Chapter 1
Preliminaries

We introduce a wide range of fundamental mathematical concepts and structures in this chapter on foundation of mathematics. Understanding their fundamental operations and attributes, we start with sets and functions. We then delve into the metric space universe, which offers a framework for comprehending distance and convergence. Moving on to algebraic structures, we examine the distinctive qualities and illustrative instances of groups, rings, and fields. Polynomial rings and their essential properties are introduced, as are matrices and their rank, trace, and determinant, all of which are highlighted as they have vital roles in the coming chapters. The latter sections of the chapter provide an overview of Euclidean space and demonstrate how to solve systems of linear equations using techniques like Cramer's rule, LU decomposition, Gauss elimination, etc. These fundamental ideas in mathematics serve as the building blocks for more complex mathematical research and have numerous applications in science and engineering.

1.1 Sets and Functions

Set theory is the core of modern mathematics and serves as a language for mathematicians to discuss and organize their ideas. It is a crucial and elegant concept at its core; a set is simply a collection of objects, similar to a bag containing multiple objects. These objects can be anything from numbers, characters, shapes, or other sets. The way set theory lets us classify, compare, and evaluate these collections is what makes it so powerful. This section will discuss some of the essential concepts in set theory. Though the notion of set is not well-defined in wide generality as it leads to paradoxes like Russell's Paradox, published by *Bertrand Russell (1872–1970)* in 1901, we start with the following simple definition for a preliminary understanding of a set.

Definition 1.1 (*Set*) A set is a well-defined collection of objects. That is, to define a set X, we must know for sure whether an element x belongs to X or not. If x is

R. K. George and A. Ajayakumar, *A Course in Linear Algebra*, University Texts in the Mathematical Sciences, https://doi.org/10.1007/978-981-99-8680-4_1

an element of X, then it is denoted by $x \in X$ and if x is not an element of X, then it is denoted by $x \notin X$. Two sets X and Y are said to be equal if they have the same elements.

Definition 1.2 (*Subset*) Let X and Y be any two sets, then X is a subset of Y, denoted by $X \subseteq Y$, if every element of X is also an element of Y. Two sets X and Y are equal if and only if $X \subseteq Y$ and $Y \subseteq X$.

A set can be defined in a number of ways. Commonly, a set is defined by either listing all the entries explicitly, called the *Roster form*, or by stating the properties that are meaningful and unambiguous for elements of the set, called the *Set builder form*.

Example 1.1 Here are some familiar collection/sets of numbers.

$\mathbb{N}-$the set of all natural numbers $-\{1, 2, 3, \ldots\}$
$\mathbb{W}-$the set of all whole numbers $-\{0, 1, 2, \ldots\}$
$\mathbb{Z}-$the set of all integers $-\{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$
$\mathbb{Q}-$the set of all rational numbers $-\{\frac{p}{q} \mid p, q \in \mathbb{Z},\ q \neq 0\}$
$\mathbb{R}-$the set of all real numbers
$\mathbb{C}-$the set of all complex numbers.

Usually, in a particular context, we have to deal with the elements and subsets of a basic set which is relevant to that particular context. This basic set is called the "Universal Set" and is denoted by $\mathcal{U}$. For example, while studying the number system, we are interested in the set of natural numbers, $\mathbb{N}$, and its subsets such as the set of all prime numbers, the set of all odd numbers, and so forth. In this case $\mathbb{N}$ is the universal set. A null set, often known as an empty set, is another fundamental object in set theory. It is a set with no elements, which means it has no objects or members. In set notation, the null set is commonly represented by Φ or $\{\}$ (an empty pair of curly braces).

Definition 1.3 (*Cardinality*) The cardinality of a set X is the number of elements in X. A set X can be finite or infinite depending on the number of elements in X. Cardinality of X is denoted by $|X|$.

Example 1.2 All the sets mentioned in Example 1.1 are infinite sets. The set of letters in the English alphabet is a finite set.

Set Operations

Set operations are fundamental mathematical methods for constructing, manipulating, and analyzing sets. They enable the combination, comparison, and modification of sets in order to acquire insights and solve various mathematical and real-life problems. Union (combining items from several sets), intersection (finding common elements between sets), complement (identifying elements not in a set), and set difference (removing elements from one set based on another) are the fundamental set operations.

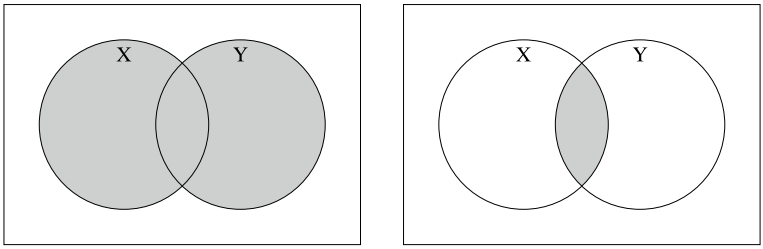


Fig. 1.1 The shaded portions in **a** and **b** represents the union and intersection of the sets X and Y, respectively

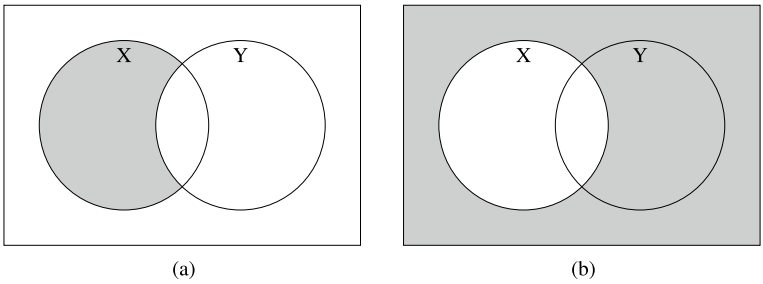


Fig. 1.2 The shaded portion in **a** represents the difference of Y related to X and the shaded portion in **b** represents the complement of a set

Definition 1.4 (*Union and Intersection*) Let X and Y be two sets. The union of X and Y, denoted by $X \cup Y$, is the set of all elements that belong to either X or Y. The intersection of X and Y, denoted by $X \cap Y$, is the set of all elements that belong to both X and Y.

The relationship between sets can be illustrated with the use of diagrams, known as *Venn diagrams*. It was popularized by the famous mathematician *John Venn (1834–1923)*. In a Venn diagram, a rectangle is used to represent the universal set and circles are used to represent its subsets. For example, the union and intersection of two sets are represented in Fig. 1.1.

Definition 1.5 (*Difference of Y related to X*) Let X and Y be two sets. The difference of Y related to X, denoted by $X \setminus Y$, is the set of all elements in X which are not in Y. The difference of a set X related to its universal set $\mathcal{U}$ is called the *complement* of X and is denoted by X^c. That is, $X^c = \mathcal{U} \setminus X$. Keep in mind that $\mathcal{U}^c = \Phi$ and $\Phi^c = \mathcal{U}$ (Fig. 1.2).

Definition 1.6 (*Cartesian Product*) Let X and Y be two sets. The Cartesian product of X and Y, denoted by $X \times Y$, is the set of all ordered pairs (x, y) such that x belongs to X and y belongs to Y. That is, $X \times Y = \{(x, y) \mid x \in X, y \in Y\}$.

Example 1.3 Let $X = \{1, 2, 3\}$ and $Y = \{3, 4, 5\}$. Then the union and intersection of X and Y are $X \cup Y = \{1, 2, 3, 4, 5\}$ and $X \cap Y = \{3\}$, respectively. The difference of Y related to X is $X \setminus Y = \{1, 2\}$, and the Cartesian product of X and Y is $X \times Y = \{(1, 3), (1, 4), (1, 5), (2, 3), (2, 4), (2, 5), (3, 3), (3, 4), (3, 5)\}$.

Remark 1.1 Two sets X and Y are said to be **disjoint**, if their intersection is empty. That is, if $X \cap Y = \Phi$.

We will now try to "connect" elements of distinct sets using the concept, "Relations". A relation between two sets allows for the exploration and quantification of links and relationships between elements of various sets. It essentially acts as a link between elements of another set and elements from another, exposing patterns, dependencies, or correspondences.

Definition 1.7 (*Relation*) A relation R from a non-empty set X to a non-empty set Y is a subset of the Cartesian product $X \times Y$. It is obtained by defining a relationship between the first element and second element (called the *"image"* of first element) of the ordered pairs in $X \times Y$.

The set of all first elements in a relation R is called the domain of the relation R, and the set of all second elements is called the range of R. As we represent sets, a relation may be represented either in the roster form or in the set builder form. In the case of finite sets, a visual representation by an arrow diagram is also possible.

Example 1.4 Consider the sets X and Y from Example 1.3 and their Cartesian product $X \times Y$. Then $R = \{(1, 3), (2, 4), (3, 5)\}$ is a relation between X and Y. The set builder form of the given relation can be given by $R = \{(x, y) \mid y = x + 2, x \in X, y \in Y\}$ (Fig. 1.3).

Remark 1.2 If $|X| = m$ and $|Y| = n$, then $|X \times Y| = mn$ and the number of possible relations from set X to set Y is 2^{mn}.

Definition 1.8 (*Equivalence Relations*) A relation R on a set X is said to be an equivalence relation if and only if the following conditions are satisfied:

(a) $(x, x) \in R$ for all $x \in X$ (*Reflexive*)
(b) $(x, y) \in R$ implies $(y, x) \in R$ (*Symmetric*)
(c) $(x, y) \in R$ and $(y, z) \in R$ implies $(x, z) \in R$ (*Transitive*).

Example 1.5 Consider $\mathbb{N}$ with the relation R, where $(x, y) \in R$ if and only if $x - y$ is divisible by n, where n is a positive integer. We will show that R is an equivalence relation on $\mathbb{N}$. For,

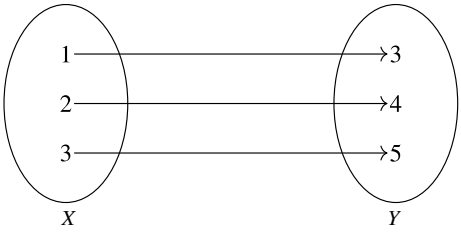


Fig. 1.3 Arrow diagram for R

(a) $(x, x) \in R$ for all $x \in \mathbb{N}$. For, $x - x = 0$ is divisible by n for all $x \in \mathbb{N}$.
(b) $(x, y) \in R$ implies $(y, x) \in R$. For,

$$\begin{aligned}(x, y) \in R &\Rightarrow x - y \text{ is divisible by } n \\ &\Rightarrow -(x - y) \text{ is divisible by } n \\ &\Rightarrow y - x \text{ is divisible by } n \\ &\Rightarrow (y, x) \in R\end{aligned}$$

(c) $(x, y) \in R$ and $(y, z) \in R$ implies $(x, z) \in R$. For,

$$\begin{aligned}(x, y), (y, z) \in R &\Rightarrow x - y \text{ and } y - z \text{ is divisible by } n \\ &\Rightarrow (x - y) + (y - z) \text{ is divisible by } n \\ &\Rightarrow x - z \text{ is divisible by } n \\ &\Rightarrow (x, z) \in R\end{aligned}$$

Thus, R is reflexive, symmetric, and transitive. Hence, R is an equivalence relation.

Example 1.6 Consider the set $X = \{1, 2, 3\}$. Define a relation R on X by $R = \{(1, 1), (2, 2), (1, 2), (2, 3)\}$. Is R an equivalence relation? Clearly, not! We can observe that R is not reflexive as $(3, 3) \notin R$. Also R is not transitive as $(1, 2), (2, 3)$ but $(1, 3) \notin R$. What if we include the elements $(3, 3)$ and $(1, 3)$ to the relation and redefine R as $\tilde{R} = \{(1, 1), (2, 2), (3, 3), (1, 2), (2, 3), (1, 3)\}$. Then $\tilde{R}$ is an equivalence relation on X.

Relations define how elements from one set correspond to elements from another, allowing for a broader range of relationships. However, there are specialized relations in which each element in the first set uniquely relates to one element in the second. This connection gives these relations mathematical precision, making them crucial for modeling precise transformations and dependencies in various mathematical disciplines, ranging from algebra to calculus. We refer to such relations as functions.

Functions

Function in mathematics is a rule or an expression that relates how a quantity (dependent variable) varies with respect to another quantity (independent variable) associated with it. They are ubiquitous in mathematics and they serve many purposes.

Definition 1.9 (*Function*) A function f from a set X to a set Y, denoted by $f : X \to Y$, is a relation that assigns to each element $x \in X$ exactly one element $y \in Y$.

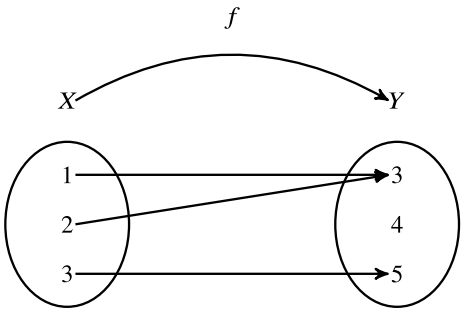
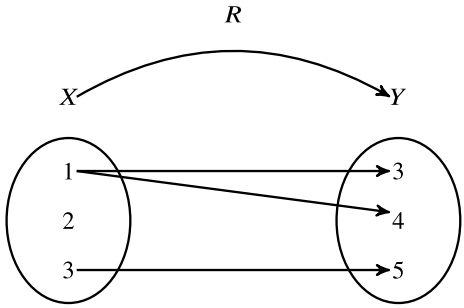


Fig. 1.4 Observe that each element from set X is mapped to exactly one element in set Y. Therefore the given relation is a function. X is called *domain* of f and Y is called the *co-domain* of f. 4 does not belong to the range set of f, as it does not have a pre-image. The range set of f is $\{3, 5\}$

Fig. 1.5 Observe that 1 is mapped to both 3 and 4. Thus $R = \{(1, 3), (1, 4), (2, 3)\}$ is not a function



Then y is called the image of x under f and is denoted by $f(x)$. The set X is called the domain of f and Y is called the co-domain of f. The collection of all images of elements in X is called the range of f.

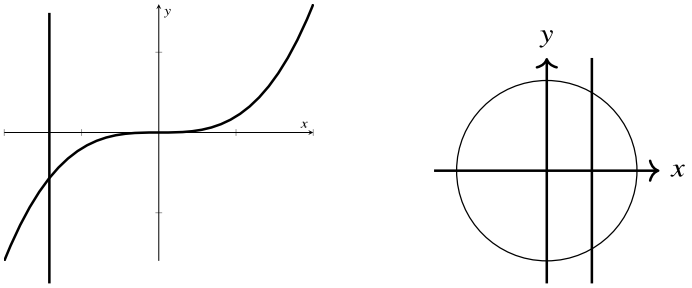
Example 1.7 Consider the sets X and Y from Example 1.3. Define a relation R from the set X to the set Y as $R = \{(1, 3), (2, 3), (3, 5)\}$. Then the relation R is a function from X to Y (Fig. 1.4).

From Definition 1.9, it is clear that any function from a set X to a set Y is a relation from X to Y. But the converse need not be true. Consider the following example.

Example 1.8 Consider the sets X and Y from Example 1.3. Then the relation $R = \{(1, 3), (1, 4), (2, 3)\}$ from the set X to the set Y is not a function as two distinct elements of the set Y are assigned to the element 1 in X (Fig. 1.5).

It would be easier to understand the dependence between the elements if we could geometrically represent a function. As a convention, the visual representation is done by plotting the elements in the domain along the horizontal axis and the corresponding images along the vertical axis.

Definition 1.10 (*Graph of a Function*) Let $f : X \to Y$ be a function. The set $\{(x, f(x)) \in X \times Y \mid x \in X\}$ is called the graph of f.



