here as it is beyond the level of this course. However, the uniqueness is not difficult to show, as you will see while solving the following exercise.

E31) Let x be a **positive** real number and $n\in\mathbb{N}$. Show that the **positive** real number r such that $r^n=x$ is unique.

(**Hint:** Let $r, s \ge 0$ be such that $r^n = x = s^n$. Suppose $r \ne s$. Then $r^n - s^n = 0$ and $r - s \ne 0$. Now you should be able to reach a contradiction.)

Now let us consider an example of obtaining all the nth roots of a complex number, where $n \in \mathbb{N}$. (This process is also called the **extraction of the nth roots** of a complex number.)

Example 9: Obtain all the fifth roots of i in $\mathbb{C}$.

Solution: Let $z = r(\cos\theta + i\sin\theta)$ be any 5th root of i. Then $z^5 = i$. The polar form of i is $i = \cos\frac{\pi}{2} + i\sin\frac{\pi}{2}$. Therefore,

$$z^5 = i$$

$$\Rightarrow \quad r^5(\cos\theta + i\sin\theta)^5 = \cos\frac{\pi}{2} + i\sin\frac{\pi}{2}$$

$$\Rightarrow \quad r^5(\cos 5\theta + i\sin 5\theta) = \cos\frac{\pi}{2} + i\sin\frac{\pi}{2}, \text{ by De Moivre's theorem.} \qquad \ldots(9)$$

Comparing the moduli and arguments of the complex numbers on both sides of (9), we get

$$r^5 = 1 \text{ and } 5\theta = \frac{\pi}{2} + 2k\pi, \text{ where } k = 0, \pm 1, \pm 2, \ldots.$$

r is the unique positive real fifth root of 1 (see Theorem 2). Since $1 \in \mathbb{R}$ is a fifth root of 1, $r = 1$, that is, $|z| = 1$. The possible values of θ are

$$\theta = \frac{1}{5}\left(\frac{\pi}{2} + 2k\pi\right), k = 0, \pm 1, \pm 2, \ldots.$$

Thus, the possible 5th roots of i are

$$z = \cos\left(\frac{\pi}{10} + 2k\frac{\pi}{5}\right) + i\sin\left(\frac{\pi}{10} + 2k\frac{\pi}{5}\right), k = 0, \pm 1, \pm 2, \ldots.$$

From this it seems that i has infinitely many 5th roots, one for each $k \in \mathbb{Z}$. Let us see if this is so.

In Remark 6 you will see why we start with $k = -2$.

When $k = -2$, $z = \cos\left(\frac{\pi}{10} - \frac{4\pi}{5}\right) + i\sin\left(\frac{\pi}{10} - \frac{4\pi}{5}\right)$

$$= \cos\frac{7\pi}{10} - i\sin\frac{7\pi}{10} = z_{-2}, \text{ say.}$$

When $k = -1$, $z = \cos\frac{3\pi}{10} - i\sin\frac{3\pi}{10} = z_{-1}$, say.

When $k = 0$, $z = \cos\frac{\pi}{10} + i\sin\frac{\pi}{10} = z_0$, say.

When $k = 1$, $z = \cos\frac{\pi}{2} + i\sin\frac{\pi}{2} = z_1$, say.

When $k = 2$, $z = \cos\frac{9\pi}{10} + i\sin\frac{9\pi}{10} = z_2$, say.

When $k = 3$, $z = \cos\frac{13\pi}{10} + i\sin\frac{13\pi}{10} = \cos\left(2\pi - \frac{7\pi}{10}\right) + i\sin\left(2\pi - \frac{7\pi}{10}\right) = z_{-2}$.

$\cos(2\pi \pm \theta) = \cos\theta$ and $\sin(2\pi \pm \theta) = \pm\sin\theta$.

When $k = 4$, $z = \cos\frac{17\pi}{10} + i\sin\frac{17\pi}{10} = \cos\left(2\pi - \frac{3\pi}{10}\right) + i\sin\left(2\pi - \frac{3\pi}{10}\right) = z_{-1}$.

Similarly, when $k = 5$, you will get z_0, and so on.

Thus, $k = 3, 4, \ldots$ don't give us new values of z.

Now, if we put $k = -3$, we get $z = \cos\left(\frac{-11\pi}{10}\right) + i\sin\left(\frac{-11\pi}{10}\right) = z_2$.

Similarly, $k = -4, -5, \ldots$ will not give us new values of z.

Therefore, the only 5th roots of i are

$\cos\left(\frac{\pi}{10} + 2k\frac{\pi}{5}\right) + i\sin\left(\frac{\pi}{10} + 2k\frac{\pi}{5}\right)$ for $k = 0, \pm 1, \pm 2$.

***

Remark 6: We also get the 5th roots of i by taking $k = 0\ 1, 2, 3, 4$ in $\cos\left(\frac{\pi}{10} + \frac{2k\pi}{5}\right) + i\sin\left(\frac{\pi}{10} + \frac{2k\pi}{5}\right)$, as you have seen. Only note that for $k = 3$ and $k = 4$, the angles θ will not lie in the range $-\pi < \theta \le \pi$. That's why we had taken $k = 0, \pm 1, \pm 2$.

Now, look at all the fifth roots of i. How are their moduli related? They have the same modulus, namely, $|i|^{1/5}\ (= 1)$. Thus, they all lie on the circle with centre $(0, 0)$ and radius 1. These points will be equally spaced along the circle, since the arguments of consecutive points differ by $\frac{2\pi}{5}$, a constant. We plot them in the Argand diagram in Fig. 10.

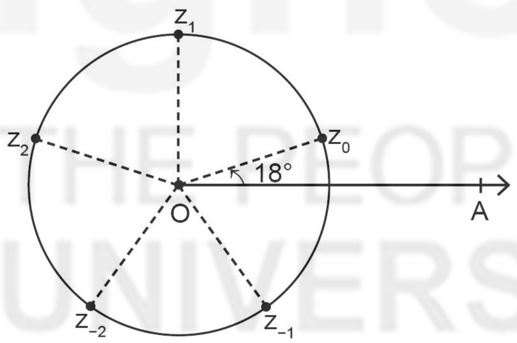


Fig. 10: The fifth roots of i.

Here's another example.

Example 10: Find all the fifth roots of unity, that is, 1.

Solution: $1 = 1(\cos 0 + i\sin 0)$.

If $[r(\cos\theta + i\sin\theta)]^5 = 1(\cos 0 + i\sin 0)$, then $r^5 = 1$ and $5\theta = 2k\pi, k = 0, \pm 1, \pm 2$. As in Example 9, $r = 1$.

Further, $\theta = \frac{2k\pi}{5}, k = 0, \pm 1, \pm 2$. Thus, the fifth roots of unity are

$\cos\frac{2k\pi}{5} + i\sin\frac{2k\pi}{5}, k = 0, \pm 1, \pm 2$.

***

Do you find any relationship between the roots in Example 9 and those in Example 10? The 5th roots of i are of the form $w\alpha_1, w\alpha_2, w\alpha_3, w\alpha_4, w\alpha_5$, where w is one 5th root of i, say, z_0 in Example 9, and $\alpha_1, \ldots, \alpha_5$ are the roots in Example 10. You should check that this is so.

Using the same procedure as in Examples 9 and 10, we can obtain the distinct nth roots of any non-zero complex number, for any $n \in \mathbb{N}$. Thus, given, any non-zero complex number z, we write it in its polar form $z = a(\cos\alpha + i\sin\alpha)$, where $a = |z|$ and $\alpha = \text{Arg } z$.

By Theorem 2, there is a unique $r \in \mathbb{R}, r > 0$, such that $r^n = a$, that is, $r = a^{1/n}$. Then the distinct nth roots of z are

$z_k = a^{1/n}\left(\cos\frac{\alpha + 2k\pi}{n} + i\sin\frac{\alpha + 2k\pi}{n}\right)$, for $k = 0, 1, \ldots, n-1$.

Geometrically, they lie on a circle of radius $a^{1/n}$ and are equally spaced along it. Note that

> a non-zero complex number has exactly n distinct nth roots for any $n \in \mathbb{N}$. If z is one root, then the others are $z\alpha_1, z\alpha_2, \ldots, z\alpha_{n-1}$, where $\alpha_1, \ldots, \alpha_{n-1}$ are the nth roots of unity.

Now you should do some related exercises.

E32) Find the cube roots of unity, that is, those $z \in \mathbb{C}$ such that $z^3 = 1$. Also represent them in an Argand diagram.

E33) Solve the equation $z^4 - 4z^2 + 4 - 2i = 0$.
(**Hint:** The equation can be rewritten as $(z^2 - 2)^2 = (1+i)^2$.)

The cube roots of unity that you obtained in E32 are very important. We usually denote the cube root $\frac{-1+i\sqrt{3}}{2}$ by the Greek letter ω (omega).

Note that $\omega^2 = \left(\frac{-1+i\sqrt{3}}{2}\right)^2 = \frac{-1-i\sqrt{3}}{2}$, the other non-real cube root of unity.

Thus, note the following:

> The three cube roots of unity are $1, \omega, \omega^2$, where $\omega = \frac{-1+i\sqrt{3}}{2}$.

Also note that

$1 + \omega + \omega^2 = 0$. …(10)

You will often find ω and the relation (10) being used in mathematics.
We will equally often use the following results, that we ask you to prove.

E34) a) Let $a \in \mathbb{R}$. Show that a has a real cube root r, and the cube roots of a are $r, r\omega, r\omega^2$.

b) Show that if $a \in \mathbb{R}, a < 0$ and n is an even positive integer, then a will not have a real nth root.

c) Let $z \in \mathbb{C} \setminus \mathbb{R}$. Show that z has three cube roots, and if any one of them is γ, the other two are $\gamma\omega, \gamma\omega^2$. Hence find the sum of the roots.

With this we come to the end of our discussion on complex numbers. This doesn't mean that you won't be dealing with them any more. In fact, you will often use whatever we have covered in this unit, while studying this course, as well as other mathematics courses.

Let us take a brief look at the points covered in this unit.

4.6 SUMMARY

In this unit on complex numbers, you have studied the following points.

1. The definition of a complex number, in algebraic (or standard) form: A complex number is a number of the form $x+iy$ where $x, y \in \mathbb{R}$ and $i=\sqrt{-1}$. Equivalently, it is a pair $(x, y) \in \mathbb{R} \times \mathbb{R}$.

2. x is the real part, and y is the imaginary part, of $x+iy$.

3. $x_1 + iy_1 = x_2 + iy_2$ iff $x_1 = x_2$ and $y_1 = y_2$.

4. The conjugate of $z = x+iy$ is $\bar{z} = x - iy$.

5. The geometric representation of the complex number $x+iy$ in an Argand diagram is the point with Cartesian coordinates (x, y).

6. The polar form of $z = x+iy$ is $z = r(\cos\theta + i\sin\theta)$, where $r = |z| = \sqrt{x^2+y^2}$ and $\theta = \operatorname{Arg} z = \tan^{-1}\left(\frac{y}{x}\right)$, where we choose θ such that it corresponds to the position of z in an Argand diagram, and $-\pi < \theta \le \pi$.

7. For $z_1, z_2 \in \mathbb{C}$,
$|z_1 z_2| = |z_1||z_2|, \operatorname{Arg}(z_1 z_2) = \operatorname{Arg} z_1 + \operatorname{Arg} z_2 + 2k\pi$
$\left|\frac{z_1}{z_2}\right| = \frac{|z_1|}{|z_2|}, \operatorname{Arg}\left(\frac{z_1}{z_2}\right) = \operatorname{Arg} z_1 - \operatorname{Arg} z_2 + 2m\pi$ (for $z_2 \ne 0$),
where $k, m \in \mathbb{Z}$ are chosen so that
$-\pi < \operatorname{Arg}(z_1 z_2) \le \pi$ and $-\pi < \operatorname{Arg}\left(\frac{z_1}{z_2}\right) \le \pi$.

8. $\forall a, b, c, d \in \mathbb{R}$
i) $(a+ib) \pm (c+id) = (a \pm c) + i(b \pm d)$,
ii) $(a+ib) \times (c+id) = (ac-bd) + i(ad+bc)$,
iii) $\frac{1}{a+ib} = \frac{a}{a^2+b^2} - \left(\frac{b}{a^2+b^2}\right) i$, for $a+ib \ne 0$,
iv) $\frac{a+ib}{c+id} = \frac{(a+ib)\,(c-id)}{c^2+d^2}$, for $c+id \ne 0$.

9. De Moivre's theorem: $(\cos\theta + i\sin\theta)^n = \cos n\theta + i\sin n\theta \,\forall\, n \in \mathbb{Z}$ and any angle θ.

10. Applying De Moivre's theorem to prove trigonometric identities and for obtaining nth roots of complex numbers, where $n \in \mathbb{N}$.

11. The cube roots of unity are $1, \omega, \omega^2$, where $\omega = \left(-\frac{1}{2}\right) + i\left(\frac{\sqrt{3}}{2}\right)$.

    Further, $1 + \omega + \omega^2 = 0$.

Now that you have gone through this unit, please **go back to the objectives** listed in Sec. 4.1. Do you think you have achieved them? As mentioned in Sec. 4.1, one way of finding out is to solve all the exercises that we have given you in this unit. If you would like to verify your solutions or answers, you can see what we have given in the following section.

4.7 SOLUTIONS/ANSWERS

E1)

z	Re z	Im z
$\frac{1+\sqrt{-23}}{2}$	$\frac{1}{2}$	$\frac{\sqrt{23}}{2}$
i	0	1
0	0	0
$\frac{-1+\sqrt{3}}{5}$	$\frac{-1+\sqrt{3}}{5}$	0

E2) Yes, because every real number x is the complex number $x + 0i$.

E3) i) $\sqrt{k} + 3i = \sqrt{3} + im \Leftrightarrow \sqrt{k} = \sqrt{3}$ and $3 = m \Leftrightarrow k = 3,\ m = 3$.

ii) $ki = m \in \mathbb{R}$ iff $k = 0$. Thus, $ki - m \in \mathbb{R} \,\forall\, m \in \mathbb{R}$ and $k = 0$.

E4) $\overline{-5} = \overline{-5 + 0.i} = -5 - 0.i = -5$.

$\sqrt{-5} = i\sqrt{5}$. Hence $\overline{\sqrt{-5}} = -i\sqrt{5} = -\sqrt{-5}$.

$\overline{2+3i} = 2 - 3i$.

$\overline{2-3i} = 2 + 3i$.

E5) Let $z = x + iy$. Then $\bar{z} = x - iy$.

