LOVELY
PROFESSIONAL
UNIVERSITY

LINEAR ALGEBRA I

Edited by
Dr. Sachin Kaushal

Printed by
EXCEL BOOKS PRIVATE LIMITED
A-45, Naraina, Phase-I,
New Delhi-110028
for
Lovely Professional University
Phagwara

SYLLABUS

Linear Algebra I

Objectives: This course is designed for theoretical study of vector spaces, bases and dimension, subspaces, linear transformations, dual spaces, Elementary Canonical forms, rational and Jordan forms, inner product spaces, spectral theory and bilinear forms. It should be noted that the successful student will be able to prove simple theorems in the subject.

Sr. No.	Description
1	Vector Space over fields, Subspaces, Bases and Dimension, Coordinates, Summary of Row-Equivalence, Computation Concerning Subspaces
2	Linear Transformations, The algebra of linear transformations, The transpose of a linear transformation, Isomorphism, Representation of Transformation by matrices
3	Linear Functional, The double dual, Introduction and Characteristic Values, Annihilating Polynomials
4	Invariant Subspaces, Simultaneous triangulation, Simultaneous diagonalization, Direct-Sum Decompositions
5	Invariant Direct Sums, The Primary Decomposition Theorem, Cyclic Subspaces and Annihilators, Cyclic Decomposition and the rational Form

CONTENT

Unit 1: Vector Space over Fields

Notes

CONTENTS

Objectives

After studying this unit, you will be able to:

- Understand the concept of abstract sets
- Explain the concept of functions
- Discuss the abstract groups and their properties
- State the properties of rings and fields
- Understand abstract vector space. This will help you to understand sub-spaces, bases and dimension in the next units
- Know that this unit is a prerequisite to understand the next few units.

Introduction

In this unit the idea of set theory is explained. The unit also deals with functions and mapping.

The ideas of rings and fields help us to study vector spaces and their structure. This unit briefly explains the properties of vector spaces which are useful in understanding the vector sub-spaces, bases dimensions and co-ordinates.

1.1 Sets

The concept of set is fundamental in all branches of mathematics. A set according to the German mathematician George Cantor, is a *collection of definite well-defined objects of perception or thought*. By a well defined collection we mean that there exists a rule with the help of which it is possible to tell whether a given object belongs or does not belong to the given collection. The objects in sets may be anything: numbers, people, animals etc. The objects constituting the set are called elements or members of the set.

One should note carefully the difference between a collection and a set. Every collection is not a set. For a collection to be a set, it must be well defined. For example the collection of "any four natural numbers" is not a set. The members of this collection are not well defined. The natural number 5 may belong or may not belong to this collection. But the collection of "the first four natural numbers" is a set. Obviously, the members of the collection are well-defined. They are 1, 2, 3 and 4.

A set is usually denoted by a capital letter, such as A, B, C, X, Y, Z etc. and an element of a set by the small letter such as a, b, c, x, y, z etc.

A set may be described by actually listing the objects belonging to it. For example, the set A of single digit positive integers is written as

$$A = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$$

Here the elements are separated by commas and are enclosed in brackets { }. This is called the *tabular form* of the set.

A set may also be specified by stating properties which its elements must satisfy. The set is then described as follows:

$A = \{x : P(x)\}$ and we say that A is the set consisting of the elements x such that x satisfies the property $P(x)$. The symbol "." is read "such that". Thus the set X of all real numbers is simply written as

$$X = \{x : x \text{ is real}\} = \{x \mid x \text{ is real}\}.$$

This way of describing a set is called *the set builder* form of a set.

When a is an element of the set A, we write $a \in A$. If a is not an element of A, we write $a \notin A$.

When three elements, a, b and c, belong to the set A, we usually write $a, b, c \in A$, instead of writing $a \in A$, $b \in A$ and $c \in A$.

Two sets A and B are said to be equal iff every element of A is an element of b and also every element of B is an element of A, i.e. when both the sets consist of identical elements. We write "$A = B$" if the sets A and B are equal and "$A \neq B$" if the sets A and B are not equal.

If two sets A and B are such that every element of A is also an element of B, then A is said to be a *subset* of B. We write this relationship by writing $A \subset B$.

If $A \subset B$, then B is called a *superset* of A and we write $B \supset A$, which is read as 'B' is a super-set of A' or 'B contains A'.

If A is not a subset of B, we write $A \not\subset B$, which is read as 'A is not a subset of B'. Similarly $B \not\subset A$ is read as 'B is not a superset of A'.

From the definition of subset, it is obvious that every set is a subset of itself, i.e., $A \subset A$. We call B a proper subset of A if, first, B is a subset of A and secondly, if B is not equal to A. More briefly, B is a proper subset of A, if

$$B \subset A \text{ and } B \neq A.$$

Another improper subset of A is the set with no element in it. Such a set is called the *null set* or the *empty set*, and is denoted by the symbol ϕ. The null set ϕ is a subset of every set, i.e., $\phi \subset A$.

If A is any set, then the family of all the subsets of A is called the *power set* of A. The power set of A is denoted by $P(A)$. Obviously ϕ and A are both elements of $P(A)$. If a finite set A has n elements, then the power set of A has 2^n elements.

Example 1: If $A = \{a, b, c\}$ then $P(A) =$

$$\{\phi, \{a\}, \{b\}, \{c\}, \{a, b\} \{b, c\}, \{a, c\}, \{a, b, c\} \}.$$

The total number of these elements of power set is 8, i.e. 2^3.

The sets A and B are equal if A is a subset of B and also B is a subset of A.

If U be the universal set, the set of those elements of U which are not the elements of A is defined to be the *complement* of A. It is denoted by A'. Thus

$$A' = \{x : x \in U \text{ and } x \notin A\}$$

Obviously, $\{A'\}' = A$, $\phi' = U$, $U' = \phi$.

It is easy to see that if $A \subset B$, then $A' \supset B'$.

The difference of two sets A and B in that order is the set of elements which belong to A but which do not belong to B. We denote the difference of A and B by $A \sim B$ or $A - B$, which is read as "A difference B" or "A minus B". Symbolically $A - B = \{x : x \in A \text{ and } x \notin B\}$. It is obvious that $A - A = \phi$, and $A - \phi = A$.

Union and Intersection

Let A and B be two sets. The union of A and B is the set of all elements which are in set A or in set B. We denote the union of A and B by $A \cup B$, which is usually read as "A union B".

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

On the other hand, the intersection of A and B is the set of all elements which are both in A and B. We denote the intersection of A and B by $A \cap B$, which is usually read as "A intersection B". Symbolically,

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

or $$A \cap B = \{x : x \in A, x \in B\}.$$

The union and intersection of sets have the following simple properties:

(i) $\left.\begin{array}{l} A \cup B = B \cup A \text{ and} \\ A \cap B = B \cap A \end{array}\right\}$ Commutative laws

(ii) $\left.\begin{array}{l} A \cup (B \cup C) = (A \cup B) \cup C \text{ and} \\ A \cap (B \cap C) = (A \cap B) \cap C \end{array}\right\}$ Associative laws

(iii) $\left.\begin{array}{l} A \cup A = A \text{ and} \\ A \cap A = A \end{array}\right\}$ Idempotent laws

(iv) $\left.\begin{array}{l} A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \text{ and} \\ A \cup (B \cap C) = (A \cup B) \cap (A \cup C) \end{array}\right\}$ Distributive laws

(v) $\left.\begin{array}{l} A - (B \cup C) = (A - B) \cap (A - C) \text{ and} \\ A - (B \cap C) = (A - B) \cup (A - C) \end{array}\right\}$ De Morgan's laws

Notes

Two results which interrelate union and intersection of sets with their complements are as follows:

(i) the complement of the union is intersection of the complements, i.e.,

$(A \cup B)' = A' \cap B'$, and

(ii) the complement of the intersection is the union of the complements, i.e.,

$(A \cap B)' = A' \cup B'$.

Suppose A and B are two sets. Then the set $(A - B) \cup (B - A)$ is called the symmetric difference of the set A and B and is denoted by $A \Delta B$.

Since $\qquad (A - B) \cup (B - A) = (B - A) \cup (A - B)$

$\therefore \qquad A \Delta B = B \Delta A$.

Product Set

Let A and B be two sets, $a \in A$ and $b \in B$. Then (a, b) denotes what we may call *an ordered pair*. The element a is called the first coordinate of the ordered pair (a, b) and the element b is called its second coordinate.

If $\quad (a, b)$ and (c, d) are two ordered pairs

then $\quad (a, b) = (c, d)$ iff $a = c$ and $b = d$.

If A and B are two sets, the set of all distinct ordered pairs whose first coordinate is an element of A and whose second coordinate is an element of B is called the *Cartesian product* of A and B (in that order) and is denoted by $A \times B$. Symbolically,

$$A \times B = \{(a, b) : a \in A \text{ and } b \in B\}.$$

In general $A \times B \neq B \times A$. If A has n elements and B has m elements, then the product set $A \times B$ has nm elements. If either A or B is a null set, the $A \times B = \phi$. If either A or B is infinite and the other is not empty, the $A \times B$ is infinite.

We may generalise the definition of the product sets. Let $A_1, A_2, \ldots, A_n$ be n given sets. The set of ordered n-tuples $(a_1, a_2, \ldots, a_n)$ where $a_1 \in A_1, a_2 \in A_2, \ldots, a_n \in A_n$ is called the Cartesian product of $A_1, A_2, \ldots, A_n$ and is denoted by $A_1 \times A_2 \times A_3 \times \ldots \times A_n$.

Functions or Mappings

Let A and B be two given sets. Suppose there is a correspondence, denoted by f, which associates to each members of A, a unique member of B. Then f is called *a function* or a *mapping* from A to B. The mapping f from A to B is denoted by

$$f : A \to B \text{ or by } A \xrightarrow{f} B.$$

Suppose f is a function from A to B. The set A is called the *domain* of the function f and B is called the *co-domain* of f. The element $y \in B$ which the mapping f associates to an element $x \in A$ is denoted by $f(x)$ and is called the f-image of x or the value of the function f for x. The element x may be referred to as a pre-image of $f(x)$. Each element of A has a unique image and each element of B need not appear as the image of an element in A. There can be more than one element of A which have the same image in B. We define the *range* of f to consist of those elements of B which appear as the image of at least one element in A. We denote the range of $f : A \to B$ by $f(A)$. Thus

$$f(A) = \{f(x) : x \in A\}.$$

Obviously, $f(x) \subset B$.

If A and B are any two non-empty sets, then a function f from A to B is a subset f of $A \times B$ satisfying the following condition:

(i) $\forall a \in A$, $(a, b) \in f$ for some $b \in B$;

(ii) $(a, b) \in f$ and $(a, b') \in f \Rightarrow b = b'$

The first condition ensures that each element in A will have image. The second condition guarantees that the image is unique.

If the domain and co-domain of a function f are both the same set say $f : A \rightarrow A$, then f is often called an *Operator* or *Transformation* of A.

Two functions f and g of $A \rightarrow B$ are said to be equal iff $f(x) = g(x)$ $\forall x \in A$ and we write $f = g$. For two unequal mappings from A to B, there must exist at least one element $x \in A$ such that $f(x) \neq g(x)$.

Types of Functions

If the function $f : A \rightarrow B$ is such that there is at least one element in B which is not the f-image of any element in A, then we say that f is a function of A '*into*' B. In this case the range of f is a proper subset of the co-domain of f.

If the function $f : A \rightarrow B$ is such that each element in B is the f-image of at least one element in A, then we say that f is a function of A '*onto*' B. In this case the range of f is equal to the co-domain of f, i.e., $f(A) = B$. Onto mapping is also sometimes known as *surjection*.

A function $f : A \rightarrow B$ is said to be *one-one* or *one-to-one* if different elements in A have different f-images in B, i.e., if

$$f(x) = f(x') \Rightarrow x = x' \ (x \text{ and } x' \in A).$$

One-to-one mapping is also sometimes known as injection.

A mapping $f : A \rightarrow B$ is said to be many-one if two (or more than two) distinct elements in A have the same f-image in B.

If $f : A \rightarrow B$ is one-one and onto B, then f is called a one-to-one correspondence between A and B. One-one onto mapping is also sometimes known as bijection.

Two sets A and B are said to be have the same number of elements iff a one-to-one correspondence of A onto B exists. Such sets are said to be cardinally equivalent and we write $A \sim B$.

Let A be any set. Let the mapping $f : A \rightarrow A$ be defined by the formula $f(x) = x, \forall x \in A$, i.e. each element of A be mapped on itself. Then f is called the identity mapping on A. We shall denote this function by I_A.

Inverse Mapping

Let f be a function from A to B and let $b \in B$. Then the inverse image of the element b under f denoted by $f^{-1}(b)$ consists of those elements in A which have b as their f-image.

Let $f : A \rightarrow B$ be a one-one onto mapping. Then the mapping $f^{-1} : B \rightarrow A$, which associates to each element $b \in B$, the element $a \in A$, such that $f(a) = b$ is called the inverse mapping of the mapping $f : A \rightarrow B$.

It must be noted that the inverse mapping of $f : A \rightarrow B$ is defined only when f is one-one onto, and it is easy to see that the inverse mapping $f^{-1} : B \rightarrow A$ is also one-one and onto.

Product or Composite of Mappings

Let $f : X \to Y$ and $g : Y \to Z$. Then the composite of the mappings f and g denoted by $(g \circ f)$, is a mapping from X to Z given by $(g \circ f) : X \to Z$ such that $(g \circ f)(x) = g[f(x)],\ \forall\ x \in X$.

If $f : X \to X$ and $g : X \to X$ then we can find both the composite mappings $g \circ f$ and $f \circ g$, but in general $f \circ g \neq g \circ f$.

The composite mapping possesses the following properties:

(i) The composite mapping $g \circ f$ is one-one onto if the mappings f and g are such.

(ii) If $f : X \to \mathrm{Y}$ is a one-one onto mapping,

then $f \circ f^{-1} = I_y$ and $f^{-1} \circ f = I_x$.

(iii) If $f : X \to Y$ and $\mathrm{g} : Y \to Z$ are two one-one onto mappings, and $f^{-1} : Y \to X$ and $g^{-1} : Z \to Y$ are their inverses, then the inverse of the mapping $g \circ f : X \to Z$ is the mapping $f^{-1} \circ g^{-1} : Z \to X$.

(iv) If $f : X \to Y$, $g : Y \to Z$, $h : Z \to U$ be any mappings, then $h \circ (g \circ f)$ and $(h \circ g) \circ f$ are equal mappings of X into U, i.e. the *composite mapping is associative.*

Relation

If a and b be two elements of a set A, a relation R between them, is symbolically written as aRb, which means a in R—related to b.

For example, if R is the relation >, the statement $a\ R\ b$ means a is greater *than* b.

A relation R is said to be well defined on the set A if for each ordered pair (a, b), where $a, b \in A$, the statement $a\ R\ b$ is either true or false. A relation in a set A is a subset of the product set $A \times A$.

Inverse Relation

Let R be a relation from A to B. The inverse relation of R denoted by R^{-1}, is a relation from B to A defined by

$$R^{-1} = \{(y,x) : y \in B, x \in A, (x,y) \in A \times B\}$$

Clearly, if R is a relation from A to B, then the domain of R is identical with the range of R^{-1} and the range of R is identical with the domain of R^{-1}.

Difference between Relations and Functions

Suppose A and B are two sets. Let f be a function from A to B. Then by the definition of function f is a subset of $A \times B$ in which each $a \in A$ appears in one and only one ordered pair belonging to f. In other words f is a subset of $A \times B$ satisfying the following two conditions:

(i) for each $a \in A$, $(a, b) \in f$ for some $b \in B$,

(ii) if $(a, b) \in f$ and $(a, b') \in f$, then $b = b'$.

On the other hand every subset of $A \times B$ is a relation from A to B. *Thus every function is a relation but every relation is not a function*. If R is a relation from A to B, then domain of R may be a subset of A. But if f is a function from A to B, then domain of f is equal to A. In a relation from A to B an element of A may be related to more than one element in B. Also there may be some elements of A which may not be related to any element in B. But in a function from A to B each element of A must be associated to one and only one element of B.

Equivalence Relation

The relation R defined on a set A is to be *reflexive* if aRa holds for every a belonging to A, i.e.,

$$(a, a) \in R, \text{ for every } a \in A.$$

The relation R is said to be symmetric if

$$a\ R\ b \Rightarrow b\ R\ a$$

for every ordered pair $(a, b) \in R$, i.e.,

$$(a, b) \in R \Rightarrow (b, a) \in R.$$

The relation R is said to be transitive if

$$(a\ R\ b, b\ R\ c) \Rightarrow a\ R\ c$$

for every a, b, c belonging to A i.e.,

$$[(a, b) \in R, (b, c) \in R] \Rightarrow (a, c) \in R.$$

A relation R defined on a set is called *an equivalence relation if it is* reflexive, symmetric and transitive.

Natural Numbers

The properties of natural numbers were developed in a logical manner for the first time by the Italian mathematician G. Peano, by starting from a minimum number of simple postulates. These simple properties, known as the *Peano's Postulates* (*Axioms*), may be stated as follows:

Let there exist a non-empty set N such that.

Postulate I: $1 \in N$, that is, 1 is a natural number.

Postulate II: For each $n \in N$ there exists a unique number $n^+ \in N$, called the successor of n.

Postulate III: For each $n \in N$, we have $n^+ \neq 1$, i.e., 1 is not the successor of any natural number.

Postulate IV: If $m, n \in N$, and $m^+ = n^+$ then $m = n$, i.e. each natural number, if it is a successor, is the successor of a unique natural number.

Postulate V: If K is any subset of N having the properties (*i*) $1 \in K$ and (*ii*) $m \in K \Rightarrow m^+ \in K$, then $K = N$.

The postulate V is known as the *Postulate of induction* or the *Axiom of induction*. The *Principle of mathematical induction* is just based on this axiom.

Addition Composition

In the set of natural numbers N, we define addition, which shall be denoted by the symbol '+' as follows:

(i) $m + 1 = m^+ \ \forall\ m, \in N$

(ii) $m + n^* = (m + n)^* \ \forall\ m, n \in N.$

The distinctive properties of the addition operation in N are the closure, associative, commutative and cancellation laws, i.e., if $m, n, p \in N$, then

(A_1) $m + n \in N$ (closure law)

(A_2) $(m + n) + p = m + (n + p)$ (associative law)

Notes

(A_3) $m + n = n + m$, (commutative law)

(A_4) $m + p = n + p \Rightarrow m = n$ (cancellation law)

All these properties can be established from the foregoing postulates and definitions only.

Multiplication Composition

In the set of natural numbers N, we define multiplication which shall be denoted by the symbol 'X' as follows:

(i) $m \times 1 = m \ \forall \ m \in N$

(ii) $m \times n^+ = m \times n + m, \ \forall \ m, n \in N$.