(a) function (b) not a function

Fig. 1.6 Observe that in the first graph any vertical line drawn in the domain will touch exactly one point of the graph. However, in the second graph it may touch more than one point

Observe that the above-defined set is exactly the same as f, by Definition 1.9. Also keep in mind that not all graphs represent a function. If any vertical line intersects a graph at more than one point, the relation represented by the graph is not a function. This is known as the *vertical line test* (Fig. 1.6).

Definition 1.11 (*One-one function and Onto function*) A function f from a set X to a set Y is called a one-one (injective) function if distinct elements in the domain have distinct images, that is, for every $x_1, x_2 \in X$, $f(x_1) = f(x_2)$ implies $x_1 = x_2$. f is called onto (surjective) if every element of Y is the image of at least one element of X, that is, for every $y \in Y$, $\exists x \in X$ such that $f(x) = y$. A function which is both one-one and onto is called a *bijective* function.

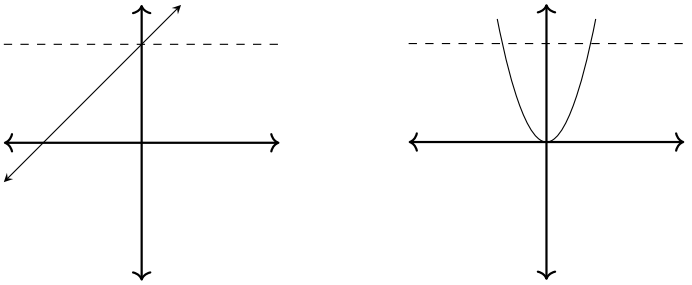
Example 1.9 Consider the function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x + 5$ for $x \in \mathbb{R}$. First, we will check whether the function is one-one or not. We will start by assuming $f(x_1) = f(x_2)$ for some $x_1, x_2 \in \mathbb{R}$. Then

$$\begin{aligned} f(x_1) = f(x_2) &\Rightarrow x_1 + 5 = x_2 + 5 \\ &\Rightarrow x_1 = x_2 \end{aligned}$$

Therefore f is one-one. Now to check whether the function is onto, take any $x \in \mathbb{R}$, then $x - 5 \in \mathbb{R}$ with $f(x - 5) = x - 5 + 5 = x$. That is, every element in $\mathbb{R}$ (co-domain) has a pre-image in $\mathbb{R}$ (domain). Thus, f is onto and hence f is a bijective function.

The graph of a function can also be used to check whether a function is one-one. If any horizontal line intersects the graph more than once, then the graph does not represent a one-one function as it implies that two different elements in the domain have the same image. This is known as the *horizontal line test* (Fig. 1.7).

Definition 1.12 (*Composition of two functions*) Let $f : X \to Y$ and $g : Y \to Z$ be any two functions, then the composition $g \circ f$ is a function from X to Z, defined by $(g \circ f)(x) = g(f(x))$ (Fig. 1.8).



(a) $f_1(x) = x+5$ (b) $f_2(x) = x^2$

Fig. 1.7 Consider the graphs of the functions f_1, f_2:$\mathbb{R} \to \mathbb{R}$ defined by $f_1(x) = x + 5$ and $f_2(x) = x^2$. Observe that if we draw a horizontal line parallel to the x-axis, it will touch exactly one point on the graph of the function f_1. But on the graph of the function f_2, it touches two points. Then by *horizontal line test*, the first function is one-one whereas the second one is not a one-one function

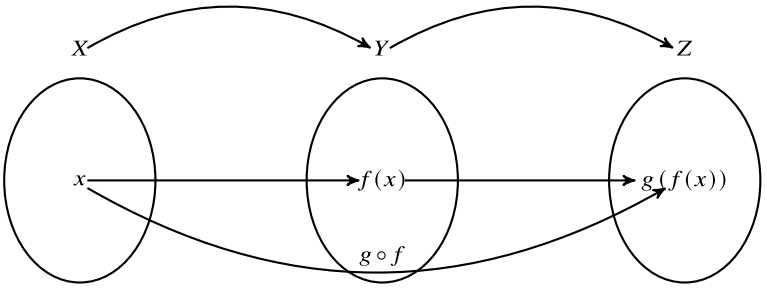


Fig. 1.8 It is clear that the range set of f must be a subset of the domain of g, for the composition function to be defined

Properties
Let $f : X \to Y$, $g : Y \to Z$, and $h : Z \to W$, then

(a) $h \circ (g \circ f) = (h \circ g) \circ f$ (Associative).
(b) If f and g are one-one, then $g \circ f$ is one-one.
(c) If f and g are onto, then $g \circ f$ is onto.

Example 1.10 Consider the functions $f, g : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2$ and $g(x) = 2x + 1$. Then $(f \circ g)(x) = f(g(x)) = f(2x + 1) = (2x + 1)^2 = 4x^2 + 4x + 1$ and $(g \circ f)(x) = g(f(x)) = g(x^2) = 2x^2 + 1$. Observe that $f \circ g \neq g \circ f$. Therefore function composition need not necessarily be commutative.

Definition 1.13 (*Inverse of a function*) A function $f : X \to Y$ is said to be invertible if there exists a function $g : Y \to X$ such that $g(f(x)) = x$ for all $x \in X$ and $f(g(y)) = y$ for all $y \in Y$. The inverse function of f is denoted by f^{-1}.

The function f is invertible if and only if f is a bijective function. For, suppose there exists an inverse function g for f. Then

$$f(x_1) = f(x_2) \Rightarrow g\,(f(x_1)) = g\,(f(x_2)) \Rightarrow x_1 = x_2$$

That is, f is injective. And $f\,(g(y)) = y$ for all $y \in Y$ implies that f is onto.

Example 1.11 Consider a function f, defined as in Fig. 1.4. Indeed, from the figure itself, it's evident that function f is not bijective. Thus f is not invertible. Observe that if we define $f(2) = 4$, then f is both one-one and onto. Then define a function, $g : Y \to X$ by $g(3) = 1$, $g(4) = 2$, and $g(5) = 3$. Now $g\,(f(1)) = g\,(3) = 1$, $g\,(f(2)) = g\,(4) = 2$, and $g\,(f(3)) = g\,(5) = 3$. That is, $g\,(f(x)) = x$ for all $x \in X$. Similarly, we can prove that $f\,(g(y)) = y$ for all $y \in Y$.

Example 1.12 Consider the function $f(x) = x + 5$, defined as in Example 1.9. We have already shown that the function is bijective. Now, we will find the inverse of f. By definition, we can say that f^{-1} is the function that will undo the operation of f. That is, if a function f maps an element x from set X to y in set Y, its inverse function f^{-1} reverses this mapping, taking y from Y back to x in X. In this case, $X = Y = \mathbb{R}$. If we consider, a $y \in \mathbb{R}$(co-domain), then there exists $x \in \mathbb{R}$(domain) such that $y = x + 5$ (Why?). Then $x = y - 5$. Thus, the function $g(y) = y - 5$ will undo the action of f. We can verify this algebraically as follows:

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

and

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

Thus $f^{-1}(x) = x - 5$.

Example 1.13 Now consider $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2$. From Fig. 1.7, we can clearly say that f is not bijective. Thus f does not have an inverse in $\mathbb{R}$. But, if we restrict the domain of f to $[0, \infty)$, f is a bijective function. Then the inverse of f is the function $f^{-1}(x) = \sqrt{x}$. For, $g\,(f(x)) = g(x^2) = \sqrt{x^2} = x$ and $f\,(g(x)) = f(\sqrt{x}) = (\sqrt{x})^2 = x$.

It is easy to check whether a real function is invertible or not, by just looking at its graph. Consider Fig. 1.9.

Now we will discuss some of the important concepts related to functions defined on the set of all real numbers to itself.

Definition 1.14 (*Continuity at a point*) Let $X \subset \mathbb{R}$ and $f : X \to \mathbb{R}$ be a function. We say that f is continuous at $x_0 \in X$, if given any $\epsilon > 0$ there exists a $\delta > 0$ such that if x is any point in X satisfying $|x - x_0| < \delta$, then $|f(x) - f(x_0)| < \epsilon$. Otherwise, f is said to be discontinuous at x_0.

A function is continuous if it is continuous at each point of its domain. In graphical terms, the continuity of a function on the set of all real numbers means that the graph does not have any gaps or breaks. From Fig. 1.7, it is clear that both the functions $f(x) = x + 5$ and $f(x) = x^2$ are continuous. Figure 1.10 gives an example for a discontinuous function.

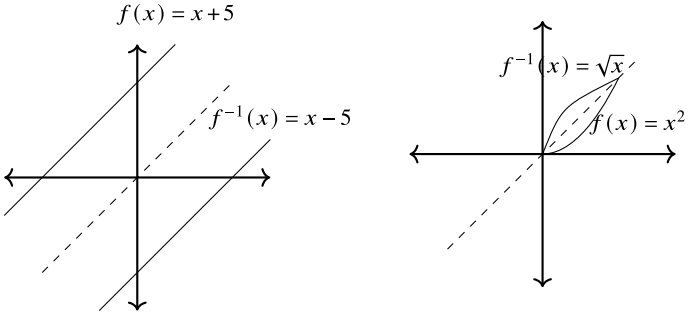
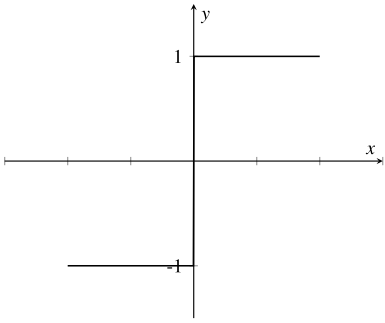


Fig. 1.9 Observe that the graph of $f^{-1}(x)$ is the graph of $f(x)$ reflected about the line $y = x$ (represented by the dotted line)

Fig. 1.10 Consider the signum function, defined by

$$f(x) = \begin{cases} 1, \ if \ x > 0 \\ 0, \ if \ x = 0 \\ -1, \ if \ x < 0 \end{cases}.$$

Clearly, f is not continuous at $x = 0$



Observe that in the definition of continuity of a function at a point, the value of δ depends on both x_0 and ϵ. If δ does not depend on the point x_0, then the continuity is called *uniform continuity*. In other words, a function f is uniformly continuous on a set X, if for every $\epsilon > 0$, there exists $\delta > 0$, such that for every element $x, y \in X$, $|f(x) - f(y)| < \epsilon$ whenever $|x - y| < \delta$. Graphically, this means that given any narrow vertical strip of width ϵ on the graph, there exists a corresponding horizontal strip of width δ such that all points in the interval within δ units of each other on the x-axis map to points within ϵ units of each other on the y-axis. Consider the following example.

Example 1.14 Consider the function $f_1(x) = x + 5$. We will show that f_1 is uniformly continuous. For, given any $\epsilon > 0$, choose $\delta = \epsilon$. Then, for any $x, y \in \mathbb{R}$ with $|x - y| < \delta$, we have

$$|f_1(x) - f_1(y)| = |x + 5 - (y + 5)| = |x - y| < \delta = \epsilon$$

Thus $f_1(x) = x + 5$ is uniformly continuous over $\mathbb{R}$. However, the function $f_2(x) = x^2$ is not uniformly continuous on $\mathbb{R}$. Suppose on the contrary that f_2 is uniformly continuous. Fix $\epsilon = 1$. Then, there exists $\delta_0 > 0$, such that for every element $x, y \in \mathbb{R}$, $|f(x) - f(y)| < 1$ whenever $|x - y| < \delta_0$. Now, take $y = x + \frac{\delta_0}{2}$. Then,

$$|f(x) - f(y)| = \left| x^2 - \left(x + \frac{\delta_0}{2} \right)^2 \right| = \left| x\delta_0 + \frac{\delta_0^2}{4} \right| < 1$$

which is a contradiction as x can be chosen arbitrarily.

Now, we will define continuity of a function using the notion of sequences of real numbers.

Definition 1.15 (*Real Sequence*) A real sequence $\{x_n\}$ is a function whose domain is the set $\mathbb{N}$ of natural numbers and co-domain is the set of all real numbers $\mathbb{R}$. In other words, a sequence in $\mathbb{R}$ assigns to each natural number $n = 1, 2, \ldots$ a uniquely determined real number. For example, the function $f : \mathbb{N} \to \mathbb{R}$ defined by $f(n) = \frac{1}{n}$ determines the sequence $\left\{1, \frac{1}{2}, \frac{1}{3}, \ldots\right\}$.

Example 1.15 The list of numbers $\{r, r, r, \ldots\}$, where r is any real number, is a sequence called *constant sequence* as we can define a function, $f : \mathbb{N} \to \mathbb{R}$, by $f(n) = r$.

Example 1.16 The list of numbers $\{r, r^2, r^3, \ldots\}$, where r is any real number, is a sequence called *geometric sequence* as we can define a function, $f : \mathbb{N} \to \mathbb{R}$, by $f(n) = r^n$.

Definition 1.16 (*Convergent Sequence*) A real sequence $\{x_n\}$ is said to converge to $x \in \mathbb{R}$, or x is said to be a limit of $\{x_n\}$, denoted by $x_n \to x$ or $\lim_{n\to\infty} x_n = x$, if for every $\epsilon > 0$, there exists a natural number N such that $|x_n - x| < \epsilon$ for all $n \geq N$. Otherwise, we say that $\{x_n\}$ is divergent.

Theorem 1.1 *A real sequence $\{x_n\}$ can have at most one limit.*

Example 1.17 Consider the sequence $\{x_n\}$, where $x_n = \frac{1}{n}$. Clearly, $x_n \to 0$. For, given any $\epsilon > 0$, we have $|x_n - 0| = \left|\frac{1}{n}\right|$. If we take $n > \frac{1}{\epsilon}$, we have $|1/n| < \epsilon$. Thus $\lim_{n\to\infty} \frac{1}{n} = 0$.

Example 1.18 Consider the sequence $\{x_n\}$, defined as in Example 1.15. It is easy to observe that $x_n \to r$ as $|x_n - r| = 0$ for all $n \in \mathbb{N}$.

Example 1.19 Consider the sequence $\{x_n\}$, defined as in Example 1.16. We can observe that the convergence of this sequence depends on the value of r. First of all, by the above example, for $r = 0$ and $r = 1$, $\{x_n\}$ converges to 0 and 1, respectively. Now let $0 < r < 1$. Then $x_n \to 0$. For any $\epsilon > 0$, if we take $N > \frac{ln\,\epsilon}{ln\,r}$ we have $|x_n - 0| = r^n < \epsilon$ for all $n > N$. Similarly, for $-1 < r < 0$, $x_n \to 0$.

Now for $r = -1$, the given sequence becomes $x_n = (-1)^n$. Take $\epsilon = \frac{1}{3}$. Then there does not exist any point $x \in \mathbb{R}$ such that $|x_n - x| < \frac{1}{3}$ as the interval $\left(x - \frac{1}{3}, x + \frac{1}{3}\right)$ must contain both 1 and -1. Therefore $\{x_n\}$ with $x_n = (-1)^n$ does not converge. Similarly, we can prove that the sequence $\{x_n\}$ with $x_n = r^n$ does not converge outside the interval $(-1, 1]$.

As we have discussed convergent sequences, Cauchy sequences must be introduced, which are a specific class of sequences in which the terms become arbitrarily close to each other as the index increases, rather than approaching a single limit.

Definition 1.17 (*Cauchy Sequence*) A real sequence $\{x_n\}$ is said to be a Cauchy sequence, if for any $\epsilon > 0$, there exists a natural number N such that $|x_m - x_n| < \epsilon$ for all $m, n \geq N$.

For a real sequence, the terms convergent sequence and Cauchy sequence do not make any difference. We have the following theorem stating this fact.

Theorem 1.2 *A real sequence $\{x_n\}$ is convergent if and only if it is Cauchy.*

However, this may not be true, if we are considering sequences in the set of rational numbers, $\mathbb{Q}$. That is, there exist sequences of rational numbers that are Cauchy but not convergent in $\mathbb{Q}$ (the sequence may not converge to a rational number). For example, consider the sequence 1.41, 1.412, 1.1421, ... This sequence will converge to $\sqrt{2}$ which is not a rational number (also, see Exercise 13 of this chapter). Now, we will introduce the sequential definition for continuity.