$\therefore\ z = \bar{z} \Rightarrow x + iy = x - iy \Rightarrow y = -y \Rightarrow y = 0$.

$\therefore\ z = \bar{z}$ iff $z \in \mathbb{R}$.

E6) Let $z = x + iy$. Then $\bar{z} = x - iy$.

$\therefore\ \bar{\bar{z}} = \overline{x - iy} = x + iy = z$.

E7) a) P, Q, R and S represent $3, -1+i, \overline{-1+i}$ and i, respectively, in Fig. 11.

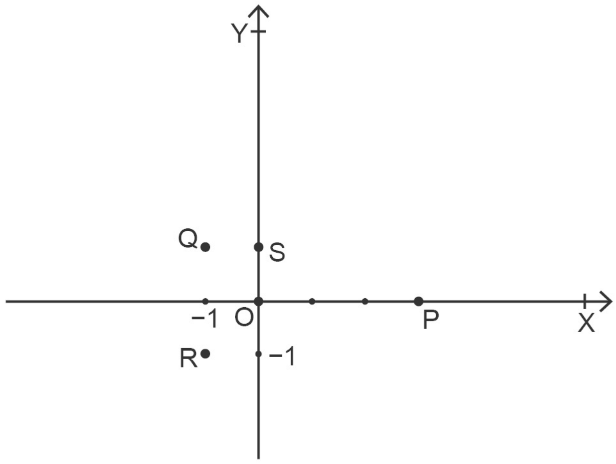


Fig. 11

b) The set $S_1 = \{(2, y) \mid y \in \mathbb{R}\}$, that is, the set of points satisfying the linear equation $x = 2$. Similarly, you can see that S_2 is represented by the line $y = 3$, and S_3 by the line $y = x$. Let L_1, L_2 and L_3 represent the sets S_1, S_2 and S_3, respectively. These are shown in Fig. 12.

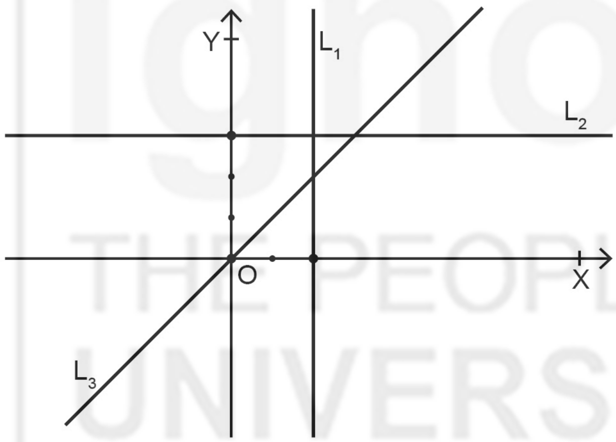


Fig. 12

E8) $\frac{-1}{2} + \frac{i}{3}$, 2 and $-2i$ are the respective elements of $\mathbb{C}$.

E9) $|3| = 3$ and $\text{Arg}(3) = \tan^{-1}(0) = 0$, since 3 lies on the positive side of the real-axis. So, $3 = 3(\cos 0 + i \sin 0)$.

Now $|-1+i| = \sqrt{1+1} = \sqrt{2}$, and

$\text{Arg}(-1+i) = \tan^{-1}(-1) = -\pi/4$ or $3\pi/4$

Since $-1+i$ corresponds to $(-1, 1)$, which lies in the 2nd quadrant,

$$\text{Arg}(-1+i) = \frac{3\pi}{4}.$$

$$\therefore \quad (-1+i) = \sqrt{2}\left(\cos\left(\frac{3\pi}{4}\right) + i \sin\left(\frac{3\pi}{4}\right)\right).$$

$$\overline{-1+i} = -1 - i = \sqrt{2}\left(\cos\left(\frac{-3\pi}{4}\right) + i \sin\left(\frac{-3\pi}{4}\right)\right)$$

$$= \sqrt{2}\left(\cos\left(\frac{3\pi}{4}\right) - i\sin\left(\frac{3\pi}{4}\right)\right).$$

$$i = \cos\frac{\pi}{2} + i\sin\frac{\pi}{2}.$$

E10) Let $z = x + iy$. Then $\bar{z} = x - iy$. So, $|z| = \sqrt{x^2 + y^2} = |\bar{z}|$.

Next, $\text{Arg } z = \tan^{-1}\left(\frac{y}{x}\right)$ and $\text{Arg } \bar{z} = -\tan^{-1}\left(\frac{y}{x}\right)$. Thus,

$\text{Arg } z = -\text{Arg } \bar{z}$.

E11) For any $z = x + iy \in \mathbb{C}, |z| = 1 \Leftrightarrow \sqrt{x^2 + y^2} = 1 \Leftrightarrow x^2 + y^2 = 1$.

E12) i) $2 + 3i + \overline{2 + 3i} = 2 + 3i + 2 - 3i = 4 + 0i = 4$.

ii) Let $z = x + iy$. Then
$z + \bar{z} = (x + iy) + (x - iy) = 2x = 2\,\text{Re}\, z$.

E13) Let $z_1 = x_1 + iy_1$ and $z_2 = x_2 + iy_2$. Then

$z_1 + z_2 = (x_1 + x_2,\ y_1 + y_2)$

$\therefore\ \overline{z_1 + z_2} = (x_1 + x_2,\ -(y_1 + y_2))$
$= (x_1 + x_2,\ -y_1 - y_2)$
$= (x_1,\ -y_1) + (x_2,\ -y_2)$
$= \bar{z}_1 + \bar{z}_2$.

E14) a) Let $z_1 = (x_1,\ y_1)$ and $z_2 = (x_2,\ y_2)$.

Then $z_1 + z_2 = (x_1,\ y_1) + (x_2,\ y_2)$
$= (x_1 + x_2,\ y_1 + y_2)$
$= (x_2 + x_1,\ y_2 + y_1)$, since $a + b = b + a\ \forall a, b \in \mathbb{R}$
$= (x_2,\ y_2) + (x_1,\ y_1)$
$= z_2 + z_1$

b) Let $z_1 = (x_1,\ y_1),\ z_2 = (x_2,\ y_2),\ z_3 = (x_3,\ y_3)$.
Then, use the fact that $(a + b) + c = a + (b + c)\ \forall a, b, c \in \mathbb{R}$, to obtain the result, on the same lines as in E14(a) above.

E15) Let $z = x + iy$.

Then $z + (a + ib) = z$
$\Leftrightarrow\ (x + iy) + (a + ib) = x + iy$
$\Leftrightarrow\ (x + a) + i(y + b) = x + iy$
$\Leftrightarrow\ x + a = x$ and $y + b = y$
$\Leftrightarrow\ a = 0,\ b = 0$
$\therefore\ a + ib = 0 + i0 = 0$ satisfies the requirement.

E16) i) $(3 - 2i) - \overline{(3 - 2i)} = (3 - 2i) - (3 + 2i) = (3 - 3) + i(-2 - 2) = -4i$.

ii) Let $z = x + iy$. Then
$z - \bar{z} = (x + iy) - (x - iy) = (x - x) + i(y + y) = 2iy$.

$z - \bar{z}$ is purely imaginary if $\text{Im}\, z \neq 0$.

$= i(2\,\mathrm{Im}\,z)$.

E17) Let $z = x + iy$. Then $-z = (-x) + i(-y)$.

i) $|z| = \sqrt{x^2 + y^2}$, and

$|-z| = \sqrt{(-x)^2 + (-y)^2} = \sqrt{x^2 + y^2} = |z|$

ii) $\mathrm{Arg}\,z = \tan^{-1}\left(\frac{y}{x}\right)$

$\mathrm{Arg}\,(-z) = \tan^{-1}\left(\frac{-y}{-x}\right) = \tan^{-1}\left(\frac{y}{x}\right) = \mathrm{Arg}\,z \pm \pi$, because $(-z)$ is the reflection of z in the origin.

The **reflection in the origin** of the point $(a, b) \in \mathbb{R} \times \mathbb{R}$ is $(-a, -b)$.

E18) Let $z = r(\cos\theta + i\sin\theta)$. Now

$1 = \cos 0 + i \sin 0, |0| = 0, i = \cos\frac{\pi}{2} + i\sin\frac{\pi}{2}$.

So $z.1 = r\{\cos(\theta + 0) + i\sin(\theta + 0)\} = r(\cos\theta + i\sin\theta) = z$

Now $|z.0| = |z||0| = 0$. Thus, $z.0 = 0$.

$z.i = r\{\cos(\theta + \pi/2) + i\sin(\theta + \pi/2)\} = r(-\sin\theta + i\cos\theta)$

$z\bar{z} = r^2(\cos^2\theta + \sin^2\theta) = r^2 = |z|^2$.

E19) a) Let $z_1 = r_1(\cos\theta_1 + i\sin\theta_1)$ and $z_2 = r_2(\cos\theta_2 + i\sin\theta_2)$. Then

$$\begin{aligned} z_1 z_2 &= r_1 r_2\{\cos(\theta_1 + \theta_2 + 2k\pi) + i\sin(\theta_1 + \theta_2 + 2k\pi)\} \\ &= r_2 r_1\{\cos(\theta_2 + \theta_1 + 2k\pi) + i\sin(\theta_2 + \theta_1 + 2k\pi)\} \\ &= z_2 z_1, \end{aligned}$$

where k is chosen so that $\theta_1 + \theta_2 + 2k\pi$, that is, $\theta_2 + \theta_1 + 2k\pi$, lies between $-\pi$ and π.

b) Let $z_m = r_m(\cos\theta_m + i\sin\theta_m)$ for $m = 1, 2, 3$.

Then you can check that $|(z_1 z_2) z_3| = (r_1 r_2) r_3$

$= r_1(r_2 r_3) = |z_1(z_2 z_3)|$.

Again, $\mathrm{Arg}[(z_1 z_2) z_3] = \{(\theta_1 + \theta_2) + 2k\pi\} + \theta_3 + 2s\pi$

$= (\theta_1 + \theta_2) + \theta_3 + 2(k + s)\pi$,

$= \theta_1 + (\theta_2 + \theta_3 + 2s\pi) + 2k\pi = \mathrm{Arg}[z_1(z_2 z_3)]$,

where integers k and s are chosen so that

$-\pi < (\theta_1 + \theta_2) + 2k\pi \le \pi$, $-\pi < (\theta_2 + \theta_3) + 2s\pi \le \pi$ and

$-\pi < (\theta_1 + \theta_2) + \theta_3 + 2(k + s)\pi \le \pi$.

E20) $z_1 = 6(\cos\pi + i\sin\pi)$, $z_2 = \sqrt{2}\left(\cos\frac{\pi}{4} + i\sin\frac{\pi}{4}\right)$

$\therefore |z_1 z_2| = 6\sqrt{2}$ and

$\mathrm{Arg}\,(z_1 z_2) = \left(\pi + \frac{\pi}{4}\right) + 2k\pi$, where $k \in \mathbb{Z}$ such that $-\pi < \mathrm{Arg}\,(z_1 z_2) \le \pi$

$\therefore \quad \mathrm{Arg}\,(z_1 z_2) = \frac{-3\pi}{4}$.

E21) $(x, y)\,(1, 0) = (x, y)$

$(x, y)\,(0, 1) = (-y, x)$

$(x, y)\,(0, 0) = (0, 0)$

$(x, 0)\,(y, 0) = (xy, 0)$

$(x, y)\,(1, 1) = (x - y, x + y)$.

$z.1 = z \;\forall z \in \mathbb{C}$

$z.0 = 0 \;\forall z \in \mathbb{C}$.

E22) $(x, y)\left(\dfrac{x}{x^2+y^2}, \dfrac{-y}{x^2+y^2}\right) = \left(\dfrac{x^2+y^2}{x^2+y^2}, \dfrac{-xy+xy}{x^2+y^2}\right) = (1, 0)$.

E23) $\dfrac{-2+i}{i\sqrt{3}+i(2i)} = \dfrac{-2+i}{-2+i\sqrt{3}}$, since $i^2 = -1$.

$$= \frac{(-2+i)\,(-2-i\sqrt{3})}{(-2)^2 + \left(\sqrt{3}\right)^2}$$

$$= \frac{4+\sqrt{3}}{7} + \frac{2}{7}\left(\sqrt{3}-1\right)i.$$

The modulus of this number is $\sqrt{\left(\dfrac{4+\sqrt{3}}{7}\right)^2 + \left(\dfrac{2}{7}\right)^2 (\sqrt{3}-1)^2} = \dfrac{\sqrt{35}}{7}$.

Its principal argument is $\theta = \tan^{-1}\left\{\dfrac{2\left(\sqrt{3}-1\right)}{4+\sqrt{3}}\right\}$.

Thus, its polar form is $\dfrac{\sqrt{35}}{7}(\cos\theta + i\sin\theta)$.

E24) $c^2 + d^2 \neq 0$ means that $c \neq 0$ or $d \neq 0$. Thus, $c + id \neq 0$. Hence $\dfrac{a+ib}{c+id}$ is meaningful.

$$\frac{a+ib}{c+id} = \frac{(a+ib)(c-id)}{(c+id)(c-id)} = \frac{(ac+bd)+i(bc-ad)}{c^2+d^2} = \left(\frac{ac+bd}{c^2+d^2}\right) + i\left(\frac{bc-ad}{c^2+d^2}\right).$$

E25) Let $z = x + iy \neq 0$. Then, from E18 we know that $z\bar{z} = |z|^2 \in \mathbb{R}$.

Therefore, $z\left(\dfrac{1}{|z|^2}\bar{z}\right) = 1$. Thus, $\dfrac{1}{|z|^2}\bar{z}$ is the multiplicative inverse of z, that is, $\dfrac{1}{z}$.

Now, $z.\dfrac{1}{z} = 1$. $\therefore |z|.\left|\dfrac{1}{z}\right| = |1| = 1$. $\therefore \left|\dfrac{1}{z}\right| = \dfrac{1}{|z|}$.

E26) The points P, Q, R, S and T in Fig. 13 represent $z_1, z_2, \bar{z}_2, \dfrac{1}{z_2}$ and $\dfrac{z_1}{z_2}$, respectively. Here $OT = \dfrac{OP}{OQ}$ and $\angle XOT = \angle XOP - \angle XOQ$.