Sometimes we often find it convenient to represent $m \times n$ by $m . n$ or simply by mn.

The following properties, which can be established from Peano's postulates, hold for multiplication.

(M_1) $m, n \in N$, or $mn \in N$, (Closure law)

(M_2) $(m . n) . p = m . (n . p)$ or $(m\, n)\, p = m\, (n\, p)$, (associative law)

(M_3) $m . n = n . m$, or $m\, n = n\, m$ (Commutative law)

(M_4) $m . p = n . p \Rightarrow m = n$, or $m\, p = n\, p \Rightarrow m = n$. (Cancellation law)

Distributive Law

The distributive property of multiplication over addition is expressed in the following two forms:

If $m, n, p \in N$, we have

(i) $m . (n + p) = m . n + n . p$ [Left distributive law]

(ii) $(m + n) . p = m . p + n . p$ [Right distributive law]

The right distributive law can also be inferred from the left distributive law, since multiplication is commutative.

Order Property

We say that a natural number m is greater than another number $n(m > n)$, if there exists a number $u \in N$, such that $m = n+ u$.

The number m is said to be less than the number $n(m < n)$, if there exists a number $v \in N$, such that $n = m + v$.

This *order relation* possesses the following property.

For any two natural numbers m and n, there exists one and only one of the following three possibilities:

(i) $m = n$

(ii) $m > n$,

(iii) $m < n$.

This is known as the *Trichotomy law* of natural numbers.

It is evident that any set of natural numbers has a smallest number, i.e., if A is a non-empty subset of N, there is a number $m \in A$, such that $m \leq n$ for every $n \in A$.

This is known as the *well ordering property* of natural numbers.

The relations between order and addition, and order and multiplication are given by the following results:

(i) $m > n \Rightarrow m + p > n + p$,

(ii) $m > n \Rightarrow m\,p > n\,p$, for all $m, n, p \in N$.

The operation of *subtracting* a number n from another number m is possible only when $m > n$, i.e., the subtraction operation is not defined for any two natural numbers. It is thus not a binary composition in N.

Similarly the operation of *dividing* one number is also not always possible, i.e., the division operation is also not a binary composition in N.

Integers

The set of integers is constructed from the set of natural numbers by defining a relation, denoted by "~" (read as wave), in $N \times N$ as follows:

$$(a,b) \sim (c,d) \text{ if } a+d = b+c,\ a,b,c,d \in N\ .$$

Since this relation is an equivalence relation it decomposes the set $N \times N$ into disjoint equivalence classes. We define the set of all these equivalence classes as the *set of integers* and denote it by Z.

The equivalence class of the pair (a, b) may be denoted by

$$(a, b) \text{ or } (a, b)^*$$

The addition and multiplication operations in Z are now defined as follows:

$(a, b)^* + (c, d)^* = (a + c, b + d)^*$

and $(a, b)^* . (c, d)^* = (ac + bd, ad + bc)^*$.

The *associative and commutative laws of addition and multiplication* hold as for natural numbers. The *cancellation law of addition* holds in general, but the cancellation law of multiplication holds with some restrictions. The *distributive law of multiplication over addition* is also valid.

The equivalence class $(1, 1)^*$ is defined as the integer *zero*, and is written as 0. Thus

$$0 = (1, 1)^* = (a, a)^* = (b, b)^*,\ a, b \in N.$$

This number 0 possesses the properties, that for any integer x,

(i) $x + 0 = x$ and

(ii) $x . 0 = 0$.

If $x = (\alpha, \beta)^*$ is an integer other than zero, we have $\alpha \neq \beta$, i.e., either $\alpha > \beta$ or $\alpha < \beta$. We say that the integer $(\alpha, \beta)^*$ is positive if $\alpha > \beta$ and *negative* if $\alpha < \beta$.

When $\alpha > \beta$, $\alpha, \beta \in N$, there exists a natural number u such that $\alpha = u + \beta$.

Therefore a positive integer x is given by

$$x = (\alpha,\beta)^*, \alpha > \beta, = (u+\beta, \beta)^* = (u+\alpha,\alpha)^*\ .$$

It is possible to identify the positive integer $(u + \alpha, \alpha)^*$ with the natural number u, and write it as $+ u$. Thus the set of positive integers may be written as

$$Z_N = \{+1, +2, +3, \ldots\}$$

Similarly, a negative integer can be identified with the number $-u$, and the set of negative integers written as

$$Z_{-N} = \{-1, -2, -3, \ldots\}$$

We define the *negative of an integer x* as the integer y, such that $x + y = 0$. It is easy to see that every integer has its negative. For, let

$$\begin{aligned} x = (a, b)^*. \text{ Then if } y &= (b, a)^*, \text{ we have} \\ x + y &= (a, b)^* + (b, a)^* = (a + b, b + a)^* \\ &= (a + b, a + b)^* = 0 \end{aligned}$$

The negative of the integer x, also called the *additive inverse* of x, is denoted by $-x$. We therefore have, for any integer x,

$$x + (-x) = 0$$

and $$x = (a, b)^* \Rightarrow -x = (b, a)^*.$$

We define *subtraction* of an integer if from an integer x as $x + (-y)$, written as $x - y$. Thus if $x = (a, b)^*$ and $y = (c, d)^*$, we have

$$\begin{aligned} x - y &= x + (-y) = (a, b) + (d, c)^* \\ &= (a + d, b + c)^* \end{aligned}$$

Order Relation in Integers

If x, y be the two integers, we define $x = y$ if $x - y$ is zero, $x > y$ if $x - y$ is positive and $x < y$ if $x - y$ is negative.

The *Trichotomy Law* for integers holds as for natural numbers. Further,

$$x > y \Rightarrow x + z > y + z,$$

and $$x > y, z > 0 \Rightarrow x z > yz, x, y, z \in Z.$$

The *cancellation law for multiplication* states that

$$xz = yz, z \neq 0 \Rightarrow x = y.$$

The addition and multiplication operations on Z satisfy the laws of natural numbers with the only modification in cancellation law of multiplication which requires $p \neq 0$. Further, the addition operation satisfies the following two properties in Z.

(i) There exists the *additive identify* 0 in the set, i.e., $0 \in Z$ such that $a + 0 = 0 + a = a$, for any $a \in Z$.

(ii) There exists the *additive inverse* of every element in Z, i.e., $a \in Z \Rightarrow -a \in Z$ such that $a + (-a) = (-a) + a = 0$.

Division

A non-zero a is said to be a *divisor* (factor) of an integer b if there exists an integer c, such that $b = ac$.

When a is divisor of b, we write "$a \mid b$". Also we say that b is an integral multiple of a. It is obvious that division is not everywhere defined in Z.

The relation of divisibility in the set of integers Z is reflexive, since $a \mid a, \forall a \in Z$. It is also transitive, since $a \mid b$ and $b \mid c \Rightarrow a \mid c$. But it is not symmetric.

The *absolute value* " $|a|$ " of an integer a is defined by

$$|a| = a \text{ when } a \geq 0$$
$$= -a \text{ when } a < 0$$

Thus, except when $a = 0$, $|a| \in Z_N$.

A non-zero integer p is called a *prime* if it is neither 1 nor –1 and if its only divisors are 1, –1, p, $-p$.

When $a = bc$ with $|b| > 1$ and $|c| > 1$, we call a *composite*. Thus every integer $a \neq 0, \pm 1$ is either a prime or composite.

The operation of division of one integer by another is carried out in accordance with the *division algorithm*, which can be stated as follows.

Given two positive integers a, b there exists uniquely two non-negative integers q, r such that

$$a = bq + r,\ 0 \leq r < b$$

The number q is called the *quotient*, and r the *remainder* obtained on dividing a by b.

Two other forms of the theorem, which are successive generalisations, are as follows:

(i) Given two integers a, b with $b > 0$, there exist unique integers q, r, such that

$$a = bq + r,\ 0 \leq r < b$$

(ii) Given two integers a, b with $b \neq 0$, there exist unique integers q, r, such that

$$a = bq + r,\ 0 \leq r < |b|.$$

Greatest Common Divisor

A *greatest common divisor* (GCD) of two integers a and b is a positive integer d such that

(i) $d \mid a$ and $d \mid b$, and

(ii) if for an integer c, $c \mid a$ and $c \mid b$, then $c \mid d$.

We shall use the notation (a, b) for the greatest common divisor of two integers a and b. The greatest common divisor is sometimes also called *highest common factor* (HCF).

Every pair of integers a and b, not both zero, has a unique greatest common divisor (a, b) which can be expressed in the form $(a, b) = ma + nb$ for some integers m and n.

Rational Numbers

Let $(a, b) \in Z \times Z_0$, where Z_0 is the set of non-zero integers. Then the equivalence class

$$\overline{(a,b)} = \{(m,n) : (m,n) \sim (a,b); m \in Z, n \in Z_0\}$$

is called a *rational number*.

The set of all equivalence classes of $Z \times Z_0$ determined by the equivalence relation ~ defined as above is called the set of rational numbers to be denoted by Q.

The addition and multiplication operations in Q are defined as follows:

$$\overline{(a,b)} + \overline{(c,d)} = \overline{(ad+bc,bd)}$$

and $$\overline{(a,b)} \cdot \overline{(c,d)} = \overline{(ac,bd)}$$

Notes

The *associative and commutative laws of addition and multiplication* hold as for integers, and so also the distributive law of multiplication over addition. The cancellation laws hold for addition and multiplication, except as for integers.

The additive identity is the number $\overline{(0,1)}$. For

$$\overline{(a,b)}+\overline{(0,1)}=\overline{(a.1+b.1)}=\overline{(a,b)}$$

The multiplicative identity is the number $\overline{(1,1)}$. For,

$$\overline{(a.b)}+\overline{(1,1)}=\overline{(a.1,b.1)}=\overline{(a,b)}$$

The additive inverse of $\overline{(a,b)}$ is $\overline{(-a,b)}$. For,

$$\overline{(a,b)}+\overline{(-a,b)}=\overline{(a.b-ba,b^2)}=\overline{(0,b^2)}=(0,1)$$

The multiplicative inverse of $\overline{(a,b)}$ is $\overline{(b,a)}$ if a ≠ 0. For,

$$\overline{(a,b)}\cdot\overline{(b,a)}=\overline{(ab,ba)}=\overline{(1,1)}$$

The additive identity $\overline{(0,1)}$, is defined as the rational number *zero* and is written as 0.

The non-zero rational number $\overline{(a,b)}$ which is such that a ≠ 0, is said to be positive or negative according as a a b is positive or negative.

The negative of a rational number z is its additive inverse; it is written as $-x$. Thus if x = $\overline{(a,b)}$ then $-x$ = $\overline{(-a,b)}$.

We define *subtraction* of a rational number y from a rational number x as $x + (-y)$, written $x - y$. Thus, if x = $\overline{(a,b)}$ and y = $\overline{(c,d)}$, we have

$$x-y=x+(-y)=\overline{(a,b)}+\overline{(-c,d)}=\overline{(ad-bc,bd)}$$

The *reciprocal* of a non-zero rational number x is its multiplicative inverse, and is written as 1/x. Thus if x = $\overline{(a,b)}$, then

$$1/x = \overline{(b,a)},\ a \neq 0,\ b \neq 0.$$

The *division* of a rational number x by a non-zero rational number y, written as $x \div y$ or $x\,|\,y$, is defined as x. $(1/y)$. Thus if x = $\overline{(a,b)}$, then

$y=\overline{(c,d)}, c \neq 0,$ we have

$$x+y=\overline{(a,b)}\cdot\overline{(d,c)}=\overline{(ad\cdot bc)}, b \neq 0, c \neq 0\ .$$

It can be shown that subtraction is a binary composition in Q, and division is also a binary composition, except for division by zero.

Order Relation

Let x, y be two rational numbers. We say that x is greater than, less than or equal to y, if $x - y$ is positive, negative or zero, and we use the usual signs to denote these relations.

if $x = \overline{(a,b)}$, $y = \overline{(c,d)}$, we have $x > y$.

if $x - y = \overline{(a,b)} + \overline{(-c,d)} = \overline{(ad-bc,bd)} > 0$,

whence we find

$(ad - bc)\, bd > 0$, i.e., $ad > bc$, $b > 0$, $d > 0$.

Similarly, $x < y$ if $ad < bc$, $b > 0$, $d > 0$.

and $\quad x = y$ if $ad = bc$.

The *Trichotomy Law* holds for rational numbers, as usual, i.e., given two rational numbers x, y either $x > y$ or $x = y$, or $x < y$.

Also the *order relation* is compatible with addition and multiplication. For,

$$x > y \Rightarrow x + z > y + z$$

and $\quad x > y, z > 0 \Rightarrow xz > yz, x, y, z \in Q$.

Representation of Rational Numbers

A rational number of the form $\overline{(a,1)}$ can be identified with the integer $a \in Z$, and written simply as a.

Further, since

$$\overline{(a,1)} \div \overline{(b,1)} = \overline{(a,1)} \cdot \overline{(1,b)} = \overline{(a,1,1\cdot b)} = \overline{(a,b)}$$

we obtain a method of representing the rational number (a, b) by means of two integers.

We have $\quad \overline{(a,b)} = \overline{(a,1)} \div \overline{(b,1)}$

$\quad = a \div b$ or $a \mid b$, $b \neq 0$.

With this notation, the sum and product of two rational numbers assume the usual meaning attached to them, viz.,

$$\frac{a}{b} + \frac{c}{d} = \frac{ad+bc}{bd}$$

and $\quad \dfrac{a}{b} \cdot \dfrac{c}{d} = \dfrac{ac}{bd}, b \neq 0, d \neq 0$

Also $\quad \dfrac{a}{b} > \dfrac{c}{d} \Rightarrow ad > bc, b > 0, d > 0$.

The system of rational numbers Q provides an extension of the system of integral Z, such that (*i*) $Q \supset Z$, (*ii*) addition and multiplication of two integers in Q have the same meanings as they have in Z and (*iii*) the subtraction and division operations are defined for any two numbers in Q, except for division by zero.

In addition to the properties described above the system of rational numbers possesses certain distinctive characteristics which distinguish it from the system of integers or natural numbers. One of these is the property of *denseness* (the density property), which is described by saying that *between any two distinct rational numbers there lies another rational number.*

Since there lies a rational number between any two rational numbers, it is clear that there lie an infinite number of rational numbers between two given rationals. This property of rational numbers make them dense every where. Evidently integral numbers or the natural numbers are not dense in this sense.

Real Numbers

We know that the equation $x^2 = 2$ has no solution in Q. Therefore if we have a square of unit length, then there exists no rational number which will give us a measure of the length of its diagonal. Thus we feel that our system of rational numbers is inadequate and we want to extend it.

The extension of rational numbers into real numbers is done by special methods two of which are due to Richard Dedekind and George Cantor. We shall not describe these methods here. We can simply say here that a real number is one which can be expressed in terms of decimals whether the decimals terminate at some state or we have a system of infinite decimals, repeating or non-repeating. We know that every repeating infinite decimals is a rational number, also every terminating decimal is a rational number.

Irrational Number

A real number which cannot be put in the form p/q where p and q are integers is called an *irrational number*. The set R of real numbers is the union of the set of rational numbers and the set of irrational numbers.

If a, b, c are real numbers, then

(i) $a + b = b + a$, $ab = ba$ (commutative of addition and multiplication)

(ii) $\left.\begin{array}{l} a+(b+c)=(a+b)+c, \\ a\,(bc)=(ab)\,c \end{array}\right\}$ Associativity of addition and multiplication

(iii) $a + 0 = 0 + a = a$, i.e., the real number 0 is the additive identity.

(iv) $a.1 = 1.a = a$, i.e., the real number 1 is the multiplicative identity.

(v) For each $a \in R$, these corresponds $-\,a \in R$ such that

$a + (-a) = -\,(a) + a = 0$

Thus every real number has an additive inverse.

(vi) Each non-zero real number has multiplicative inverse.

(vii) Multiplication composition distributes addition, i.e.,

$$a\,(b + c) = ab + ac$$

(viii) The cancellation law invariably holds good for addition. For multiplication, if $a \neq 0$, then

$$ab = ac \Rightarrow b = c$$

(ix) The order relations satisfy the *trichotomy law*.

Complex Numbers

An ordered pair (a, b) of real numbers is called a complex number. The product set $R \times R$ consisting of the ordered pairs of real numbers is called the *set of complex numbers*. We shall denote the set of complex numbers by C.

Thus

$$C = \{z : z = (a, b),\ a, b \in R\}.$$

Two complex numbers (a, b) and (c, d) are equal if and only if

$$a = c \text{ and } b = d.$$

The *sum* of two complex numbers (a, b) and (c, d) is defined to be the complex number $(a + c, b + d)$ and symbolically, we write

$$(a, b) + (c, d) = (a + c, b + d)$$

The addition of complex numbers is commutative, associative, admits of identity element and every complex number possesses additive inverse.

If u and v are two complex numbers, then $u - v = u + (-v)$.

The cancellation law for addition in C is

$$(a, b) + (c, d) = (a, b) + (e, f) \Rightarrow (c, d) = (e, f) \ \forall (a, b), (c, d), (e, f) \in C.$$

The *product* of the complex numbers (a, b) and (c, d) is defined to be the complex number $(ac - bd, ad + bc)$ and symbolically we write

$$(a, b)\,(c, d) = (ac - bd, ad + bc).$$

The multiplication of complex numbers is commutative, associative admits of identity element and every non-zero complex number possesses multiplicative inverse.

Cancellation law for multiplication in C is

$$[(a, b)\,(c, d) = (a, b)\,(e, f) \text{ and } (a, b) \neq (0, 0)] \Rightarrow (c, d) = (e, f)$$

In C multiplication distributes addition.

A complex number (a, b) is said to be divided by a complex number (c, d) if there exists a complex number (x, y) such that $(x, y)\,(c, d) = (a, b)$.

The division, except by $(0, 0)$, is always possible in the set of complex numbers.

Usual Representation of Complex Numbers

Let (a, b) be any complex number.

We have
$$\begin{aligned}(a, b) &= (a, 0) + (0, b)\\ &= (a, 0) + (0, 1)\,(b, 0)\end{aligned}$$

Also, we have $(0, 1)\,(0, 1) = (-1, 0) = -1$. If we denote the complex number $(0, 1)$ by i, we have $i^2 = -1$. Also we have $(a, b) = a + ib$, which is the usual notation for a complex number.

In the notation $Z = a + ib$ for a complex number, a is called the real part and b is called the imaginary parts. A complex number is said to be purely real if its imaginary part is zero, and purely imaginary if its part is zero but its imaginary part is not zero.

For each complex number $z = (a, b)$, we define the complex number $z = (a, -b)$ to be the conjugate of z. In our usual notation, if

$$z = a + ib$$

then
$$\bar{z} = a - ib$$

If $z = (a, b)$ be any complex number, then the non-negative real number $\sqrt{(a^2 + b^2)}$ is called the *modulus of the complex number z* and is denoted by $|z|$.