Definition 1.18 (*Sequential Continuity*) A function $f : X \subseteq \mathbb{R} \to \mathbb{R}$ is said to be sequentially continuous at point $x_0 \in X$ if for every $\{x_n\}$ in X with $x_n \to x_0$, we have $f(x_n) \to f(x_0)$. That is if, $\lim_{n\to\infty} x_n = x_0 \Rightarrow \lim_{n\to\infty} f(x_n) = f(x_0)$.

Then, we have the following result which asserts that sequential continuity and continuity of a real function are the same.

Theorem 1.3 *A function $f : X \subseteq \mathbb{R} \to \mathbb{R}$ is continuous if and only if it is sequentially continuous.*

Example 1.20 Consider the signum function as defined in Fig. 1.10. We know that f is not continuous at $x = 0$. We can use the definition of sequential continuity to prove this fact. Consider the sequence $\left\{\frac{1}{n}\right\}$. In Example 1.17, we have seen that $\frac{1}{n} \to 0$. However, observe that $f\left(\frac{1}{n}\right) = 1 \to 1 \neq f(0)$. Thus f is not sequentially continuous at 0 and hence f is not continuous at 0.

Now, consider the function $f(x) = x + 5$. We have already seen that f is continuous on $\mathbb{R}$ as its graph does not have any gaps or breaks. Let us check whether f is sequentially continuous or not. Consider any real number $r \in \mathbb{R}$ and a sequence $\{r_n\}$ with $r_n \to r$ as $n \to \infty$. For sequential continuity $f(r_n)$ must converge to $f(r)$. Observe that $f(r_n) = r_n + 5 \to r + 5$ as $n \to \infty$. Thus f is sequential continuous.

Remark 1.3 A set S is said to be countably infinite if there exists a bijective function from $\mathbb{N}$ to S. A set which is empty, finite, or countably infinite is called a countable set. Otherwise it is called uncountable set. For example $\mathbb{Z}$ is countable and $\mathbb{R}$ is uncountable.

Sequence of Functions

Now, we will combine the ideas of functions and sequences discussed so far and define "sequence of functions".

Definition 1.19 (*Sequence of Functions*) Let f_n be real-valued functions defined on an interval $[a, b]$ for each $n \in \mathbb{N}$. Then $\{f_1, f_2, f_3, \ldots\}$ is called a sequence of real-valued functions on $[0, 1]$, and is denoted by $\{f_n\}$.

Example 1.21 For each $n \in \mathbb{N}$, let f_n be defined on $[0, 1]$ by $f_n(x) = x^n$. Then $\{x, x^2, x^3, \ldots\}$ is a sequence of real-valued functions on $[a, b]$.

For a sequence of functions, we have two types of convergences, namely *point-wise convergence* and *uniform convergence*. We will discuss these concepts briefly in this section.

Let $\{f_n\}$ be a sequence of functions on $[a, b]$ and $x_0 \in [a, b]$. Then the sequence of real numbers, $\{f_n(x_0)\}$, may be convergent. In fact, this may be true for all points in $[a, b]$. The limiting values of the sequence of real numbers corresponding to each point $x \in X$ define a function called the limit function or simply the limit of the sequence $\{f_n\}$ of functions on $[a, b]$.

Definition 1.20 (*Point-wise convergence*) Let $\{f_n\}$ be a sequence of real-valued function defined on an interval $[a, b]$. If for each $x \in [a, b]$ and each $\epsilon > 0$, there exists an $N \in \mathbb{N}$ such that $|f_n(x) - f(x)| < \epsilon$ for all $n > N$, then we say that $\{f_n\}$ converges point-wise to the function f on $[a, b]$ and is denoted by $\lim_{n\to\infty} f_n(x) = f(x),\ \forall\, x \in [a, b]$.

Example 1.22 Let $f_n(x) = x^n$ be defined on $[0, 1]$. By Example 1.19, the limit function $f(x)$ is given by

$$f(x) = \lim_{n\to\infty} f_n(x) = \begin{cases} 0, & x \in [0, 1) \\ 1, & x = 1 \end{cases}$$

Let $\epsilon = \frac{1}{2}$. Then for each $x \in [0, 1]$, there exists a positive integer N such that $|f_n(x) - f(x)| < \frac{1}{2}$ for all $n > N$. If $x = 0$, $f(x) = 0$ and $f_n(x) = 0$ for all n. $|f_n(x) - f(x)| < \frac{1}{2}$ is true for all $n > 1$.
If $x = 1$, $f(x) = 1$ and $f_n(x) = 1$ for all n. $|f_n(x) - f(x)| < \frac{1}{2}$ is true for all $n > 1$.
If $x = \frac{3}{4}$, $f(x) = 0$ and $f_n(x) = \left(\frac{3}{4}\right)^n$ for all n. Then

$$|f_n(x) - f(x)| = \left(\frac{3}{4}\right)^n < \frac{1}{2}$$

is true for all $n > 2$.

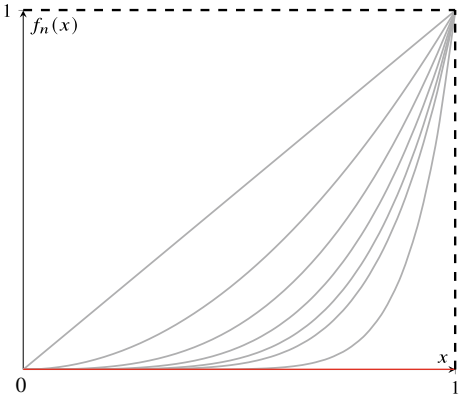


Fig. 1.11 Point-wise convergence of $\{f_n\}$, where $f_n(x) = x^n, x \in [0, 1]$

If $x = \frac{9}{10}$, $f(x) = 0$ and $f_n(x) = \left(\frac{9}{10}\right)^n$ for all n. Then

$$|f_n(x) - f(x)| = \left(\frac{9}{10}\right)^n < \frac{1}{2}$$

is true for all $n > 6$ (Fig. 1.11).

Observe that there is no value of N for which $|f_n(x) - f(x)| < \frac{1}{2}$ is true for all $x \in [0, 1]$. N depends on both x and ϵ. But, this is not the case for the following example.

Example 1.23 Consider $f_n(x) = \dfrac{x}{1 + nx}, x \geq 0$. Clearly,

$$\lim_{n \to \infty} f_n(x) = f(x) = 0, \ \forall\, x \geq 0$$

Also, we have

$$0 \leq \frac{x}{1 + nx} \leq \frac{x}{nx} = \frac{1}{n}$$

Therefore, $|f_n(x) - f(x)| = |f_n(x)| \leq \frac{1}{n} < \epsilon$ for all $x \geq 0$, provided $N > \frac{1}{\epsilon}$. That is, if $N > \frac{1}{\epsilon}$, then $|f_n(x) - f(x)| < \epsilon$ for all $n > N$ and for all $x \geq 0$. Here N depends only on ϵ. Such type of convergence is called uniform convergence (Fig. 1.12).

Definition 1.21 (*Uniform convergence*) Let $\{f_n\}$ be a real-valued function defined on an interval $[a, b]$. Then $\{f_n\}$ is said to converge uniformly to the function f on $[a, b]$, if for each $\epsilon > 0$, there exists an integer N (dependent on ϵ and independent of x) such that for all $x \in [a, b]$, $|f_n(x) - f(x)| < \epsilon$ for all $n > N$ (Fig. 1.13).

Clearly, we can observe that uniform convergence implies point-wise convergence, but the converse does not hold true always. Also observe that, in Example 1.22, all the functions in $\{f_n\}$ were continuous. However, their point-wise limit was not continuous. In the case of uniform convergence, this is not possible. That is, if $\{f_n\}$ is a sequence of continuous functions and $f_n \to f$ uniformly then f is continuous.

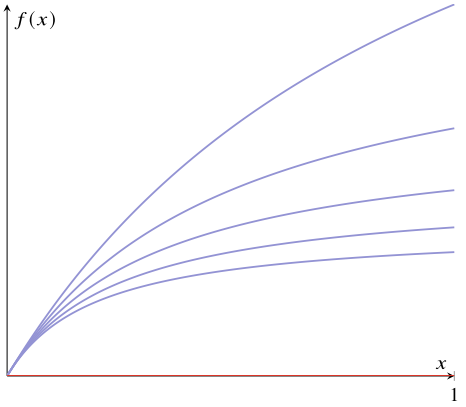


Fig. 1.12 Uniform convergence of $\{f_n\}$, where $f_n(x) = \dfrac{x}{1+nx}, x \geq 0$

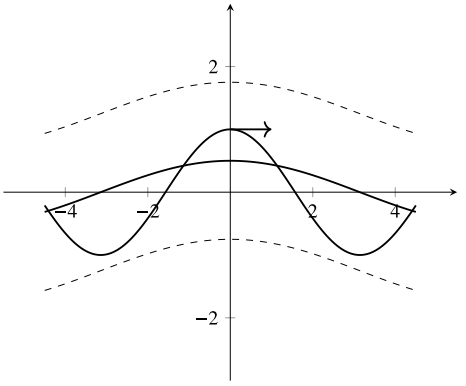


Fig. 1.13 If $\{f_n\}$ converges uniformly to a function f on $[a, b]$, for a given $\epsilon > 0$, there exists a positive integer N such that the graph of $f_n(x)$ for all $n > N$ and for all $x \in [a, b]$ lies between the graphs of $f(x) - \epsilon$ and $f(x) + \epsilon$

1.2 Metric Spaces

In $\mathbb{R}$, we have the notion of usual distance provided by the modulus function, to discuss the ideas like continuity of a function, convergence of a sequence, etc. These concepts can also be extended to a wide range of sets by generalizing the notion of "distance" to these sets by means of a function, called *metric*. A set with such a distance notion defined on it is called as a *metric space*. Consider the following definition.

Definition 1.22 (*Metric Space*) Let X be any non-empty set. A metric (or distance function) on X is a function $d : X \times X \to \mathbb{R}^+$ which satisfies the following properties for all $x, y, z \in X$:

$(M1)$ $d(x, y) \geq 0$ and $d(x, y) = 0$ if and only if $x = y$. (*Non-negativity*)
$(M2)$ $d(x, y) = d(y, x)$. (*Symmetry*)
$(M3)$ $d(x, z) \leq d(x, y) + d(y, z)$. (*Triangle Inequality*)

If d is a metric on X, we say that (X, d) is a metric space.

Example 1.24 Consider the set of all real numbers, $\mathbb{R}$. For $x, y \in \mathbb{R}$, the function defined by

$$d(x, y) = |x - y|$$

is the usual distance between two points on the real line.

($M1$) Clearly $d(x, y) = |x - y| \geq 0$ and $d(x, y) = |x - y| = 0$ if and only if $x - y = 0$. That is, if and only if $x = y$.
($M2$) $d(x, y) = |x - y| = |y - x| = d(y, x)$
($M3$) Also, by the properties of modulus

$$\begin{aligned} d(x, z) &= |x - z| \\ &= |x - y + y - z| \\ &\leq |x - y| + |y - z| \\ &= d(x, y) + d(y, z) \end{aligned}$$

Thus all the conditions for a metric are satisfied and hence $(\mathbb{R}, |.|)$ is a metric space. This metric is known as the *usual metric* or *Euclidean Distance*.

Example 1.25 For any non-empty set X, define a function d by

$$d(x, y) = \begin{cases} 1\,, & x \neq y \\ 0\,, & x = y \end{cases}$$

Clearly conditions ($M1$) and ($M2$) are satisfied. Now we will check ($M3$),

Case 1 $x \neq y = z$
Then $d(x, y) = 1, d(x, z) = 1$ and $d(y, z) = 0$
Case 2 $x = y \neq z$
Then $d(x, y) = 0, d(x, z) = 1$ and $d(y, z) = 1$
Case 3 $x = y = z$
Then $d(x, y) = 0, d(x, z) = 0$ and $d(y, z) = 0$
Case 4 $x \neq y \neq z$
Then $d(x, y) = 1, d(x, z) = 1$ and $d(y, z) = 1$.

In all four cases, condition ($M3$) is clearly satisfied. Hence (X, d) is a metric space for any non-empty set X. The given metric d is known as a *discrete metric*.

Definition 1.23 (*Open Ball*) Let (X, d) be a metric space. For any point $x_0 \in X$ and $\epsilon \in \mathbb{R}^+$,

$$B_\epsilon(x_0) = \{x \in X \mid d(x, x_0) < \epsilon\}$$

is called an *open ball* centered at x_0 with radius ϵ.

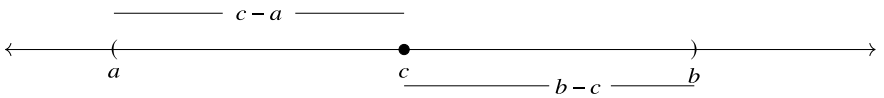


Fig. 1.14 Observe that if we take ϵ less than both $c - a$ and $b - c$, $B_\epsilon(c) \subset (a, b)$

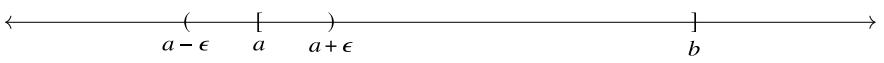


Fig. 1.15 Clearly $B_\epsilon(a) \not\subset [a, b]$ for any $\epsilon > 0$. Also, any open interval containing b is not a subset of $[a, b]$

Definition 1.24 (*Open Set and Closed Set*) Let (X, d) be a metric space. A subset $Y \subseteq X$ is said to be *open* if it contains an open ball about each of its elements. $Y \subseteq X$ is said to be *closed* if its complement Y^c is open.

Example 1.26 Consider the metric space $(\mathbb{R}, |.|)$. Then we can verify that every open interval in the real line is an open set (see Exercise 8 of this chapter). Consider an arbitrary open interval $(a, b) \subset \mathbb{R}$ and choose an arbitrary element $c \in (a, b)$. We have to show that there exists $\epsilon > 0$ such that $B_\epsilon(c) \subset (a, b)$ (Fig. 1.14).

From Fig. 1.14, if we take $\epsilon < min\{c - a, b - c\}$, it is clear that $B_\epsilon(c) \subset (a, b)$ for any $c \in (a, b)$. Similarly, we can prove that the union of open intervals is also an open set in $\mathbb{R}$. But a closed interval $[a, b] \subset \mathbb{R}$ is not an open set as $B_\epsilon(a) \nsubseteq [a, b]$ for any $\epsilon > 0$ (Fig. 1.15).

As $[a, b]^c = (-\infty, a) \cup (b, \infty)$ is an open set, $[a, b]$ is a closed set.

Example 1.27 Every singleton set in a discrete metric space X is an open set. It is obvious from the fact that for any $x \in X$, we have $B_\epsilon(x) = \{x\}$ when $\epsilon < 1$. Also it is interesting to observe that every subset of a discrete metric space is open as every open set can be written as a union of singleton sets. Therefore, every subset of a discrete metric space X is a closed set also.

As we have defined sequences on $\mathbb{R}$, we can define sequences on an arbitrary metric space (X, d) as a function from the set of all natural numbers taking values in X, and we can discuss their convergence based on the distance function d.

Definition 1.25 (*Convergent Sequence*) Sequence $\{x_n\}$ in a metric space (X, d) converges to $x \in X$ if for every $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that $x_n \in B_\epsilon(x)$ for all $n > N$ and x is called the limit of the sequence $\{x_n\}$. We denote this by $x_n \to x$ or $\lim_{n\to\infty} x_n = x$. In other words, we can say that $d(x_n, x) \to 0$ as $n \to \infty$.