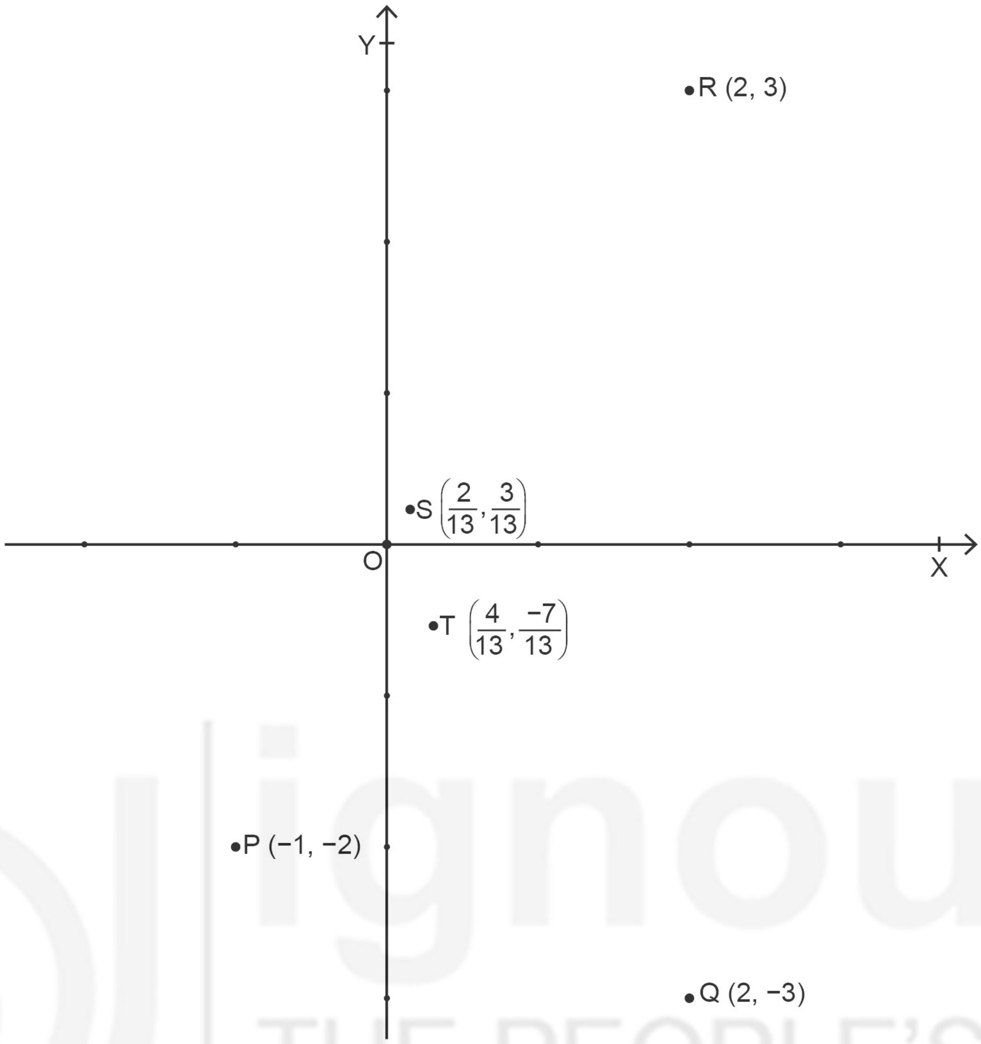


Fig. 13

E27) i) LHS $= (1+i)\,[(\sqrt{2}+5)-2i] = (7+\sqrt{2})+i(3+\sqrt{2})$

RHS $= [(\sqrt{2}+3)+i\,(\sqrt{2}-3)]+(4+6i) = (\sqrt{2}+7)+i\,(\sqrt{2}+3)$

Thus, LHS=RHS.

ii) Let $z_1 = x_1 + iy_1,\ z_2 = x_2 + iy_2,\ z_3 = x_3 + iy_3$

Then $z_1(z_2 + z_3) = (x_1 + iy_1)[(x_2 + x_3) + i(y_2 + y_3)]$

$= [x_1(x_2 + x_3) - y_1(y_2 + y_3)] + i[x_1(y_2 + y_3) + y_1(x_2 + x_3)]$

$= (x_1x_2 - y_1y_2) + (x_1x_3 - y_1y_3) + i[(x_1y_2 + x_2y_1) + (x_1y_3 + x_3y_1)]$

$= [(x_1x_2 - y_1y_2) + i\,(x_1y_2 + x_2y_1)] + [x_1x_3 - y_1y_3) + i(x_1y_3 + x_3y_1)]$

$= z_1z_2 + z_1z_3$.

You can also solve this by writing $z_1,\ z_2$ and z_3 in polar form. If you do, you must remember to be careful about $z_i = 0$ for any i.

E28) $(\cos\theta + i\sin\theta)^3 = \cos 3\theta + i\sin 3\theta$. …(11)

Also,

$(\cos\theta + i\sin\theta)^3 = \cos^3\theta + 3\cos^2\theta(i\sin\theta) + 3\cos\theta(i\sin\theta)^2 + (i\sin\theta)^3$

$= (\cos^3\theta - 3\cos\theta\sin^2\theta) + i\,(3\sin\theta\cos^2\theta - \sin^3\theta)$. …(12)

Thus, comparing real parts of (11) and (12), we get

$\cos 3\theta = \cos^3\theta - 3\cos\theta\sin^2\theta = \cos^3\theta - 3\cos\theta(1-\cos^2\theta)$

$\quad = 4\cos^3\theta - 3\cos\theta$.

Similarly, comparing the imaginary parts, we get

$\sin 3\theta = 3\sin\theta(1-\sin^2\theta) - \sin^2\theta = 3\sin\theta - 4\sin^3\theta$.

E29) $(\cos\theta + i\sin\theta)^2 = \cos 2\theta + i\sin 2\theta$, and

$(\cos\theta + i\sin\theta)^2 = \cos^2\theta + 2i\cos\theta\sin\theta - \sin^2\theta$.

$\therefore\ \ \cos\theta = \cos^2\theta - \sin^2\theta$, and

$\sin 2\theta = 2\sin\theta\cos\theta$.

E30) Let $z = \cos\theta + i\sin\theta$. Then, using (7) and (8) of Example 8, we get

$$(2\cos\theta)^6 = \left(z + \frac{1}{z}\right)^6 = \left(z^6 + \frac{1}{z^6}\right) + 6\left(z^4 + \frac{1}{z^4}\right) + 15\left(z^2 + \frac{1}{z^2}\right) + 20, \text{ and}$$

$$(2i\sin\theta)^6 = \left(z^6 + \frac{1}{z^6}\right) - 6\left(z^4 + \frac{1}{z^4}\right) + 15\left(z^2 + \frac{1}{z^2}\right) - 20.$$

$$\therefore\ \ 2^6(\cos^6\theta - \sin^6\theta) = 2\left(z^6 + \frac{1}{z^6}\right) + 30\left(z^2 + \frac{1}{z^2}\right)$$

$$= 4\cos 6\theta + 60\cos 2\theta, \text{ using (6).}$$

$$\Rightarrow\ \ \cos^6\theta - \sin^6\theta = \frac{1}{16}(\cos 6\theta + 15\cos 2\theta).$$

E31) Let $r, s \in \mathbb{R}$, $r, s > 0$ and $r^n = x = s^n$. Then

$r^n - s^n = (r-s)(r^{n-1} + r^{n-2}s + \cdots + rs^{n-2} + s^{n-1}) = 0$

Now, $r > 0$, $s > 0$, so that $r^{n-1} + r^{n-2}s + \cdots + rs^{n-2} + s^{n-1} > 0$.

Thus, $(r-s)\ (r^{n-1} + r^{n-2}s + \cdots + s^{n-1}) = 0$ only if $r - s = 0$, i.e., $r = s$.

E32) Let $z = r(\cos\theta + i\sin\theta)$ be a cube root of $1 = \cos 0 + i\sin 0$.

Then $r = 1^{1/3} = 1,\ \theta = \dfrac{0 + 2k\pi}{3} = \dfrac{2k\pi}{3}$ for $k = 0, 1, -1$.

Thus, the roots are $1, \dfrac{-1}{2} + i\dfrac{\sqrt{3}}{2}$ and $\dfrac{-1}{2} - i\dfrac{\sqrt{3}}{2}$.

They are represented in Fig. 14 by z_0, z_1 and z_{-1}.

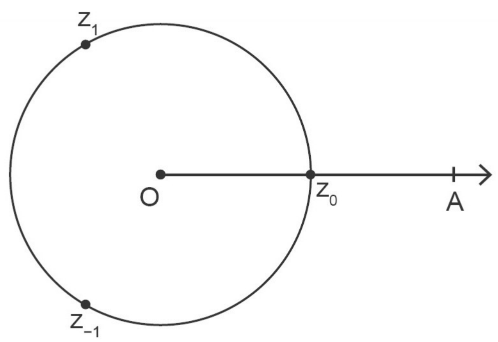


Fig. 14: Cube roots of unity.

E33) We want to obtain those $z \in \mathbb{C}$ for which

$(z^2 - 2) = \pm(1+i)$, that is,

$z^2 - 2 = 1 + i$ and $z^2 - 2 = -(1+i)$, that is,

$z^2 = 3 + i$ and $z^2 = 1 - i$.

Thus, we want to find the square roots of $3 + i$ and $1 - i$.

Now, $3+i=\sqrt{10}\left\{\cos\left(\tan^{-1}\frac{1}{3}\right)+i\sin\left(\tan^{-1}\frac{1}{3}\right)\right\}$.

Thus, the square roots of $3+i$ are

$10^{1/4}\left(\cos\frac{\theta}{2}+i\sin\frac{\theta}{2}\right)$ and $10^{1/4}\left\{\cos\left(\frac{\theta}{2}+\pi\right)+i\sin\left(\frac{\theta}{2}+\pi\right)\right\}$,

where $\theta=\tan^{-1}\frac{1}{3}$.

Also $1-i=\sqrt{2}\left\{\cos\left(\frac{-\pi}{4}\right)+i\sin\left(\frac{-\pi}{4}\right)\right\}$, so that the square roots of $1-i$ are $2^{1/4}\left(\cos\frac{\pi}{8}-i\sin\frac{\pi}{8}\right)$ and $2^{1/4}\left(\cos\frac{7\pi}{8}-i\sin\frac{7\pi}{8}\right)$.

These 4 square roots are the 4 roots of the given equation.

E34) a) If $a\geq 0$, then by Theorem 2, a has a real cube root, $a^{1/3}$. Now, $a=a(\cos 0+i\sin 0)$.

Thus, the cube roots of a are

$$a^{1/3}\left(\cos\frac{2k\pi}{3}+i\sin\frac{2k\pi}{3}\right),\ k=0,1,2$$

that is, $a^{1/3}, a^{1/3}\omega, a^{1/3}\omega^2$.

If $a<0$, then $-a>0$. Thus, $-a$ has a real cube root, say b.

Then $r=-b$ is a real cube root for a. And $|r|=|a|^{1/3}$, that is, $r=-|r|=-|a|^{1/3}$ (since r is negative).

Now $a=|a|(\cos\pi+i\sin\pi)$. Therefore, the cube roots of a are

$$|a|^{1/3}\left(\cos\frac{(2k+1)\pi}{3}+i\sin\frac{(2k+1)\pi}{3}\right),\ k=0,1,2.$$

$$=r(\cos\pi+i\sin\pi)\left(\cos\frac{(2k+1)\pi}{3}+i\sin\frac{(2k+1)\pi}{3}\right),\ k=0,1,2$$

(since $-1=\cos\pi+i\sin\pi$).

$$=r\left(\cos\frac{(2k+4)\pi}{3}+i\sin\frac{(2k+4)\pi}{3}\right),\ k=0,1,2.$$

Thus, the cube roots of a are $r, r\omega, r\omega^2$.

b) Let $n=2m, m\in\mathbb{N}$. Then, for any $b\in\mathbb{R}$,

$b^n=b^{2m}=(b^2)^m\geq 0$.

Thus, $b^n\neq a$ for any $b\in\mathbb{R}$. Hence, a can't have a real nth root.

c) Let $z=r(\cos\theta+i\sin\theta)$, in polar form.

Then its cube roots are

$$r^{1/3}\left(\cos\frac{\theta+2k\pi}{3}+i\sin\frac{\theta+2k\pi}{3}\right),\ k=0,1,2.$$

Thus, if $\gamma=r^{1/3}\left(\cos\frac{\theta}{3}+i\sin\frac{\theta}{3}\right)$, then the other roots are

$r^{1/3}\left(\cos\frac{\theta+2\pi}{3}+i\sin\frac{\theta+2\pi}{3}\right)=\gamma\left(\cos\frac{2\pi}{3}+i\sin\frac{2\pi}{3}\right)=\gamma\omega$, and

$r^{1/3}\left(\cos\frac{\theta+4\pi}{3}+i\sin\frac{\theta+4\pi}{3}\right)=\gamma\omega^2$.

UNIT 5

POLYNOMIAL EQUATIONS AND THEIR SOLUTIONS

Structure **Page Nos.**

5.1 INTRODUCTION

In your school studies, and in the units of this course, you have come across equations of the form $ax + b = 0$, or $ax^2 + bx + c = 0$, where $a, b, c \in \mathbb{R}$. These are examples of polynomial equations over $\mathbb{R}$, as you will see in this unit. Finding solutions of such equations has exercised the minds of several mathematicians through the ages. The ancient Indian, Arabic and Babylonian mathematicians had discovered methods of solving linear and quadratic equations. The ancient Babylonians and Greeks had also discovered methods of solving some cubic equations, that is, equations like $ax^3 + bx^2 + cx + d = 0$, $a, b, c, d \in \mathbb{R}$. But, as we have said in Unit 4, they had not thought of complex numbers. So, for them, many quadratic and cubic equations had no solutions.

Here, in Sec. 5.2, you will get a chance to recall what you have studied about linear and quadratic equations, and their roots. Then, in Sec. 5.3, we will introduce you to the general polynomial equation over $\mathbb{R}$, and its roots in $\mathbb{C}$.

Next, in Sec. 5.4, you will find some very interesting, and maybe unexpected, relations between the roots and coefficients of polynomial equations. You will also see how these relations can be exploited to find solutions of the equations concerned.

Finally, in Sec. 5.5, we will discuss the types of roots of polynomial equations, whether there are any real roots, how many could be non-real, etc. Here you would also apply a very interesting rule, noticed by Descartes, giving a relationship between the types of real roots of a polynomial equation and the sign of its coefficients.

In this unit, when we talk of polynomial equations, we will always assume them to be in one variable, and with coefficients in $\mathbb{R}$, unless otherwise mentioned.