1.2 Groups

The theory of groups, an important part in present day mathematics, started early in nineteenth century in connection with the solutions of algebraic equations. Originally a group was the set of all permutations of the roots of an algebraic equation which has the property that combination of any two of these permutations again belongs to the set. Later the idea was generalized to the concept of an abstract group. An abstract group is essentially the study of a set with an operation defined on it. Group theory has many useful applications both within and outside mathematics. Group arise in a number of apparently unconnected subjects. In fact they appear in crystallography and quantum mechanics, in geometry and topology, in analysis and algebra and even in biology. Before we start talking of a group it will be fruitful to discuss the binary operation on a set because these are sets on whose elements algebraic operations can be made. We can obtain a third element of the set by combining two elements of a set. It is not true always. That is why this concept needs attention.

Binary Operation on a Set

The concept of binary operation on a set is a generalization of the standard operations like *addition* and *multiplication* on the set of numbers. For instance we know that the operation of addition (+) gives for any two natural numbers m, n another natural number $m + n$, similarly the multiplication operation gives for the pair m, n the number $m.n$ in N again. These types of operations are found to exist in many other sets. Thus we give the following definition.

Definition

A binary operation to be denoted by 'o' on a non-empty set G is a *rule* which associates to each pair of elements a, b in G a unique element $a \; o \; b$ of G.

Alternatively a binary operation 'o' on G is a mapping from $G \times G$ to G i.e. $o : G \times G \to G$ where the image of (a, b) of $G \times G$ under 'o', i.e., $o \, (a, b)$, is denoted by $a \; o \; b$.

Thus in simple language we may say that a binary operation on a set tells us how to combine any two elements of the set to get a unique element, again of the same set.

If an operation 'o' is binary on a set G, we say that G is *closed or closure property* is satisfied in G, with respect to the operation 'o'.

Examples:

(i) Usual addition (+) is binary operation on N, because if $m, n \in N$ then $m + n \in N$ as we know that sum of two natural numbers is again a natural number. But the usual substraction (–) is not binary operation on N because if $m, n \in N$ then $m - n$ may not belongs to N. For example if $m = 5$ and $n = 6$ their $m - n = 5 - 6 = -1$ which does not belong to N.

(ii) Usual addition (+) and usual substraction (–) both are binary operations on Z because if $m, n \in Z$ then $m + n \in Z$ and $m - n \in Z$.

(iii) Union, intersection and difference are *binary operations* on $P(A)$, the power set of A.

(iv) Vector product is a binary operation on the set of all 3-dimensional Vectors but the dot product is not a binary operation as the dot product is not a vector but a scalar.

Types of Binary Operations

Binary operations have the following types:

1. ***Commutative Operation:*** A binary operation o over a set G is said to be commutative, if for every pair of elements $a, b \in G$,

$$a \ o \ b = b \ o \ a$$

Thus addition and multiplication are commutative binary operations for natural numbers whereas subtraction and division are not commutative because, for $a - b = b - a$ and $a \div b = b \div a$ cannot be true for every pair of natural numbers a and b.

For example 5 – 4 ≠ 4 – 5 and 5 ÷ 4 = 4 ÷ 5.

2. ***Associative Operation:*** A binary operation o on a set G is called *associative* if $a \ o \ (b \ o \ c) = (a \ o \ b) \ o \ c$ for all $a, b, c \in G$.

Evidently ordinary addition and multiplication are associative binary operations on the set of natural numbers, integers, rational numbers and real numbers. However, if we define $a \ o \ b = a - 2b \ \forall \ , a, b \in R$

then $\quad (a \ o \ b) \ oc = (a \ o \ b) - 2c = (a - 2b) - 2c = a - 2b - 2c$

and $\quad a \ o \ (b \ o \ c) = a - 2(b \ o \ c) = a - 2(b - 2c)$

$$= a - 2b + 4c.$$

Thus the operation defined as above is not associative.

3. ***Distributive Operation:*** Let o and o' be two binary operations defined on a set, G. Then the operation o' is said to be *left distributive* with respect to operation o if

$$a \ o' \ (b \ o \ c) = (a \ o' \ b) \ o \ (a \ o' \ c) \text{ for all } a, b, c \in G$$

and is said to be right distributive with respect to o if,

$$(b \ o \ c) \ o' \ a = (b \ o' \ a) \ o \ (c \ o' \ c) \text{ for } a, b, c, \in G.$$

Whenever the operation o is left as well as right distributive, we simply say that o is distributive with respect to o.

Identity and Inverse

Identity: A composition o in a set G is said to admit of an identity if these exists an element $e \in G$ such that

$$a \ o \ e = a = e \ o \ a \ \forall \ a \in G.$$

Moreover, the element e, if it exists is called an identity element and the algebraic structure (G, o) is said to have an identity element with respect to o.

Examples:

(i) If a ∈ R, the set of real numbers then 0 (zero) is an additive identity of R because

$$a + 0 = a = 0 + a \ \forall \ a \in R$$

N the set of natural numbers, has no identity element with respect to addition because $0 \in N$.

(ii) 1 is the multiplicative identity of N as

$$a.1 = 1.a = a \,\forall\, a \in N.$$

Evidently 1 is identity of multiplication for I (set of integers), Q (set of rational numbers, R (set of real numbers).

Inverse: An element $a \in G$ is said to have its inverse with respect to certain operation o if there exists $b \in G$ such that

$$a \; o \; b = e = b \; o \; a.$$

e being the identity in G with respect to o.

Such an element b, usually denoted by a^{-1} is called the inverse of a. Thus $a^{-1} \; o \; a = e = a \; o \; a^{-1}$ for $a \in G$.

In the set of integers the inverse of an integer a with respect to ordinary addition operation is $-a$ and in the set of non-zero rational numbers, the inverse of a with respect to multiplication is $1/a$ which belongs to the set.

Algebraic Structure

A non-empty set G together with at least one binary operation defined on it is called an *algebraic structure*. Thus if G is a non-empty set and 'o' is a binary operation on G, then (G, o) is an algebraic structure.

$$(n, +), (I, +), (I, -), (R, +, .)$$

are all algebraic structures. Since addition and multiplication are both binary operations on the set R of real numbers, $(R, +, .)$ is an algebraic structure equipped with two operations.

Illustrative Examples

Example 2: If the binary operation o on Q the set of rational numbers is defined by

$$a \; o \; b = a + b - a\,b, \text{ for every } a, b \in Q$$

show that Q is commutative and associative.

Solution:

(i) 'o' is commutative in Q because if $a, b \in Q$, then

$a \; o \; b = a + b - a\,b = b + a - b\,a = b \; o \; a.$

(ii) 'o' is associative in Q because if $a, b, c \in Q$ then

$$\begin{aligned} a \; o \; (b \; o \; c) &= a \; o \; (b + c - b\,c) \\ &= a + (b + c - b\,c) - a\,(b + c - b\,c) \\ &= a + b - a\,b + c - (a + b - a\,b)\,c \\ &= (a \; o \; b) \; oc. \end{aligned}$$

Example 3: Given that $S = \{A, B, C, D\}$ where $A = \phi$, $B = \{a\}$, and $C = \{a, b\}$. $D = \{a, b, c\}$ show that S is closed under the binary operations $\cup$ (union of sets) and $\cap$ (intersection of sets) on S.

Solution:

(i) $A \cap B = \phi \cap \{a\} = \{a\} = B$

Similarly, $A \cap C = C, A \cap D$ and $A \cap A = A$.

Also, $B \cap B = B, B \cap C = \{a\} \cap \{a, b\} = \{a, b\} = C,$

$B \cap D = \{a\} \cap \{a, b, c\} = \{a, b, c\} = D$

$C \cap C = C, C \cap D = \{a, b\} \cap \{a, b, c\} = \{a, b, c\} = D$

Hence $\cap$ is a binary operation on S.

(ii) Again, $A \cup A = A, A \cup B = \phi \cup \{a\} = \phi = A$

$A \cup C = A, A \cup D = A$

and $B \cup B = B, B \cup C = \{a\} \cup \{a, b\} = \{a\} = B$

$B \cup D = \{a\} \cup \{a, b, c\} = \{a\} = B$

$C \cup C = C, C \cup D = \{a, b\} \cup \{a, b, c\}$

$= \{a, b\} = C.$

Hence $\cup$ is a binary operation on S.

Self Assessment

1. Show that multiplication is a binary operation on the set $A = \{1, -1\}$ but not on $B = \{1, 3\}$.
2. If $A = \{1, -1\}$ and $B = \{1, 2\}$, then show that multiplication is a binary operation on A but not on B.
3. If $S = \{A, B, C, D\}$ where $A = \phi$, $B = \{a, b\}$, $C = \{a, c\}$, $D = \{a, b, c\}$ show that $\cap$ is a binary operation on S but $\cup$ is not.

Group

Definition: *An algebraic structure (G, o) where G is a non-empty set with a binary operation 'o' defined on it is said to be a group, if the binary operation satisfies the following axioms (called group axioms).*

(G_1) ***Closure Axiom:*** G is closed under the operation o, i.e., $a\ o\ b \in G$, for all $a, b \in G$.

(G_2) ***Associative Axiom:*** The binary operation o is associative, i.e.,

$(a\ o\ b)\ o\ c = a\ o\ (b\ o\ c)\ \forall\ a, b, \in G.$

(G_3) ***Identity Axiom:*** There exists an element $e \in G$ such that

$$e\ o\ a = a\ o\ e = a\ \forall\ a \in G.$$

The element e is called the identity of 'o' in G.

(G_4) ***Inverse Axiom:*** Each element of G possesses inverse, i.e., for each element $a \in G$, there exists an element $b \in G$ such that

$$b\ o\ a = a\ o\ b = e.$$

The element b is then called the inverse of a with respect to 'o' and we write $b = a^{-1}$. Thus a^{-1} is an element of G such that

$$a^{-1}\ o\ a = a\ o\ a^{-1} = e.$$

Abelian Group of Commutative Group

A group (G, o) is said to be abelian or commutative if the composition 'o' is commutative, i.e., if

$$a \; o \; b = b \; o \; a \;\; \forall \; a, b \in G$$

A group which is not abelian is called non-abelian.

Examples:

(i) The structures (N, +) and (N, ×) are not groups i.e., the set of natural numbers considered with the addition composition or the multiplication composition, does not form a group. For, the postulate (G_3) and (G_4) in the former case, and (G_4) in the latter case, are not satisfied.

(ii) The structure (Z, +) is a group, i.e., the set of integers with the addition composition is a group. This is so because addition in numbers is associative, the additive identity O belongs to Z, and the inverse of every element a, viz., $-a$ belongs to Z. This is known as *additive group of integers*.

The structure (Z, ×), i.e., the set of integers with the multiplication composition does not form a group, as the axiom (G_4) is not satisfied.

(iii) The structures (Q, +), (R, +), (C, +) are all groups i.e., the sets of rational numbers, real numbers, complex numbers, each with the additive composition, form a group.

But the same sets with the multiplication composition do not form a group, for the multiplicative inverse of the number zero does not exist in any of them.

(iv) The structure (Q_0, x) is a group, where Q_0 is the set of non-zero rational numbers. This is so because the operation is associative, the multiplicative identity 1 belongs to Q_0, and the multiplicative inverse of every element a in the set is $1/a$, which also belongs to Q_0. This is known as the *multiplicative group of non-zero rationals*.

Obviously (R_0, X) and (C_0, X) are groups, where R_0 and C_0 are respectively the sets of non-zero real numbers and non-zero complex numbers.

(v) The structure (Q^+, ×) is a group, where Q^+ is the set of positive rational numbers. It can easily be seen that all the postulates of a group are satisfied.

Similarly, the structure (R^+, ×) is a group, where R^+ is the set of positive real numbers.

(vi) The groups in (*ii*), (*iii*), (*iv*) and (*v*) above are all *abelian groups,* since addition and multiplication are both commutative operations in numbers.

Finite and Infinite Groups

If a group contains a finite number of distinct elements, it is called *finite group* otherwise an *infinite group*.

In other words, a group (G, 0) is said to be finite or infinite according as the underlying set G is finite or infinite.

Order of a Group

The number of elements in a finite group is called the order of the group. An infinite group is said to be of infinite order.

Notes

Note: It should be noted that the smallest group for a given composition is the set $\{e\}$ consisting of the identity element e alone.

Illustrative Examples

Example 4: Show that the set of all integers ……, –4, –3, –2, –1, 0, 1, 2, 3, 4, … is an infinite abelian group with respect to the operation of addition of integers.

Solution: Let us test all the group axioms for abelian group.

(G_1) ***Closure Axiom:*** We know that the sum of any two integers is also an integer, i.e., for all $a, b \in I$, $a + b \in I$. Thus I is *closed* with respect to addition.

(G_2) ***Associativity:*** Since the addition of integers is associative, the associative—axiom is satisfied, i.e., for $a, b, c \in I$.

$$a + (b + c) = (a + b) + c$$

(G_3) ***Existence of Identity:*** We know that O is the additive identity and $O \in I$, i.e.,

$$O + a = a = a + O \,\forall\, a \in \mathrm{I}$$

Hence additive identity exists.

(G_4) ***Existence of Inverse:*** If $a \in I$, then $-a \in I$. Also,

$$(-a) + a = O = a + (-a)$$

Thus every integer possesses additive inverse.

Therefore I is a group with respect to addition.

Since addition of integers is a *commutative operation*, therefore $a + b = b + a \,\forall\, a, b \in I$.

Hence $(I, +)$ is an *abelian group*. Also, I contains an infinite number of elements. Therefore $(I, +)$ is an abelian group of infinite order.

Example 5: Show that the set of all even integers (including zero) with additive property is an abelian group.

Solution: The set of all even integers (including zero) is

$$I = \{0, \pm 2, \pm 4, \pm 6\ldots\}$$

Now, we will discuss the group axioms one by one:

(G_1) The sum of two even integers is always an even integer, therefore *closure axiom* is satisfied.

(G_2) The addition is associative for even integers, hence *associative axiom* is satisfied.

(G_3) $O \in I$, which is an additive identity in I, hence identity axiom is satisfied.

(G_4) Inverse of an even integer a is the even integer $-a$ in the set, so *axiom of inverse* is satisfied.

(G_5) Commutative law is also satisfied for addition of even integers. Hence the set forms an abelian group.

Example 6: Show that the set of all non-zero rational numbers with respect to binary operation of multiplication is a group.

Solution: Let the given set be denoted by Q_0. Then by group axioms, we have—

(G_1) We know that the product of two non-zero rational numbers is also a non-zero rational number. Therefore Q_0 is closed with respect to multiplication. Hence, *closure axiom* is satisfied.

(G_2) We know for rational numbers.

$$(a \cdot b) \cdot c = a \cdot (b \cdot c) \text{ for all } a, b, c \in Q_0$$

Hence, associative axiom is satisfied.

(G_3) Since, 1 the multiplicative identity is a rational number hence identity axiom is satisfied.

(G_4) If $a \in Q_0$, then obviously, $1/a \in Q_0$. Also

$$1/a \, . \, a = 1 = a \, . \, 1/a$$

so that $1/a$ is the multiplicative inverse of a. Thus *inverse axiom* is also satisfied.

Hence Q_0 is a group with respect to multiplication.

Example 7: Show that C, the set of all non-zero complex numbers is a multiplicative group.

Solution: Let $C = \{z : z = x + i\, y, x, y \in R\}$

Hence R is the set of all real numbers are $i = \sqrt{(-1)}$.

(G_1) ***Closure Axiom:*** If $a + i\, b \in C$ and $c + id \in c$, then by definition of multiplication of complex numbers

$$(a + i\, b)\,\{(c + i\, d) = (a\, c - b\, d) + i\,(a\, d + b\, c) \in C,$$

since $\quad a\, c - b\, d, a\, d + b\, c \in R, \ \textit{for} \ \ a, b, c, d \in R$.

Therefore, C is closed under multiplication.

(G_2) ***Associative Axiom:***

$$(a + i\, b)\,\{(c + i\, d) \cdot (e + i\, f)\} = (a\, c\, e - a\, d\, f - b\, c\, f - b\, d\, e) + i\,(a\, c\, f + a\, d\, e + b\, c\, e - b\, d\, f)$$

$$= \{(a + i\, b) \cdot (c + i\, d)\} \cdot (e + i\, f)$$

for $a, b, c, d \in R$.

(G_3) ***Identity Axiom:*** $e = 1$ ($= 1 + i_0$) is the identity in C.

(G_4) ***Inverse Axiom:*** Let $(a + i\, b)$ ($\neq 0$) $\in C$, then

$$(a + ib)^{-1} = \frac{1}{a + ib} = \frac{a - ib}{a^2 + b^2}$$

$$= \left(\frac{a}{a^2 + b^2}\right) + i\left(\frac{b}{a^2 + b^2}\right)$$

$$= m + i\, n \in C, \text{ Where } m = \left(\frac{a}{a^2 + b^2}\right),$$

$$n = -\frac{b}{a^2+b^2} \in R.$$

Hence C is a multiplicative group.

Self Assessment

4. Show that the set of all odd integers with addition as operation is not a group.

5. Verify that the totality of all positive rationals form a group under the composition defined by

$$a \; o \; b = ab/2$$

6. Show that the set of all numbers cos θ + *i* sin θ forms an infinite abelian group with respect to ordinary multiplication; where θ runs over all rational numbers.

Composition (Operation) Table

A binary operation in a finite set can completely be described by means of a table. This table is known as *composition table*. The composition table helps us to verify most of the properties satisfied by the binary operations.

This table can be formed as follows:

(i) Write the elements of the set (which are finite in number) in a row as well as in a column.

(ii) Write the element associated to the ordered pair (a_i, a_j) at the intersection of the row headed by a_i and the column headed by a_j. Thus (i^{th} entry on the left). (j^{th} entry on the top) = entry where the i^{th} row and j^{th} column intersect.

For example, the composition table for the group {0, 1, 2, 3, 4} for the operation of addition is given below:

	0	*1*	*2*	*3*	*4*
0	0	1	2	3	4
1	1	2	3	4	5
2	2	3	4	5	6
3	3	4	5	6	7
4	4	5	6	7	8

In the above example, the first element of the first row in the body of the table, 0 is obtained by adding the first element 0 of head row and the first element 0 of the head column. Similarly the third element of 4th row (5) is obtained by adding the third element 2 of the head row and the fourth element of the head column and so on.

An operation represented by the composition table will be binary, if every entry of the composition table belongs to the given set. It is to be noted that composition table contains all possible combinations of two elements of the set will respect to the operation.

Notes:

(i) It should be noted that the elements of the set should be written in the same order both in top border and left border of the table, while preparing the composition table.

(ii) Generally a table which defines a binary operation '.' on a set is called *multiplication table*, when the operation is '+' the table is called an addition table.