Example 1.28 Consider the sequence $\{x_n\}$, where $x_n = r + \frac{1}{n}$, $n \in \mathbb{N}$ in the metric space $(\mathbb{R}, |.|)$ for some $r \in \mathbb{R}$. We will show that $x_n \to r$ in $(\mathbb{R}, |.|)$. For any $\epsilon > 0$, if we take $N > \frac{1}{\epsilon}$

$$d(x_n, r) = \left| r + \frac{1}{n} - r \right| = \left| \frac{1}{n} \right| < \epsilon \; \forall \, n > N$$

That is, $x_n \in B_\epsilon(r)$ for all $n > N$. Therefore $x_n \to r$ in $(\mathbb{R}, |.|)$.

Example 1.29 Let $\{x_n\}$ be a sequence in a metric space (X, d), where d is the discrete metric. We have seen in Example 1.27 that every singleton set in a discrete metric space is open. Therefore for the sequence $\{x_n\}$ to converge to a point $x \in X$, the open set $\{x\}$ must contain almost all terms of the sequence. In other words, a sequence $\{x_n\}$ in a discrete metric space converges if and only if it is of the form $x_1, x_2, \ldots, x_N, x, x, \ldots$. That is, if and only if $\{x_n\}$ is *eventually constant*.

Definition 1.26 (*Cauchy Sequence*) Sequence of points $\{x_n\}$ in a metric space (X, d) is said to be a Cauchy sequence if for every $\epsilon > 0$, there exists an $N_\epsilon \in \mathbb{N}$ such that $d(x_n, x_m) < \epsilon$ for every $m, n > N_\epsilon$.

Theorem 1.4 *In a metric space, every convergent sequence is Cauchy.*

The converse of the above theorem need not be true. That is, there exists metric spaces where every Cauchy sequence may not be convergent.

Example 1.30 Consider the sequence $\{x_n\}$ with $x_n = a + \frac{1}{n}$ in the metric space $((a, b), |.|)$ where (a, b) is any open interval in $\mathbb{R}$. We will show that this sequence is Cauchy but not convergent. For an $\epsilon > 0$, if we choose $N > \frac{2}{\epsilon}$

$$d(x_n, x_m) = \left|\frac{1}{n} - \frac{1}{m}\right| \leq \left|\frac{1}{n}\right| + \left|\frac{1}{m}\right| \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon \ \forall\, m, n > N$$

That is, the given sequence is a Cauchy sequence. As we have seen in Example 1.28, the given sequence converges to a as $n \to \infty$. As $a \notin (a, b)$, $\{x_n\}$ with $x_n = a + \frac{1}{n}$ is not convergent in $((a, b), |.|)$.

Definition 1.27 (*Complete Metric Space*) A metric space in which every Cauchy sequence is convergent is called a complete metric space.

Example 1.31 By Theorem 1.2, $(\mathbb{R}, |.|)$ is a complete metric space and from Example 1.30, $((a, b), |.|)$ is an incomplete metric space.

Definition 1.28 (*Continuous Function*) Let (X, d_1) and (Y, d_2) be two metric spaces. A function $f : X \to Y$ is said to be continuous at a point $x_0 \in X$ if for every $\epsilon > 0$ there is a $\delta > 0$ such that $d_2(f(x), f(x_0)) < \epsilon$ whenever $d_1(x, x_0) < \delta$. f is said to be continuous on X if f is continuous at every point of X.

Theorem 1.5 *Let (X, d_1) and (Y, d_2) be two metric spaces. Then a function $f : X \to Y$ is said to be continuous if and only if the inverse image of any open set of (Y, d_2) is open in (X, d_1).*

The continuity of a function in metric spaces can also be discussed in terms of sequences. Consider the following definition.

Definition 1.29 (*Sequential Continuity*) Let (X, d_1) and (Y, d_2) be two metric spaces. A function $f : X \to Y$ is said to be sequentially continuous at a point $x_0 \in X$ if $\{x_n\}$ is any sequence in (X, d_1) with $x_n \to x_0$, then $f(x_n) \to f(x_0)$ in (Y, d_2).

Theorem 1.6 *Let (X, d_1) and (Y, d_2) be two metric spaces. Then a function $f : X \to Y$ is continuous on X, if and only if it is sequentially continuous.*

1.3 Some Important Algebraic Structures

An algebraic structure consists of a non-empty set together with a collection of operations defined on it satisfying certain conditions or axioms which are defined as per the context under discussion. The operations are of great importance when the resultant obtained by combining two elements in the set belongs to the same set.

Definition 1.30 (*Binary Operation*) Let G be any set. A binary operation $'*'$ on G is a function $* : G \times G \to G$ defined by

$$*(g_1, g_2) = g_1 * g_2$$

Example 1.32 Let $G = \mathbb{R}$, the set of all real numbers, and let $+$ be the usual addition of real numbers. Now $+ : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ such that $+(a, b) = a + b \in \mathbb{R}$ defines a binary operation. Similarly, the usual multiplication and subtraction of real numbers are also binary operations on $\mathbb{R}$. But as the division of a real number with 0 is not defined, division is not a binary operation.

Definition 1.31 (*Group*) A non-empty set G together with a binary operation $'*'$ is said to be a group, denoted by $(G, *)$, if $'*'$ satisfies the following properties:

(a) $g_1 * (g_2 * g_3) = (g_1 * g_2) * g_3 \ \forall \ g_1, g_2, g_3 \in G$ (Associative property)
(b) There exists $e \in G$, such that $e * g = g = g * e \ \forall \ g \in G$ (Existence of Identity)
(c) For each $g \in G$, there exists $g^{-1} \in G$ such that $g * g^{-1} = e = g^{-1} * g$. (Existence of Inverse)

If $'*'$ satisfies $g_1 * g_2 = g_2 * g_1 \ \forall \ g_1, g_2 \in G$ (Commutative property) also, then $(G, *)$ is called an Abelian group.

Example 1.33 Consider $\mathbb{R}$ together with the binary operation $'+'$. Then $\mathbb{R}$ is an Abelian group under the operation $'+'$. For,

(a) Addition is associative over $\mathbb{R}$.
(b) For all $r \in \mathbb{R}$, there exists $0 \in \mathbb{R}$ such that $r + 0 = r = 0 + r$.
(c) For all $r \in \mathbb{R}$, there exists $-r \in \mathbb{R}$ such that $r + (-r) = 0 = (-r) + r$.
(d) Addition is commutative over $\mathbb{R}$.

Similarly, $\mathbb{C}$, the set of all complex numbers, $\mathbb{Q}$, the set of all rational numbers, and $\mathbb{Z}$, the set of all integers together with the binary operation $'+'$ is an Abelian group. But $(\mathbb{R}, .)$ is not a group, where '.' denotes usual multiplication as there does not exist any inverse element for 0.

Example 1.34 Consider $\mathbb{R}^* = \mathbb{R} \setminus \{0\}$ under usual multiplication. We can show that $(\mathbb{R}^*, .)$ is an Abelian group. Similarly, we can show that $(\mathbb{Q}^*, .)$ and $(\mathbb{C}^*, .)$ are also Abelian groups where $\mathbb{Q}^* = \mathbb{Q} \setminus \{0\}$ and $\mathbb{C}^* = \mathbb{C} \setminus \{0\}$. Observe that $\mathbb{Z}^*$ with usual multiplication is not a group as the inverse of every element does not exist in $\mathbb{Z}^*$.

Example 1.35 Consider $\mathbb{R}^+$, the set of all positive real numbers under usual multiplication. We can show that $\left(\mathbb{R}^+, .\right)$ is an Abelian group. Similarly, we can show that $\left(\mathbb{Q}^+, .\right)$ and $\left(\mathbb{C}^+, .\right)$ are also Abelian groups.

Example 1.36 The set $\mathbb{Z}_n = \{0, 1, 2, \ldots, n-1\}$, for $n \geq 1$, is a group under the operation *addition modulo n*, denoted by $+_n$. The basic operation is usual addition of elements, which ends by reducing the sum of the elements modulo n, that is, taking the integer remainder when the sum of the elements is divided by n. This group is usually referred to as the *group of integers modulo n*. Consider the following examples:

$+_2$	0	1
0	0	1
1	1	0

$+_3$	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1

$+_4$	0	1	2	3
0	0	1	2	3
1	1	2	3	0
2	2	3	0	1
3	3	0	1	2

$+_5$	0	1	2	3	4
0	0	1	2	3	4
1	1	2	3	4	0
2	2	3	4	0	1
3	3	4	0	1	2
4	4	0	1	2	3

The above group multiplication table is called Cayley table. A Cayley table, named after the British mathematician *Arthur Cayley (1821–1895)* of the nineteenth century, illustrates the structure of a finite group by arranging all the possible products of all the group's members in a square table resembling an addition or multiplication table.

Example 1.37 A one-one function from a set S onto itself is called a permutation. Consider the set $S = \{1, 2, \ldots, n\}$. Let S_n denote the set of all permutations on S to itself. Then S_n is a non-Abelian group under the operation function composition, called *symmetric group on n letters*. Permutations of finite sets are represented by an explicit listing of each element of the domain and its corresponding image. For example, the elements of S_3 can be listed as $\left\{\rho_0 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}, \rho_1 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}, \rho_2 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}, \mu_1 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}, \mu_2 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}, \mu_3 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}\right\}$

Theorem 1.7 *Let $(G, *)$ be a group. Then*

(a) the identity element is unique.
(b) each element in G has a unique inverse.

Definition 1.32 (*Subgroup*) A subset H of a group $(G, *)$ is said to be a subgroup of G, if H is a group with respect to the operation $*$ in G. Let $H \leq G$ denote that H is a subgroup of G and $H < G$ denote that H is a subgroup of G, but $H \neq G$.

Example 1.38 We have $(\mathbb{Z}, +) < (\mathbb{Q}, +) < (\mathbb{R}, +)$. But $(\mathbb{Z}_n, +_n)$ is a not a subgroup of $(\mathbb{R}, +)$ even though as sets $\mathbb{Z}_n \subset \mathbb{R}$, as the operations used are different.

Example 1.39 Consider the permutation group S_3. Then $\{\rho_0\}, \{\rho_0, \mu_1\}$, $\{\rho_0, \mu_2\}, \{\rho_0, \mu_3\}$ and $\{\rho_0, \rho_1, \rho_2\}$ are subgroups of S_3.

Definition 1.33 (*Order of a Group*) Let $(G, *)$ be a group, then the order of G is the number of elements in G.

Example 1.40 Observe that $(\mathbb{Z}, +)$, $(\mathbb{Q}, +)$, $(\mathbb{R}, +)$, and $(\mathbb{C}, +)$ are groups of order infinity and $(\mathbb{Z}_n, +_n)$ is a group of order n. Also observe that S_n has order $n!$.

Definition 1.34 (*Order of an element*) Let $(G, *)$ be a group, then the order of an element $g \in G$, denoted by $\mathcal{O}(g)$, is the least positive integer n such that $g^n = e$, where e is the identity in G. Clearly, identity element in a group G has order 1.

Example 1.41 Consider the group $(\mathbb{R}, +)$. Then we get that no element other than 0 in $\mathbb{R}$ has finite order. This is because of the fact that repeated addition of a real number will never give us 0.

Example 1.42 Consider a finite group, say $(\mathbb{Z}_4, +_4)$. Then $\mathcal{O}(0) = 1, \mathcal{O}(1) = 4, \mathcal{O}(2) = 2$, and $\mathcal{O}(3) = 4$. It is easy to observe that, in a finite group G, every element has finite order. Consider another example, S_3. Then $\mathcal{O}(\rho_0) = 1, \mathcal{O}(\rho_1) = \mathcal{O}(\rho_2) = 3$, and $\mathcal{O}(\mu_1) = \mathcal{O}(\mu_2) = \mathcal{O}(\mu_3) = 2$.

Remark 1.4 A set G together with a binary operation $'*'$ defined on it is called a Groupoid or Magma. If $'*'$ satisfies associative property also, then $(G, *)$ is called a Semi-group. A semi-group containing an identity element is called a Monoid.

Definition 1.35 (*Group Homomorphism*) Let $(G, *)$ and $\left(G', *'\right)$ be any two groups. A map ϕ from G to G' satisfying $\phi(g_1 * g_2) = \phi(g_1) *' \phi(g_2)$, $\forall\ g_1, g_2 \in G$ is called a group homomorphism. If ϕ is one-one and onto, we say that ϕ is an isomorphism or $(G, *)$ and $\left(G', *'\right)$ are isomorphic, denoted by $G \cong G'$.

Definition 1.36 (*Kernel of a Homomorphism*) The kernel of a homomorphism of a group G to a group G' with identity e' is the set of all elements in G which are mapped to e'. That is, $Ker(\phi) = \{g \in G \mid \phi(g) = e'\}$.

Example 1.43 Consider the groups $(\mathbb{R}, +)$ and $(\mathbb{R}^*, .)$. We will show that they are isomorphic. Define $\phi : \mathbb{R} \to \mathbb{R}^*$ by $\phi(x) = e^x$. Then for $x_1, x_2 \in \mathbb{R}$,

$$\phi(x_1 + x_2) = e^{x_1 + x_2} = e^{x_1}.e^{x_2} = \phi(x_1).\phi(x_2)$$

Therefore ϕ is a homomorphism from $\mathbb{R}$ to $\mathbb{R}^*$. Also we can easily verify that ϕ is both one-one and onto. Thus $(\mathbb{R}, +) \cong (\mathbb{R}^*, .)$. Now let us find the Kernel of ϕ. By definition, $Ker(\phi)$ is the set of all elements of the domain which are mapped to the identity element in the co-domain, in this case, 1. Therefore $Ker(\phi) = \{x \in \mathbb{R} \mid \phi(x) = e^x = 1\} = \{0\}$.

Example 1.44 Consider $(\mathbb{Z}, +)$ and $(\mathbb{Z}_n, +_n)$. Define $\phi : \mathbb{Z} \to \mathbb{Z}_n$ by $\phi(m) = r$, where r is the remainder when m is divided by n. Let us check whether ϕ is a homomorphism or not. Take two elements $m_1, m_2 \in \mathbb{Z}$. By division algorithm, we can write $m_i = q_i n + r_i$ with $0 \leq r_i < n$, where $i = 1, 2$ and hence $\phi(m_1) = r_1$ and $\phi(m_2) = r_2$. Observe that $m_1 + m_2 = (q_1 + q_2)n + r_1 + r_2$. Therefore, we can say that $\phi(m_1 + m_2)$ is the remainder when $r_1 + r_2$ is divided by n. That is, $\phi(m_1 + m_2) = r_1 +_n r_2$. Also $\phi(m_1) +_n \phi(m_2) = r_1 +_n r_2$. Thus ϕ is a homomorphism. Now the set of all elements mapped to $0 \in \mathbb{Z}_n$ are integer multiples of n. That is, $Ker(\phi) =< n >$.

Example 1.45 Consider the map $\phi : (\mathbb{R}, +) \to (\mathbb{R}^*, .)$ defined by $\phi(x) = x^2$. Then for $x_1, x_2 \in \mathbb{R}$, we have

$$\phi(x_1 + x_2) = (x_1 + x_2)^2 \neq x_1^2.x_2^2 = \phi(x_1).\phi(x_2)$$

Thus ϕ is not a homomorphism.

Example 1.46 Consider $(\mathbb{R}^*, .)$. Define a map $\phi : \mathbb{R}^* \to \mathbb{R}^*$ by $\phi(x) = |x|$. Then for $x_1, x_2 \in \mathbb{R}^*$, we have

$$\phi(x_1 x_2) = |x_1 x_2| = |x_1||x_2| = \phi(x_1)\phi(x_2)$$

Thus ϕ is a homomorphism from $\mathbb{R}^*$ to itself. Observe that $Ker(\phi) = \{x \in \mathbb{R}^* \mid |x| = 1\} = \{-1, 1\}$. Thus ϕ is not one-one (Why?). Also ϕ is not onto as only positive real numbers have pre-images. Therefore ϕ is not an isomorphism.