There are several reasons, apart from a mathematician's natural curiosity, for studying polynomial equations. The material covered in this unit is also useful for mathematicians, physicists, chemists and social scientists.

After going through the unit, please check to see if you have achieved the following objectives.

Objectives

After studying this unit, you should be able to:

- solve any linear or quadratic equation over $\mathbb{R}$;
- apply the procedure for obtaining one or more roots in $\mathbb{C}$ of a polynomial over $\mathbb{R}$, by inspection;
- give the relations between the roots and coefficients of a polynomial equation over $\mathbb{R}$;
- use the relations between the roots and coefficients of a cubic or quartic polynomial to solve such equations;
- apply Descartes' rule of signs, and the discriminant of a polynomial equation, for finding the nature of the roots of a polynomial over $\mathbb{R}$.

5.2 LINEAR AND QUADRATIC EQUATIONS

Let us begin the unit by recalling what you have studied about linear and quadratic equations.

5.2.1 Linear Equations

As you know, $2x+3, -\pi x, \sqrt{2}x-5$ are all linear polynomials over $\mathbb{R}$. You also know that $2x+3=0$, or $-\pi x=0$, are examples of linear equations. More generally, we have the following definitions.

Definition: i) An expression of the form $ax+b$, with $a, b \in \mathbb{R}$ and $a \neq 0$, is a **linear polynomial over $\mathbb{R}$ in one variable x**.

ii) An equation of the form $ax+b=0, a, b \in \mathbb{R}, a \neq 0$, is called a **linear equation over $\mathbb{R}$**.

iii) A **solution of the linear equation** $ax+b=0$ is a complex number α for which $a\alpha+b=0$. α is also called **a root** of $ax+b=0$.

Now, you can find a solution of a linear equation just by looking at it, that is, **by inspection**. For instance, if $\sqrt{2}x-7=0$ is the equation, you know that $x=\frac{7}{\sqrt{2}}$ is a solution, as $\sqrt{2}\left(\frac{7}{\sqrt{2}}\right)-7=0$. Are there any more solutions?

Let's see. Let $\alpha \in \mathbb{C}$ be such that $\sqrt{2}\alpha - 7 = 0$. Then, you can see that $\alpha = \frac{7}{\sqrt{2}}$, which is the same as the earlier solution.

Thus, more generally, we have the following theorem.

Theorem 1: The linear equation $ax + b = 0,\ a,\ b \in \mathbb{R},\ a \neq 0$, has one and only one solution, viz., $x = \frac{-b}{a}$.

In earlier units, and in school, you would have solved several linear equations. You may also recall, from Unit 2, that a function $f : \mathbb{R} \to \mathbb{R} : f(x) = ax + b,\ a,\ b \in \mathbb{R},\ a \neq 0$, is called a **linear function**. So a linear polynomial gives rise to a linear function.

Let us now look at an example of the use of linear equations in daily life.

Example 1: Suppose I bought two plots of land for Rs. 1,20,000, and then sold them. Also suppose that I have made a profit of 15% on the first plot and a loss of 10% on the second plot. If my total profit is Rs. 5500, how much did I pay for each piece of land?

Solution: Suppose the first piece of land cost Rs. x. Then the second piece cost Rs. $(1,20,000 - x)$. Thus, my profit is Rs. $\frac{15}{100}x$ and my loss is Rs. $\frac{10}{100}(1,20,000 - x)$.

$\therefore \frac{15}{100}x - \frac{10}{100}(1,20,000 - x) = 5500$

i.e., $25x - 17,50,000 = 0$

i.e., $x = 70,000$.

Thus, the first piece cost Rs.70,000 and the second plot cost Rs.50,000.

***

Sometimes you may come across equations that do not appear to be linear but, after simplification, they become linear. Let us consider an example.

Example 2: Solve $\frac{3p-1}{3} - \frac{2p}{p-1} = p$. (Here we must assume $p \neq 1$.)

Solution: At first glance, this equation in p does not appear to be linear. But, by cross-multiplying, we get the following equivalent equation:

$(3p-1)\ (p-1) - 3(2p) = 3(p-1)p$.

On simplifying this we get

$3p^2 - 4p + 1 - 6p = 3p^2 - 3p$, that is, $7p - 1 = 0$.

The solution of this equation is $\frac{1}{7}$. Thus, this is the solution of the equation we started with.

***

You may like to try these exercises now.

E1) Solve each of the following equations for the variable indicated. Assume that all denominators are non-zero.

i) $J\left(\frac{x}{k}+a\right)=x$ for x, where J, k and a are constants, $J \neq k$.

ii) $\frac{1}{R}=\frac{1}{r_1}+\frac{1}{r_2}$ for R, keeping r_1 and r_2 constant, $r_1+r_2 \neq 0$.

iii) $C=\frac{5}{9}(F-32)$ for F, keeping C constant.

E2) A student cycles from her home to the study centre in 20 minutes. The return journey is uphill and takes her half an hour. If her rate is 8 km per hour slower on the return trip, how far does she live from the study centre?

Now that we have looked at linear equations, let us consider quadratic equations.

5.2.2 Quadratic Equations

The word 'quadratic' comes from the Latin word 'quadratum', meaning 'square'.

In the earlier units you have seen several quadratic equations. One is $x^2=5$, which is the same as $x^2-5=0$. Another is the equation Cardano tried to solve, namely, $x^2-10x+40=0$ (see Sec. 4.1). We are sure you can think of several others. Let us define the term, in general.

Definitions: i) An expression of the form ax^2+bx+c, where $a, b, c \in \mathbb{R}, a \neq 0$, is called **a quadratic polynomial** over $\mathbb{R}$ in one variable x.

ii) On equating a quadratic polynomial to zero, we get **a quadratic equation over $\mathbb{R}$** in standard form.

iii) A **solution** of the quadratic equation $ax^2+bx+c=0$ is a complex number α such that $a\alpha^2+b\alpha+c=0$.

Now, you know that $x^2-5=0$ has two solutions in $\mathbb{C}$, i.e., $x=\sqrt{5}$ and $x=-\sqrt{5}$. These are called the roots of the given equation.

Various methods for solving quadratic equations have been known since Babylonian times (2000 B.C.). Brahmagupta, in 628 A.D. approximately, also gave a rule for solving quadratic equations. As you may recall, the method that can be used for any quadratic equation is "completing the square". Using it, we get the quadratic formula given in the box.

Quadratic Formula: The two solutions of the quadratic equation $ax^2+bx+c=0$, where $a, b, c \in \mathbb{R}$ and $a \neq 0$, are $x=\frac{-b \pm \sqrt{b^2-4ac}}{2a}$.

The expression b^2-4ac is called the **discriminant** of $ax^2+bx+c=0$. Note that this formula tells us a quadratic equation has only two roots. These roots may be equal or they may be distinct, they may be real or complex.

Convention: We call a root that lies in $\mathbb{C} \setminus \mathbb{R}$ **a complex root**. Note that such a root is of the form $a+ib, a, b \in \mathbb{R}, b \neq 0$.

Let us consider some examples, which show us the different possibilities of the nature of the roots of a quadratic equation.

Example 3: Solve the following quadratic equations:

i) $x^2 - 4x + 1 = 0$,

ii) $4x^2 + 25 = 20x$,

iii) $x^2 - 10x + 40 = 0$.

Solution: i) This equation is in standard form. So we can apply the quadratic formula immediately. Here $a = 1,\ b = -4,\ c = 1$. Substituting these values in the quadratic formula, we get the two roots of the equation to be

$$x = \frac{-(-4) + \sqrt{(-4)^2 - 4(1)\ (1)}}{2(1)} = \frac{4 + \sqrt{12}}{2} = 2 + \sqrt{3}, \text{ and}$$

$$x = \frac{-4(-4) - \sqrt{(4)^2 - 4(1)\ (1)}}{2(1)} = 2 - \sqrt{3}.$$

Thus, the solutions are $2 + \sqrt{3}$ and $2 - \sqrt{3}$, two distinct elements of $\mathbb{R}$. Note that in this case the discriminant is positive.

ii) In this case let us first rewrite the equation in standard form as $4x^2 - 20x + 25 = 0$.
Now, putting $a = 4,\ b = -20,\ c = 25$ in the quadratic formula, we find that

$$x = \frac{20 + \sqrt{400 - 4(4)\ (25)}}{2(4)} = \frac{20 + \sqrt{0}}{8} = \frac{5}{2}, \text{ and}$$

$$x = \frac{20 - \sqrt{400 - 4(4)\ (25)}}{2(4)} = \frac{5}{2}.$$

Here we find that both the roots coincide and are real.
Note that in this case the discriminant is 0.

iii) Using the quadratic formula, we find that the solutions are

$$x = \frac{10 \pm \sqrt{100 - 160}}{2} = 5 \pm \frac{\sqrt{-60}}{2}$$

$$= 5 \pm i\sqrt{15}.$$

Thus, in this case we get two distinct complex roots, $5 + i\sqrt{15}$ and $5 - i\sqrt{15}$.
Note that in this case the discriminant is negative.

***

In the example above do you see a relationship between the types of roots of a quadratic equation and the value of its discriminant? There is such a relationship, which we now state.

Theorem 2: The equation $ax^2 + bx + c = 0,\ a \neq 0,\ a,\ b,\ c \in \mathbb{R}$, has two roots in $\mathbb{C}$. They are:

i) real and distinct if $b^2 - 4ac > 0$;

ii) real and equal if $b^2 - 4ac = 0$;

iii) complex and distinct if $b^2 - 4ac < 0$.

Now let us consider some important remarks which would be useful to you while solving quadratic equations.

Remark 1: α and β are roots of a quadratic equation $ax^2 + bx + c = 0$ if and only if $ax^2 + bx + c = a(x-\alpha)\,(x-\beta)$.
Thus, $\alpha \in \mathbb{C}$ is a root of $ax^2 + bx + c = 0$ if and only if $(x-\alpha) \mid (ax^2 + bx + c)$.

Remark 2: From the quadratic formula, you can see that if $b^2 - 4ac < 0$, then the quadratic equation $ax^2 + bx + c = 0$ has 2 complex roots which are each other's conjugates. Thus**, if $\mathbf{b^2 - 4ac < 0}$, and one root is $\boldsymbol{\alpha + i\beta}$, then the other root must be $\boldsymbol{\alpha - i\beta}$.**

Remark 3: Sometimes a quadratic equation can be solved without resorting to the quadratic formula, just by inspection. For example, the equation $x^2 = 9$ clearly has 3 and -3 as its roots. Similarly, the equation $(x-1)^2 = 0$ has two coincident roots, both equal to 1 (see Remark 1).

Let us now consider an equation which is not quadratic, but whose solutions can be obtained from related quadratic equations.

Example 4: Solve $x = \sqrt{15-2x}$.

Solution: $x = \sqrt{15-2x}$ is not a polynomial equation. But, if we square both sides, we obtain the polynomial equation $x^2 = 15 - 2x$.
Now, any root of $x = \sqrt{15-2x}$ is also a root of the equation $x^2 = 15 - 2x$.
(But the converse need not be true, since $x^2 = 15 - 2x$ can also mean $x = -\sqrt{15-2x}$.) So we will obtain both the roots of $x^2 = 15 - 2x$, and see which of these satisfy $x = \sqrt{15-2x}$.
Now, the roots of the quadratic equation $x^2 = 15 - 2x$ are $x = -5$ and $x = 3$.
We must put these values in the given equation to see if they satisfy it.
Now, for $x = -5$, $x - \sqrt{15-2x} = (-5) - \sqrt{15+10} = (-5) - 5 = -10 \neq 0$.
So $x = -5$ is not a solution of the given equation. But it is a solution of $x^2 = 15 - 2x$. We call such a root **an extraneous solution**.
Next, what happens when we put $x = 3$ in the given equation? We get $3 = \sqrt{15-6}$, i.e., $3 = 3$, which is true. Thus, $x = 3$ is the solution of the given equation.

***

Using what we have said so far, try and solve the following exercises.

E3) A quadratic equation over $\mathbb{R}$ can have complex roots while a linear equation over $\mathbb{R}$ can only have a real root. True or false? Why?

E4) Solve the following equations:

i) $x^2 + 5 = 0$,

ii) $(x+9)\,(x-1) = 0$,

iii) $x^2 - \sqrt{5}x = 1$.

E5) For which values of k will the equation $kx^2 + (2k+6)x + 16 = 0$ have coincident roots?

E6) If α and β are roots of $ax^2+bx+c=0$, then show that $\alpha+\beta=-\frac{b}{a}$ and $\alpha\beta=\frac{c}{a}$.

E7) Let $\alpha, \beta \in \mathbb{C}$ such that $\alpha+\beta=p\in\mathbb{R}$ and $\alpha\beta=q\in\mathbb{R}$. Show that α and β are the roots of $x^2-px+q=0$.

E8) Reduce $\sqrt{2x+3}-\sqrt{x+1}=1$ to a quadratic equation, and hence, solve it.

E9) Ameena walks 1 km per hour faster than Alka. Both walked from their village to the nearest library, a distance of 24 km. Alka took 2 hours more than Ameena. What was Alka's average speed?

Did you notice that E7 is the converse of E6? In fact, you will see this relationship clearly in the next section. In this section, our aim was to help you recall the methods of solving linear and quadratic equations. Let us now discuss polynomial equations in general.

5.3 POLYNOMIAL EQUATIONS

You have already studied linear and quadratic polynomials in one variable with coefficients in $\mathbb{R}$. You have also seen expressions like $2x^3+5x^2$, or $\frac{1}{2}x^4+\frac{1}{9}x^3+\pi x^2+\sqrt{2}$. These are examples of what we shall now define.

Definition: An expression of the form $a_0x^0+a_1x^1+a_2x^2+\cdots+a_nx^n$, where $n\in\mathbb{N}$ and $a_i\in\mathbb{R}\ \forall\ i=1, \ldots, n$, is called a **polynomial** over $\mathbb{R}$ in the variable x. $a_0, a_1, \ldots, a_n$ are the **coefficients** of the polynomial.
Further, if $a_n\neq 0$, we say that the **degree of the polynomial** is n and the **leading term** is a_nx^n.

We usually denote polynomials in x by $f(x), g(x)$, etc. If the variable x is understood, then we often only write f instead of $f(x)$. We denote the degree of a polynomial $f(x)$ by **deg f(x)**, or **deg f**.

While discussing polynomials, we will observe the following conventions.