Group Tables

The composition tables are useful in examining the following axioms in the manner explained below:

1. ***Closure Property:*** If all the elements of the table belong to the set *G* (say) then *G* is closed under the Composition *o* (say). If any of the elements of the table does not belong to the set, the set is not closed.

2. ***Existence of Identity:*** The element (in the vertical column) to the left of the row identical to the top row (border row) is called an identity element in the*G* with respect to operation '*o*'.

3. ***Existence of Inverse:*** If we mark the identity elements in the table then the element at the top of the column passing through the identity element is the inverse of the element in the extreme left of the row passing through the identity element and vice versa.

4. ***Commutativity:*** If the table is such that the entries in every row coincide with the corresponding entries in the corresponding column i.e., the composition table is symmetrical about the principal or main diagonal, the composition is said to have satisfied the commutative axiom otherwise it is not commutative.

The process will be more clear with the help of following illustrative examples.

Illustrative Examples

Example 8: Prove that the set of cube roots of unity is an abelian finite group with respect to multiplication.

Solution: The set of cube roots of unity is $G = \{1, \omega, \omega^2\}$. Let us form the composition table as given below:

	I	ω	ω^2
1	1	ω	ω^2
ω	ω	ω^2	$\omega^3 = 1$
ω^2	ω^2	$\omega^3 = 1$	$\omega^4 = \omega$

(G_1) ***Closure Axiom:*** Since each element obtained in the table is a unique element of the given set *G*, multiplication is a binary operation. Thus the closure axiom is satisfied.

(G_2) ***Associative Axiom:*** The elements of *G* are all complex numbers and we know that multiplication of complex number is always associative. Hence associative axiom is also satisfied.

(G_3) ***Identity Axiom:*** Since row 1 of the table is identical with the top border row of elements of the set, 1 (the element to the extreme left of this row) is the identity element in *G*.

(G_4) ***Inverse Axiom:*** The inverse of 1, ω, ω^2 are 1, ω^2 and ω respectively.

(G_5) ***Commutative Axiom:*** Multiplication is commutative in *G* because the elements equidistant with the main diagonal are equal to each other.

The number of elements in *G* is 3. Hence (*G*,.) is a finite group of order 3.

Example 9: Prove that the set {1, –1, i, –i} is abelian multiplicative finite group of order 4.

Solution: Let G = {1, –1, i, –i}. The following will be the composition table for (G,.)

	1	–1	i	–i
1	1	–1	i	–i
–1	–i	1	–i	–i
i	i	–i	–1	1
–i	–i	i	1	–1

(G_1) ***Closure Axiom:*** Since all the entries in the composition table are elements of the set G, the set G is closed under the operation multiplication. Hence closure axiom is satisfied.

(G_2) ***Associative Axiom:*** Multiplication for complex numbers is always associative.

(G_3) ***Identity Axiom:*** Row 1 of the table is identical with that at the top border, hence the element 1 in the extreme left column heading row 1 is the identity element.

(G_4) ***Inverse Axiom:*** Inverse of 1 is 1. Inverse of –1 is –1. Inverse of i is –i and of –i is i. Hence inverse axiom is satisfied in G.

(G_5) ***Commutative Axiom:*** Since in the table the 1st row is identical with 1st column, 2nd row is identical with the 2nd column, 3rd row is identical with the 3rd column and 4th row is identical with the 4th column, hence the multiplication in G is commutative.

The number of elements in G is 4. Hence G is an abelian finite group of order 4 with respect to multiplication.

General Properties of Groups

Theorem 1: The identity element of a group is unique.

Proof: Let us suppose e and e' are two identity elements of group G, with respect to operation o.

Then $e \; o \; e' = e$ if e' is identity.

and $e \; o \; e' = e'$ if e is identity.

But $e \; o \; e'$ is unique element of G, therefore,

$e \; o \; e' = e$ and $e \; o \; e' = e \Rightarrow e = e'$

Hence the identity element in a group is unique.

Theorem 2: The inverse of each element of a group is unique, i.e., in a group G with operation o' for every $a \in G$, there is only one element a^{-1} such that $a^{-1} \; oa = a \; o \; a^{-1} = e$, e being the identity.

Proof: Let a be any element of a group G and let e be the identity element. Suppose there exist a^{-1} and a' two inverses of a in G then

$$a^{-1} \; o \; a \;= e = a \; o \; a^{-1}$$

and

$$a' \; o \; a \;= e = a' \; o \; a$$

Now, we have

$$a^{-1} \; o \; (a \; o \; a') \;= a^{-1} \; o \; e \text{ (since } a \; o \; a' = e)$$

Notes

$= a^{-1}$ (because e is identity)

Also, $(a^{-1}\ o\ a)\ o\ a' = e\ o\ a'$ (because $a^{-1}\ o\ a = e$)

$= a'$ (because e is identity)

But $a^{-1}\ o\ (ao\ a') = (a^{-1}\ oa)\ o\ a'$ as in a group composition is associative

$\therefore \quad a^{-1} = a'.$

Theorem 3: If the inverse of a is a^{-1} then the inverse of a^{-1} is a, i.e., $(a^{-1})^{-1} = a$.

Proof: If e is the identity element, we have

$a^{-1}\ o\ a = e$ (by definition of inverse)

$\Rightarrow (a^{-1})^{1}\ o\ (a^{-1}\ o\ a) = (a^{-1})^{-1}\ o\ e$

[because $a^{-1} \in G \Rightarrow (a^{-1})^{-1} \in G$]

$\Rightarrow [(a^{-1})^{-1}\ o\ a^{-1}]\ o\ a = (a^{-1})^{-1}$

[because Composition in G is associative and e is identity element]

$\Rightarrow e\ o\ a = (a^{-1})^{-1}$

$\Rightarrow a = (a^{-1})^{-1}$

$\Rightarrow (a^{-1})^{-1} = a.$

Theorem 4: The inverse of the product of two elements of a group G is the product of the inverse taken in the reverse order i.e.,

$$(a\ o\ b)^{-1} = b^{-1}\ o\ a^{-1} \ \forall\ a, b \in G.$$

Proof: Let us suppose a and b are any two elements of G. If a^{-1} and b^{-1} are inverses of a and b respectively, then

$a^{-1}\ o\ a = e = a\ o\ b^{-1}$ (e being the identity element)

and $\quad b^{-1}\ o\ b = e = b\ o\ b^{-1}$

Now, $(a\ o\ b)\ o\ b^{-1}\ o\ a^{-1} = [(a\ o\ b)\ ob^{-1}]\ o\ a^{-1}$ (by associativity)

$= [a\ o\ (b\ o\ b^{-1})]\ o\ a^{-1}$ (by associativity)

$= (a\ o\ e)\ o\ a^{-1}$ [because $b\ o\ b^{-1} = e$]

$= a\ o\ a^{-1}$ [because $a\ o\ e = a$]

$= e$ [because $a\ o\ a^{-1} = e$]

Also $(b^{-1}\ o\ a^{-1})\ o\ (aob) = b^{-1}\ o\ [a^{-1}\ o\ (a\ o\ b)]$ (by associativity)

$= b^{-1}\ o\ [(a^{-1}\ o\ a)\ ob]$

$= b^{-1}\ o\ (e\ o\ b)$ [because $a^{-1}\ o\ a = e$]

$= b^{-1}\ o\ b$ [because $e\ o\ b = b$]

$= e.$

Hence, we have

$(b^{-1}\ o\ a^{-1}\ o\ (a\ o\ b) = e = (a\ o\ b)\ o\ (b^{-1}\ o\ a^{-1})$

Therefore, by definition of inverse, we have

$(a\ o\ b)^{-1} = b^{-1}\ o\ a^{-1}$

This theorem can be generalised as:

if $a, b, c, \dots k, l, m \in G$, then

$$(a \text{ o } b \text{ o } c \text{ o } \dots k \text{ o } l \text{ o } m)^{-1} = m^{-1} \text{ o } l^{-1} \text{ o } k^{-1} \text{ o. .} c^{-1} \text{ o } b^{-1} \text{ o } a^{-1}.$$

Theorem 5: Cancellation laws hold good in a group, i.e., if $a, b, c,$ are any elements of G, then

$$a \text{ o } b = a \text{ o } c \Rightarrow b = c \qquad \text{(left cancellation law)}$$

and $$b \text{ o } a = c \text{ o } a \Rightarrow b = c \qquad \text{(right cancellation law)}$$

Proof: Let $a \in G$. Then

$a \in G \Rightarrow a^{-1} \in G$ such that $a^{-1} \text{ o } a = e$

$= a \text{ o } a^{-1}$, where e is the identity element

Now, let us assume that

$a \text{ o } b = a \text{ o } c$

then $a \text{ o } b = a \text{ o } c \Rightarrow a^{-1} \text{ o } (a \text{ o } b) = a^{-1} \text{ o } aoc$

$\Rightarrow (a^{-1} \text{ } oa) \text{ } ob = (a^{-1} \text{ } oa) \text{ } oc$ (by associative law)

$\Rightarrow eob = eoc$ (because $a^{-1} \text{ } oa = e$)

$\Rightarrow b = c.$

Similarly, $b \text{ o } a = c \text{ o } a$

$\Rightarrow (boa) \text{ o } a^{-1} = (coa) \text{ o } a^{-1}$

$\Rightarrow bo \text{ } (ao \text{ } a^{-1}) = co \text{ } (ao \text{ } a^{-1})$

$\Rightarrow boe = coe$

$\Rightarrow b = c.$

Theorem 6: If G is a group with binary operation o and if a and b are any elements of G, then the linear equations

$$aox = b \text{ and } yoa = b$$

have unique solutions in G.

Proof: Now $a \in G \Rightarrow a^{-1} \in G,$

and $a^{-1} \in G, b \in G \Rightarrow a^{-1} \text{ } ob \in G.$

Substituting $a^{-1} \text{ } ob$ for x in the equation $aox = b$, we obtain

$a \text{ o } (a^{-1} \text{ o } b) = b$

$\Rightarrow (a \text{ o } a^{-1}) \text{ o } b = b$

$\Rightarrow e \text{ o } b = b$

$\Rightarrow b = b$ [because e is identity]

Thus $x = a^{-1} \text{ } ob$ is a solution of the equation $aox = b$.

To show that the solution is unique let us suppose that the equation $aox = b$ has two solutions given by

$$x = x_1 \text{ and } x = x_2$$

Notes

Then $aox_1 = b$ and aox_2

$$\Rightarrow aox_1 = aox_2 = b$$

$$\Rightarrow x_1 = x_2 \quad \text{(by left cancellation law)}$$

In a similar manner, we can prove that the equation

$$y\ o\ a = b$$

has the unique solution

$$y = b\ o\ a^{-1}.$$

Theorem 7: If corresponding to any element a ∈ G; there is an element O_a which satisfies one of the conditions

$$a + O_a = a \text{ or } O_a + a = a$$

then it is necessary that $O_a = o$, where O_a is the identity element of the group.

Proof: Since o is the identity element,

We have

$$a + o = a \quad \dots \text{(i)}$$

also, it is given that

$$a + O_a = a \quad \dots \text{(ii)}$$

Hence, from (i) and (ii)

$$a + O_a = a + O$$

or $O_a = o$ (by left cancellation law)

Again, we have

$$o + a = a \quad \dots \text{(iii)}$$

and $O_a + a = a$ (given) … (iv)

Hence, from (iii) and (iv), we get

$$O_a + a = o + a$$

so that $O_a = o$ (by right cancellation law.)

Modulo System

It is of common experience that railway time-table is fixed with the provision of 24 hours in a day and night. When we say that a particular train is arriving at 15 hours, it implies that the train will arrive at 3 p.m. according to our watch.

Thus all the timing starting from 12 to 23 hours correspond to one of 0, 1, 3… 11 o'clock as indicated in watches. In other words all integers from 12 to 23 one equivalent to one or the other of integers 0, 1, 2, 3, …, 11 with modulo 12. In saying like this the integers in question are divided into 12 classes.

In the manner described above the integer could be divided into 2 classes, or 5 classes or m (m being a positive integer) classes and then we would have written mod 2 or mod 5 or mod m. This system of representing integers is called modulo system.

Addition Modulo m

We shall now define a new type of addition known as "addition modulo m" and written as $a +_m b$ where a and b are any integers and m is a fixed positive integer.

By definition, we have

$$a + \widehat{m}b = r, \quad 0 \le r \le m$$

where r is the least non-negative remainder when $a + b$, i.e., the ordinary sum of the a and b, is divided by m.

For example $5 +_6 3 = 2$, since $5 + 3 = 8 = 1\,(6) + 2$, i.e., 2 is the least non-negative remainder when $5 + 3$ is divided by 6.

Similarly, $5 +_7 2 = 0$, $4 +_3 2 = 0$; $3 +_3 1 = 1$, $15 +_5 7 = 2$.

Thus to find $a +_m b$, we add a and b in the ordinary way and then from the sum, we remove integral multiples of m in such a way that the remainder r is either 0 or a positive integer less than m.

When a and b are two integers such that $a - b$ is divisible by a fixed positive integer m, then we write

$$a = b \text{ (mod } m)$$

which is read as "a is concurrent to b modulo m".

Thus $a = b$ (mod m) if $a - b$ is divisible by m. For example 13 = 3 (mod 5) since 13 – 3 = 10 is divisible by 5, 5 = 5 (mod 5), 16 = 4 (mod 6); –20 = 4 (mod 6)

Multiplication Modulo p

We shall now define a new type of multiplication known as "multiplication modulo p" and written as $a \times_p b$ where a and b are any integers and p is a fixed positive integer.

$$a \times_p b = r, \; 0 \le r \le p,$$

where r is the least non-negative remainder when ab, i.e., the ordinary product of a and b, is divided by p. For example $4 \times_7 2 = 1$, since $4 \times 2 = 8 = 1(7) + 1$.

It can be easily shown that if $a = b$ (mod p) then $a \times_p C = b \times_p C$.

Additive Group of Integers Modulo m

The set $G = \{0, 1, 2, \ldots m - 1\}$ of first m non-negative integers is a group, the composition being addition reduced modulo m.

Closure Property: We have by definition of addition modulo m,

$$a +_m b = r$$

where r is the least non-negative remainder when the ordinary sum $a + b$ is divided by m. Obviously $0 \le r \le m - 1$. Therefore for all $a, b \in G$, we have $a +_m b \in G$ and thus G is closed with respect to the composition addition modulo m.

Associative Property: Let a, b, c be any arbitrary elements in G.

Then $\quad (a + b) +_m c = (a +_m b) +_m c$

$\therefore \quad b +_m c = b + c \text{ (mod } m)$

= least non-negative remainder when $a + (b + c)$ is divisible by m

= least non-negative remainder when $(a + b) + c$ divided by m.

since $a + (b + c) = (a + b) + c$

$= (a + b) +_m c$ [by definition of ^+m]

$= (a +_m b) +_m c$ [$\therefore a + b = a +_m b \pmod m$]

'$+_{\text{m}}$' is an associative composition.

Existence of Identity Element: We have $0 \in G$. Also, if a is any element of G, then $0 +_m a = a + m^0$. Therefore 0 is the identity element.

Existence of Inverse: The inverse of 0 is 0 itself. If $r \in G$ and $r \neq 0$, then $m - r \in G$. Also $(m - r) +_m r = 0 = r + \widehat{m}\,(m - r)$. Therefore $(m - r)$ is the inverse of r.

Commutative Property: The composition '^+m' is commutative also.

Since

$a +_m b$ = least non-negative remainder when $a + b$ is divided by m

= least non-negative remainder when $b + a$ is divided by m

$= b +_m a$.

The set G contains m elements.

Hence $(G, {}^+m)$ is a finite abelian group of order m.

Multiplicative Group of Integers Modulo p where p is Prime

The set G of $(p - 1)$ integers $1, 2, 3, \ldots, p - 1$, p being prime, is a finite abelian group of order $p - 1$, the composition being multiplication modulo p.

Let $G = \{1, 2, 3, \ldots p - 1\}$ where p is prime.

Closure Property: Let a and b be any elements of G. Then $1 < a < p - 1$, $1 < b < p - 1$. Now by definition $a \times_p b = r$ where r is the least non-negative remainder when the ordinary product $a\,b$ is divided by p. Since p is prime, therefore $a\,b$ is not exactly divisible by p. Therefore r cannot be zero and w shall have $1 \leq r \leq p - 1$. Thus $a \times_p b \in G \;\forall\; a, b \in G$. Hence the closure axiom is satisfied.

Associative Law: a, b, c, be any arbitrary elements of G.

Then $a \times p^b \times p^c = a \times p^{(bc)}$ [$\therefore b \times_p C = bc \pmod p$]

= least non-negative remainder when $a\,(bc)$ is divided by p

= least non-negative remainder when $ab\,(c)$ is divided by p

$= (ab) \times_p C$

$= (a \times_p b) \times_p C$ [$\therefore ab = a \times pb \pmod p$]

$\therefore$ 'X'$_p$ is an associative composition.

Existence of left identity: We have $1 \in G$. Also if a is any element of G, then $1 \times_p a = a$. Therefore 1 is the left identity.

Existence of left inverse: Let s be any member of G. Then $1 < s < p - 1$.

Let us consider the following (p – 1) products:

$$1 \times_p s, 2 \times_p s, 3 \times_p s, \ldots (p - 1) \times_p s.$$

All these are elements of G. Also no two of these can be equal as shown below:

Let i and j be two unequal integers such that

$$1 \le i \le p-1, 1 \le j \le p-1 \text{ and } i > j$$

Then $\quad i \times_p s = j \times_p s$

$\Rightarrow$ $i\,s$ and $j\,s$ leave the same least non-negative remainder when divided by p

$\Rightarrow$ $i\,s - j\,s$ is divisible by p

$\Rightarrow$ $(i - j)\,s$ is divisible by p.

Since $1 \le (i - j) < p - 1$; $1 \le s \le p - 1$ and p is prime, therefore $(i - j)\,s$ cannot be divided by p.

$\therefore$ $\quad i \times_p s \times j \times_p s.$

Thus $1 \times_p s, 2 \times_p s, \ldots (p - 1) \times_p S$ are (p – 1) distinct elements of the set G. Therefore one of these elements must be equal to 1.

Let $\quad s' \times_p s = 1$. The s' is the left inverse of s.

Commutative Law: The composition 'Xp' is commutative, since

$a \times_p b$ = least non-negative remainder when ab is divisible by p

= least non-negative remainder when ba is divided by p

$= b \times_p a$

$\therefore$ $\quad (G, X^p)$ is a finite abelian group of order p – 1.

Theorem 8: The residue classes modulo form a finite group with respect to addition of residue classes

Proof: Let G be the set of residue classes (mod m), then

$$G = \{ \{0\}, \{1\}, \{2\}, \ldots \{r_1\}, \ldots \{r_2\}, \ldots \{m - 1\} \}$$

or $\quad G = \{0, 1, 2, \ldots \{r_1\} \ldots \{r_2\} \ldots m - 1 \text{ (mod } m)\}$

Closure axiom: $\quad (r_1) + \{r_2\} = \{r_1 + r_2\}$

$= \{r\} \in G$ where r is the least positive integer obtained as remainder when $r_1 + r_2$ is divided by m $(0 \le r \le m)$.