Theorem 1.8 *Let ϕ be a homomorphism from a group $(G, *)$ to $(G', *')$. Then*

(a) if e is the identity element in G, $\phi(e)$ is the identity element in G'.
(b) $Ker(\phi)$ is a subgroup of G.
(c) for any $g \in G$, if $\mathcal{O}(g)$ is finite $\mathcal{O}(\phi(g))$ divides $\mathcal{O}(g)$.
(d) for any subgroup H of G, $\phi(H)$ is a subgroup of $\phi(G)$ and if H is Abelian, $\phi(H)$ is also Abelian.

Two algebraic structures $(G, *)$ and $(G', *')$ are isomorphic, if there exists a one-one, onto homomorphism from G to G'. But it will be difficult to show that $(G, *)$ and $(G', *')$ are not isomorphic, following the definition as it means that there is no one-one homomorphism from G onto G'. It is not possible to check whether such a function exists or not. In such cases, we could use the idea of structural properties of an algebraic structure, which are properties that must be shared by any isomorphic structure. Cardinality is an example for structural property.

Example 1.47 In Remark 1.3, we have seen that $\mathbb{R}$ is an uncountable set and $\mathbb{Z}$ is a countable set. Hence $(\mathbb{R}, +)$ and $(\mathbb{Z}, +)$ are not isomorphic.

Theorem 1.9 (Cyclic subgroup) *Let $(G, *)$ be a group. Then the set $\{g^n \mid g \in G, n \in \mathbb{Z}\}$ is a subgroup of G called cyclic subgroup of G generated by g, denoted by $< g >$.*

If the group $G =< g >$ for some $g \in G$, then G is called a *cyclic group* and g is called a generator of G.

Example 1.48 $(\mathbb{Z}, +)$ is a cyclic group with two generators $\{1, -1\}$.

Example 1.49 $(\mathbb{Z}_n, +_n)$ is a cyclic group. The generators are the elements $m \in \mathbb{Z}_n$ with $gcd(m, n) = 1$, where $gcd(m, n)$ denotes the greatest common divisor for m and n (verify).

Theorem 1.10 *Let $(G, *)$ be a cyclic group with generator g. If $\mathcal{O}(G)$ is finite, then $(G, *) \cong (\mathbb{Z}_n, +_n)$ and if $\mathcal{O}(G)$ is infinite, then $(G, *) \cong (\mathbb{Z}, +)$.*

Example 1.50 By Example 1.47, $(\mathbb{R}, +)$ is not a cyclic group.

Definition 1.37 (*Coset*) Let $(G, *)$ be a group and H be a non-trivial subgroup of G. Then $gH = \{g * h \mid h \in H\}$ is called left coset of H in G containing g and $Hg = \{h * g \mid h \in H\}$ is called right coset of H in G containing g.

Example 1.51 Consider $(\mathbb{Z}_8, +_8)$ and the subgroup $H = \{0, 2, 4, 6\}$ of $\mathbb{Z}_8$. Then

$$0H = \{0, 2, 4, 6\} = 2H = 4H = 6H$$

$$1H = \{1, 3, 5, 7\} = 3H = 5H = 7H$$

Also observe that as $(\mathbb{Z}_8, +_8)$ is an Abelian group, the left and right cosets of each element coincide.

Example 1.52 Consider the subgroup $H = \{\rho_0, \mu_1\}$ in S_3. Then

$$\rho_0 H = \{\rho_0, \mu_1\} = \mu_1 H$$

$$\rho_1 H = \{\rho_1, \mu_3\} = \mu_3 H$$

$$\rho_2 H = \{\rho_2, \mu_2\} = \mu_2 H$$

are the distinct left cosets of H in G and

$$H\rho_0 = \{\rho_0, \mu_1\} = H\mu_1$$

$$H\rho_1 = \{\rho_1, \mu_2\} = H\mu_2$$

$$H\rho_2 = \{\rho_2, \mu_3\} = H\mu_3$$

are the distinct right cosets of H in G

Theorem 1.11 (Lagrange's Theorem) *Let G be a finite group and H be a subgroup of G, then $\mathcal{O}(H)$ divides $\mathcal{O}(G)$. Moreover, the number of distinct left/right cosets of H in G is* $\dfrac{\mathcal{O}(G)}{\mathcal{O}(H)}$.

Example 1.53 In Example 1.51, $H = \{0, 2, 4, 6\}$ and $G = \mathbb{Z}_8$. We have $\mathcal{O}(H) = 4$ and $\mathcal{O}(G) = 8$. Clearly, $\mathcal{O}(H)$ divides $\mathcal{O}(G)$ and the number of distinct left/right cosets of H in G is $\dfrac{\mathcal{O}(G)}{\mathcal{O}(H)} = 2$

Example 1.54 In Example 1.52, $H = \{\rho_0, \mu_1\}$ and $G = S_3$. We have $\mathcal{O}(H) = 2$ and $\mathcal{O}(G) = 6$. Clearly, $\mathcal{O}(H)$ divides $\mathcal{O}(G)$ and the number of distinct left/right cosets of H in G is $\dfrac{\mathcal{O}(G)}{\mathcal{O}(H)} = 3$.

Definition 1.38 (*Normal Subgroup*) A subgroup H of G is called a normal subgroup of G if $gH = Hg$ for all $g \in G$.

Example 1.55 From Example 1.51, $H = \{0, 2, 4, 6\}$ is a normal subgroup of $(\mathbb{Z}_8, +_8)$. In fact, every subgroup of an Abelian group is a normal subgroup (verify).

Example 1.56 From Example 1.52, $H = \{\rho_0, \mu_1\}$ is not a normal subgroup of S_3.

Theorem 1.12 (Factor Group) *Let $(G, *)$ be a group and H be a normal subgroup. Then the set $G/H = \{gH \mid g \in G\}$ is a group under the operation $*'$, where $*'$ is defined by $(g_1H) *' (g_2H) = (g_1 * g_2)H$.*

Example 1.57 In Example 1.55 we have seen that $H = \{0, 2, 4, 6\}$ is a normal subgroup of $(\mathbb{Z}_8, +_8)$. From Example 1.51, $G/H = \{0H, 1H\}$. Then G/H is a group, with the operation $*'$ defined as $(0H) *' (0H) = (0H), (0H) *' (1H) = (1H) *' (0H) = (1H)$, and $(1H) *' (1H) = (0H)$.

Example 1.58 Consider the group $(\mathbb{Z}, +)$. Clearly $3\mathbb{Z} = \{\ldots, -6, -3, 0, 3, 6\}$ is a normal subgroup of $\mathbb{Z}$. Then $G/H = \{0\,(3\mathbb{Z}), 1\,(3\mathbb{Z}), 2\,(3\mathbb{Z})\}$ is a group, with the operation $*$ defined as $0\,(3\mathbb{Z}) *' 0\,(3\mathbb{Z}) = 0\,(3\mathbb{Z}), 0\,(3\mathbb{Z}) *' 1\,(3\mathbb{Z}) = 1\,(3\mathbb{Z}) *' 0\,(3\mathbb{Z}) = 1\,(3\mathbb{Z}), 0\,(3\mathbb{Z}) *' 2\,(3\mathbb{Z}) = 2\,(3\mathbb{Z}) *' 0\,(3\mathbb{Z}) = 2\,(3\mathbb{Z}), 2\,(3\mathbb{Z}) *' 1\,(3\mathbb{Z}) = 0\,(3\mathbb{Z}), 1\,(3\mathbb{Z}) *' 2\,(3\mathbb{Z}) = 0\,(3\mathbb{Z}), 1\,(3\mathbb{Z}) *' 1\,(3\mathbb{Z}) = 0\,(3\mathbb{Z})$ and $2\,(3\mathbb{Z}) *' 2\,(3\mathbb{Z}) = 1\,(3\mathbb{Z})$.

Theorem 1.13 (First Isomorphism Theorem) *Let ϕ be a homomorphism from a group G to a group G'. Then the mapping $\Psi : G/Ker\,(\phi) \to G'$ given by $\Psi\,(gKer(\phi)) = \phi\,(g)$ is an isomorphism. That is, $G/Ker\,(\phi) \cong \phi\,(G)$.*

Example 1.59 In Example 1.44, we have seen that $\phi(m) = m \bmod n$ is a homomorphism from $(\mathbb{Z}, +)$ and $(\mathbb{Z}_n, +_n)$ with $Ker\,(\phi) =< n >$. Therefore by Theorem 1.13, $\mathbb{Z}/ < n > \cong \mathbb{Z}_n$.

Definition 1.39 (*Ring*) A non-empty set R together with two operations $'+'$ and $'.'$, known as addition and multiplication, respectively, is called a ring (denoted by $\langle \mathcal{R}, +, . \rangle$) if the following conditions are satisfied:

(a) $(\mathcal{R}, +)$ is an Abelian group.
(b) $(\mathcal{R}, .)$ is a semi-group.
(c) For all $r_1, r_2, r_3 \in \mathcal{R}$

$$r_1.(r_2 + r_3) = r_1.r_2 + r_1.r_3 \text{ (left distributive law)}$$

$$(r_1 + r_2).r_3 = r_1.r_3 + r_2.r_3 \text{ (right distributive law)}$$

If there exists a non-zero element $1 \in \mathcal{R}$ such that for every element $r \in \mathcal{R}, r.1 = r = 1.r$, then $\langle \mathcal{R}, +, . \rangle$ is called a ring with unity and if multiplication is also commutative, then the ring is called a commutative ring.

Example 1.60 The set of all real numbers under usual addition and multiplication is a commutative ring with unity. From Example 1.33, we have $(\mathbb{R}, +)$ is an Abelian group. Clearly, the usual multiplication $'.'$ is closed, associative, and commutative over $\mathbb{R}$. Also $1 \in \mathbb{R}$ acts as unity and the distributive laws are satisfied. Similarly $\langle \mathbb{C}, +, . \rangle$,$\langle \mathbb{Q}, +, . \rangle$, and $\langle \mathbb{Z}, +, . \rangle$ are commutative rings with unity.

Example 1.61 The set $\mathbb{Z}_n = \{0, 1, 2, \ldots, n-1\}$, for $n \geq 1$, under the operations addition and multiplication modulo n (taking the integer remainder when the product is divided by n) is a ring with unity 1.

Definition 1.40 (*Sub-Ring*) A sub-ring of a ring $\mathcal{R}$ is a subset of the $\mathcal{R}$ that is a ring under the induced operations from $\mathcal{R}$.

Example 1.62 Clearly $\langle \mathbb{Q}, +, . \rangle$ is a sub-ring of $\langle \mathbb{Q}, +, . \rangle$. Also $\langle \mathbb{Q}, +, . \rangle$ is a sub-ring of $\langle \mathbb{R}, +, . \rangle$ which is again a sub-ring of $\langle \mathbb{C}, +, . \rangle$

Example 1.63 $\mathbb{Z}_n$, for $n \geq 1$, is a ring under the operation *addition modulo n* and *multiplication modulo n* (denoted by $\times_n$). The basic operation in $\times_n$ is multiplication, which ends by reducing the result modulo n; that is, taking the integer remainder when the result is divided by n as in $+_n$.

Definition 1.41 (*Division Ring*) Let $\langle \mathcal{R}, +, . \rangle$ be a ring with unity $'1'$. An element $r \in \mathcal{R}$ is a unit of $\mathcal{R}$ if it has multiplicative inverse in $\mathcal{R}$. That is, if there exists an element $r^{-1} \in \mathcal{R}$ such that $r.r^{-1} = 1 = r^{-1}.r$. If every non-zero element in $\mathcal{R}$ is a unit, then $\mathcal{R}$ is called a division ring or skew-field.

Example 1.64 $\langle \mathbb{R}, +, . \rangle$ is a division ring as for any $r (\neq 0) \in \mathbb{R}$, there exists $\frac{1}{r} \in \mathbb{R}$ such that $r \cdot \frac{1}{r} = 1 = \frac{1}{r} \cdot r$.

Theorem 1.14 *An element $m \in \mathbb{Z}_n$ is a unit if and only if $gcd(m, n) = 1$.*

Corollary 1.1 *$\mathbb{Z}_n$ is a division ring only if n is a prime.*

Definition 1.42 (*Field*) A field is a commutative division ring. In other words, $\langle \mathcal{R}, +, . \rangle$ is a field if the following conditions are satisfied:

(a) $(\mathcal{R}, +)$ is an Abelian group.
(b) $(\mathcal{R} \setminus \{0\}, .)$ is an Abelian group.

Example 1.65 The set of all real numbers $\mathbb{R}$ under usual addition and multiplication is a field. Similarly, the set of all complex numbers $\mathbb{C}$ and the set of all rational numbers $\mathbb{Q}$ under usual addition and multiplication are fields.

Example 1.66 From Corollary 1.1, the set $\mathbb{Z}_n$ is a field under the operations addition and multiplication modulo n, if and only if n is a prime (Why?). Clearly, $\langle \mathbb{Z}_n, +_n, \times_n \rangle$ is an example for a finite field.

Example 1.67 The set of all integers $\mathbb{Z}$ under usual addition and multiplication is not a field as it is not a division ring. But $\mathbb{Z}$ is a commutative ring with unity.

Definition 1.43 (*Sub-Field*) A sub-field of a field is a subset of the field that is a field under the induced operations from the field.

Example 1.68 Clearly $\langle \mathbb{Q}, +, . \rangle$ is a sub-field of $\langle \mathbb{R}, +, . \rangle$ which is again a sub-field of $\langle \mathbb{C}, +, . \rangle$.

1.4 Polynomials

Polynomials are a type of mathematical expression built by combining variables by the operations addition, subtraction, and multiplication. They are an important tool in mathematics as many mathematical problems can be encoded into polynomial equations. In this section, we will discuss some of the important properties of polynomials in one variable.

Definition 1.44 (*Ring of polynomials*) Let $\mathbb{K}$ be a field. Consider the set

$$\mathbb{K}[x] = \{a_0 + a_1 x + \cdots + a_{n-1}x^{n-1} + a_n x^n \mid a_i \in \mathbb{K}, n \in \mathbb{Z}^+\}$$

$a_i \in \mathbb{K}$ are called coefficients of the polynomial, and the order of the highest power of x with non-zero coefficient is called the degree of the polynomial. For $f(x) = a_0 + a_1 x + \cdots + a_n x^n$, $g(x) = b_0 + b_1 x + \cdots + b_m x^m \in \mathbb{K}[x]$, define

$$f(x) + g(x) = (a_0 + b_0) + (a_1 + b_1)x + \cdots + (a_{k-1} + b_{k-1})x^{k-1} + (a_k + b_k)x^k$$

where $k = max(m, n)$, $a_i = 0$ for $i > n$ and $b_i = 0$ for $i > m$. Also

$$f(x)g(x) = c_0 + c_1 x + \cdots + c_{m+n-1}x^{m+n-1} + c_{m+n}x^{m+n}$$

where $c_k = a_k b_0 + a_{k-1}b_1 + \cdots + a_1 b_{k-1} + a_0 b_k$ for $k = 0, 1, \ldots, m + n$. Then $\mathbb{K}[x]$ forms a ring with respect to the operations defined above, called the ring of polynomials over $\mathbb{K}$ in the indeterminate x.

Remark 1.5 If the coefficient of the highest power of x is the multiplicative identity of $\mathbb{K}$, then the polynomial is called a monic polynomial. Two elements in $\mathbb{K}[x]$ are equal if and only they have the same coefficients for all powers of x.