Conventions: We will

i) write x^0 as 1, so that we will write a_0 for a_0x^0;

ii) write x^1 as x, so that a_1x^1 is a_1x;

iii) write x^m instead of $1.x^m$ (i.e., when $a_m=1$);

iv) omit terms of the type $0.x^m$;

v) **define the degree of the zero polynomial, 0, to be $-\infty$.**

Thus, the polynomial $2+3x^2-x^3$ is actually $2x^0+0.x^1+3x^2+(-1)x^3$.

Note that the degree of $f(x)$ is the highest power of x occurring in $f(x)$. For example,

Any non-zero element of $\mathbb{R}$ is a polynomial of degree 0 over $\mathbb{R}$.

i) $3x + 6x^2 + \frac{5}{2}x^3$ is a polynomial of degree 3,

ii) x^5 is a polynomial of degree 5, and

iii) 2 is a polynomial of degree 0, since $2 = 2x^0$.

You will often come across polynomials of degree 3 and 4. They have special names.

Definition: i) A polynomial of degree 3 is called a **cubic polynomial.**

ii) A polynomial of degree 4 is called a **quartic polynomial** (or **biquadratic polynomial**).

The degree of a polynomial has some properties, which we shall now state.

Theorem 3: If $f(x)$ and $g(x)$ are two polynomials over $\mathbb{R}$,

i) $f(x) \pm g(x)$ is a polynomial over $\mathbb{R}$, and
$\deg(f(x) \pm g(x)) \le \max(\deg f(x), \deg g(x))$

ii) $f(x)g(x)$ is a polynomial over $\mathbb{R}$, and
$\deg(f(x).g(x)) = \deg f(x) + \deg g(x)$.

Throughout this unit, we have been talking of polynomials over $\mathbb{R}$. In the same vein, we say that $f(x)$ is **a polynomial over** $\mathbb{C}$ if its coefficients are complex numbers, and **$f(x)$ is over** $\mathbb{Q}$ if its coefficients are rational numbers. Of course, every polynomial over $\mathbb{R}$ is a polynomial over $\mathbb{C}$. For example, $2x + 3$ and $x^2 + 3$ are polynomials over $\mathbb{Q}$ (as well as $\mathbb{R}$ and $\mathbb{C}$). On the other hand, $\sqrt{3}$ is a polynomial over $\mathbb{R}$ but not over $\mathbb{Q}$. In this course **we shall almost always be dealing with polynomials over $\mathbb{R}$.**

Let us now define a related term.

Definition: If we put a polynomial of degree n equal to zero, we get a **polynomial equation of degree n**, or an nth degree equation.

For example,

i) $2x + 3 = 0$ is a polynomial equation of degree 1, i.e., a linear equation.

ii) $3x^2 + \sqrt{2}x - 1 = 0$ is a polynomial equation of degree 2, i.e., a quadratic equation.

iii) $1 + \sqrt{5}x^3 = 0$ is a polynomial equation of degree 3, i.e., a cubic equation.

However, $(\sin x)^4 + 1 = 0$ and $x^3 + \sqrt{x} = 0$ are not polynomial equations since they cannot be written in the form $f(x) = 0$, where f is a polynomial.

Now, if $f(x) = a_0 + a_1x + \cdots + a_nx^n$ is a polynomial, and $a \in \mathbb{C}$, we can substitute x by a to get $f(a)$, **the value of the polynomial at $x = a$**. Thus, $f(a) = a_0 + a_1a + a_2a^2 + \cdots + a_na^n$.

For example, if $f(x) = 2x + 3$, then $f(1) = 2.1 + 3 = 5$, $f(i) = 2i + 3$, and $f\left(\frac{-3}{2}\right) = 2\left(\frac{-3}{2}\right) + 3 = 0$.

Since $f\left(\frac{-3}{2}\right) = 0$, you know that $\frac{-3}{2}$ is a root of $f(x)$.

Now, as you have seen in the case of quadratic equations, a polynomial can have several roots. Here are some related definitions.

Definitions: i) Let $f(x)$ be a non-zero polynomial over $\mathbb{R}$. $\alpha \in \mathbb{C}$ is called a **root** (or a **zero**) of $f(x)$ if $f(\alpha) = 0$. In this case we also say that α is **a solution** (or a **root**) of the equation $f(x) = 0$.

ii) The set of solutions of an equation is called its **solution set**.

So, for example, the solution set of $x^2 + 1 = 0$ is $\{i, -i\}$, and of $x^3 - 3x = 0$ is $\{0, \sqrt{3}, -\sqrt{3}\}$.

There are several situations in which one needs to solve cubic and quartic equations. For example, many problems in the social, physical and biological sciences reduce to obtaining the eigenvalues of a square matrix of order 3 or 4 (which you can study about in the Linear Algebra course). And for this you need to know how to obtain the solutions of such equations.

For obtaining solutions of a polynomial equation, we need some results about the roots of polynomial equations. You have already seen, and used, them in the context of linear and quadratic equations. We will present them here, without proof.

Theorem 4: The polynomial equation of degree n, $a_0 + a_1x + \cdots + a_nx^n = 0$, where $a_0, a_1, \ldots, a_n \in \mathbb{R}$ and $a_n \neq 0$, has n roots in $\mathbb{C}$. Further, if $x_1, \ldots, x_n$ are the n roots of the equation, then
$a_0 + a_1x + \cdots + a_nx^n = a_n(x - x_1)\,(x - x_2)\ldots(x - x_n)$.
(Note that the roots need not be distinct. For example, $1 + 2x + x^2 = (x+1)^2$.)

Though we will not prove this theorem here, we will now state a very important result which is used in the proof. This is analogous to the division algorithm for integers, that you studied in school.

Theorem 5 (Division algorithm): Given polynomials $f(x)$ and $g(x)$ over $\mathbb{R}$, with $g(x) \neq 0$, $\exists$ unique polynomials $q(x)$ and $r(x)$ over $\mathbb{R}$ such that $f(x) = g(x)q(x) + r(x)$ and $\deg r(x) < \deg g(x)$.

Now, if you go back to Remark 2, you find that if $b^2 - 4ac < 0$, then $ax^2 + bx + c = 0$ has complex roots $\alpha + i\beta$ and $\alpha - i\beta$, that is, they **occur in conjugate pairs**. This is not only true for quadratic equations, as you can see from the following theorem.

Theorem 6: If a polynomial equation over $\mathbb{R}$ has complex roots, they occur in conjugate pairs. In fact, if $a + ib \in \mathbb{C}$ is a root, then $a - ib$ is also a root.

Proof: Let $a_0 + a_1x + \cdots + a_nx^n = 0$ be a polynomial equation over $\mathbb{R}$, and

$a+ib$ be a root of this polynomial. Then, by definition,
$a_0 + a_1(a+ib) + \cdots + a_n(a+ib)^n = 0$.
Taking conjugates on both sides of the equation, we get
$a_0 + a_1(a-ib) + \cdots + a_n(a-ib)^n = 0$.
This implies $a-ib$ is a root of the polynomial too. ■

Here is an important point related to this theorem.

Remark 4: Note that Theorem 6 **does not say** that $f(x)=0$ must have a complex root. It only says that **if it has a complex root**, then the conjugate of the root is also a root. For instance, $x^2-1=0$ has no complex roots, but $x^2+1=0$ has two complex roots, i and its conjugate, $-i$.

So, what do Theorems 4 and 6 say in the context of cubic equations?
Consider the general cubic equation over $\mathbb{R}$,
$ax^3+bx^2+cx+d=0, a \neq 0$.
Theorem 4 says that this equation has 3 roots in $\mathbb{C}$. Theorem 6 says that either all 3 roots are real or one is real and two are complex.

Next, what do Theorems 4 and 6 say in the context of quartic equations? Theorem 4 tells us that a quartic has 4 roots, which may be real or complex. By Theorem 6, the possibilities are

i) all the roots are real, or

ii) two are real and two are complex conjugates of each other, or

iii) the roots are two pairs of complex conjugates, that is, $a+ib, a-ib, c+id, c-id$ for some $a, b, c, d \in \mathbb{R}, b, d \neq 0$.

Let us consider an example.

Example 5: Obtain the roots of $2x^4+x^2+1=0$.

Solution: $2x^4+x^2+1=0$ can be written as $2y^2+y+1=0$, where $y=x^2$.

Then, solving this for y, we get $y=\dfrac{-1\pm i\sqrt{7}}{4}$, that is, $x^2=\dfrac{-1\pm i\sqrt{7}}{4}$, two polynomials over $\mathbb{C}$. By applying De Moivre's theorem, you can see the roots of $x^2=\dfrac{-1+i\sqrt{7}}{4}$ and $x^2=\dfrac{-1-i\sqrt{7}}{4}$ are

$z_1 = 2^{-1/4}\left(\cos\frac{\theta}{2} - i\sin\frac{\theta}{2}\right)$, $z_2 = 2^{-1/4}\left(-\cos\frac{\theta}{2} + i\sin\frac{\theta}{2}\right)$ and

$z_3 = 2^{1/4}\left(\cos\frac{\theta}{2} + i\sin\frac{\theta}{2}\right)$, $z_4 = 2^{-1/4}\left(-\cos\frac{\theta}{2} - i\sin\frac{\theta}{2}\right)$ respectively, where

$\theta = \tan^{-1}\sqrt{7}$.
Note that z_1 and z_3 are conjugate pairs, as are z_2 and z_4.

***

You can now try some related exercises.

E10) Give an example, with justification, of an equation over $\mathbb{R}$ which **is not** a polynomial equation.

E11) Obtain the solution set of the cubic equation $2x^3 - \sqrt{5}x^2 = 0$.

E12) Obtain the solution set of

i) $x^7 - 1 = 0$;

ii) $5 = x^7$.

E13) Find all the roots of $x^5 + 4x^3 = 5x$.

So far you have been obtaining solutions of equations, just by inspecting them, or using the quadratic formula, or De Moivre's theorem. Let us look at some rules that help us locate solutions.

5.4 RELATIONS BETWEEN ROOTS AND COEFFICIENTS

In this section we shall first look at what E6 and E7 say. Over there, we saw how closely the roots of a quadratic equation are linked with its coefficients. In fact, the same is true for a cubic equation. For showing this we first need a definition.

Definition: Two polynomials $a_0 + a_1x + \cdots + a_nx^n$ and $b_0 + b_1x \ldots + b_mx^m$ are called **equal** if $n = m$ and $a_i = b_i \; \forall \; i = 0, 1, \ldots, n$.

Thus, two polynomials are equal if they have the same degree and their corresponding coefficients are equal. Thus, $2x^3 + 3 = ax^3 + bx^2 + cx + d$ iff $a = 2, b = 0, c = 0, d = 3$.

Why don't you try and prove the relationship that we give in the following exercise, using this definition?

E14) Show that α, β and γ are the roots of the cubic equation $ax^3 + bx^2 + cx + d = 0, a \neq 0$, if and only if

i) $\alpha + \beta + \gamma = -\dfrac{b}{a}$;

ii) $\alpha\beta + \beta\gamma + \alpha\gamma = \dfrac{c}{a}$;

iii) $\alpha\beta\gamma = -\dfrac{d}{a}$.

(**Hint:** Note that, by Theorem 4, the given cubic equation is equivalent to $a(x - \alpha)(x - \beta)(x - \gamma) = 0$.)

What E14 tells us is that for a cubic equation

i) sum of the roots $= -\dfrac{\text{coefficient of } x^2}{\text{coefficient of } x^3}$

ii) sum of the product of the roots taken two at a time $= \dfrac{\text{coeff. of } x}{\text{coeff. of } x^3}$

iii) product of the roots $= -\dfrac{\text{coeff. of } x^0}{\text{coeff. of } x^3} = -\dfrac{\text{constant term}}{\text{coeff. of } x^3}$.

'coeff' is an abbreviation of 'coefficient'.

Let us use these three relations now to create equations with 'designed' roots!

Example 6: If α, β, γ are the roots of the equation $x^3 - 7x^2 + x - 5 = 0$, find the equation whose roots are $\alpha+\beta, \beta+\gamma, \alpha+\gamma$.

Solution: By the relations in E14, we know that

$$\left.\begin{aligned} &\alpha+\beta+\gamma = 7 \\ &\alpha\beta+\beta\gamma+\alpha\gamma = 1 \\ &\alpha\beta\gamma = 5 \end{aligned}\right] \qquad \ldots(1)$$

Therefore, $(\alpha+\beta)+(\beta+\gamma)+(\alpha+\gamma) = 2(\alpha+\beta+\gamma) = 14$ …(2)

Also, $\alpha+\beta = 7-\gamma, (\beta+\gamma) = 7-\alpha, \gamma+\alpha = 7-\beta$, so that

$(\alpha+\beta)(\beta+\gamma)+(\beta+\gamma)(\alpha+\gamma)+(\alpha+\gamma)(\alpha+\beta)$

$= \{49 - 7(\gamma+\alpha)+\gamma\alpha\} + \{49 - 7(\alpha+\beta)+\alpha\beta\} + \{49 - 7(\beta+\gamma)+\beta\gamma\}$

$= 147 - 98 + 1$, using (1) and (2).

$= 50$, and …(3)

$(\alpha+\beta)(\beta+\gamma)(\alpha+\gamma) = (7-\gamma)(7-\beta)(7-\alpha)$ …(4)

To evaluate the expression on the right hand side of (4), we can use (1) or we can use the fact that

$x^3 - 7x^2 + x - 5 = (x-\alpha)(x-\beta)(x-\gamma)$. So, putting $x = 7$, we get

$7^3 - 7.7^2 + 7 - 5 = (7-\alpha)(7-\beta)(7-\gamma)$, i.e., $(7-\alpha)(7-\beta)(7-\gamma) = 2$

Therefore, (4) gives $(\alpha+\beta)(\beta+\gamma)(\alpha+\gamma) = 2$. …(5)

Now, E14, (2), (3) and (5) give us the required equation, which is $x^3 - 14x^2 + 50x - 2 = 0$.

***

Why don't you try the following exercises now?

E15) Find the sum of the cubes of the roots of the equation $x^3 - 6x^2 + 11x - 6 = 0$. Hence find the sum of the fourth powers of the roots.

E16) Let $a, b, c \in \mathbb{R}, S_1 = a+b+c, S_2 = ab+bc+ca, S_3 = abc$. Show that a, b, c are positive if and only if S_1, S_2, S_3 are positive.

E17) Find $a \in \mathbb{R}$ if you know that $x^3 - 3x^2 + ax - 1 = 0$ has three positive solutions. Also solve the equation.