Thus the closure axiom is satisfied.

Associative axiom: The addition is associative.

Identity axiom: $\{0\} \in G)$ and $\{0\} + \{r\} = \{r\}$. Hence the identity for addition is $\{0\}$.

Inverse axiom: Since $\{m - r\} + \{r\} = \{m\} = \{0\}$, the additive inverse of the element $\{r\}$ is $\{m - r\}$.

Hence G is a finite group with respect to addition modulo m.

Theorem 9: The set of non-zero residue classes modulo p, where p is a prime, forms a group with respect to multiplication of residue classes.

Proof: Let I = {…, –3, –2, –1, 0, 1, 2, 3, …} be the set of integers. Let $a \in$ then $\{a\}$ is residue class modulo p of I, if $\{a\} = x : x \in I$ and $x - a$ is divisible by $p\}$.

If $p \mid a$ then $\{a\} = \{0\}$ which is called the zero residue class. Let G be the set of non-zero residue classes mod p (p being prime) then

$$G = \{1, 2, 3, \ldots (p - 1)\}$$

Closure axiom: Let $r_1, r_2 \in G$ then $r_1 \cdot r_2 = r \pmod{p}$

where r is the least non-negative integer such that $0 < r < p - 1$ obtained after dividing r_1, r_2 by p.

Also, since p is prime, r_1, r_2 is not divisible by p. Hence r cannot be zero.

Hence, $r_1 \cdot r_2 = r \in G$.

Thus closure axiom is satisfied.

Associative axiom: Multiplication of residue classes is associative.

Existence of Identity: $1 \in G$ and $a.\,1 = 1\,a = a \;\forall\; a \in G$.

Therefore 1 is the identity element in G with respect to multiplication.

Existence of Inverse: Let $s \in G$ then $1 \le s \le p - 1$. Let us consider following $(p - 1)$ elements.

$$1 \cdot s, 2 \cdot s, 3 \cdot s, \ldots, (p-1) \cdot s\,.$$

All these elements are elements of G because the closure law is true. All these elements are distinct as otherwise if

$$i \cdot s = j \cdot s \text{ for } i \ne j \text{ and } i, j \in G$$

the $\qquad i \cdot s = j \cdot s \;\Rightarrow\; i \cdot s - j \cdot s$ is divisible by p

$\Rightarrow\; (i - j) \cdot s$ is divisible by p

$\Rightarrow (i - j)$ is divisible by p [because $1 \le s < (p - 1)$]

$\Rightarrow i - j$ which is contrary to our assumption that $i \ne j$.

Therefore above $(p - 1)$ elements are the same as the elements of G. Hence some one of them should be 1 also, let $s' \cdot s = 1$ where $1 \le s' \le p - 1$. Hence s' is inverse of s. Hence inverse axiom is also satisfied.

$\therefore \qquad G$ is a group under multiplication mod p.

Note: Since $r \cdot s = s \cdot r \;\forall\; r, s \in G$.

G is finite abelian group of order $(p - 1)$.

Illustrative Examples

Example 10: Prove that the set $G = \{0, 1, 2, 3, 4\}$ is a finite abelian group of order 5 with respect to addition modulo 5.

Solution: Let us prepare a composition table as given below:

+ *(mod 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

Closure Property: All the entries in the composition table are elements of the set *G*. Hence G is closed under addition modulo 5.

Associative Property: Addition modulo 5 is associative always.

Identity: 0 ∈ *G* is the identity element.

Inverse: It is clear from composition table.

Element	–	0	1	2	3	4
Inverse	–	0	4	3	2	1

∴ Inverse exists for every element of *G*.

Commutative Law: The composition is commutative as the corresponding rows and columns is *G* are 5.

Hence {*G*, + (mod 5)} is a finite abelian group of order 5.

Example 11: Prove that the set *G* = {1, 2, 3, 4, 5, 6} is a finite abelian group of order 5 with respect to multiplication modulo 7.

Solution: Let us prepare the following composition table:

X_7	*1*	*2*	*3*	*4*	*5*	*6*
1	1	2	3	4	5	6
2	2	4	6	1	3	5
3	3	6	2	5	1	4
4	4	1	5	2	6	3
5	5	3	1	6	4	2
6	6	5	4	3	2	1

Closure Property: All the entries in the table are elements of *G*. Therefore *G* is closed with respect to multiplication modulo 7.

Associative Property: Multiplication modulo 7 is associative always.

Identity: Since first row of the table is identical to the row of elements of *G* in the horizontal border, the element to the left of first row in vertical border is identity element, i.e., 1 is identity element in *G* with respect to multiplication modulo 7.

Inverse: From the table it is obvious that inverses of 1, 2, 3, 4, 5, 6 are 1, 4, 5, 2, 3 and 6 respectively. Hence inverse of each element in *G* exists.

Commutative Property: The composition is commutative because the elements equidistant from principal diagonal are equal each to each.

The set G has 6 elements. Hence (G, X_7) is a finite abelian group of order 6.

Self Assessment

7. Show that the set {1, 2, 3, 4} does not form a group under 'addition modulo 5', but it forms a group under 'multiplication modulo 5'.

8. Prove that the set {0, 1, 2, 3} is a finite abelian group of order 4 under addition modulo 4 as composition.

1.3 Rings

The concept of a group has its origin in the set of mappings or permutations, of a set onto itself. So far we have considered sets with one binary operation only. But rings are the outcome of the motivation which arises from the fact that integers follow a definite pattern with respect to the addition and multiplication. Thus we now aim at studying *rings which are algebraic systems with two suitably restricted and related binary operations*.

Definition: An algebraic structure $(R, +, .)$ where R is a non-empty set and + and . are two defined operations in R, is called a ring if for all a, b, c in R, the following axioms are satisfied:

R_1 . $(R, +)$ is an abelian group, i.e.,

(R_{11}) $a + b \in R$ (closure law for addition)

(R_{12}) $(a + b) + c = a + (b + c)$ (associative law for addition)

(R_{13}) R has an identity, to be denoted by O, with respect to addition,

i.e., $a + 0 = a \ \forall \ a \in R$ (Existence of additive identity)

(R_{14}) There exists an additive inverse for every element in R, i.e., there exists an element $-a$ in R such that

$a + (-a) = 0 \ \forall \ a \in R$ (Existence of additive inverse)

(R_{15}) $a + b = b + a$ (Commutative law for addition)

R_2 $(R, .)$ is a semigroup, i.e.,

(R_{21}) $a . b \in R$ (Closure law for multiplication)

(R_{22}) $(a . b) . c = a . (b . c)$ (associative law for multiplication)

R_3 Multiplication is left as well as right distributive over addition, i.e.,

$$a . (b + c) = a . b + a . c$$

and $$(b + c) . a = b . a + c . a$$

Elementary Properties of a Ring

Theorem 10: If R is a ring, then for all $a, b \in R$.

(a) $a . 0 = 0 . a = 0$

(b) $a (-b) = (-a) b = -(ab)$

(c) $(-a) (-b) = ab$

Proof: (*a*) We know that

$$a0 = a\,(0 + 0) = a0 + a0 \ \forall\ a \in R \qquad \text{(using distributive law)}$$

Since R is a group under addition, applying right cancellation law,

$$a0 = a0 + a0 \Rightarrow 0 + a0 = a0 + a0 \Rightarrow a0 = 0$$

Similarly, $\quad 0a = (0 + 0)\,a = 0a + 0a \qquad$ (using distributive law)

$\therefore \quad 0 + 0a = 0a + 0a \qquad$ (because $0 + 0a = 0a$)

Applying right cancellation law for addition, we get

$$0 = 0a \text{ i.e., } 0a = 0$$

Thus $\quad a0 = 0a = 0.$

(b) To prove that $a\,(-b) = -ab$ we would show that

$$ab + a\,(-b) = 0$$

We know that $\quad a\,[b + (-b)] = a0 \qquad$ [because $b + (-b) = 0$]

$= 0 \qquad$ (with the virtue of result (*a*) above)

or $\quad ab + a\,(-b) = 0 \qquad$ (by distributive law)

$\therefore \quad a\,(-b) = -\,(ab).$

Similarly, to show $(-a)\,b = -\,ab$, we must show that

$$ab + (-a)\,b = 0$$

But $\quad ab + (-a)\,b = [a + (-a)]\,b = 0b = 0$

$\therefore \quad -\,(a)\,b = -\,(ab)$

Hence the result.

(c) Actually to prove $(-a)\,(-b) = ab$ is a special case of foregoing article. However its proof is given as under:

$(-a)\,(-b) = -\,[a\,(-b)] \qquad$ [by result *b*]

$= [-\,(ab)] \qquad$ [because $a\,(-b) = -ab$]

$= ab$

because $-\,(-x) = x$ is a consequence of the fact that in a group inverse of the inverse of an element is element itself.

Illustrative Examples

Example 12: Prove that the set of all rational numbers is a ring with respect to ordinary addition and multiplication.

Let Q be the set of all rational numbers.

R_1 $\quad (Q, +)$ is abelian.

(R_{11}) Let $a, b \in Q$ then $a + b \in Q$ because sum of two rational numbers is a rational number.

(R$_{12}$) Let $a, b, c \in Q$ then

$$(a + b) + c = a + (b + c)$$

because associative law for addition holds.

(R$_{13}$) $0 \in Q$ and $0 + a = a + 0 = a \;\forall\; a \in Q$, i.e., 0 is the additive identity in Q.

(R$_{14}$) $\forall\; a \in Q, -a \in Q$ and $a + (-a) = 0$. Hence additive inverse in Q exists for each element in Q.

(R$_{15}$) Let $a, b \in Q$ then $a + b = b + a$ because addition is commutative for rationals.

R$_2$ $(Q, .)$ is a semi group.

(R$_{21}$) Since the product of two rational numbers is a rational number, $a, b \in Q \Rightarrow a \cdot b \in Q$.

(R$_{22}$) Multiplication in Q is associative.

R$_3$ Multiplication is left as well as right distributive over addition in the set of rational numbers, i.e.,

$$\begin{aligned} a\cdot(b+c) &= a\cdot b + a\cdot c \\ (b+c)\cdot a &= b\cdot a + c\cdot a, \end{aligned}$$

for $a, b, c, \in Q$.

Hence $(Q, +, .)$ is a ring.

Example 13: A Gaussian integer is a complex number $a + ib$, where a and b are integers. Show that the set $J(i)$ of Gaussian integers forms a ring under ordinary addition and multiplication of complex numbers.

Solution: Let $a_1 + ib_1$ and $a_2 + ib_2$ be any two elements of $J(i)$ then

$$\begin{aligned} (a_1 + ib_1) + (a_2 + ib_2) &= (a_1 + a_2) + i\,(b_1 + b_2) \\ &= A + iB \text{ (say)} \end{aligned}$$

and

$$\begin{aligned} (a_1 + ib_1)\,.\,(a_2 + ib_2) &= (a_1a_2 - b_1b_2) + i\,(a_1b_2 + b_1a_2) \\ &= C + i\,D \text{ (say)} \end{aligned}$$

These are Gaussian integers and therefore $J(i)$ is closed under addition as well as multiplication of complex numbers.

Addition and multiplication are both associative and commutative compositions for complex numbers.

Also, multiplication distributes with respect to addition.

$0\ (= 0 + 0i) \in J(i)$ is the additive identity.

The additive inverse of $a + ib \in, J(i)$ is

$(-a) + (-b)\,i \in J(i)$ is

$$\begin{aligned} (a + ib) + (-a) + (-b)\,i & \\ &= (a - a) + (b - b)\,i \\ &= 0 + 0i = 0. \end{aligned}$$

The Gaussian integer $1 + 0.i$ is multiplicative identity.

Therefore, the set of Gaussian integers is a commutative ring with unity as multiplicative identity.

Example 14: Prove that the set of all real numbers of the form $m+n\sqrt{2}$ where *m, n* are rational numbers is a ring under the usual addition and multiplication.

Solution: Let $R = \{m+n\sqrt{2}$: *m, n* are real numbers}.

R_1 (*R*, +) is abelian group.

(R_{11}) Let , $m_1+n_1\sqrt{2}, m_2+n_2\sqrt{2} \in R$ then

$$(m_1+n_1\sqrt{2})+(m_2+n_2\sqrt{2}) = (m_1+m_2)+(n_1+n_2)\sqrt{2} \in R$$

because sum of two real numbers is a real number.

$$(R_{12})\,(m_1+n_1\sqrt{2})+(m_2+n_2\sqrt{2}) = (m_1+n_2\sqrt{2})\,(m_1+n_2\sqrt{2})\,(m_1+n_2\sqrt{2})$$

because addition of real numbers is a real number

(R_{13}) Associative law for addition of real numbers holds, i.e.,

$$(m_1+n_1\sqrt{2})+\{(m_2+n_2\sqrt{2})+(m_3+m_3\sqrt{2})\}$$

$$= \{(m_1+n_1\sqrt{2}+(m_2+n_2\sqrt{2})\}+(m_3+m_3\sqrt{2})$$

for $m_n, n_1, m_2, n_2, m_3, n_3$ to be rational numbers.

(R_{14}) 0 $(=0+0.\sqrt{2})$ Î *R* is the identity of addition in *R*.

(R$_{15}$) Let $m+n\sqrt{2} \in R$, then – $(m+n\sqrt{2})$

= $-m-n\sqrt{2} \in R$ and also

$$(m+n\sqrt{2})+(-m-n\sqrt{2}) = (m-n)+(n-n)\sqrt{2} = 0$$

Hence additive inverse for each element in *R* exists in *R*.

R_2 (*R*, .) is a semi-group.

(R_{21}) $(m_1+n_1\sqrt{2})\cdot(m_2+n_2\sqrt{2})$

= $(m_1m_2+2n_1n_2)+(m_1n_2+m_2n_1)\sqrt{2}$

= $a+b\sqrt{2} \in R$

as *a* and *b* being the sums of products of rational numbers are rational.

(R_{22}) Multiplication is associative in *R*, i.e.,

$$\left\{(m_1+n_1\sqrt{2})\,(m_2+n_2\sqrt{2})\right\}\cdot(m_3+n_3\sqrt{2})$$

$$= (m_1+n_1\sqrt{2})\,(m_2+n_2\sqrt{2})\cdot(m_3+n_3\sqrt{2})$$

R_3 Multiplication is left as well as right distributive over addition in *R*. Hence *R* is a ring under usual addition and multiplication.

Example 15: Prove that the set of residues {0, 1, 2, 3, 4} modulo 5 is using with respect to addition and multiplication of residue classes (mod 5).

Solution: Let R = {0, 1, 2, 3, 4}.

Addition and multiplication tables for the given set R, are as under

+ *mod 5*	*0*	*1*	*2*	*3*	*4*	*mod 5*	*0*	*1*	*2*	*3*	*4*
0	0	1	2	3	4	0	0	0	0	0	0
1	1	2	3	4	0	1	0	1	2	3	4
2	2	3	4	0	1	2	0	2	4	1	3
3	3	4	0	1	2	3	0				
4	4	0	1	2	3	4					

From the addition composition table following is clear:

(i) Since all the elements of the table belong to the set, it is closed under addition (mod 5).

(ii) Addition (mod 5) is always associative.

(iii) $0 \in R$ is the identity of addition.

(iv) The additive inverse of the elements 0, 1, 2, 3, 4 are 0, 4, 3, 2, 1 respectively.

(v) Since the elements equidistant from the principal diagonal are equal to each other, the addition (mod 5) is commutative.

Hence (R, +) is an abelian group.

From the multiplication composition table, we see that (R, .) is semi group, i.e., following axioms hold good.

(vi) Since all the elements of the table are in R, the set R is closed under multiplication (mod 5).

(vii) Multiplication (mod 5) is always associative.

(viii) The multiplication (mod 5) is left as well as right distributive over addition (mod 5).

Hence (R, +, .) is a ring.

Example 16: Prove that the set of residue classes modulo the positive integer m is a ring with respect to addition and multiplication of residue classes (mod m).

Solution: Let $R = \{0, 1, 2, \ldots, r_1, \ldots, r_2, \ldots (m - 1) \pmod{m}\}$

R_1 (R, +) is an abelian group.

(i) Let $r_1, r_2 \in R$ then

where r is the remainder obtained after dividing $r_1 + r_2$ by m.

$\therefore$ R is closed under addition (mod m).

(ii) Addition is associative.

(iii) $O \in R$ is the identity element for addition in R.

(iv) Since $(m - r) + r = m = 0$, the additive inverse of $r \in R$ is $(m - r) \in R$.

(v) Addition is commutative.

R_2 $(R, .)$ is a semigroup, *i.e.,*

(vi) $r_1 r_2 = r' \pmod{m} \in R$

r being the remainder obtained after dividing $r_1 r_2$ by m if $r_1 r_2 \geq m$.

(vii) $(r_1 \cdot r_2) \cdot r_3 = r_1 \cdot (r_2 \cdot r_3) \; \forall \; r_1, r_2, r_3, \in R$ i.e., multiplication is associative. R_3 Distributive axiom is satisfied, *i.e.,*

(viii) $r_1 (r_2 + r_3) = r_1 r_2 + r_1 r_3$ and $(r_2 + r_3) r_1 = r_2 r_1 + r_3 r_1$ for $r_1 r_2, r_3 \in R$.

Hence $(R, +, .)$ is a ring.

Special Types of Rings

Some special types of rings are discussed below:

1. ***Commutative Rings:*** A ring R is said to be a commutative, if the multiplication composition in R is commutative, i.e.,

$$ab = ba \; \forall \; a, b \in R.$$

2. ***Rings with Unit Element:*** A ring R is said to be a ring with unit element if R has a multiplicative identity, i.e., if there exists an element R denoted by 1, such that

$$1 \cdot a = a \cdot 1 = a \; \forall \; a \in R.$$

The ring of all $n \times n$ matrices with elements as integers (rational, real or complex numbers) is a ring with unity. The unity matrix

$$I_n = \begin{bmatrix} 1 & 0 & 0 & \ldots & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ 0 & 0 & 0 & \ldots & 1 \end{bmatrix}$$

is the unity element of the ring.

3. ***Rings with or without Zero Divisors:*** While dealing with an arbitrary ring R, we may find elements a and b in R neither of which is zero, and their product may be zero. We call such elements *divisors of zero* or *zero divisors*.

Definition: A *ring element* a ($\neq 0$) is called a divisor of zero if there exists an element b ($\neq 0$) in the ring such that either

$$ab = 0 \text{ or } ba = 0$$

We also say that a ring R is without zero divisors if the product of no. two non-zero elements of same is zero, i.e., if

$$ab = 0 \Rightarrow \text{either } a = 0 \text{ or } b = 0 \text{ or both } a = 0 \text{ and } b = 0.$$

Cancellation Laws in a Ring

We say that cancellation laws hold in a ring R if

$$ab = ac \; (a \neq 0) \Rightarrow b = c$$

and $ba = ca$ $(a \neq 0) \Rightarrow b = c$ where $a, b, c,$ are in R.