Theorem 1.15 (Division Algorithm) *Let $\mathbb{K}$ be a field and let $f(x), g(x) \in \mathbb{K}[x]$ with $g(x) \neq 0$. Then there exists unique polynomials $q(x), r(x) \in \mathbb{K}[x]$ such that $f(x) = g(x)q(x) + r(x)$ and either $r(x) = 0$ or $deg[r(x)] < deg[g(x)]$. If $r(x) = 0$ we have $f(x) = g(x)q(x)$ and we say that $g(x)$ is a factor $f(x)$.*

Theorem 1.16 *Let $\mathbb{K}$ be a field and let $f(x), g(x) \in \mathbb{K}[x]$. The greatest common divisor of $f(x)$ and $g(x)$, denoted by $(f(x), g(x))$, is the unique monic polynomial $r(x) \in \mathbb{K}[x]$ such that*

1. *$r(x)$ is a factor of both $f(x)$ and $g(x)$.*
2. *if $q(x) \in \mathbb{K}[x]$ is a factor of both $f(x)$ and $g(x)$, then $r(x)$ is a factor of $q(x)$.*

Moreover, there exists polynomials $l(x), m(x) \in \mathbb{K}[x]$ *such that*

$$r(x) = l(x)f(x) + m(x)g(x)$$

Remark 1.6 If $(f(x), g(x)) = 1$, then we say that $f(x), g(x) \in \mathbb{K}[x]$ are relatively prime.

Definition 1.45 (*Zero of a polynomial*) Let $f(x) \in \mathbb{K}[x]$; an element $\mu \in \mathbb{K}$ is called a zero (or a root) of $f(x)$ if $f(\mu) = 0$.

Theorem 1.17 (Factor Theorem) *Let* $\mathbb{K}$ *be a field and* $f(x) \in \mathbb{K}[x]$. *Then* $\mu \in \mathbb{K}$ *is a zero of* $f(x)$ *if and only if* $x - \mu$ *is a factor of* $f(x)$.

Definition 1.46 (*Algebraically Closed Field*) A field $\mathbb{K}$ is said to be an algebraically closed field, if every non-constant polynomial in $\mathbb{K}[x]$ has a root in $\mathbb{K}$.

Theorem 1.18 (Fundamental Theorem of Algebra) *The field of complex numbers is algebraically closed. In other words, every non-constant polynomial in* $\mathbb{C}[x]$ *has at least one root in* $\mathbb{C}$.

From the above theorem, we can infer that every polynomial of degree n in $\mathbb{C}[x]$ has exactly n roots in $\mathbb{C}$.

Example 1.69 Consider $x^2 + 1 \in \mathbb{R}[x]$. As the given polynomial has no root in $\mathbb{R}$, the field of real numbers is not algebraically closed, whereas if we consider $x^2 + 1$ as a polynomial in $\mathbb{C}[x]$, it has roots in $\mathbb{C}$.

Remark 1.7 (*Vieta's Formula*) Let $f(x) = a_0 + a_1x + \cdots + a_nx^n \in \mathbb{K}[x]$ with roots $x_1, x_2, \ldots, x_n$, then

$$x_1 + x_2 + \cdots + x_n = -\frac{a_{n-1}}{a_n}$$

$$x_1x_2 \cdots x_n = (-1)^n \frac{a_0}{a_n}$$

It is named after the French mathematician *Francois Viete (1540–1603).*

1.5 Matrices

A matrix in mathematics is a rectangular arrangement of numbers, symbols, or functions in rows and columns. They are of great importance in mathematics and are widely used in linear algebra to study linear transformations which will be discussed in later chapters.

Definition 1.47 An $m \times n$ matrix A over a field $\mathbb{K}$ is a rectangular array of m rows and n columns of entries from $\mathbb{K}$:

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

Such a matrix, written as $A = \left(a_{ij}\right)$, where $1 \leq i \leq m$, $1 \leq j \leq n$ is said to be of *size* (or order) $m \times n$. Two matrices are considered to be equal if they have the same size and same corresponding entries in all positions. $\mathbb{M}_{m\times n}(\mathbb{K})$ denotes the set of all $m \times n$ matrices with entries from $\mathbb{K}$.

Matrix Operations

Let us discuss some of the important operations that are used in the collection of all matrices.

Definition 1.48 (*Matrix Addition*) Let $A = \left(a_{ij}\right)$ and $B = \left(b_{ij}\right)$, where $1 \leq i \leq m$, $1 \leq j \leq n$ be any two elements of $\mathbb{M}_{m\times n}(\mathbb{K})$. Then $A + B = \left(a_{ij} + b_{ij}\right) \in \mathbb{M}_{m\times n}(\mathbb{K})$. Two matrices must have an equal number of rows and columns to be added.

Properties
For any matrices A, B and $C \in \mathbb{M}_{m\times n}(\mathbb{K})$

1. $A + B = B + A$. (Commutativity)
2. $A + (B + C) = (A + B) + C$. (Associativity)
3. There exists a matrix $O \in \mathbb{M}_{m\times n}(\mathbb{K})$ with all entries 0 such that $A + O = A$. (Existence of Identity)
4. There exists a matrix $-A$ such that $A + (-A) = O$. (Existence of Inverse)

Remark 1.8 $\mathbb{M}_{m\times n}(\mathbb{K})$ with matrix addition defined on it forms an Abelian group.

Definition 1.49 (*Matrix Multiplication*) Let $A = \left(a_{ij}\right)_{m\times n}$ and $B = \left(b_{ij}\right)_{n\times p}$. Then their product $AB \in \mathbb{M}_{m\times p}$ and its (i, j)th entry is given by

$$a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj}$$

For AB to make sense, the number of columns of A must equal the number of rows of B. Then we say that the size of matrices A and B are compatible for multiplication.

Properties
For any matrices A, B and $C \in \mathbb{M}_{n\times n}(\mathbb{K})$

1. $A(BC) = (AB)C$ (Associativity)
2. $A(B + C) = AB + AC$ and $(A + B)C = AC + BC$. (Distributive laws)

Remark 1.9 1. Matrix multiplication need not be commutative. For example, if $A = \begin{pmatrix} 1 & -1 \\ 0 & 2 \end{pmatrix}$ and $B = \begin{pmatrix} 3 & 4 & 5 \\ 6 & 0 & 8 \end{pmatrix}$ then $AB = \begin{pmatrix} -3 & 4 & -3 \\ 12 & 0 & 16 \end{pmatrix}$. Note that BA is undefined. It need not be commutative even if BA is defined. For example, if $A = \begin{pmatrix} 1 & -1 \\ 0 & 2 \end{pmatrix}$ and $B = \begin{pmatrix} 3 & 4 \\ 6 & 0 \end{pmatrix}$ then $AB = \begin{pmatrix} -3 & 4 \\ 12 & 0 \end{pmatrix}$ and $AB = \begin{pmatrix} 3 & 5 \\ 6 & -6 \end{pmatrix}$.
2. The set of all invertible matrices over the field $\mathbb{K}$ under matrix multiplication forms a non-Abelian group, denoted by $GL_n(\mathbb{K})$. Also observe that $\mathbb{M}_{n\times n}(\mathbb{K})$ forms a ring under the operations matrix addition and multiplication.

Definition 1.50 (*Scalar Multiplication*) Let $A = \left[a_{ij}\right] \in \mathbb{M}_{m\times n}(\mathbb{K})$ and $\lambda \in \mathbb{K}$, then $\lambda A = \left[\lambda a_{ij}\right] \in \mathbb{M}_{m\times n}(\mathbb{K})$.

Properties
For any matrices A, $B \in \mathbb{M}_{m\times n}(\mathbb{K})$ and λ, $\mu \in \mathbb{K}$

1. $\lambda(A + B) = \lambda A + \lambda B$
2. $(\lambda + \mu)A = \lambda A + \mu A$
3. $\lambda(\mu A) = (\lambda\mu)A$
4. $A(\lambda B) = \lambda(AB) = (\lambda A)B$.

Definition 1.51 (*Transpose of a matrix*) The transpose of an $m \times n$ matrix $A = \left[a_{ij}\right]$ is the $n \times m$ matrix (denoted by A^T), given by $A^T = \left[a_{ji}\right]$.

Properties
Let A and B be matrices of appropriate order, then

1. $\left(A^T\right)^T = A$
2. $(A + B)^T = A^T + B^T$
3. $(AB)^T = B^T A^T$
4. $(kA)^T = kA^T$.

Definition 1.52 (*Conjugate transpose of a matrix*) The conjugate transpose of an $m \times n$ matrix $A = \left[a_{ij}\right]$ is the $n \times m$ matrix (denoted by A^*) given by $A^* = \left[\overline{a_{ji}}\right]$ where bar denotes complex conjugation (if $a_{ij} = c + id$, then $\overline{a_{ij}} = c - id$).

Properties
Let A and B be matrices of appropriate orders and λ be a scalar, then

1. $(A^*)^* = A$
2. $(A + B)^* = A^* + B^*$
3. $(AB)^* = B^* A^*$
4. $(\lambda A)^* = \overline{\lambda} A^*$, where $\overline{\lambda}$ is the conjugate of λ.

Definition 1.53 (*Trace of a matrix*) Let $A = \left[a_{ij}\right]$ be an $n \times n$ matrix. The trace of A, denoted by $tr(A)$, is the sum of diagonal entries; that is $tr(A) = \sum_{i=1}^{n} a_{ii}$.

Properties
For any $n \times n$ matrices $A, B, C,$ *and* D and $\lambda \in \mathbb{R}$, we have the following properties:

1. Trace is a linear function.
   $tr(A + B) = tr(A) + tr(B)$
   $tr(\lambda A) = \lambda\, tr(A)$
2. $tr(A^T) = tr(A)$ and $tr(A^*) = \overline{(tr A)}$
3. $tr(AB) = tr(BA)$
4. $tr(ABCD) = tr(DABC) = tr(CDAB) = tr(BCDA)$
5. $tr(ABC) \neq tr(ACB)$ in general.
6. $tr(AB) \neq tr(A).tr(B)$ in general.

Definition 1.54 (*Determinant of a matrix*) For each square matrix A with entries in $\mathbb{K}$ ($\mathbb{K} = \mathbb{R}$ *or* $\mathbb{C}$), we can associate a single element of $\mathbb{K}$ called determinant of A, denoted by $det\ (A)$.
If A is a 1×1 matrix, i.e., $A = [a_{11}]$, then its determinant is defined by $det\,(A) = a_{11}$.
If A is a 2×2 matrix, say $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, then its determinant is defined by

$$det(A) = a_{11}a_{22} - a_{21}a_{12}$$

The determinant for a square matrix with higher dimension n may be defined inductively as follows:

$$det\ (A) = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} M_{ij}$$

for a fixed j, where M_{ij} is the determinant of the $(n-1) \times (n-1)$ matrix obtained from A by deleting ith row and jth column, called *minor* of the element a_{ij}.

Properties
Let A and B be any $n \times n$ matrices and λ be any scalar, then

1. $det\ (I_n) = 1$, where I_n is the $n \times n$ identity matrix.
2. $det\ (A^T) = det\ (A)$ and $det\ (A^*) = \overline{det\ (A)}$.
3. $det\ (AB) = det\ (A)\ det\ (B)$.
4. $det\ (\lambda A) = \lambda^n\ det\ (A)$.
5. If B is a matrix obtained from A by multiplying one row (or column) by a scalar λ, then $det\ (B) = \lambda\ det\ (A)$.
6. If B is a matrix obtained from A by interchanging any two rows (or columns) of A then $det\ (B) = -\ det\ (A)$.
7. If two rows of a matrix are identical then the matrix has determinant zero.
8. If B is a matrix obtained from A by adding λ times one row (or column) of A to another row (or column) of A, then $det\ (B) = det\ (A)$.

Remark 1.10 An $n \times n$ matrix with determinant zero is called *singular matrix*, otherwise it is called a *non-singular matrix.*

Definition 1.55 (*Adjoint of a Matrix*) The adjoint of a matrix $A = \left[a_{ij}\right]_{n\times n}$ (denoted by $adj\ (A)$) is the transpose of the co-factor matrix, where co-factor matrix of $A = \left[a_{ij}\right]_{n\times n}$ is $\left[(-1)^{i+j} M_{ij}\right]_{n\times n}$, where M_{ij} is the determinant of the $(n-1) \times (n-1)$ matrix obtained from A by deleting ith row and jth column, called minor of the ijth element.

Properties
Let A and B be any $n \times n$ matrices, then

1. $adj(I_n) = I_n$
2. $adj(AB) = adj(B)\ adj(A)$
3. $adj(kA) = k^{n-1} adj(A)$
4. $adj(A^m) = (adj(A))^m$
5. $adj(A^T) = (adj(A))^T$
6. $A\ adj(A) = det(A)\ I = adj(A)\ A$
7. $det\ (adj(A)) = (det(A))^{n-1}$
8. $adj\ (adj(A)) = (det(A))^{n-2}\ A$.

Definition 1.56 (*Inverse of a matrix*) The inverse of a square matrix $A_{n\times n}$ if it exists is the matrix $A^{-1}_{n\times n}$ such that $AA^{-1} = I_n = A^{-1}A$ and is given by $A^{-1} = \frac{1}{det(A)} adj(A)$.

Properties
Let A and B be any $n \times n$ matrices and λ be any scalar, then

1. The inverse of a matrix if it exists is unique.
2. A is invertible if and only if $det\ A \neq 0$.
3. $\left(A^{-1}\right)^{-1} = A$.
4. $(kA)^{-1} = k^{-1}A^{-1}$, where $k \neq 0$ is any scalar.
5. $det\ (A^{-1}) = \frac{1}{det\ (A)}$.
6. $(AB)^{-1} = B^{-1}A^{-1}$.
7. $\left(A^T\right)^{-1} = \left(A^{-1}\right)^T$.

Remark 1.11 1. There are matrices for which $AB = I$ but $BA \neq I$. For example take
$A = \begin{bmatrix}1 & 2\end{bmatrix}$ and $B = \begin{bmatrix}1\\0\end{bmatrix}$. Then $AB = \begin{bmatrix}1\end{bmatrix} = I$ and $BA = \begin{bmatrix}1 & 2\\0 & 0\end{bmatrix} \neq I$.
2. If $A = \begin{bmatrix}a & b\\c & d\end{bmatrix}$ is invertible, then A^{-1} is given by $A^{-1} = \frac{1}{ad-bc}\begin{bmatrix}d & -b\\-c & a\end{bmatrix}$.
3. Set of all $n \times n$ non-singular matrices with entries from the field $\mathbb{K}$ under matrix multiplication forms a non-Abelian group called *general linear group*, and is denoted by $GL_n(\mathbb{K})$.
    1. For any matrices $A, B \in GL_n(\mathbb{K})$, $AB \in GL_n(\mathbb{K})$ $(det(A), det(B) \neq 0 \Rightarrow det(AB) \neq 0)$. (Closure property)

2. Matrix multiplication is associative.
3. $I_n \in GL_n(\mathbb{K})$ acts as identity matrix.
4. For each $A \in GL_n(\mathbb{K})$, we have $det(A) \neq 0$ and hence A^{-1} exists. Also, $det\ (A^{-1}) = \frac{1}{det\ (A)}$, and thus $A^{-1} \in GL_n(\mathbb{K})$.

Definition 1.57 (*Rank of a matrix*) The rank of a matrix is the order of the highest order sub-matrix having non-zero determinant.

Properties

1. Let A be an $m \times n$ matrix. Then $Rank(A) \leq min\{m, n\}$.
2. Only zero matrix has rank zero.
3. A square matrix $A_{n\times n}$ is invertible if and only if $Rank(A) = n$.
4. *Sylvester's Inequality:* If A is an $m \times n$ matrix and B is an $n \times p$ matrix, then

$$Rank(A) + Rank(B) - n \leq Rank(AB) \leq min\{Rank(A), Rank(B)\}$$

   This result is named after the famous English mathematician *James Joseph Sylvester (1814–1897).*
5. *Frobenius Inequality:* Let A, B, and C be any matrices such that AB, BC, and ABC exists, then

$$Rank(AB) + Rank(BC) \leq Rank(ABC) + Rank(B)$$

   This result is named after the famous German mathematician *Ferdinand Georg Frobenius (1849–1917).*
6. Rank is sub-additive. That is, $Rank(A + B) \leq Rank(A) + Rank(B)$.
7. $Rank(A) = Rank(A^T) = Rank(A^T A)$.
8. $Rank(kA) = Rank(A)$ if $k \neq 0$.