Now, just as in the case of quadratic and cubic equations, if r_1, r_2, r_3, r_4 are the roots of the quartic $ax^4 + bx^3 + cx^2 + dx + e = 0$, then we can find their relationship with the coefficients of the equation.
Let us see what they are. We know that

$ax^4 + bx^3 + cx^2 + dx + e = a(x-r_1)(x-r_2)(x-r_3)(x-r_4)$

$\Leftrightarrow x^4 + \frac{b}{a}x^3 + \frac{c}{a}x^2 + \frac{d}{a}x + \frac{e}{a} = (x-r_1)(x-r_2)(x-r_3)(x-r_4)$

$= x^4 - (r_1+r_2+r_3+r_4)x^3 + (r_1r_2+r_1r_3+r_1r_4+r_2r_3+r_2r_4+r_3r_4)x^2$

$-(r_1r_2r_3+r_1r_2r_4+r_1r_3r_4+r_2r_3r_4)x + r_1r_2r_3r_4$.

Comparing the coefficients of x^3, x^2, x^1 and x^0, we see that

$r_1 + r_2 + r_3 + r_4 = -\frac{b}{a}$,

$r_1 r_2 + r_1 r_3 + r_1 r_4 + r_2 r_3 + r_2 r_4 + r_3 r_4 = \frac{c}{a}$,

$r_1 r_2 r_3 + r_1 r_2 r_4 + r_1 r_3 r_4 + r_2 r_3 r_4 = -\frac{d}{a}$,

$r_1 r_2 r_3 r_3 = \frac{c}{a}$.

This means that

sum of the roots $= -\frac{\text{coeff. of } x^3}{\text{coeff. of } x^4}$,

sum of the roots taken two at a time $= \frac{\text{coeff. of } x^2}{\text{coeff. of } x^4}$,

sum of the roots taken three at a time $= -\frac{\text{coeff. of } x}{\text{coeff. of } x^4}$,

product of the roots $= \frac{\text{coeff. of } x^0}{\text{coeff. of } x^4} = \frac{\text{constant term}}{\text{coeff. of } x^4}$.

These four equations are a particular case of the following result that relates the roots of a polynomial equation with its coefficients. It is due to the French mathematician, Viète.

Fig. 1: François Viète (1504-1603)

Theorem 7: Let $\alpha_1, \ldots, \alpha_n$ be the n roots of the equation $a_0 x^n + a_1 x^{n-1} + \cdots + a_n = 0,\ a_i \in \mathbb{R}\ \forall i = 0, 1, \ldots, n,\ a_0 \neq 0$. Then

$$\sum_{i=1}^{n} \alpha_i = -\frac{a_1}{a_0}$$

$$\sum_{\substack{i,\, j=1 \\ i<j}}^{n} \alpha_i \alpha_j = \frac{a_2}{a_0}$$

$\vdots$

$$\sum_{i_1 < i_2 < \ldots < i_t}^{n} \alpha_{i_1} \alpha_{i_2} \ldots \alpha_{i_t} = (-1)^t \frac{a_t}{a_n}$$

$\vdots$

$$\prod_{i=1}^{n} \alpha_i = (-1)^n \frac{a_n}{a_0}.$$

$\sum_{i=1}^{n} A_i = A_1 + A_2 + \cdots + A_n$

$\prod_{i=1}^{n} A_i = A_1 A_2 \cdots A_n$

In E6, E7 and E14 you have already seen that this result is true for $n = 2$ and 3. Theorem 7 is very useful in several ways. Let us consider an application in the case $n = 4$.

Example 7: If the sum of two roots of the equation $4x^4 - 24x^3 + 31x^2 + 6x - 8 = 0$ is zero, find all the roots of the equation.

Solution: Let the roots be a, b, c, d, where $a + b = 0$.

Then $a + b + c + d = \frac{24}{4} = 6$.

$\therefore\ c + d = 6$ …(6)

Also, $ab + ac + ad + bc + bd + cd = (a+b)\,(c+d) + ab + cd = \frac{31}{4}$

$\therefore\ ab + cd = \dfrac{31}{4}$ …(7)

Further, $(a+b)cd + ab(c+d) = acd + bcd + abc + abd = \dfrac{-6}{4} = -\dfrac{3}{2}$

$\therefore\ (6) \Rightarrow ab = -\dfrac{1}{4}$ …(8)

Finally, $abcd = \dfrac{-8}{4} = -2$

$\therefore\ (8) \Rightarrow cd = 8$ …(9)

Now, E7, (6) and (9) tell us that c and d are roots of $x^2 - 6x + 8 = 0$. Thus, by the quadratic formula, $c = 2,\ d = 4$.

Similarly, you should check that a and b are roots of $x^2 - \dfrac{1}{4} = 0$.

$\therefore a = \dfrac{1}{2},\ b = -\dfrac{1}{2}$.

Thus, the roots of the given quartic are $\dfrac{1}{2}, -\dfrac{1}{2}, 2, 4$.

***

Try the following problems now.

E18) Solve the equation $x^4 + 15x^3 + 70x^2 + 120x + 64 = 0$ given that the roots are in G.P., i.e., geometrical progression. (**Hint:** If four numbers a, b, c, d are in G.P., then $ad = bc$.)

E19) Show that if the sum of two roots of $x^4 - px^3 + qx^2 - rx + s = 0$ (where $p, q, r, s \in \mathbb{R}$) equals the sum of the other two, then $p^3 - 4pq + 8r = 0$.

Fig. 2: Niels Henrik Abel

We have touched upon relations between roots and coefficients for $n = 2, 3, 4$. But you can apply Theorem 7 for any $n \in \mathbb{N}$. You have seen how these relations can also be used for solving the equations. You may also know that there are formulae, like the quadratic formula, for solving cubic and quartic equations, which are due to Cardano, Ferrari and others. However, in 1824 the Norwegian algebraist, Abel (1802-1829), published a proof of the following result:

> There can be no general formula, expressed in explicit algebraic operations on the coefficients of a polynomial equation, for the roots of the equation, if the degree of the equation is greater than 4.

This result says that polynomial equations of degree > 4 do not have a general algebraic solution. But, there are methods that can give us approximate values of real roots, which you will study in our course on Numerical Analysis. There are, of course, special polynomial equations of degree ≥ 5 that can be solved (as in E13).

Now, given a polynomial equation, is there some way of knowing the type of roots it has without actually solving it? We discuss this in the next section.

5.5 NATURE OF ROOTS

In Sec. 5.3, you saw that a polynomial equation of degree n has n roots, which may be real or complex. Theorem 6 also tells you that the complex roots occur in conjugate pairs, so that the number of such roots must be even. You also know that the discriminant of a quadratic equation tells you what the roots of such an equation are like. So, can we generalise this concept? Let's see.

5.5.1 Discriminant

In the case of a quadratic equation $x^2 + bx + c = 0$, you know that the discriminant is $b^2 - 4c$. Also, if α and β are the two roots of the equation, then $\alpha + \beta = -b, \alpha\beta = c$. Therefore, $(\alpha - \beta)^2 = (\alpha + \beta)^2 - 4\alpha\beta = b^2 - 4c$. Thus, the discriminant $= (\alpha - \beta)^2$, where α and β are the roots of the quadratic equation.

Now consider the general quadratic equation, $ax^2 + bx + c = 0$. Let its roots be α and β. Then its discriminant is $b^2 - 4ac = a^2(\alpha - \beta)^2$.
We generalise this relationship to define the discriminant of any polynomial equation.

The discriminant of a polynomial equation of degree $n \in \mathbb{N}$ is zero iff it has at least two equal roots.

Definition: The **discriminant** of $a_0 + a_1x + a_2x^2 + \cdots + a_nx^n = 0$ is $a_n^{2(n-1)} \prod_{1 \le i < j \le n} (\alpha_i - \alpha_j)^2$, where $\alpha_1, \ldots, \alpha_n$ are the roots of the polynomial equation.

In particular, if we consider the case $n = 3$ and $a_n = 1$, we find that the discriminant of a cubic equation is a complicated expression. This actually arises from Cardano's solution of cubics, which we shall not be doing in this course. However, we give the following:

> The discriminant of the cubic $x^3 + px^2 + qx + r = 0$ is $D = -(27B^2 + 4A^3)$, where $A = q - \frac{p^2}{3}$, $B = \frac{2p^3}{27} - \frac{pq}{3} + r$.

Now, we know that $D = (\alpha - \beta)^2(\beta - \gamma)^2(\alpha - \gamma)^2$, where α, β, γ are the roots of the cubic. As in the case of quadratic equations, does the sign of the discriminant tell us anything about the nature of the roots of the equation? Let us look at the different possibility for the roots α, β and γ of the cubic.

1) If the roots are all real and distinct, then $(\alpha - \beta)^2(\beta - \gamma)^2(\alpha - \gamma)^2 = D > 0$.

2) If only one root is real, say α, then β and γ must be complex conjugates.
$\therefore \beta - \gamma$ is purely imaginary, so that $(\beta - \gamma)^2 < 0$.
Also, $\overline{\alpha - \beta} = \alpha - \gamma$, so that $(\alpha - \beta)\,(\alpha - \gamma) > 0$.
Hence, in this case $D < 0$.

3) Suppose $\alpha = \beta$ and $\gamma \ne \alpha$. Since $\alpha - \beta = 0, D = 0$.

Also, $B \neq 0$. Why? Because if $B = 0$, then $A = 0$ (since $D = 0$).

But $A = 0 \Rightarrow q = \frac{p^2}{3}$, that is $\alpha(\alpha + 2\gamma) = \frac{(2\alpha+\gamma)^2}{3}$,

[over here we have used the relationship between the roots, since $p = -(\alpha+\beta+\gamma) = -(2\alpha+\gamma)$ and $q = \alpha\beta + \beta\gamma + \alpha\gamma = \alpha(\alpha+2\gamma)$]

On simplifying we get $\alpha = \gamma$, a contradiction.

Thus, $B \neq 0$.

So, if exactly two roots of the cubic are equal, then $D = 0$ and $B \neq 0$, and hence, $A \neq 0$.

4) If $\alpha = \beta = \gamma$, then $D = 0, B = 0$, and hence $A = 0$.

Let us summarize the different possibilities for the nature of the roots of a cubic equation now.

Consider the cubic equation $x^3 + px^2 + qx + r = 0,\ p, q, r \in \mathbb{R}$, and let $B = \frac{2p^3}{27} - \frac{pq}{3} + r$ and $A = q - \frac{p^2}{3}$. Then $D = -(27B^2 + 4A^3)$, and

1. all its roots are real and distinct iff $D > 0$.
2. exactly one root is real iff $D < 0$.
3. exactly two roots are equal iff $D = 0$ and $B \neq 0$. Further, in this case all the roots are real.
4. all three roots are equal iff $D = 0$ and $B = 0$.

Let us consider an example of the use of the discriminant of a polynomial equation for analysing the type of its roots.

Example 8: Obtain the discriminant of $x^3 + 2x - 2\sqrt{5} = \sqrt{5}x^2$. Hence, examine the nature of its roots.

Solution: Here $p = -\sqrt{5},\ q = 2,\ r = -2\sqrt{5}$. Therefore, $A = \frac{1}{3},\ B = \frac{-10\sqrt{5}}{27}$ and $D = \frac{-56}{3}$. Since $D < 0$, exactly one root is real, and hence the other two are complex conjugates.

***

You may now like to try the following problems to see if you've understood what we have just discussed.

E20) Under what condition on the coefficients of
$ax^3 + 3bx^2 + 3cx + d = 0,\ a \neq 0$,
will the equation have complex roots?

E21) Will all the roots of $x^3 = 15x + 126$ be real? Why, or why not?

So far we have introduced you to one way of determining the type of roots of a polynomial equation. Let us now look at a method that tells us the signs of the real roots of such an equation.

5.5.2 Descartes' Rule of Signs

Let us begin by taking the polynomial $2x^5+3x^4-\sqrt{2}x^3-\frac{1}{\sqrt{3}}x^2+x+\pi=0$.

Just by looking at it, can you say how many positive and negative real roots it could have? It is possible, due to the following remarkable theorem by Descartes.

Theorem 8 (Descartes' Rule of Signs): The number of positive real roots of a polynomial is bounded by the number of changes of sign in its coefficients.

So, if we apply Theorem 8 to the polynomial equation above, then the signs of the coefficients of $x^5, x^4, x^3, x^2, x^1, x^0$ are $++--++$. Thus, there are only two changes of sign, one from x^4 to x^3, and one from x^2 to x^1. Thus, it can have a maximum of two positive real roots.

Remark 5: Note that we do not say that the polynomial actually has any real roots. We are saying that **if it has real roots**, then at most two of these can be positive. In fact, it may not have any real roots, or positive real roots at all! (See E24.)

Now, you may be wondering if there is a similar rule for negative real roots too. Note that if $\mathbf{a}$ **is a root of a polynomial** $\mathbf{f(x)}$ **over** $\mathbb{R}$**, then** $-\mathbf{a}$ **is a root of** $\mathbf{f(-x)}$. For example, if $p(x)=x^2-3x+2$, then
$p(-x)=(-x)^2-3(-x)+2=x^2+3x+2$. Now 1 and 2 are the roots of $p(x)$, so that -1 and -2 are the roots of $p(-x)$.
So, using this relationship, we can write down the following corollary to Theorem 8.

Corollary 1: The maximum number of negative roots that the polynomial $f(x)$, over $\mathbb{R}$, can have is the number of changes of sign of the coefficients of $f(-x)$.

A **corollary to a theorem** is a statement that immediately follows from the theorem.

Let us consider an example of the use of Descartes' Rule.

Example 9: Find the nature of the roots of the equation $3x^4+12x^2+5x-4=0$.

Solution: Let $f(x)=3x^4+12x^2+5x-4$. Using Descartes' Rule, we can see that $f(x)$ has only one change of sign. Hence, it can have at most one positive root.
Now consider $f(-x)=3x^4+12x^2-5x-4$. This too has only one change of sign. Thus, $f(x)$ has at most one negative root.
We also know that $f(x)$ must have 4 roots in $\mathbb{C}$. Thus, at least two of these roots must be non-real, and hence complex conjugates.

***

Why don't you try some exercises now?

E22) Find the nature of the roots of x^2+3x, using Descartes' Rule. Also check this by actually obtaining the roots.

E23) What is the possible nature of the roots of the equation
$x^{10} - 4x^6 + x^4 - 2x - 3 = 0$?

E24) Consider $p(x) = x^2 - 2x + 3$.

i) By Descartes' Rule, how many positive roots can $p(x)$ have?

ii) Obtain the roots of $p(x)$.