Thus in a ring with zero divisors, it is impossible to define a cancellation law.

Theorem 11: A ring has no divisor of zero if and only if the cancellation law s holds in R.

Proof: Suppose that R has no zero divisors. Let $a, b, c,$ be any three elements of R such that $a \neq 0$, $ab = ac$.

Now, $ab = ac \Rightarrow ab - ac = 0$

$\Rightarrow a\,(b - c) = 0$

$\Rightarrow b - c = 0$ (because R is without zero divisors and $a \neq 0$)

$\Rightarrow b = c$.

Thus the left cancellation law holds in R. Similarly, it can be shown that right cancellation law also holds in R.

Conversely, suppose that the cancellation laws hold in R.

Let $a, b \in R$ and if possible let $ab = 0$ with $a \neq 0$, $b \neq 0$ then $ab = a \cdot 0$ (because $a \cdot 0 = 0$)

Since $a \neq 0$, $ab = a \cdot 0 \Rightarrow b = 0$ (by left cancellation law)

Hence we get a contradiction to our assumption that $b \neq 0$ and therefore the theorem is established.

Division Ring

A ring is called a division ring if its non-zero elements form a group under multiplication.

Pseudo ring: A non-empty set R with binary operations '+' and '.' satisfying all the postulates of a ring except right and left distributive laws, is called a pseudo ring if

$$(a+b)\cdot(c+d) = a\cdot c + a\cdot d + b\cdot c + b\cdot d \text{ for all } a,b,c,d \in R$$

Subrings

Definition: *Let R be a ring. A non-empty subset S of the set R is said to be a subring of R if S is closed under addition and multiplication in R and S itself is a ring, for those operations.*

If R is any ring, then $\{0\}$ and R are always subrings of R. These are said to be improper subrings. The subrings of R other than these two, if any, are said to be proper subrings of R.

Evidently, if S is a subring of a ring R, it is a sub group of the additive group R.

Theorem 12: The necessary and sufficient condition for a non-empty subset S of a ring R to be a subring of R are

(i) $a, b \in S \Rightarrow a - b \in S$,

(ii) $a, b \in S \Rightarrow ab \in S$.

Proof: To prove that the conditions are necessary let us suppose that S is a subring of R.

Obviously S is a group with respect to addition, therefore,

$$b \in S \Rightarrow -b \in S$$

Since S is closed under addition

$$a \in S, b \in S \Rightarrow a \in S, -b \in S \Rightarrow a + (-b) \in S$$

$$\Rightarrow a - b \in S.$$

Also S is closed with respect to multiplication,

$$a \in S, b \in S \Rightarrow ab \in S.$$

Now to prove that the conditions are sufficient suppose S is a non-empty subset of R for which the conditions (*i*) and (*ii*) are satisfied.

From condition (*i*)

$$a \in S \Rightarrow a - a \in S \Rightarrow 0 \in S.$$

Hence additive identity is in S.

Now $\quad 0 \in S, a \in S \Rightarrow -a \in S$

i.e., each element of S possesses additive inverse.

Let $a, b \in S$ then $-b \in S$ and then from condition (*i*)

$$a \in S, -b \in S \Rightarrow a - (-b) \in S \Rightarrow (a + b) \in S$$

Thus S is closed under addition. S being subset of R, associative and commutative laws hold in S. Therefore, $(S, +)$ is an abelian group.

From condition (ii) S is closed under multiplication.

Since S is a subset of R, the associative law for multiplication and distributive laws of multiplication over addition hold in S. Thus S is a subring of R.

Intersection of Subrings

Theorem 13: The intersection of two subrings is a subring.

Proof: Let S_1 and S_2 be two subrings of ring R.

Since $0 \in S_1$ and $0 \in S_2$ at least $0 \in S_1 \cap S_2$. Therefore $S_1 \cap S_2$ is non-empty.

Let $a, b \in S_1 \cap S_2$, then

$$a \in S_1 \cap S_2 \Rightarrow a \in S_1 \text{ and } a \in S_2$$

and $b \in S_1 \cap S_2 \Rightarrow b \in S_1$ and $b \in S_2$.

But S_1 and S_2 are subrings of R, therefore

$$a, b \in S_1 \Rightarrow a - b \in S_1 \text{ and } a\,b \in S_1.$$

and $\quad a, b \in S_2 \Rightarrow a - b \in S_2$ and $a\,b \in S_2$.

Consequently, $a, b \in S_1 \cap S_2 \Rightarrow a - b \in S_1 \cap S_2$ and $a\,b \in S_1 \cap S_2$.

Hence $S_1 \cap S_2$ is a subring of R.

Illustrative Examples

Example 17: If R is a ring with additive identity 0, then for all $a, b \in R$, prove that

$$a\,(b - c) = ab - ac$$

and $\quad (b - c)\,a = ba - ca.$

Solution: We have, $\quad a\,(b - c) = a\,[b + (-c)]$

$= ab + a\,(-c)$ $\quad$ [left distributive law]

$$= ab + [-(ac)]$$

$$= ab - ac.$$

Also, $$(b - c)\,a = [b + (-c)]\,a$$

$$= ba + (-c)\,a \qquad \text{(right distributive law)}$$

$$= ba + [-(ca)] = ba - ca.$$

Example 18: Suppose M is a ring of all 2 × 2 matrices with their elements as integers, the addition and multiplication of matrices being the two ring compositions. Then M is a ring with left zero-divisor.

Solution: The null matrix $O = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ is the zero element of ring M.

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \text{ and } B = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \text{ are two non-zero elements of } M.$$

Now $AB = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = 0.$

Hence M is a ring with left zero divisor.

Example 19: Prove that the ring of integers is a ring without zero divisors.

Solution: Since the product of two non-zero integers is never zero, it is the ring without zero divisors.

Example 20: Prove that the ring of residue classes modulo *a* composite integer m possess proper zero divisors.

Solution: Let $m = ab$ i.e., a and b are two factors of m.

Then $ab \not\equiv 0 \pmod{m}$

But $a \not\equiv \pmod{m}$ and $b \not\equiv 0 \pmod{m}$.

Hence the residue classes $\{a\}$ and $\{b\}$ are proper zero-divisors.

Example 21: Prove that the totality R of all ordered pairs (a, b) of real numbers is a ring with zero divisors under the addition and multiplication defined as

$$(a, b) + (c, d) = (a + c, b + d),$$

$$(a, b)\,(c, d) = (ac, bd), \ \forall\ (a, b), (c, d) \in R.$$

Solution: First of all, we prove that R is ring. We have

$$(R_1) : [(a, b) + (c, d)] = (a + c, b + d) \in R$$

Hence R is closed for addition.

$$(R_1) : [(a, b) + (c, d)] + (e, f) = (a + c, b + d) + (e, f)$$

$$= ((a+c)+e, (b+d)+f)$$

Notes

$$= [a+(c+e), b+(d+f)]$$

[addition is associative in real numbers]

$$= (a,b)+(c+e, d+f)$$

$$= (a,b)+[(c+d)+(e,f)]$$

So the addition is associative in R

(R_3) : (0, 0) + (a, b) = (0 + a, 0 + b) = (a, b) $\forall$ $(a, b) \in R$, so that (0, 0) is the additive identity in R.

(R_4) : $(-a,-b)+(a,b)=(-a+a,-b+b)=(0,0)$ so the additive inverse of (a, b), is $(-a, -b)$ $\forall$ $(a, b) \in R$.

(R_5) : $(a, b) + (c, d) = (a + c, b + d)$

$$= (c+a, d+b)$$

[because addition is commutative in real numbers]

$$= (c,d)+(a,b) \;\forall\; (a, b), (c, d) \in R.$$

(R_6): $[(a, b), (c, d)]\ [e, f]$

$$= (ac,bd)(e,f)$$

$$= \{(ac)e,(bd)f\}$$

$$= \{a(c,e,b(d\,f)\}$$

[because ordinary multiplication is associative]

$$= (a,b)(c,e,d\,f)$$

$$= (a,b)\,[(c,d)(e,f)] \;\forall\; (a,b)\,(c,d)(e,f) \in R\,.$$

(R_7) : $(a,b)\,[(c,d)+(e,f)]$

$$= (a,b)\,(c+e, d+f)$$

$$= (a\,c+a\,e, b\,d+b\,f) \qquad \text{(by distributive law of reals)}$$

$$= (a\,c\,b\,d),(a\,e, b\,f)$$

$$= (a\,b),(c,d)+(a,b)(e,f)\,.$$

Similarly

$$[(c\,d)+(e,f)]\,(a,b)$$

$$= (c\,d)+(a,b)+(e,f)\,(a,b)\,.$$

Hence R is a ring.

Now, in order to show that R is a ring with zero divisors we must produce at least two non-zero elements whose product is zero. Clearly neither $(a, 0)$ with $a \neq 0$ nor $(0, b)$ with $b \neq 0$ is the zero element (additive identity) or R yet their product

$$(a,0)(0,b) = (a.0, 0.b) = (0,0)$$

which is zero element in R.

Thus R is a ring with zero divisors.

It can also be verified that R is also a commutative ring with unity element (1, 1).

Example 22: Prove that M the set of all 2 × 2 matrices of the form

$$\begin{bmatrix} a+bi & c+di \\ -c+di & a-bi \end{bmatrix}, i = \sqrt{-1}$$

where a, b, c, d are real numbers, form a division ring.

Solution: Since I = $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \in M$ is a ring with unity under matrix addition and multiplication.

Let A be a non-zero matrix in M, and let

$$A = \begin{bmatrix} a+bi & c+di \\ -c+di & a-bi \end{bmatrix}$$

where a, b, c, d are not all zero. Consider

$$B = \begin{bmatrix} \dfrac{a-bi}{a^2+b^2+c^2+d^2} & -\dfrac{c+di}{a^2+^2+c^2+d^2} \\ \dfrac{c-di}{a^2+b^2+c^2+d^2} & \dfrac{a+bi}{a^2+b^2+c^2+d^2} \end{bmatrix}$$

Evidently $B \in M$. Also $A\,B$ = $B\,A$ = $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.

Thus every non-zero matrix of M is invertible. Hence M is a division ring.

Example 23: Prove that the set of integers is a subring of the ring of rational numbers.

Solution: Let I be the set of integers and Q the set of rational numbers.

Clearly $I \subset Q$ and $a,b \in I \Rightarrow a-b \in I$ and $a\,b \in I$

Therefore, I is a subring of Q.

Example 24: Show that the set of matrices $\begin{bmatrix} a & b \\ 0 & c \end{bmatrix}$ is a subring of the ring of 2 × 2 matrices with integral elements.

Solution: Let M be the set of matrices of the type $\begin{bmatrix} a & b \\ 0 & c \end{bmatrix}$

Clearly $M \subset R$

Let $A = \begin{bmatrix} a_1 & b_1 \\ 0 & c_1 \end{bmatrix}, B = \begin{bmatrix} a_2 & b_2 \\ 0 & c_2 \end{bmatrix} \in M$ then

$A - B = \begin{bmatrix} a_1 - a_2 & b_1 - b_2 \\ 0 & c_1 - c_2 \end{bmatrix} \in M$ and also

$$A\,B = \begin{bmatrix} a_1 & b_1 \\ 0 & c_1 \end{bmatrix}\begin{bmatrix} a_2 & b_2 \\ 0 & c_2 \end{bmatrix} = \begin{bmatrix} a_1a_2 & a_1b_2\;b_2c_1 \\ 0 & c_1\,c_2 \end{bmatrix} \in M$$

$\therefore$ M is subring of R.

Self Assessment

9. Show that the set of even integers including zero is a commutative ring with zero-divisors under the usual addition and multiplication.

10. Prove that the ring $R = \{0, 1, 2, 3, 4, 5, 6, 7\}$ under the addition and multiplication modulo 8 is a commutative ring without zero divisors.

11. Prove that set I of integers is a subring of R, the set of real numbers.

12. If a, b belong to a ring R and $(a + b)^2 = a^2 + 2ab + b^2$, then show that R is a commutative ring.

Ideals

Definition: Let $(R, +,.)$ be any ring and S a subring of R, then S is said to be right ideal of R if $a \in S, b \in R \Rightarrow a\,b \in S$ and left ideal of R if $a \in S, b \in R \Rightarrow b\,a \in S$.

Thus a non-empty subset S or R is said to be a ideal of R if:

(i) S is a subgroup of R under addition.

(ii) $\forall\ a \in S$ and $b \in R$, both $a\,b$ and $ba \in S$.

Principal Ideals: If R is a commutative ring with unity and $a \in R$, the ideal $\{ax : x \in R\}$ is called the principal ideal generated by a and is denoted by (a), thus (a) stands for the ideal generated by a.

Principal Ideal Ring: A commutative ring with unity for which every ideal is a principal ideal is said to be a principal ideal ring.

Prime Ideal: Let R be a commutative ring. An ideal P of ring R is said to be a prime ideal of R if

$$ab \in P, a, b \in R \Rightarrow a \in P \text{ or } b \in P\,.$$

Example 25: In the commutative ring of integers I, the ideal $P = \{5r : r \in I\}$ is a prime ideal since if $ab \in P$, then $5 \mid ab$ and consequently $5 \mid a$ or $5 \mid b$ as 5 is prime.

Integral Domain

Definition: A commutative ring with unity is said to be an integral domain if it has no zero-divisors. Alternatively a commutative ring R with unity is called an integral domain if for all $a, b \in R$, $a\,b = 0 \Rightarrow a = 0$ *or* $b = 0$.

Examples:

(i) The set I of integers under usual addition and multiplication is an integral domain as for any two integers $a, b; ab = 0 \Rightarrow a = 0 \text{ or } b = 0$.

(ii) Consider a ring $R = \{0, 1, 2, 3, 4, 5, 6, 7\}$ under the addition and multiplication modulo 8. This ring is commutative but it is not integral domain because $2 \in R, 4 \in R$ are two non-zero elements such that 2.4 = 0 (mod 8).

Euclidean Rings

An integral domain R is said to be a Euclidean ring if for every $a \neq 0$ in R there is defined a non-negative integer, to be denoted by d (a), such that:

(i) for all $a, b \in R$, both non-zero, $d(a) \leq d(ab)$,

(ii) for any $a, b \in R$, both non-zero, there exists $q, r \in R$ such that $a = q\,b + r$ when either $r = 0$ or $d(r) < d(b)$.

Illustrative Examples

Example 26: Prove that the ring of complex numbers C is an integral domain.

Solution: Let J (i) = $\{a + bi : a, b \in I\}$.

It is easy to prove that J (i) is a commutative ring with unity.

The zero element 0 + 0.i and unit element 1 + 0.i.

Also this ring is free from zero-divisors because the product of two non-zero complex numbers cannot be zero. Hence J (i) is an integral domain.

Example 27: Prove that set of numbers of the form $a + b\sqrt{2}$ with a and b as integers is an integral domain with respect to ordinary addition and multiplication.

Solution: Let $D = \{a + b\sqrt{2} : a, b \in I\}$

(I_1) $(D, +)$ is an abelian group.

(I_{11}) Let $a_1 + b_1\sqrt{2} \in D$ and $a_2 + b_2\sqrt{2} \in D$, then $a_1, b_1, a_2, b_2 \in I$

Now, $(a_1 + b_1\sqrt{2}) + (a_2 + b_2\sqrt{2}) = (a_1 + a_2) + (b_1 + b_2)\sqrt{2} \in D$

as $a_1 + a_2, b_1 + b_2 \in I$.

Hence D is closed under addition.

(I_{12}) Addition is associative in the set of real numbers.

(I_{13}) $0 = (0 + 0\sqrt{2}) \in D$ is the additive identity in D because 0 ∈ I.

(I_{14}) If $(a + b\sqrt{2}) \in D$

Then $(-a) + (-b)\sqrt{2} \in D$ and $(a + b\sqrt{2}) + [(-a) + (-b)\sqrt{2}] = 0 + 0\sqrt{2} = 0$ the additive identity. Hence each element in D possesses additive inverse.

(I_{15}) Addition is commutative in the set of real numbers.

I_2 (D, .) is semi-abelian group with unity.

$(I_{21})(a_1 + b_2\sqrt{2})(a_2 + b_2\sqrt{2}) = (a_1 a_2 + 2b_1 b_2) + (a_1 b_2 + a_2 b_1)\sqrt{2} \in D$ as $a_1 + b_2 + 2b_1 b_2, a_1 b_1 \in I$ for $a_1, b_1, a_2, b_2 \in I$.

Hence D is closed under multiplication.

(I_{12}) Multiplication is commutative in the set of real numbers.

(I_{23}) Multiplication is associative in the set of real numbers.

(I_{24}) $1+0\sqrt{2}=1\in D$ and for $a+b\sqrt{2}\in D$, we have

$$(1+0\sqrt{2}(a+b\sqrt{2})=(a+b\sqrt{2})(I+0\sqrt{2})+a+b\sqrt{2}$$

$\therefore$ 1 is the multiplicative identity in D.

I_3. In the set of real numbers multiplication is distributive over addition.

I_4. Now, to prove that this ring is without zero divisors let $a+b\sqrt{2}$ and $c+d\sqrt{2}$ be two arbitrary elements of D. Then

$$(a+b\sqrt{2})\,(c+d\sqrt{2})=0 \;\Rightarrow\; ac+2bd=0 \text{ and } bc+ad=0$$

$$= \text{either } a = 0 \text{ and } b = 0 \text{ or } c = 0 \text{ and } d = 0$$

$$= \text{either } a+b\sqrt{2} \text{ or } c+d\sqrt{2} = 0.$$

Thus the given set is a commutative ring with unity and without zero-divisors, i.e., it is an integral domain.

1.4 Fields

Definition: A commutative ring with unity is called a field if its every non-zero element possesses a multiplicative inverse.

Thus a ring R in which the elements of R different from 0 form an abelian group under multiplication is a field.

Hence, a set F, having at least two distinct elements together with two operations '+' and '.' is said to form a field if the following axioms are satisfied:

(F_1) (F, +) is an abelian group.

(F_{11}) F is closed under addition, i.e., $\forall\ a,b\in F\Rightarrow a+b\in F$.

(F_{12}) Addition is commutative in F i.e., $(a+b)+c=a+(b+c)$

for all $a,b,c\in F$.

(F_{14}) Identity element with respect to addition exists in F, i.e., $\exists$, $0\in$ F such that $a + 0 = 0 + a = a$ $\forall\ a\in F$.

(F_{15}) There exists inverse of every element of F, i.e., $\forall\ a\in F$, there exists an element $-a$ in F such that

$$a+(-a)=(-a)+a=0\,.$$

(F_{20}) Properties of (F,.)