Definition 1.58 (*Block Matrix*) A block matrix or a partitioned matrix is a matrix that is defined using smaller matrices called blocks.

Example 1.70 Consider $X = \begin{bmatrix} A & B \\ C & D \end{bmatrix}_{5\times 5}$ where $A = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}_{2\times 2}$, $B = \begin{bmatrix} 2 & 1 & 3 \\ 6 & 2 & 7 \end{bmatrix}_{2\times 3}$, $C = \begin{bmatrix} 1 & 0 \\ 5 & 2 \\ 7 & 3 \end{bmatrix}_{3\times 2}$, and $D = \begin{bmatrix} 1 & 9 & 8 \\ 4 & 2 & 1 \\ 7 & 0 & 1 \end{bmatrix}_{3\times 3}$.

Properties

1. Let $X = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$ where $A_{n\times n}$, $B_{n\times m}$, $C_{m\times n}$, and $D_{m\times m}$ are matrices.
   If A is invertible, then

$$det\ (X) = (det\ (A))\left(det\ (D - CA^{-1}B)\right)$$

Definition 1.59 (*Block Diagonal Matrix*) A block diagonal matrix is a block matrix which is a square matrix such that all blocks except the diagonal ones are zero.

Properties

1. Consider a block diagonal matrix of the form
$A = \begin{bmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_n \end{bmatrix}$, where each $A_i's$ is a square matrix. Then
   (a) $det(A) = det(A_1)det(A_2)\cdots det(A_n)$
   (b) $Tr(A) = Tr(A_1) + Tr(A_2) + \cdots + Tr(A_n)$
   (c) $Rank(A) = Rank(A_1) + Rank(A_2) + \cdots + Rank(A_n)$.

Definition 1.60 (*Elementary Operations*) There are three kinds of elementary matrix operations:

(1) Interchanging two rows (or columns).
(2) Multiplying each element in a row (or column) by a non-zero number.
(3) Multiplying a row (or column) by a non-zero number and adding the result to another row (or column).

When these operations are performed on rows, they are called *elementary row operations;* and when they are performed on columns, they are called *elementary column operations.*

Definition 1.61 (*Equivalent matrices*) Two matrices A and B are said to be row(column) equivalent if there is a sequence of elementary row(column) operations that transforms A into B and is denoted by $A \sim B$.

Definition 1.62 (*Row Echelon form of a matrix*) A matrix is said to be in row echelon form when it satisfies the following conditions:

(a) Each leading entry (the first non-zero entry in a row) is in a column to the right of the leading entry in the previous row.
(b) Rows with all zero elements, if any, are below rows having a non-zero element.

If the matrix also satisfies the condition

(c) The first non-zero element in each row, called the leading entry or pivot, is 1.

Then the matrix is in *reduced row echelon form*.

Example 1.71 Consider the matrix $A = \begin{bmatrix} 3 & 2 & 1 & 4 \\ 1 & 2 & 3 & 4 \\ 1 & 6 & 11 & 12 \end{bmatrix}$. Now

$$
\begin{aligned}
A &= \begin{bmatrix} 3 & 2 & 1 & 4 \\ 1 & 2 & 3 & 4 \\ 1 & 6 & 11 & 12 \end{bmatrix} && R_1 \leftrightarrow R_2 \\
&\sim \begin{bmatrix} 1 & 2 & 3 & 4 \\ 3 & 2 & 1 & 4 \\ 1 & 6 & 11 & 12 \end{bmatrix} && \begin{matrix} R_2 \to R_2 - 3R_1 \\ R_3 \to R_3 - R_1 \end{matrix} \\
&\sim \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & -4 & -8 & -8 \\ 0 & 4 & 8 & 8 \end{bmatrix} && R_3 \to R_3 + R_2 \\
&\sim \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & -4 & -8 & -8 \\ 0 & 0 & 0 & 0 \end{bmatrix} && R_2 \to -\tfrac{1}{4} R_2 \\
&\sim \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix} = B
\end{aligned}
$$

Then B is called the reduced row echelon form of A.

Remark 1.12 1. A matrix is equivalent to any of its row echelon form and reduced row echelon form. The reduced row echelon form of A is unique.
2. The rank of a matrix is equal to the number of non-zero rows in its row echelon form. For example, the matrix $A = \begin{bmatrix} 3 & 2 & 1 & 4 \\ 1 & 2 & 3 & 4 \\ 1 & 6 & 11 & 12 \end{bmatrix}$ has rank 2 as it is equivalent to $B = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$, which is in the row echelon form.

1.6 Euclidean Space $\mathbb{R}^n$

In a mathematical environment, Euclidean space is a geometric concept that contains all conceivable positions and locations. It provides the theoretical framework for many other mathematical fields, including classical geometry. We can use well-defined connections and rules to describe points, lines, angles, and distances inside this space. It acts as a foundational tool and gives a framework for comprehending spatial relationships. Any point in $\mathbb{R}^n$ is a list of n real numbers, denoted as $v = (v_1, v_2, \ldots, v_n)$. For convenience, we may use this list as a matrix with one column or one row called *column vector* and *row vector*, respectively. In the physical world, a vector is a quantity which has both magnitude and direction, which can be easily visualized when we work on $\mathbb{R}^2$ or $\mathbb{R}^3$.

Vectors in $\mathbb{R}^2$

Algebraically, a vector in $\mathbb{R}^2$ is simply an ordered pair of real numbers. That is $\mathbb{R}^2 = \{(v_1, v_2) \mid v_1, v_2 \in \mathbb{R}\}$. Two vectors (u_1, u_2) and (v_1, v_2) are equal if and only if the corresponding components are equal. That is, if and only if $u_1 = v_1$ and $u_2 = v_2$. Now we can define some operations on $\mathbb{R}^2$.

Definition 1.63 (*Vector Addition*) The sum of two vectors $u = (u_1, u_2)$ and $v = (v_1, v_2)$, denoted by $u + v$, is given by $u + v = (u_1 + v_1, u_2 + v_2) \in \mathbb{R}^2$.

Properties
Let $u = (u_1, u_2)$, $v = (v_1, v_2)$, $w = (w_1, w_2) \in \mathbb{R}^2$. Then

1. $u + v = (u_1 + v_1, u_2 + v_2) = (v_1 + u_1, v_2 + u_2) = v + u$. (Commutative)
2. $u + (v + w) = (u_1 + (v_1 + w_1), u_2 + (v_2 + w_2)) = ((u_1 + v_1) + w_1, (u_2 + v_2) + w_2) = (u + v) + w$. (Associative)
3. There exists $\mathbf{0} = (0, 0)$ such that $v + \mathbf{0} = v$ for all v. (Existence of identity element)
4. For each $v \in \mathbb{R}^2$, there exists $-v = (-v_1, -v_2) \in \mathbb{R}^2$ such that $v + (-v) = \mathbf{0}$. (Existence of inverse)

Remark 1.13 The set $\mathbb{R}^2$ with vector addition forms an Abelian group.

Definition 1.64 (*Scalar Multiplication*) Let $v = (v_1, v_2) \in \mathbb{R}^2$ and $\lambda \in \mathbb{R}$, then $\lambda v = (\lambda v_1, \lambda v_2) \in \mathbb{R}^2$.

Properties
Let $u = (u_1, u_2)$, $v = (v_1, v_2) \in \mathbb{R}^2$ and $\lambda, \mu \in \mathbb{R}$. Then

1. $\lambda(u + v) = (\lambda(u_1 + v_1), \lambda(u_2 + v_2)) = \lambda(u_1, u_2) + \lambda(v_1, v_2) = \lambda u + \lambda v$
2. $(\lambda + \mu)v = ((\lambda + \mu)v_1, (\lambda + \mu)v_2) = \lambda(v_1, v_2) + \mu(v_1, v_2) = \lambda v + \mu v$
3. $\lambda(\mu v) = (\lambda\mu)v = \mu(\lambda v)$.

From the above properties, it is clear that $0v = 0$ for any $v \in V$ and $0 \in \mathbb{R}$. Also, $(-1)v = -v$ for any $v \in V$ and $-1 \in \mathbb{R}$.

The Geometric Notion of Vectors in $\mathbb{R}^2$

Corresponding to every vector in $\mathbb{R}^2$, there exists a point in the Cartesian plane, and each point in the Cartesian plane represents a vector in $\mathbb{R}^2$. But the representation of vectors in $\mathbb{R}^2$ as points of Cartesian plane does not provide much information about the operations like vector addition and scalar multiplication. So it is better to represent a vector in $\mathbb{R}^2$ as a directed line segment which begins at the origin and ends at the point. Such a visualization of a vector v is called position vector of v. Then as in the physical world, the vector possess both magnitude and direction. However, to represent a vector in $\mathbb{R}^2$, the directed line segment need not start from the origin;

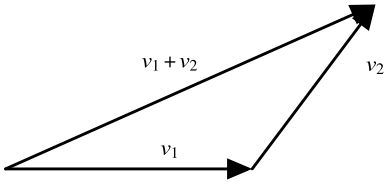


Fig. 1.16 Triangle law of vector addition

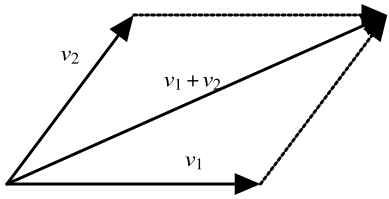


Fig. 1.17 Parallelogram law of vector addition

it may start at some point in $\mathbb{R}^2$, but the magnitude and direction cannot vary. For convenience, the directed line segment is considered to be starting from the origin.

Theorem 1.19 (Triangle Law of Vector Addition) *If two vectors are represented in magnitude and direction by the two sides of a triangle, taken in order, then their sum is represented in magnitude and direction by the third side of the triangle, taken in the reverse order (Fig. 1.16).*

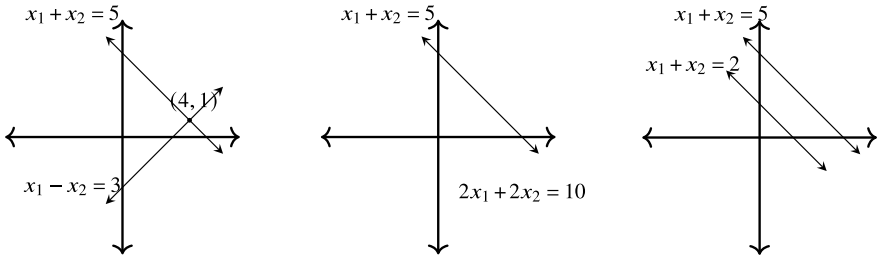
Theorem 1.20 (Parallelogram Law of vector Addition) *If two vectors are represented in magnitude and direction by the two adjacent sides of a parallelogram, then their sum is represented in magnitude and direction by the diagonal of the parallelogram through their common point (Fig. 1.17).*

These ideas of vectors and vector operations in $\mathbb{R}^2$ can be extended to general Euclidean space $\mathbb{R}^n$.

1.7 System of Linear Equations

Solving simultaneous linear equations is one among the central problems in algebra. In this section, we will get to know some of the methods that are used to solve the system of linear equations. Let us start by discussing the solution of a system having n equations in n unknowns. Consider the basic problem with $n = 1$, i.e., consider an equation of the form, $ax = b$. We know that there are three possible numerical realizations for this equation:

(1) $a \neq 0$: In this case, we know that the equation have a unique solution, which is $x = \frac{b}{a}$.
(2) $a, b = 0$: Any numerical value for x will be a solution for this equation. That is, there are infinite number of solutions.



(a) unique solution (b) infinely many solutions (c) no solution

Fig. 1.18 Observe that in **a**, the lines $x_1 + x_2 = 5$ and $x_1 - x_2 = 3$ have a unique intersection point $(4, 1)$, in **b** both the equations $x_1 + x_2 = 5$ and $2x_1 + 2x_2 = 10$ represent the same line and in **c**, the lines $x_1 + x_2 = 5$ and $x_1 + x_2 = 2$ are parallel to each other

(3) $a = 0, b \neq 0$: Then it is clear that no numerical value of x would satisfy the equation. That is, the system has no solutions.

Now consider a set of two equations in 2 unknowns x_1 and x_2:

$$a_1x_1 + a_2x_2 = b_1$$

$$a_3x_1 + a_4x_2 = b_2$$

We know that these equations represent two lines on a plane and solution of this system, if it exists, are the intersecting points of these two lines. If the lines are intersecting, either there will be a unique intersection point or there will be an infinite number of intersection points and if the lines are non-intersecting, they must be parallel to each other. Thus, here also, there are only three possibilities. The possibilities will be the same in the case of a system of n equations with n unknowns. The three possibilities are demonstrated in the Fig. 1.18.

Now that we have seen the possibilities for the number of solutions of a system of equations, we have to find a method to solve a system of linear equations. Consider a system of n equations in n unknowns $x_1, x_2, \ldots, x_n$ given by

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ \vdots \quad \vdots \quad \vdots \quad \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= b_n \end{aligned}$$

The system can be written in the form $Ax = b$, where

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}, x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \text{ and } b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}.$$

The matrix A is called the *coefficient matrix*. A method to solve this system is given by *Gabriel Cramer (1704–1752)*, using the determinants of the coefficient matrix and matrices obtained from it by replacing one column by the column vector of right-hand sides of the equations. *Cramer's rule* states that if $x = (x_1, x_2, \ldots, x_n)$ is a solution of the system, $x_i = \frac{det(A_i)}{det(A)}$, $i = 1, 2, \ldots, n$, where A_i is the matrix obtained by replacing the ith column of A by the column vector b. Observe that this rule is applicable only if $det(A) \neq 0$. For example, consider the equations $x_1 + x_2 = 5$ and $x_1 - x_2 = 3$. The system can be expressed in the form,

$$\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 5 \\ 3 \end{bmatrix}$$

As $det(A) = -2 \neq 0$, we have

$$x = \frac{det\left(\begin{bmatrix} 5 & 1 \\ 3 & -1 \end{bmatrix}\right)}{det\left(\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\right)} = 4 \text{ and } y = \frac{det\left(\begin{bmatrix} 1 & 5 \\ 1 & 3 \end{bmatrix}\right)}{det\left(\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\right)} = 1$$

As we can see, *Cramer's rule* is applicable only if the determinant of A is non-zero. Even if the determinant of A is non-zero, this rule may cause computational difficulties for higher values of n. Also it cannot be applied to a system of m equations in n unknowns. Another method to find the solution of a system of equations is *elimination*, in which multiples of one equation is added or subtracted to other equations so as to remove the unknowns from the equations till only one equation in one by unknown remains, which can be solved easily. We can use the value of this unknown to find the value of the remaining ones.

Consider a system of m equations in n unknowns $x_1, x_2, \ldots, x_n$ given by

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

The system can be written in the form $Ax = b$, where $A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$, $x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ and $b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$. The matrix A is called the *coefficient matrix*, and the matrix $[A \mid b] = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_n \end{bmatrix}$ is called the *augmented matrix* of the system. If $b = 0$, then the system is called a *homogeneous system.* Otherwise, it is called *non-homogeneous system.* A system is said to be *consistent*, if it has a solution. Otherwise, it is called *inconsistent.* We will see that a homogeneous system is always consistent, whereas a non-homogeneous system can be inconsistent (as given in Fig. 1.18c).

Gauss Elimination Method

Consider a system of equations given by $Ax = b$. We can solve the system using the following method called *Gauss elimination method*, named after the famous German mathematician *Carl Friedrich Gauss (1777–1855)*.