What can you infer from (i) and (ii)?

With this discussion on trying to understand the character of the roots of a polynomial, we end our discussion on polynomial equations. Let us briefly summarise the discussion in the unit.

5.6 SUMMARY

In this unit, we have covered the following points.

1. A quick recall of linear and quadratic equations, their solutions and what the discriminant of a quadratic equation tells us about the nature of the roots of the equation.

2. A polynomial over $\mathbb{R}$ in the variable x is an expression of the form $a_0 + a_1x + \cdots + a_nx^n$, $n \in \mathbb{N}$, $a_i \in \mathbb{R}$, $a_n \neq 0$. Its degree is n. The corresponding polynomial equation is $a_0 + a_1x + \cdots + a_nx^n = 0$, which is of degree n.

3. $\deg 0 = -\infty$.

4. A polynomial of degree n, $f(x)$, has n roots in $\mathbb{C}$. $a \in \mathbb{C}$ is a root of $f(x)$ if and only if $f(a) = 0$ if and only if $(x-a) \mid f(x)$.

5. If $f(x)$ and $g(x)$ are polynomials over $\mathbb{R}$, then so are $f(x) \pm g(x)$ and $f(x)g(x)$.
Further, $\deg\,(f(x) \pm g(x)) \leq \max(\deg f(x), \deg g(x))$, and
$\deg(f(x)g(x)) = \deg f(x) + \deg g(x)$.

6. A polynomial equation of degree 3 is called a cubic equation, and of degree 4 is called a quartic (or biquadratic) equation.

7. The division algorithm states that given polynomials $f(x)$ and $g(x)$ over $\mathbb{R}$, with $g(x) \neq 0$, $\exists$ unique polynomials $q(x)$ and $r(x)$ over $\mathbb{R}$, such that $f(x) = g(x)q(x) + r(x)$, with $\deg r(x) < \deg g(x)$.

8. If a polynomial equation has complex roots, then they occur in conjugate pairs.

9. Applications of the following theorem:
Let $\alpha_1, \ldots, \alpha_n$ be the n roots of the equation
$a_0x^n + a_1x^{n-1} + \cdots + a_n = 0$, $a_i \in \mathbb{R}\, \forall\, i = 0, 1, \ldots, n$, $a_0 \neq 0$. Then
$$\sum_{i=1}^{n} \alpha_i = -\frac{a_1}{a_0}$$

$$\sum_{\substack{i,\, j=1 \\ i<j}}^{n} \alpha_i \alpha_j = \frac{a_2}{a_0}$$

$\vdots$

$$\sum_{i_1<i_2<\ldots<i_t}^{n} \alpha_{i_1} \alpha_{i_2} \ldots \alpha_{i_t} = (-1)^t \frac{a_t}{a_n}$$

$\vdots$

$$\prod_{i=1}^{n} \alpha_i = (-1)^n \frac{a_n}{a_0}$$

10. If the polynomial equation $a_0 + a_1 x + \cdots + a_n x^n = 0,\ a_n \neq 0$, has roots $\alpha_1, \alpha_2, \ldots, \alpha_n$, then its discriminant is $D = a_n^{2(n-1)} \prod_{i<j} (\alpha_i - \alpha_j)^2$.

    We also discussed the way D can be used for understanding the nature of the roots of the equation, particularly in the case of cubics.

11. If $f(x)$ is a polynomial over $\mathbb{R}$, Descartes' Rule of Signs tells us that the maximum number of positive roots $f(x)=0$ can have is the number of changes of sign in its coefficients. Applying this, we also obtained the bound on the number of negative roots. Hence we tried to gauge the nature of all the roots of $f(x)=0$.

5.7 SOLUTIONS/ANSWERS

E1) i) $J\left(\dfrac{x}{k} + a\right) = x \Leftrightarrow (J-k)x = -Jak \Leftrightarrow x = \dfrac{-Jak}{J-k}$,

which is well-defined since $J \neq k$.

ii) $\dfrac{1}{R} = \dfrac{1}{r_1} + \dfrac{1}{r_2} = \dfrac{r_1 + r_2}{r_1 r_2}$

Hence, $R = \dfrac{r_1 r_2}{r_1 + r_2}$.

iii) $F = \dfrac{9}{5}C + 32$

E2) 20 min $= \dfrac{1}{3}$ hr.

Let the student's rate to the study centre be x km per hour.

Then the distance travelled by her is $\dfrac{x}{3}$ km.

Similarly, since the rate of travel back is $(x-8)$ km per hour, the same distance, i.e., from the study centre to her home, is $\dfrac{1}{2}(x-8)$ km.

So $\dfrac{x}{3} = \dfrac{1}{2}(x-8)$.

Solving this for x, gives $x = 24$.

Thus, the student lives at a distance of $\dfrac{24}{3} = 8$ km.

E3) The statement is true.
Consider a linear equation, say $ax + b = 0$.
Here $a, b \in \mathbb{R}, a \neq 0$. Its solution is $-\frac{b}{a}$, which is in $\mathbb{R}$.
Now, consider the quadratic equation $x^2 + 1 = 0$. This has complex roots $i, -i$. Thus, a quadratic equation over $\mathbb{R}$ can have complex roots.

E4) i) $x^2 + 5 = 0 \Rightarrow x^2 = -5 = 5i^2$. Thus, $x = \pm i\sqrt{5}$.

ii) $(x+9)\ (x-1) = 0 \Rightarrow x + 9 = 0$ or $x - 1 = 0 \Rightarrow x = -9, 1$.

iii) $x^2 - \sqrt{5}x = 1 \Rightarrow x^2 - \sqrt{5}x - 1 = 0$, in standard form. Applying the quadratic formula, we get $x = \frac{\sqrt{5} \pm \sqrt{5+4}}{2} = \frac{\sqrt{5}+3}{2}, \frac{\sqrt{5}-3}{2}$.

E5) The roots will be coincident, that is, both will be equal, iff the discriminant is zero, i.e., $(2k+6)^2 - 64k = 0$
$\Rightarrow\ 4k^2 - 40k + 36 = 0 \Rightarrow\ k^2 - 10k + 9 = 0$.
By the quadratic formula, we get
$k = \frac{10 \pm \sqrt{100-36}}{2} = 5 \pm 4 = 9, 1$, which are the required values.

E6) By Remark 1,
$ax^2 + bx + c = a(x-\alpha)\ (x-\beta) = a\{x^2 - (\alpha+\beta)x + \alpha\beta\}$.
Equating the coefficient of x, and the constant term, on both sides of the equality, we get
$b = -a(\alpha+\beta)$ and $c = a\alpha\beta$,
that is, $\alpha + \beta = -\frac{b}{a},\ \alpha\beta = \frac{c}{a}$.

E7) α is a root of $x^2 - (\alpha+\beta)x + \alpha\beta = 0$ iff $\alpha^2 - (\alpha+\beta)\alpha + \alpha\beta = 0$, which is true. Hence, α is a root of the given equation.
Similarly, β is a root of the given equation.

E8) $\sqrt{2x+3} - \sqrt{x+1} = 1$.
Squaring both sides, we get
$(2x+3) + (x+1) - 2\sqrt{(2x+3)\ (x+1)} = 1$
$\Leftrightarrow\ 3x + 3 = 2\sqrt{(2x+3)\ (x+1)}$.
Again, squaring this, we get
$9x^2 + 18x + 9 = 4(2x+3)\ (x+1)$
$\Leftrightarrow\ 9x^2 + 18x + 9 = 8x^2 + 20x + 12$
$\Leftrightarrow\ x^2 - 2x - 3 = 0$.
The roots of this equation are 3 and -1.
Now, putting these values in the original equation, we find that
$\sqrt{2(3)+3} - \sqrt{3+1} = 1$, and $\sqrt{2(-1)+3} - \sqrt{(-1)+1} = 1$.
Thus, both 3 and -1 are solutions of the given equation.

E9) Let Alka's average speed be x km per hour, and suppose she took t

hours to walk 24 km.

Then $24 = xt$, that is, $t = \frac{24}{x}$.

Also, Ameena's speed is $(x+1)$ km per hour, and time taken is $(t-2)$ hrs.

So $24 = (x+1)\,(t-2) = (x+1)\left(\frac{24}{x} - 2\right)$.

$\Leftrightarrow\ x^2 + x - 12 = 0$

$\Leftrightarrow\ x = 3, -4$.

Since average speed is non-negative, $x = -4$ is not acceptable in the given situation. Therefore, $x = 3$, that is, Alka's average speed is 3km per hour.

E10) There are infinitely many examples.
For instance, $1 + x + x^2 + x^3 + \cdots + x^n + \cdots$, containing infinitely many terms, is not a polynomial.

E11) $2x^3 - \sqrt{5}x^2 = 0 \ \Leftrightarrow\ x^2(2x - \sqrt{5}) = 0$. By inspection, we note that two roots are zero. So, we get $x = 0$ or $2x - \sqrt{5} = 0$.

Thus, the solution set is $\left\{0, \frac{\sqrt{5}}{2}\right\}$.

E12) i) The required set is the set of the seventh roots of unity (see Unit 4, Sec. 4.5). Thus, the solution set is
$$\left\{\cos\frac{2k\pi}{7} + i\sin\frac{2k\pi}{7},\ k = 0, \pm 1, \pm 2, \pm 3\right\}.$$

ii) Again, from Sec. 4.5, you know that if $x^7 = 5$ has one root α, then the solution set is $\left\{\alpha\left(\cos\frac{2k\pi}{7} + i\sin\frac{2k\pi}{7}\right),\ k = 0, \pm 1, \pm 2, \pm 3\right\}$.

Here we choose α to be the real number $(5)^{1/7}$.

E13) $x^5 + 4x^3 = 5x$

$\Leftrightarrow\ x(x^4 + 4x^2 - 5) = 0$.

By inspection, we see that one root is 0, and the other roots are those of $x^4 + 4x^2 - 5 = 0$.

Now, to solve $x^4 + 4x^2 - 5 = 0$, put $x^2 = t$. Then the equation is $t^2 + 4t - 5 = 0$.

Its roots are $t = 1, -5$.

Thus, $x^2 = 1$ and $x^2 = -5$ give us the roots of the given equation.

Now, $x^2 = 1 \Rightarrow x = \pm 1$.

Also, $x^2 = -5 \Rightarrow x = \pm i\sqrt{5}$.

So the 5 roots of the given equation are $0, \pm 1, \pm i\sqrt{5}$.

E14) Using Theorem 4, we get
$$ax^3 + bx^2 + cx + d = a(x-\alpha)\,(x-\beta)\,(x-\gamma)$$
$$= a\{x^3 - (\alpha+\beta+\gamma)x^2 + (\alpha\beta + \beta\gamma + \alpha\gamma)x - \alpha\beta\gamma\}$$

Now, equating the coefficients of x^2, x^1, x^0 on both sides we get (i), (ii) and (iii), respectively.

E15) If α, β, γ are the roots of the equation, we know that

$\alpha+\beta+\gamma=6$, …(10)

$\alpha\beta+\beta\gamma+\alpha\gamma=11$, …(11)

$\alpha\beta\gamma=6$. …(12)

We need to find $\alpha^3+\beta^3+\gamma^3$, and then $\alpha^4+\beta^4+\gamma^4$.

Now, $(\alpha+\beta+\gamma)^2=\alpha^2+\beta^2+\gamma^2+2(\alpha\beta+\beta\gamma+\alpha\gamma)$.

Thus, from (10) and (11) we find that

$\alpha^2+\beta^2+\gamma^2=(6)^2-2(11)=14$ …(13)

Now, since α, β, γ are roots of the given equation, we know that

$\alpha^3-6\alpha^2+11\alpha-6=0$,

$\beta^3-6\beta^2+11\beta-6=0$,

$\gamma^3-6\gamma^2+11\gamma-6=0$.

Adding these three equations, we get

$(\alpha^3+\beta^3+\gamma^3)-6(\alpha^2+\beta^2+\gamma^2)+11(\alpha+\beta+\gamma)-18=0$.

Now, using (10) and (13), we find

$\alpha^3+\beta^3+\gamma^3=6(14)-11(6)+18=36$. …(14)

Now, note that multiplying the given equation by x, we get

$x^4-6x^3+11x^2-6x=0$

As α, β, γ are also roots of this equation, using the procedure above, we get three equations. Adding them, we get

$(\alpha^4+\beta^4+\gamma^4)-6(\alpha^3+\beta^3+\gamma^3)+11(\alpha^2+\beta^2+\gamma^2)-6(\alpha+\beta+\gamma)=0$

So, using (14), (13) and (10), we get

$\alpha^4+\beta^4+\gamma^4=6(36)-11(14)+6(6)=98$.

E16) Firstly, if a, b, c are positive, S_1, S_2 and S_3 have to be positive. (Why?)

Now, we need to prove the converse, that is, if S_1, S_2, S_3 are positive, then a, b, c must be positive. For this, note that a, b, c must be the roots of the cubic equation

$x^3-S_1x^2+S_2x-S_3=0$

So, $a^3-S_1a^2+S_2a-S_3=0$

$\Rightarrow\ a(a^2-S_1a+S_2)=S_3>0$

Thus, $a\neq 0$.

Also, if $a<0$, then $a^2-S_1a+S_2>0$, so that $a(a^2-S_1a+S_2)<0$, which is a contradiction. So, a must be positive.

Similarly, b and c must be positive.

E17) If α, β, γ are the roots of the equation, then

$\alpha+\beta+\gamma=3,\ \alpha\beta+\beta\gamma+\alpha\gamma=a,\ \alpha\beta\gamma=1$.

By inspection, we see that $\alpha=\beta=\gamma=1$ satisfy all the relations.

Then $a=3$.

E18) Let a, b, c, d be the roots in GP, then

$ad=bc$ …(15)

Now, we know that

$a+b+c+d=-15$ …(16)

$ab+ac+ad+bc+bd+cd=70$ …(17)

$(a+b)cd+ab(c+d)=-120$

$\Rightarrow\ ad(b+c)+bc(a+d)=-120$ …(18)

$abcd=64$ …(19)

Now (15), (16) and (18) give us $ad(-15)=-10$, i.e., $ad=8=bc$.
Next, (17) gives us

$(a+d)\ (b+c)+2bc=70$

$\Rightarrow\ \{-15-(b+c)\}\ (b+c)+2bc=70$, using (16)

$\Rightarrow\ (b+c)^2+15(b+c)+54=0$

$\Rightarrow\ b+c=-6$ or $b+c=-9$

We take $b+c=-6$ …(20)

Then $(b-c)^2=(b+c)^2-4bc=4$

$\Rightarrow\ b-c=2$ …(21)

(20) and (21) give us $b=-2,\ c=-4$.
Also, (16) and (20) give us $a+d=-9$.
As above, we find $a=-1,\ d=-8$.
So, the roots, given in order, are $-1,\ -2,\ -4,\ -8$, with the common ratio being 2.
(Note that in (20) we could have taken $b+c=-9$, then in the further calculations you would have got the same roots ultimately.)