(F_{21}) F is closed under multiplication, i.e., $\forall\ a,b\in F\Rightarrow a,b\in F$.

(F_{22}) Multiplication is commutative in F, i.e., $a\cdot b=b\cdot a$ for all $a, b, \in F$.

(F_{23}) Multiplication is associative in F, i.e., $(a\cdot b)\cdot c = a\cdot(b\cdot c)$ for all $a, b, c, \in F$.

(F_{24}) There exists an identity element 1 for multiplication F such that

$$a \cdot 1 = 1 \cdot a = a \ \ \forall \ a \in F.$$

(F_{25}) For all $a \in F$, $a \neq 0$, there exists an element a^{-1} (multiplicative inverse) in F such that $a \cdot a^{-1} = a^{-1} \cdot a = 1$.

F_3. Distributive laws of multiplication over addition for all $a, b, c \in F$,

$$a \cdot (b + c) = a \cdot b + a \cdot c$$

and $$(b + c) \cdot a = b \cdot a + c \cdot a$$

The above properties can be summarised as:

(1) (F, +) is an abelian group.

(2) (F, .) is a semi-abelian group and (F – {0}, .) is an abelian group.

(3) Multiplication is distributive over addition.

Examples:

(i) The set of real numbers is a field under usual addition and multiplication compositions.

(ii) The set of rational numbers is a field under usual addition and multiplication operations.

(iii) The set of integers is not a field.

Some Theorems

Theorem 14: The multiplicative inverse of a non-zero element of a field is unique.

Proof: Let there be two multiplicative inverse a^{-1} and a' for a non-zero element $a \in F$.

Let (1) be the unity of the field F.

$\therefore$ $aa^{-1} = 1$ and $a \cdot a' = 1$ so that $a \cdot a^{-1} = a \cdot a'$.

Since F – {0} is a multiplicative group, applying left cancellation, we get $a^{-1} = a'$.

Theorem 15: A field is necessarily an integral domain.

Proof: Since a field is a commutative ring with unity, therefore, in order to show that every field is an integral domain we only need proving that a field is without zero divisors.

Let F be any field let $a, b \in F$ with $a \neq 0$ such that $ab = 0$. Let 1 be the unity of F. Since $a \neq 0$, a^{-1} exists in F and therefore,

$$ab = 0 \Rightarrow a^{-1}(ab) = a^{-1}\,0$$

$$\Rightarrow (a^{-1}a)\,b = 0 \qquad \text{(because } a^{-1}a = 1\text{)}$$

$$\Rightarrow 1 \,.\, b = 0$$

$$\Rightarrow b = 0 \qquad \text{(because 1. } b = \text{b)}$$

Similarly if $b \neq 0$ then it can be shown that

$$ab = 0 \Rightarrow a = 0$$

Thus $ab = 0 \Rightarrow a = 0$ or $b = 0$.

Hence, a field is necessarily an integral domain.

Corollary: Since integral domain has no zero divisors and field is necessarily an integral domain, therefore, field has no zero-divisor.

Theorem 16: If a, b are any two elements of a field F and $a \neq 0$, there exists a unique element x such that $a\,.\,x = b$.

Proof: Let 1 be the unity of F and a^{-1}, the inverse of a in F then

$$a\,.\,(a^{-1}\,b) = (aa^{-1})\,.\,b = 1\,.\,b = b$$

$$\therefore \qquad ax = b \Rightarrow a\,.\,x = a\,.\,(a^{-1}\,b)$$

$$\Rightarrow x = a^{-1}\,b \qquad \text{(by left cancellation)}$$

Thus $\qquad x = a^{-1}\,b \in F.$

Now, suppose there are two such elements x_1, x_2 (say) then

$$a\,.\,x_1 = b \text{ and } a\,.\,x_2 = b$$

Hence $a\,.\,x_1 = a\,.x_2$

On applying left cancellation, we get

$$x_1 = x_2$$

Hence the uniqueness is established.

Theorem 17: Every finite integral domain is a field.

or

A finite commutative ring with no zero divisor is a field.

Proof: Let D be an integral domain with a finite number of distinct elements $a_1, a_2, ..., a_n$. In order to prove that D is a field, we have to prove that there exists $1 \in D$ such that $1\,.\,a = a \;\forall\; a \in D$ and for every $a\ (\neq 0) \in D$ there exists an element $a^{-1} \in D$ such that $a^{-1}\,a = 1$.

Let $a \neq 0$ and $a \in D$. Now the elements $aa_1 = aa_2, ..., aa_n$ are the elements of D.

All of them are distinct because otherwise if $aa_i = aa_j$, for $i \neq j$ then

$$aa_i = aa_j \Rightarrow a(a_i - a_j) = 0$$

$$\Rightarrow a_i - a_j = 0$$

(because $a \neq 0$ and D is without zero divisors)

$$\Rightarrow a_i = a_j \text{ contradicting } i \neq j.$$

Let one of these elements be a. Thus there exists an element, say $1 \in D$ such that

$a\,.\,1 = a = 1\,.\,a$ (because multiplication is commutative)

Let y be any element of D then for some $x \in D$ we should have

$$ax = y = xa$$

$$\begin{aligned} \text{Therefore,} \qquad 1y &= 1\,(ax) && \text{(because } ax = y\text{)} \\ &= (1a)\,x = ax && \text{(because } 1a = a\text{)} \end{aligned}$$

Notes

Thus $1y = y = y1 \forall\ y \in D = y1.$ (because multiplication is commutative) Therefore 1 is the unit element of D.

Now $1 \in D$ and as such one of the elements $aa_1, aa_2, ..., aa_n$ is equal to 1, i.e.,

$$aa_s = 1 = a_s\ a \text{ for some } s \text{ such that } 1 \leq s \leq n.$$

Thus $a_s \in D$ is the multiplicative inverse of the non-zero element a in D. Since a is arbitrary element in D, we conclude that each non-zero element of D possesses multiplicative inverse.

Hence D is a field.

Illustrative Examples

Example 28: Prove that the set of complex numbers is a field with respect to addition and multiplication operation.

or

Let C be the set of ordered pairs (a, b) of real numbers. Define addition and multiplication in C by the equations

$$\begin{aligned}(a, b) + (c, d) &= (a + c, b + d)\\(a, b)\,(c, d) &= (ac - bd, bc + ad)\end{aligned}$$

Prove that C is a field.

Solution: C is closed under addition and multiplication since $a + c, b + d, ac - bd, bc + ad$ are all real numbers.

Let $(a, b), (c, d), (e, f) \in C$

then

$$\begin{aligned}[(a, b) + (c, d)] + (e, f) &= (a + c, b + d) + (e, f)\\&= [(a + c) + e, (b + d) + f]\\&= [a + (c + e), b + (d + f)]\\&= (a, b) + (c + e, d + f)\\&= (a, b) + [(c, d) + (e, f)]\end{aligned}$$

Hence addition is associative

Since
$$\begin{aligned}(a, b) + (c, d) &= (a + c, b + d)\\&= (c + a, d + b) = (c, d) + (a, b)\end{aligned}$$

addition is commutative in C.

$(0, 0) \in C$ is additive identity in C as

$$(0, 0) + (a, b) = (0 + a, 0 + b) = (a, b) \forall\ (a, b) \in C.$$

If $(a, b) \in C$ then $(-a, -b) \in C$

and also $$(-a, -b) + (a, b) = (-a + a, b + b) = (0, 0)$$

Hence [(–*a*, –*b*) is the additive inverse, of (*a*, *b*)]

Also
$$\begin{aligned}[(a, b)(c, d)](e, f) &= [(ac - bd, bc + ad)\,(e, f)] \\ &= [(ac - bd)e - (bc + ad)f, (bc + ad)e + (ac - bd)f] \\ &= [a(ce - df) - b(de + cf), b(ce - df) + a(de + cf)] \\ &= (a,b)(ce - df, de + cf) \\ &= (a,b)[(c,d)(e,f)]\end{aligned}$$

Hence multiplication is associative in *C*.

Distributive laws also hold in *C* because,

$$\begin{aligned}(a, b)[(c, d) + (e, f)] &= (a, b)(c + e, d + f) \\ &= [a(c + e) - b(d + f), b(c + e) + a(d + f)] \\ &= [(ac - bd) + (ac - bf), (bf + ad) + (be + af)] \\ &= (ac - bd, bc + ad) + (ae - bf, be + af) \\ &= (a,b)\,(c,d) + (a,b)(e,f)\end{aligned}$$

Similarly, it can be proved that multiplication is distributive over addition in *C* from right too.

Multiplication is commutative in *C* because

$$\begin{aligned}(a, b)(c, d) &= (ac - bd, bc + ad) \\ &= (ca - db\ cb + da) \\ &= (c, d)(a, b)\end{aligned}$$

Since (1, 0) ∈ *C* and also (1, 0) (*a*, *b*)

= (*a*, *b*) (1, 0) is multiplicative identity in *C*.

Multiplicative inverse for non-zero elements in *C* exists because if (*a*, *b*) is non-zero elements in *C* then *a* and *b* are not zero at a time.

Let (*c*, *d*) be the multiplicative inverse of (*a*, *b*) then

$$(a, b)(c, d) = (1, 0)$$

i.e. $$[(ac - bd), (bc + ad)] = (1, 0)$$

so that $$ac - bd = 1, bc + ad = 0$$

i.e., $$c = \frac{a}{a^2 + b^2}, d = \frac{-b}{a^2 + b^2}$$

Since $a \neq 0$ or $b \neq 0$, $a^2 + b^2 \neq 0$, i.e., C or d or both are non-zero real numbers.

Hence C is a field.

Note: The question could have been done by assuming the elements of C as $a + ib$ etc. also.

Example 29: Show that the set of numbers of the form $a+b\sqrt{2}$ with a and b as rational numbers is a field.

Solution: Let $R = \{a+b\sqrt{2} : a,\ b \in Q\}$

$F_1(R, +)$ is a abelian group.

(F_{11}) Let $a_1 + b_1\sqrt{2} \in R$ and $a_2 + b_2\sqrt{2} \in R,$ then a_1, b_1, a_2, b_2 are the elements of Q, the set of rational numbers.

Now $(a_1 + b_1\sqrt{2}) + (a_2 + b_2\sqrt{2}) = (a_1 + a_2) + (b_1 + b_2)\sqrt{2} \in R$ since $a_1 + a_2,\ b_1 + b_2 \in Q.$

Hence closure axiom for addition is satisfied.

(F_{12}) Addition is commutative for real numbers.

(F_{13}) Addition is associative for real numbers.

$(F_{14})\ 0 + 0\sqrt{2} = 0 \in R$ as $0 \in Q$, hence 0 is the identity of addition in R because

$$\begin{aligned}(0+0\sqrt{2})+(a+b\sqrt{2}) &= (0+a)+(0+b\sqrt{2})\\ &= a+b\sqrt{2}\ \forall\ a, b \in Q.\end{aligned}$$

(F_{15}) If $a+b\sqrt{2} \in R$ then $(-a)+(-b)\sqrt{2} \in R$ and also

$$\begin{aligned}&[(-a)+(-b)\sqrt{2}]+(a+b\sqrt{2})\\ &= (-a+a)+(-b+b)\sqrt{2} = 0+0\sqrt{2}\\ &= 0\end{aligned}$$

$\therefore$ each element of R possesses additive inverse .

(F_2) Properties of field for $(F,)$

$(F_{21})\ (a_1 + b_1\sqrt{2})(a_2 + b_2\sqrt{2})$

$$(a_1a_2 + 2b_1b_2) + (a_1b_2 + a_2b_1\sqrt{2}) \in R$$

Since $a_1\, a_2 + 2b_1b_2,\ a_1b_2 + a_2b_1 \in Q$ for $a_1, a_2, b_1, b_2 \in Q.$

Thus R is closed under multiplication

(F_{22}) Multiplication in R is commutative

(F_{23}) Multiplication in R is associative

$(F_{24})\ 1 + 0\sqrt{2}\ = 1 \in R$ and $1\ (a+b\sqrt{2})$

$$= a+b\sqrt{2}\ \forall\ a, b \in Q.$$

$\therefore$ 1 is multiplicative identity in R.

(F_{22}) Let $a+b\sqrt{2} \neq 0,$ i.e., at least one of a and b is non-zero then

$$\frac{1}{a+b\sqrt{2}} = \frac{a-b\sqrt{2}}{(a+b\sqrt{2})(a-b\sqrt{2})} = \frac{a}{a^2-2b^2} + \frac{-b}{a^2-2b^2}\sqrt{2}$$

$$= A+B\sqrt{2} \in R, \text{ Where } A, B \in \mathrm{Q}$$

and

$$A = \frac{a}{a^2-2b^2}, B = \frac{-b}{a^2-2b^2}$$

$a^2-2b^2 \neq 0$ as otherwise if $a = 0$, $b = 0$ which is impossible due to our assumption for non-zero element $a+b\sqrt{2}.$

Thus at least one of A and B is non-zero. Hence inverse of $a+b\sqrt{2}$ is a non-zero element $A+B\sqrt{2}$ in $R,$ because

$$\left(A+B\sqrt{2}\right)\left(a+b\sqrt{2}\right) = \frac{1}{(a+b\sqrt{2})}(a+b\sqrt{2}) = 1.$$

Thus every non-zero element in R possesses multiplicative inverse.

Hence R is a field.

Example 30: If the operations be addition and multiplication (mod p), prove that the set {0, 1, 2, ..., p – 1}, (mod p) where p is prime, is a field.

Solution: Let this set be denoted by I | (p) which has already be shown a commutative ring with unity. To prove that I | (p) is a field we will have to show that every non-zero element of I | (p) is invertible. Let $r \in I$ (p) and $r \neq 0$.

$$\text{Now } r \neq 0 \Rightarrow \cong 0 \text{ (mod } p)$$

$$\Rightarrow r \text{ is not divisible by p}$$

$$\Rightarrow r \text{ and } p \text{ are relatively prime.}$$

i.e., there exist integers x, y such that $rx + py = 1$ implying that

$$rx \in 1 \text{ (mod } p) \text{ as } py \equiv 0 \text{ (mod } p).$$

Thus x is inverse of r in I | (p).

Hence I | (p)is a field.

Self Assessment

13. With addition and multiplication as operation prove that

    (i) The set {0, 1} (mod 2) is a field.

    (ii) The set {0, 1, 2} (mod 3) is a field.

14. Prove that the set of all real numbers of the form $u+v\sqrt{3}$ where u and v are of the form $a+b\sqrt{2}$ in which a and b are rational numbers, is a field.

15. Prove that the set E of all even integers is a commutative ring but not a field.

16. Show that a finite commutative ring without zero divisors is a field.

1.5 Vector Spaces

Before giving a formal definition of an abstract vector space we define what is known as an external composition in one set over another. We have already defined a binary composition in a set A as a mapping of $A \times A$ to A. This may be referred to as an *internal composition* in A. Let now A and B be two non-empty sets. Then a mapping

$$f : A \times B \to B$$

is called an external composition in B over A.

Definition: Let $(F, +, .)$ be a field. Then *a* set V is called a vector space over the field F, if V is an abelian group under an operation which is denoted by +, and if for every $a \in F$, $u \in V$ there is defined an element $a\,u$ in V such that:

(i) $a(u+v) = au + av,$ for all $a \in F,\ u,\ v \in V.$

(ii) $(a+b)u = au + bu,$ for all $a, b \in F,\ u \in V.$

(iii) $a(bu) = (ab)u,$ for all $a, b \in F,\ u \in V.$

(iv) $1\,.\,u = u\,.\,1$ represents the unity element of F under multiplication.

The following notations will be constantly used in the forthcoming discussions.

(i) Generally F will be field whose elements shall often be referred to as *scalars*.

(ii) V will denote vector space over F whose elements shall be called as *vectors*.

Thus to test that V is a vector space over F, the following axioms should be satisfied.

V_1 $(V, +)$ is an abelian group.

(V_{11}) ***Closure law:*** $u, v \in V \Rightarrow u + v \in V.$

(V_{12}) ***Associative law:*** For all $u, v, w \in V \Rightarrow (u+v)+w = u+(v+w)$

(V_{13}) ***Existence of identity:*** There exists an element of zero vector.

(V_{14}) ***Existence of Inverse:*** For all $u \in V$, there exists a unique vector $-u \in V$ such that

$$u + (-u) = 0$$

(V_{15}) ***Commutative Law:***

$$u + v = v + u \text{ for } u, v \in V$$

V_2 scalar multiplication is distributive over addition in V, i.e.,

$$a(u+v) = au + av,\ a \in F,\ v \in V$$

V_3 distributivity of scalar multiplication over addition in F, i.e.,

$$(a+b)u = au + bu,\ a, b \in F, u \in V.$$

V$_4$ Scalar multiplication is associative i.e.,

$$a(bu) = (ab)u \forall\ a, b \in F \text{ and } u \in V$$

V$_5$ ***Property of Unity:*** Let 1 ∈ *F* be the unity of *F*, then

$$1u = u = u \,|\, \forall\ u \in V$$

A vector space *V* over a field *F* is expressed by writing *V*(*F*). Sometimes writing only *V* is sufficient provided the context makes it clear that which field has been considered.

If the field is *R*, the set of real numbers, then *V* is said to be real vector space. If the field is *Q*, the set of rational numbers, then *V* is said to be a rational vector space and if the field is *C*, the set of complex numbers *V* is called a complex vector space.

Illustrative Examples

Example 31: Show that the set of all vectors in a plane over the field of real numbers is a vector space.

Solution: Let *V* be the set of all Vectors in a plane and *R* be the field of real numbers.

(V_1) (*V*, +) is an abelian group.

(V_{11}) $u, \nu \in V \Rightarrow u + \nu \in V$ (Closure axiom)

(V_{12}) $(u + \nu) + w = u + (\nu + w)$, for $u, \nu, w \in V$ (associative axiom)

(V_{13}) There is a null vector $O \in V$ such that

$u + 0 = u \forall\ u \in V$ (additive identity)

(V_{14}) If $u \in V, -u \in V$ and also $u + (-u) = 0$

Hence –*u* is inverse of *u* in *V*, i.e., inverse axiom is satisfied for each element in *V*.

(V$_{15}$) $u + \nu = \nu + u$ for all $u, \nu, \in V$.

$V_2\ a(u + \nu) = au + a\nu,\ a \in R, u, \nu \in V$

$V_3(a + b)u = au + bu, a, b \in R, u \in V.$

$V_4 a(bu) = (ab)u,\ a,\ b \in R, u \in V.$

$V_5\ 1\ u = u, \forall\ u \in V,$ where 1 is unity of *R*.

Hence *V* is a vector space over *R*.

Example 32: Let *C* be the field of complex numbers and *R* be the field of real numbers, then prove that

(i) *R* is a vector space over *R*.

(ii) *C* is a vector space over *C*.