1. Construct the augmented matrix for the given system of equations.
2. Use elementary row operations to transform the augmented matrix to its row echelon form.
3. The system
   - is consistent if and only if $Rank\,[A \mid b] = Rank(A)$.
     - ◇ has unique solution if and only if $Rank\,[A \mid b] = Rank(A) = n$.
     - ◇ has an infinite number of solutions if $Rank\ [A \mid b] = Rank(A) = r < n$.
   - is inconsistent if and only if $Rank\,[A \mid b] \neq Rank\ (A)$.
4. If the system is consistent, write and solve the new set of equations corresponding to the row echelon form of the augmented matrix.

If reduced row echelon form is used, the method is called *Gauss–Jordan* method.

Remark 1.14 A homogeneous system $Ax = 0$ is always consistent (since $Rank\,[A \mid 0] = Rank(A)$ always). The system

- has a unique solution if $Rank(A) = n$.
- has infinite number of solutions if and only if $Rank(A) = r < n$.

Example 1.72 Consider the system of equations

$$2x_1 + 3x_2 + 5x_3 = 9$$

$$7x_1 + 3x_2 - 2x_3 = 8$$

$$2x_1 + 3x_2 + \lambda_1 x_3 = \lambda_2$$

where λ_1 and λ_2 are some real numbers.
The above system can be written in the matrix form $Ax = b$ as

$$\begin{bmatrix} 2 & 3 & 5 \\ 7 & 3 & -2 \\ 2 & 3 & \lambda_1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 9 \\ 8 \\ \lambda_2 \end{bmatrix}$$

Now the augmented matrix $[A \mid b]$ is given by

$$[A \mid b] = \begin{bmatrix} 2 & 3 & 5 & 9 \\ 7 & 3 & -2 & 8 \\ 2 & 3 & \lambda_1 & \lambda_2 \end{bmatrix} \qquad \begin{matrix} R_2 \to R_2 - \frac{7}{2}R_1 \\ R_3 \to R_3 - R_1 \end{matrix}$$

$$\sim \begin{bmatrix} 2 & 3 & 5 & 9 \\ 0 & \frac{-15}{2} & \frac{-39}{2} & \frac{-47}{2} \\ 0 & 0 & \lambda_1 - 5 & \lambda_2 - 9 \end{bmatrix}$$

As the first two rows in the reduced form are non-zero, both $Rank(A)$ and $Rank\,[A \mid b]$ are greater than or equal to 2.

- ◇ The system has unique solution if and only if $Rank\,[A \mid b] = Rank(A) = 3$. That is, if $\lambda_1 \neq 5$ and for any arbitrary values λ_2.
- ◇ The system has an infinite number of solutions if $Rank\,[A \mid b] = Rank(A) < 3$. If $\lambda_1 = 5$ and $\lambda_2 = 9$, we have $Rank\,[A \mid b] = Rank(A) = 2 < 3$.
- ◇ The system has no solution when $Rank\,[A \mid b] \neq Rank(A)$. That is, if $\lambda_1 = 5$ and $\lambda_2 \neq 9$.

If $b = 0$ in the above system, then

- ◇ The homogeneous system has a unique solution if and only if $Rank(A) = 3$. That is, if $\lambda_1 \neq 5$ the given system has only the zero vector as solution.
- ◇ If $\lambda_1 = 5$, then $Rank(A) = 2 < 3$ and hence the given system has an infinite number of solutions.

As we have identified the values of λ_1 and λ_2 for which the given system is consistent, let us try to compute the solutions of the given system for some particular values of λ_1 and λ_2. Take $\lambda_1 = 1$ and $\lambda_2 = 9$. Then,

$$[A \mid b] \sim \begin{bmatrix} 2 & 3 & 5 & 9 \\ 0 & \frac{-15}{2} & \frac{-39}{2} & \frac{-47}{2} \\ 0 & 0 & -4 & 0 \end{bmatrix}$$

That is, the given system is reduced to the following equivalent form:

$$\begin{aligned} 2x_1 + 3x_2 + 5x_3 &= 9 \\ \frac{15}{2}x_2 + \frac{39}{2}x_3 &= \frac{47}{2} \\ -4x_3 &= 0 \end{aligned}$$

Thus, we have $x = \begin{bmatrix} \frac{-1}{5} \\ \frac{47}{15} \\ 0 \end{bmatrix}$ as the unique solution for the given system. Similarly, if we take $\lambda_1 = 5$ and $\lambda_2 = 9$, we can show that set of all solutions of the given system is $\left\{(x_1, x_2, x_3) \mid x_3 \in \mathbb{R}, x_1 = \frac{14x_3 - 2}{10} \text{ and } x_2 = \frac{47 - 39x_3}{15}\right\}$ (Verify!).

Remark 1.15 If the coefficient matrix A is an $n \times n$ non-singular matrix, then the system $Ax = b$ has a unique solution $x = A^{-1}b$.

LU Decomposition

The LU decomposition method consists of factorizing A into a product of two triangular matrices

$$A = LU$$

where L is the lower triangular and U is the upper triangular. We use the *Doolittle method* to convert A into the form $A = LU$, where L and U are as mentioned above. We initialize this process by setting $A = IA$ and use Gaussian elimination procedure to achieve the desired form. The pivot element is identified in each column during this procedure, and if necessary, the rows are switched. We update the entries of both I and A on the right-hand side in accordance with each column, using row operations to remove elements below the main diagonal and multipliers to generate L. We get a lower triangular matrix L with ones on its principal diagonals and an upper triangular matrix U after iterating over all the columns. This decomposition allows us to reduce the solution of the system $Ax = b$ to solving two triangular systems $Ly = b$ and $Ux = y$. Generally, there are many such factorizations. If L is required to have all diagonal elements equal to 1, then the decomposition, when it exists, is unique. This method was introduced by the Polish mathematician *Tadeusz Julian Banachiewicz (1882–1954)*.

Example 1.73 Consider the system of equations

$$2x_1 - x_2 + 3x_3 = 9$$

$$4x_1 + 2x_2 + x_3 = 9$$

$$-6x_1 - x_2 + 2x_3 = 12$$

The above system can be written in the matrix form $Ax = b$ as

$$\begin{bmatrix} 2 & -1 & 3 \\ 4 & 2 & 1 \\ -6 & -1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 9 \\ 9 \\ 12 \end{bmatrix}$$

Consider the coefficient matrix A. We will use elementary row transformations to convert A into the form LU. We have

$$\begin{aligned} A = \begin{bmatrix} 2 & -1 & 3 \\ 4 & 2 & 1 \\ -6 & -1 & 2 \end{bmatrix} &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & -1 & 3 \\ 4 & 2 & 1 \\ -6 & -1 & 2 \end{bmatrix} \quad \begin{matrix} R_2 \to R_2 - (2)R_1 \\ R_3 \to R_3 - (-3)R_1 \end{matrix} \\ &= \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -3 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & -1 & 3 \\ 0 & 4 & -5 \\ 0 & -4 & 11 \end{bmatrix} \quad R_3 \to R_3 - (-1)R_1 \\ &= \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -3 & -1 & 1 \end{bmatrix} \begin{bmatrix} 2 & -1 & 3 \\ 0 & 4 & -5 \\ 0 & 0 & 6 \end{bmatrix} = LU \end{aligned}$$

Now $Ly = b$ implies

$$\begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -3 & -1 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 9 \\ 9 \\ 12 \end{bmatrix}$$

Solving the system, we get $y_1 = 9$, $y_2 = -9$, and $y_3 = 30$. Now consider the system $Ux = y$

$$\begin{bmatrix} 2 & -1 & 3 \\ 0 & 4 & -5 \\ 0 & 0 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 9 \\ -9 \\ 30 \end{bmatrix}$$

Solving the system, we get $x_1 = -1$, $x_2 = 4$, and $x_3 = 5$.

Theorem 1.21 *If y and z are two distinct solutions of $Ax = b$, then $\lambda y + \mu z$ is also a solution of $Ax = b$, for any scalars $\lambda, \mu \in \mathbb{K}$ with $\lambda + \mu = 1$. If $b = 0$, $\lambda y + \mu z$ is a solution of $Ax = 0$, for any scalars $\lambda, \mu \in \mathbb{K}$.*

Proof Suppose that $b \neq 0$ and y and z are two given solutions of $Ax = b$, then $Ay = b$ and $Az = b$. Let $\lambda, \mu \in \mathbb{K}$ be such that $\lambda + \mu = 1$. Then

$$A(\lambda y + \mu z) = \lambda Ay + \mu Az = \lambda b + \mu b = (\lambda + \mu)b = b$$

Now let $b = 0$. If y and z are two given solutions of $Ax = 0$, then $Ay = 0$ and $Az = 0$. Then

$$A(\lambda y + \mu z) = \lambda Ay + \mu Az = 0$$

Hence the proof.

1.8 Exercises

1. For any sets A and B, show that
   (a) $A \cap B \subseteq A, B \subseteq A \cup B$.
   (b) $A \subseteq B$ if and only if $A \cap B = A$.
2. Consider the relation $R = \{(0, 1), (0, 2), (1, 2)\}$ on $X = \{0, 1, 2\}$. Check whether R is an equivalence relation.
3. Let $f : X \to Y$ and $g : Y \to Z$ be any two functions. Then show that
   (a) if f and g are one-one, then $g \circ f$ is one-one.
   (b) if f and g are onto, then $g \circ f$ is onto.
4. Check whether the following functions are bijective or not.
   (a) $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2 + 1$
   (b) $f : [0, \pi] \to [-1, 1]$ defined by $f(x) = sin\ x$
   (c) $f : \mathbb{R}^* \to \mathbb{R}^*$ defined by $f(x) = \frac{1}{x}$
   (d) $f : \mathbb{C} \to \mathbb{C}$ defined by $f(z) = \overline{z}$.
5. Let $\lambda_i, \mu_i \in \mathbb{K}, i \in \mathbb{N}$. Then show that
   (a) for $1 < p < \infty$ and $\frac{1}{p} + \frac{1}{q} = 1$, we have

$$\sum_{i=1}^{\infty} |\lambda_i \mu_i| \leq \left(\sum_{i=1}^{\infty} |\lambda_i|^p \right)^{\frac{1}{p}} \left(\sum_{i=1}^{\infty} |\mu_i|^q \right)^{\frac{1}{q}}$$

   (b) for $1 < p < \infty$, we have

$$\left(\sum_{i=1}^{\infty} |\lambda_i + \mu_i|^p \right)^{\frac{1}{p}} \leq \left(\sum_{i=1}^{\infty} |\lambda_i|^p \right)^{\frac{1}{p}} + \left(\sum_{i=1}^{\infty} |\mu_i|^p \right)^{\frac{1}{p}}$$

   These inequalities are called *Holder's inequality* and *Minkowski's inequality*, respectively.
6. For $1 < p < \infty$, consider the following collections of sequences.

$$l^p = \left\{ v = (v_1, v_2, \ldots) \mid v_i \in \mathbb{K} \text{ and } \sum_{i=1}^{\infty} |v_i|^p < \infty \right\}$$

and

$$l^\infty = \left\{ v = (v_1, v_2, \ldots) \mid v_i \in \mathbb{K} \text{ and } \sup_{i \in \mathbb{N}} |v_i| < \infty \right\}$$

Show that for $u = (u_1, u_2, \ldots), v = (v_1, v_2, \ldots) \in l^p$

$$d_p(u, v) = \left(\sum_{i=1}^{\infty} |u_i - v_i|^p \right)^{\frac{1}{p}}$$

defines a metric on l^p and for $u = (u_1, u_2, \ldots), v = (v_1, v_2, \ldots) \in l^\infty$,

$$d_\infty(u, v) = \sup_{i \in \mathbb{N}} |u_i - v_i|$$

defines a metric on l^∞.

7. Let X be a metric space with respect to the metrics d_1 and d_2. Then show that each of the following:

   (a) $d(x, y) = d_1(x, y) + d_2(x, y)$
   (b) $d(x, y) = \dfrac{d_1(x, y)}{1 + d_1(x, y)}$
   (c) $d(x, y) = max\{d_1(x, y) + d_2(x, y)\}$

   also defines a metric on X.
8. Let (X, d) be a metric space. Show that

   (a) union of any number of open sets is open.
   (b) finite intersection of open sets is open.

   Also give an example to show that arbitrary intersection of open sets need not necessarily be open.
9. Show that a set is closed if and only if it contains all its limit points.
10. Show that $\left(l^p, d_p\right)$ and (l^∞, d_∞) are complete metric spaces.
11. Show that a closed subspace of a complete metric space is complete.
12. Prove that if a sequence of continuous functions on $[a, b]$ converges on $[a, b]$ and the convergence is uniform on $[a, b]$, then the limit function f is continuous on $[a, b]$.
13. Let $x \in \mathbb{R}$. Show that the sequence $\{x_n\}$, where $x_n = \frac{\lfloor nx \rfloor}{n}$, is a rational sequence that converges to x. ($\lfloor x \rfloor$ denotes the greatest integer less than or equal to x.)
14. Let $(G, *)$ be a group. Then show that

   (a) the identity element in G is unique.
   (b) each element in G has a unique inverse.

15. **Center of a group**: Let $(G, *)$ be group. The center of G, denoted by $\mathcal{Z}(G)$, is the set of all elements of G that commute with every other element of G.
    (a) Show that $\mathcal{Z}(G)$ is a subgroup of G.
    (b) Show that $\mathcal{Z}(G) = G$ for an Abelian group.
    (c) Find the center of $GL_2(\mathbb{K})$ and S_3.
16. Find the order of the following elements in $GL_2(\mathbb{K})$
    (a) $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$
    (b) $\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$.
17. Let $\phi : (G, *) \to (G', *')$ be a homomorphism. Then, prove the following:
    (a) if e is the identity element in G, $\phi(e)$ is the identity element in G'.
    (b) $Ker(\phi)$ is a subgroup of G.
    (c) for any $g \in G$, if $\mathcal{O}(g)$ is finite $\mathcal{O}(\phi(g))$ divides $\mathcal{O}(g)$.
    (d) for any subgroup H of G, $\phi(H)$ is a subgroup of $\phi(G)$ and if H is Abelian, $\phi(H)$ is also Abelian.
18. Consider $\phi : GL_n(\mathbb{K}) \to (\mathbb{R}^*, .)$, defined by $\phi(A) = det(A)$.
    (a) Show that ϕ is a homomorphism.
    (b) Find $Ker(\phi)$.
19. Show that every cyclic group is Abelian.
20. Find the normal subgroups of S_3.
21. Prove that $\langle \mathbb{Q}, +, . \rangle$, $\langle \mathbb{R}, +, . \rangle$, and $\langle \mathbb{C}, +, . \rangle$ are fields with respect to the given algebraic operations. Also show that $\langle \mathbb{Z}, +, . \rangle$ is not a field.
22. Give an example of a finite field.
23. Show that $\mathbb{K}[x] = \{a_0 + a_1 x + \cdots + a_{n-1}x^{n-1} + a_n x^n \mid a_i \in \mathbb{K}, n \in \mathbb{Z}^+\}$ forms a ring with respect to the operations defined in Definition 1.44.
24. Prove *the Fundamental Theorem of Algebra*.
25. Show that the set of all $n \times n$ matrices with entries in $\mathbb{K}$, denoted by $M_n(\mathbb{K})$ with matrix addition and scalar multiplication, forms a ring with unity.
26. Find the rank of the matrix $A = \begin{bmatrix} 1 & 2 & -1 & 3 \\ 4 & 5 & 3 & 6 \\ 0 & 1 & 2 & -1 \end{bmatrix}$ using row reduced echelon form.
27. Show that the set of all solutions of a homogeneous system of equations forms a group with respect to coordinate-wise addition and scalar multiplication.
28. Consider the system of equations

$$2x_1 + x_2 + 3x_3 = 9$$

$$3x_1 + 2x_2 + 5x_3 = 15$$

$$4x_1 - 2x_2 + 7x_3 = 16$$

Solve the above system of equations using

(a) Gauss Elimination method
(b) LU Decomposition method.

Is it possible to solve this system using *Cramer's rule*? If yes, find the solution using *Cramer's rule*.

Solved Questions related to this chapter are provided in Chap. 8.