E19) Let the roots be a, b, c, d, with

$a+b=c+d$. …(22)

Now $a+b+c+d=p$ …(23)

$ab+ac+ad+bc+bd+cd=q$ …(24)

$(a+b)cd+ab(c+d)=r$ …(25)

$abcd=s$ …(26)

(22) and (23) $\Rightarrow\ a+b=\dfrac{p}{2}=c+d$

Then (25) $\Rightarrow\ ab+cd=\dfrac{2r}{p}$

Also (24) $\Rightarrow\ (a+b)c+(a+b)d+ab+cd=q$

So $\dfrac{p^2}{4}+\dfrac{2r}{p}=q \Rightarrow p^3+8r=4pq$

$\Rightarrow\ p^3-4pq+8r=0$.

E20) The only case will be when $D<0$, i.e.,

$27B^2+4A^3>0$,

where $B=2\dfrac{(3b)^3}{a^3}.\dfrac{1}{27}-\dfrac{9bc}{a^2}.\dfrac{1}{3}+\dfrac{d}{a}=\dfrac{2b^3}{a^3}-\dfrac{3bc}{a^2}+\dfrac{d}{a}$,

and $A=\dfrac{3c}{a}-\dfrac{1}{3}\left(\dfrac{3b}{a}\right)^2=\dfrac{3c}{a}-3\dfrac{b^2}{a^2}$.

Thus, the requirement is

$$27\left(\frac{2b^3-3abc+ad^2}{a^3}\right)^2+4(27)\left(\frac{ac-b^2}{a^2}\right)^3>0$$

$\Rightarrow\ (2b^3-3abc+a^2d)^2+4(ac-b^2)^3>0$.

E21) Here $p = 0,\ q = -15,\ r = -126$.
So, $B = -126,\ A = -15$,
$D = -\{27(-126)^2 + 4(-15)^3\} < 0$
Therefore, only one root will be real, and two will be complex conjugates.

E22) By Descartes' Rule, this has no positive roots, and at most one negative root. So, it will have one root equal to zero and one negative root.
Now, by inspection we see it has two roots, 0 and -3, which matches the nature suggested by Descartes' Rule.

E23) Using Descartes' Rule, we see $f(x)$ can have at most three positive roots, since the signs are
$\underbrace{+\quad -}_{1}\underbrace{\quad +}_{1}\underbrace{\quad -}_{1}\quad -$, i.e., a total of 3 sign changes.
Now, $f(-x) = x^{10} - 4x^6 + x^4 + 2x - 3$.
So, there are two sign changes.
Therefore, the equation can have at most two negative roots.
Now, this equation has 10 roots, and we see that at most 5 can be real. So, at least 5 roots are complex. But, complex roots occur in pairs. So, at least 6 roots are complex. Then 4 roots would be real.
Another possibility is it could have 8 complex roots, and two real roots, one of which is positive and one negative.
A third possibility is that it has 5 pairs of complex roots and no real roots.

E24) i) By the Rule, it can have two positive roots.

ii) The roots are $\dfrac{2 \pm \sqrt{4-12}}{2} = 1 \pm i\sqrt{2}$.

Thus, even though the Rule shows the maximum possibility of real roots, there need not be any real roots at all. Here both the roots are complex.

MISCELLANEOUS EXERCISES

The exercises given below cover the concepts and processes you have studied in this block. Doing them will give you a better understanding of the concepts concerned, as well as practice in solving such problems.

1. If $a, b, c \in \mathbb{R}$ such that $(2ax+b) \mid (ax^2+bx+c)$, what is the nature of the roots of $ax^2+bx+c=0$?

2. Give a geometrical representation of the following sets:
   i) $\{-3, 0.75, \sqrt{2}\} \subseteq \mathbb{R}$,
   ii) $\{x \mid x < -3\} \cup \{3\} \subseteq \mathbb{R}$,
   iii) $\{(x, y) \mid y = \sqrt{3}x - 2\} \subseteq \mathbb{R} \times \mathbb{R}$.

3. Graph the function f, defined by
   i) $f: \mathbb{R} \to \mathbb{R} : f(x) = \begin{cases} 1-x, & \text{if } x \le 1, \\ x^2, & \text{if } x > 1. \end{cases}$
   ii) $f: \mathbb{R} \to \mathbb{R} : f(x) = -|x|$,
   iii) $f: \mathbb{N} \to \mathbb{N} : f(x) = 2^x$.

4. Is $\left(\frac{4-i}{1+i} - \frac{2i}{2+i}\right)4i$ a purely imaginary number? Give reasons for your answer.

5. Solve the equation $\frac{x+1}{x-3} = \frac{4}{x+3} + 6, \ x \neq \pm 3$.

6. If two people lay tiles on the floor of a room, it takes them 4 hours to do the job together. If each works alone, one of them could do the job in 1 hour less than the other. How long would it take each of them to tile the floor alone?

7. Evaluate the following, where ω is a cube root of unity:
   i) $(1-\omega+\omega^2)\,(1+\omega-\omega^2)$,
   ii) $(1-\omega)\,(1-\omega^2)\,(1-\omega^4)\,(1-\omega^5)$.

8. Find the equation whose roots are 4 less in value than the roots of $x^4-5x^3+7x^2-17x+11=0$.
   (**Hint:** You can re-write the equation as an equation in $x-4$.)

9. Solve $x^4+9x^3+16x^2+9x+1=0$.
   (**Hint:** Note that the coefficients of x^r and x^{4-r} are the same in this, for $r = 0, 1, 2$. Also $x=0$ is not a root. Dividing throughout by x^2, you can rewrite the equation as one in $x + \frac{1}{x} = y$, say. Then solve for y.)

10 Find the least possible number of imaginary roots of $x^9 - x^5 + x^4 + x^2 + 1 = 0$.

11. Find the square root of $4ab - 2i(a^2-b^2)$, where $a, b \in \mathbb{R}$.

SOLUTIONS/ANSWERS

1. $ax^2 + bx + c = 0$ has two roots. By the given condition, one root is $\frac{-b}{2a}$.

   Suppose the other root is $r \in \mathbb{R}$. Then

   $a\left(x - \frac{b}{2a}\right)(x - r) = ax^2 + bx + c$.

   On equating the coefficients, we get $r = \frac{-b}{2a}$.

   Thus, both the roots are equal and real.

2. i)

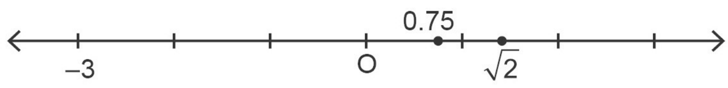


Fig. 1

ii)

Fig. 2: The portion of the number line to the left of -3, together with the point representing 3, is the representation of the set.

iii)

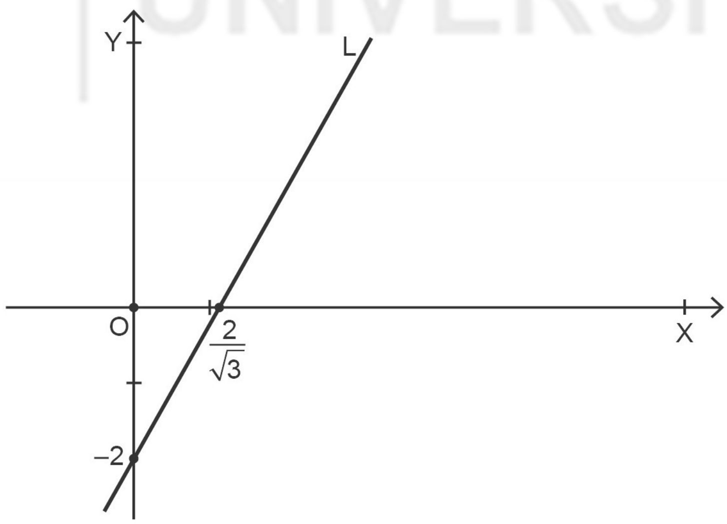
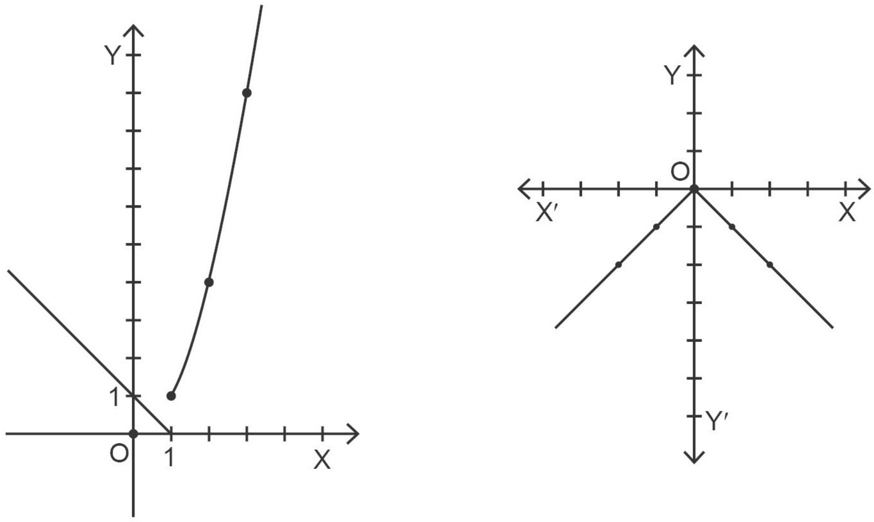


Fig. 3: L represents $y = \sqrt{3}x - 2$.

3. (i) and (ii)



(i) **(ii)**

Fig. 4: (i) $\mathbf{y} = \begin{cases} \mathbf{1 - x}, & \mathbf{x \le 1} \\ \mathbf{x^2}, & \mathbf{x > 1} \end{cases}$, (ii) $\mathbf{y = -|x|}$.

iii)

Fig. 5: $\mathbf{y = 2^x \ \forall x \in \mathbb{N}}$.

4. The number is $4i\left\{\dfrac{(4-i)\ (2+i) - 2i(1+i)}{(1+i)\ (2+i)}\right\} = \dfrac{44i}{1+3i} = \dfrac{44i\,(1-3i)}{10}$.

   Thus, its real part is non-zero, and hence the number is not purely imaginary.

5. $\dfrac{x+1}{x-3} = \dfrac{4}{x+3} + 6$

$\Rightarrow 5x^2 = -69$

$\Rightarrow x = \pm i\sqrt{\frac{69}{5}}$.

6. Suppose the slower worker completes the job alone in x hours.
The faster worker completes the job alone in $(x-1)$ hours.

The rate of job done per hour of each is $\frac{1}{x}$ and $\frac{1}{x-1}$, respectively.

Together they complete the job in 4 hours.

$\therefore 4\left(\frac{1}{x}+\frac{1}{x-1}\right)=1$

$\Rightarrow x^2 - 9x + 4 = 0$

$\Rightarrow x = \frac{9+\sqrt{65}}{2}, \frac{9-\sqrt{65}}{2} = 8.5, 0.5$ (approximately).

$x = 0.5$ is not possible, since $x - 1 = -0.5$, which is negative, and hence cannot be the time taken by a worker.

Thus, $x \approx 8.5$ and $x - 1 \approx 7.5$.

So, the workers would each take 7.5 hours and 8.5 hours, respectively, to tile the floor alone.

7. i) We know that $1+\omega+\omega^2 = 0$ and $\omega^3 = 1$.

$\therefore (1-\omega+\omega^2)\ (1+\omega-\omega^2) = (-2\omega)\ (-2\omega^2) = 4$.

ii) $(1-\omega)\ (1-\omega^2)\ (1-\omega^4)\ (1-\omega^5)$

$= (1-\omega)\ (1-\omega^2)\ (1-\omega)\ (1-\omega^2)$

$= (1-\omega)^2\ (1-\omega^2)^2$

$= (1-2\omega+\omega^2)\ (1-2\omega^2+\omega^4) = 9$.

8. Let $y = x - 4$. Then the given equation becomes

$(y+4)^4 - 5(y+4)^3 + 7(y+4)^2 - 17(y+4) + 11 = 0$

$\Leftrightarrow y^4 + 11y^3 + 43y^2 + 55y - 9 = 0$ the required equation.

9. $x^4 + 9x^3 + 16x^2 + 9x + 1 = 0$

$\Leftrightarrow \left(x^2+\frac{1}{x^2}\right)+9\left(x+\frac{1}{x}\right)+16=0$, dividing throughout by x^2.

$\Leftrightarrow \left(x+\frac{1}{x}\right)^2+9\left(x+\frac{1}{x}\right)+14=0$.

Put $x+\frac{1}{x}=y$, to get $y^2+9y+14=0$.

Then $y = -2, -7$.

So $x+\frac{1}{x}=-2$ and $x+\frac{1}{x}=-7$ give us $x^2+2x+1=0$ and $x^2+7x+1=0$.

Solving these equations, we get the four roots of the original equation, namely, $1, 1, \frac{-7\pm\sqrt{45}}{2}$.

10. Let $f(x) = x^9 - x^5 + x^4 + x^2 + 1$. Then $f(-x) = -x^9 + x^5 + x^4 + x^2 + 1$.
Since $f(x)$ has only two changes of sign, it has at most two positive roots. Since $f(-x)$ has at most one change of sign, $f(x)$ has at most one negative root.
Since $f(x)$ has 9 roots, of which at most 3 are real, at least 6 are complex roots.

11. $4ab - 2i(a^2 - b^2) = 2(a^2 + b^2)(\cos\theta + i\sin\theta)$, where $\theta = \tan^{-1}\left(\frac{b^2 - a^2}{2ab}\right)$.

Therefore, its square roots are
$\sqrt{2}\sqrt{a^2 + b^2}\left(\cos\frac{\theta}{2} + i\sin\frac{\theta}{2}\right)$, $\sqrt{2}\sqrt{a^2 + b^2}\left(-\cos\frac{\theta}{2} - i\sin\frac{\theta}{2}\right)$, where
$\theta = \tan^{-1}\left(\frac{b^2 - a^2}{2ab}\right)$.