Notes

Solution:

(i) $V_1(R, +)$ is an abelian group as $(R, +)$ is a field.

$V_2\ \alpha(a+b) = \alpha\, a + \alpha\, b \;\forall\; \alpha \in R$ and $\forall\; a, b \in R$.

$V_3(\alpha+\beta)\, a = \alpha\, a + \beta\, a \;\forall\; \alpha, \beta \in R$ and $\forall\; a \in R$.

$V_4 \alpha(\beta\, a) = (\alpha\, \beta) a, \;\forall\; \alpha, \beta \in R$ and $\forall\; a \in R$.

$V_5 1 \,.\, a = a \,.\, 1 = a, 1 \in R$ and $\forall\; a \in R$.

Hence R is a vector space over R.

(ii) $V_1(C, +)$ is an abelian group because C is a field

$V_2\ \alpha(u+\nu) = \alpha\, u + \alpha\, \nu \;\forall\; \alpha \in C$ and $\forall\; u, \nu \in C$

(using left distributive law of multiplication over addition in C.)

$V_3.\ (\alpha+\beta)u = \alpha\, u + \beta\, u, \;\forall\; \alpha, \beta \in C$ and $\forall\; u \in C$.

(using right distributive law in C)

$V_4 \alpha(\beta u) = (\alpha\, \beta) u, \;\forall\; \alpha, \beta \in C$ and $\forall\; u \in C$

(associative law of multiplication in C)

$V_5 1 \,.\, u = u$ for $1 \in C$ for $\forall\; u \in C$.

Hence C is a vector space over the field C.

Example 33: A field K can be regarded as a vector space over any subfield H or K.

Solution: We consider K as a set of vectors. Let us regard the elements of the satisfied H as scalars.

Let addition of vectors be the composition in the field K. Let us define the scalar multiplication as follows:

If $a \in H$ and $\alpha \in K$, $a\,\alpha$ is the product of these two elements in the field K.

V_1 Since K is a field, therefore $(K, +)$ is an abelian group.

$$V_2 a(\alpha+\beta) = a\,\alpha + a\,\beta \;\forall\; a \in H \text{ and } \forall, \alpha, \beta \in K.$$

This is a consequence of the left distributive law in K because

$$a, \alpha, \beta \in K \qquad\qquad \text{(because } H < K \text{ and } a \in H\text{)}$$

$V_3(a+b)\alpha = a\,\alpha + b\,\alpha \;\forall\; a, b \in H$ and $\forall\; \alpha \in K$. This is due to the right distributive law in K.

$V_4(ab)\alpha = a(b\alpha) \;\forall\; a, b \in H$ and $\forall\; \alpha \in K$. This result is due to associativity of multiplication in K.

$V_5.\ 1 \,.\, \alpha = \alpha \;\forall\; \alpha \in K$ where 1 is the unity of the subfield H. But $H \subset K$ and as such 1 is also the unity of the field K.

Hence K is a vector space over H.

Notes

General Properties of Vector Spaces

Let V be a vector space over a field F then

1. $a\,O = O$ for $a \in F, O \in V$
2. $O\nu = O$ for $O \in F, \nu \in V$
3. $a(-\nu) = (-\nu)a = -(a\nu)$ for $a \in F, \nu \in V$
4. $a(u-\nu) = au - a\nu$ for $a \in F, u$ and $\nu \in V$
5. If $a\nu = 0$ then either $a = 0$ or $V = 0$ for $a \in F, \nu \in V$.

Proof:

1.
$$\begin{aligned} \text{L.H.S} &= a\,O \\ &= a\,(O + O) && \text{(because } O = O + O\text{)} \\ &= a\,O + a\,O && \text{(distributive law)} \end{aligned}$$

Thus $aO = aO + aO$ or $aO + O = aO + aO$

Hence by cancellation law we get

$$aO = O.$$

2.
$$\begin{aligned} \text{L.H.S.} &= O\nu = (O+O)\nu && \text{(because } O = 0 + 0\text{)} \\ &= 0\nu + 0\nu && \text{(distributive law)} \end{aligned}$$

Thus $0\nu = 0\nu + 0\nu$

or $0 + 0\nu = 0\nu + 0\nu$

Hence by cancellation law

$$= 0\nu = 0.$$

3.
$$a\nu + a(-\nu) = a(\nu - \nu) = a\,0 = 0$$

Therefore $a\,\nu$ is additive inverse of $a(-\nu)$.

Again $a\,\nu + (-a)\nu = (\nu - \nu)a = Oa = 0.$

Therefore $a\,\nu$ is additive inverse of $(-\nu)\,a$.

i.e. $(-v)\,a = -av$

4.
$$\begin{aligned} \text{L.H.S.} &= a(u - \nu) \\ &= a\,[u + (-\nu)] \\ &= au + a(-\nu) && \text{[by property (3)]} \\ &= a\,u - a\nu \\ &= \text{R.H.S.} \end{aligned}$$

5. If $a = 0$ then the proposition is true.

But if $a \neq 0$ then a^{-1} exists in F.

$$\begin{aligned} a\nu &= 0 \Rightarrow a^{-1}(a\nu) = a^{-1}0 \Rightarrow (a^{-1}\ a)\nu \\ &= 0 \Rightarrow 1\,.\,\nu = 0 \Rightarrow \nu = 0. \end{aligned}$$

Notes

Cancellation

Let V be a vector space over a field F, then

(i) $a\nu = b\nu \Rightarrow a = b$ for $a, b \in F$ and $\nu \in V, \nu \neq 0.$

(ii) $au = a\nu \Rightarrow \nu = u$ for $a \in F\ a \neq 0,$ and $u, \nu \in V.$

Proof:

(i) L.H.S. = $a\nu = b\nu$ or $a\nu - b\nu = 0$

or $(a-b)\ \nu = 0.$

Since ν ≠ 0, therefore, we must have

$a - b = 0$ or $a = b$

(ii) L.H.S. $au = a\nu$

or $a(u - \nu) = 0$

Since $a \neq 0$, we must have

$\nu - u = 0 \Rightarrow u = \nu$

Example 34: Let F be a field and let V be the totality of all ordered n-tuples $(\alpha_1, \alpha_2, ...\alpha_n)$ where $\alpha_1\ \Sigma\ F.$ Two elements $(\alpha_1, \alpha_2, ...\alpha_n)$ and $(\beta_1, \beta_2, ...\beta_n)$ of V are declared to be equal if and only if $\alpha_i = \beta_i$ for each $i = 1, 2, ..., n$. We now introduce the requisite operations in V to make of it a vector space by defining:

1. $(\alpha_1, \alpha_2, ...\alpha_n) + (\beta_1, \beta_2, ...\beta_n) = (\alpha_1 + \beta_1, \alpha_2 + \beta_2, ...,\alpha_n + \beta_n)$

2. $a\ (\alpha_1, \alpha_2, ...\alpha_n) = (a\alpha_1, a\alpha_2, ...a\alpha_n)$ for $a \in F$

It is easy to verify that with these operations, V is a vector space over F.

Example 35: Let F be any field and let $V = F(x)$, the set of polynomials in x over F. We merely choose the fact that two polynomials can be added to get again a polynomial and that a polynomial can always be multiplied by an element of F. With these natural operations $F(x)$ is a vector space over F.

Example 36: The set of continuous real-valued functions on the real line is a real vector space with addition of functions $f + g$ and multiplication by real numbers as its laws of composition.

Example 37: The set of solution of the differential equation $\frac{d^2y}{dx^2} = -y$ is a real vector space.

Self Assessment

17. Show that the set W of ordered tried $(a_1, a_2, 0)$ where $a_1, a_2 \in F$ is a vector space.

18. Prove that the set $W = \{(x, 2y, 4z) : x, y, z \in R\}$ is a vector space.

19. In $F(x)$ let V_n be the set of all polynomials of degree less than n. Using the natural operations for polynomials of addition and multiplication by $a \in F$, show that V_n is a vector space over $F(x)$.

1.6 Summary

- The concept of set is fundamental in all branches of mathematics. A set according to the German mathematician George Cantor, is a *collection of definite well-defined objects of perception or thought*. By a well defined collection we mean that there exists a rule with the help of which it is possible to tell whether a given object belongs or does not belong to the given collection.
- Let A and B be two sets. The union of A and B is the set of all elements which are in set A or in set B. We denote the union of A and B by $A \cup B$, which is usually read as "A union B". On the other hand, the intersection of A and B is the set of all elements which are both in A and B. We denote the intersection of A and B by $A \cap B$, which is usually read as "A intersection B".
- The properties of natural numbers were developed in a logical manner for the first time by the Italian mathematician G. Peano, by starting from a minimum number of simple postulates. These simple properties, are known as the *Peano's Postulates* (*Axioms*).
- The system of rational numbers Q provides an extension of the system of integral Z, such that (*i*) $Q \supset Z$, (*ii*) addition and multiplication of two integers in Q have the same meanings as they have in Z and (*iii*) the subtraction and division operations are defined for any two numbers in Q, except for division by zero.

1.7 Keywords

Complex Number: The product set $R \times R$ consisting of the ordered pairs of real numbers.

Fields: A commutative ring with unity is called a field if its every non-zero element possesses a multiplicative inverse.

Irrational Number: A real number which cannot be put in the form p/q where p and q are integers.

Modulus of the Complex Number z: If $z = (a, b)$ be any complex number, then the non-negative real number $\sqrt{(a^2 + b^2)}$.

Operator or Transformation of A: If the domain and co-domain of a function f are both the same set say $f : A \to A$, then f is often called the operator.

Tabular form of the Set: Here the elements are separated by commas and are enclosed in brackets { }

1.8 Review Questions

1. Let S be a set of all real numbers of the form $(m + \sqrt{2}n)$ where $m, n \in Q$, a set of rational number, prove that S is a multiplication or additive group, m, n not vanishing simultaneously.

2. Prove that the four matrices

$$\begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}, \begin{pmatrix}-1 & 0\\0 & 1\end{pmatrix}, \begin{pmatrix}1 & 0\\0 & -1\end{pmatrix}, \begin{pmatrix}-1 & 0\\0 & -1\end{pmatrix}$$

form a multiplicative group.

3. If addition and multiplication modulo 10 is defined on the set of integers $R = \{0, 2, 4, 6, 8\}$.

Prove that the resulting system is a ring, Is it an integral domain?

4. Prove that the field has no proper ideals.

5. Show that the complex field C is a vector space over the real field R.

1.9 Further Readings

Books

I.N. Herstein *Topics in Algebra*.

Kenneth Hoffman and Ray Kunze, *Linear Algebra*

Michael Artin, *Algebra*

Unit 2: Vector Subspaces

CONTENTS

Objectives

After studying this unit, you will be able to:

- Understand the concept of a vector subspace
- Know more about subspaces by worked out examples
- See that a subspace has all the properties of a vector space.

Introduction

The unit one is the basis of the next five units. This unit is also based on the ideas of a vector space.

The subspace idea will help us in understanding the concepts of basis and dimension as well as how to set up the co-ordinates of a vector.

2.1 Vector Subspace

Let *V* be a vector space over a field *F*. Then a non-empty subset *W* of *V* is called a vector subspace of *V* if under the operations of *V*, *W* itself, is a vector space of *F*. In other words, *W* is a subspace of *V* whenever $w_1, w_2 \in W, \alpha, \beta \in F \Rightarrow \alpha w_1 + \beta w_2 \in W$.

Algebra of Subspaces

Theorem 1: The intersection of any two subspaces w_1 and w_2 of a vector space $V(F)$ is also a subspace of $V(F)$.

Proof: $w_1 \cap w_2$ is non-empty because at least $o \in w_1$ and w_2 both.

Let $u, v \in w_1 w_2$ and $\alpha, \beta \in F$

Then $u \in w_1 \cap w_2 \Rightarrow u \in w_1$ and $u \in w_2$

and $v \in w_1 \cap w_2 \Rightarrow v \in w_1$ and $v \in w_2$ since w_1 is subspace, hence

$$\alpha, \beta \in F \text{ and } u, v \in w_1 \Rightarrow \alpha u + \beta v \in w_1 \text{ and with the same argument}$$

$$\alpha, \beta \in F \text{ and } u, v \in w_2 \Rightarrow \alpha u + \beta v \in w_2.$$

Therefore $\alpha u + \beta v \in w_1$ and $\alpha u + \beta v \in w_2$.

$\Rightarrow \alpha u + \beta v \in w_1 \cap w_2$.

Thus $w_1 \cap w_2$ is a subspace of $V(F)$.

Theorem 2: The union of two subspaces is a subspace if one is contained in the other.

Proof: Let W_1 and W_2 be two subspaces of a vector space V.

Let $W_1 \subset W_2$ or $W_2 \subset W_1$. Then $W_1 \cup W_2$ or W_1 (whichever is the case). Since W_1, W_2 are subspaces of $V, W_1 \cap W_2$ is also a subspace of V.

Conversely, suppose $W_1 \cup W_2$ is a subspace of V then we have to prove $W_1 < W_2$ or $W_2 < W_1$. Suppose it is not so, i.e., let us assume that W_1 is not a subset of W_2 and W_2 is also not a subset of W_1.

If W_1 is not a subset of W_2 then it implies that there exists

$$\alpha \in W_1 \text{ and } \alpha \notin W_2 \qquad \text{...(i)}$$

Similarly if W_2 is not a subset of W_1 then there exists

$$\beta \in W_2 \text{ and } \beta \notin W_1 \qquad \text{...(ii)}$$

From (i) and (ii) we see that

$\alpha \in W_1 \cup W_2$ and $\beta \in W_1 \cup W_2$ since $W_1 \cup W_2$ is a subspace of $V, \alpha + \beta \in W_1 \cup W_2$

But $\alpha + \beta \in W_1 \cup W_2 \Rightarrow (\alpha + \beta) \in W_1$ or W_2.

Suppose it belongs to W_1 then since $\alpha \in W_1$ and W_1 is a subspace of $V, (\alpha + \beta) - \alpha = \beta \in W_1$ which is contradiction. Similar contradiction is arrived by assuming $\alpha + \beta \in W_2$.

Therefore, either $W_1 \subset W_2$ or $W_2 \subset W_1$.

2.2 Illustrative Examples

Example 1: Prove that the set W of ordered tried (a_1, a_2, 0) where $a_1, a_2 \in F$ is a subspace of $V_3(F)$,

Solution: Let $a = (a_1, a_2, 0)$ and $b = (b_1, b_2, 0)$ be two elements of W.

Therefore $a_1, a_2, b_1, b_2 \in F$. Let $a, b \in F$ then

$$a\alpha + b\beta \quad = a(a_1, a_2, 0) + b(b_1, b_2, 0)$$

$$= (aa_1, aa_2, 0) + (bb_1, bb_2, 0)$$

$$= (aa_1 + bb_1, aa_2 + bb_2, 0) \in W$$

because $aa_1 + bb_1, aa_2 + bb_2 \in F$.

Therefore, *W* is a subspace of $V_3(F)$.

Example 2: Let *R* be the field of real numbers. Show that

$\{x, 2y, 3z) : x, y, z \in R\}$ is a subspace of $V_3(R)$.

Solution: Let $W = \{(x, 2y, 3z) : x, y, z \in R\}$.

Let $\alpha = (x_1, 2y_1, 3z_1), \beta = (x_2, 2y_2, 3z_2)$ be any two elements of W then $x_1, y_1, z_1, x_2, y_2, z_2$ are obviously real numbers. If *a*, *b* are two real numbers, then

$$a\alpha + b\beta \quad = a(x_1 + 2y_1 + 3z_1) + b(x_2 + 2y_2 + 3z_2)$$

$$= (ax_1 + bx_2, 2ay_1 + 2ay_2, 3az_1 + 3az_2)$$

which belongs of $W, ax_1 + bx_2, ay_1 + by_2$ and $az_1 + bz_2$ being real numbers.

Thus $\alpha, \beta \in R$ and $b \in W$

$$a\alpha + b\beta \in W.$$

i.e., *W* is subspace of $V_3(R)$.

Example 3: If *V* is any vector space, *V* is a subspace of *V*; the subset consisting of the zero vector alone is a space of *V*, and is called the zero subspace.

Example 4: An *n* × *n* matrix A over the field F is symmetric if $A_{ij} = A_{ji}$, for each i and j. The symmetric matrices form a subspace of the square of all *n* × *n* matrices over the field *F*.

Example 5: The space of polynomial functions over the field *F* is a subspace of the space of all functions from *F* into *F*.

Example 6: Let *F* be a subfield of the field C of complex numbers, and let *V* be the vector space of all 2 × 2 matrices over *F*. Let W_1 be the subset of *V* consisting of all matrices of the form

$$\begin{pmatrix} a & b \\ c & 0 \end{pmatrix}$$

where a, b, c are arbitrary scalars in F. Finally let W_2 be the subset of V consisting of all matrices of the form

$$\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}$$

where a, b are arbitrary scalars in F. Then W_1, W_2 are subspaces of V.

Example 7: The solution space of a system of homogeneous linear equations. Let us consider the simultaneous equations involving n unknown x_i's.

$$a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = 0$$

$$a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = 0$$

$$\begin{matrix} . & . & . \\ . & . & . \\ . & . & . \\ . & . & . \\ . & . & . \end{matrix}$$

$$a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = 0$$

In matrix form we write the equation as

$$AX = 0$$

where A is a $m \times n$ matrix over the field F as all $a_{ij} \in A$ for i = 1 to m and j = 1 to n. Then the set of all $n \times 1$ matrices X over the field such that

$$AX = 0$$

is a subspace of the space of all $n \times 1$ matrices over F. To prove this we must show that

$$A(ax + y) = 0$$

when $AX = 0$ and $AY = 0$.

and C is an arbitrary scalar in F.

Consider a matrix A an $m \times n$ matrix over F and B and C are $n \times p$ matrices over F, then

$$A(a\,B + C) = a\,(AB) + AC$$

for each scalar a in F. Now

$$\begin{aligned} \left[A(aB+C)\right]_{ij} &= \sum_k A_{ik}(aB+C)_{kj} \\ &= \sum_k \left(aA_{ik}B_{kj} + A_{ik}C_{kj}\right) \\ &= a\sum_k A_{ik}B_{kj} + \sum_k A_{ik}C_{kj} \\ &= a(AB)_{ij} + (AC)_{ij} \\ &= (aAB + AC)_{ij} \end{aligned}$$

Similarly one can show that

$(aB+C)A = a(BA)+CA$, if the matrix sums and products are defined.

Thus $A(aX+Y) = a(AX)+AY = a(0)+0 = 0$

Theorem 3: Let *V* be a vector space over the field *F*. The intersection of any collection of subspaces of *V* is a subspace of *V*.

Proof: As shown in theorem 1, here let $\{W_a\}$ be a collection of subspaces of *V*, and let $W = \underset{a}{\cap} W_a$ be their intersection. Remember that W is defined as the set of all elements belonging to every W_a. Also since each W_a is a subspace, each contains the zero vector. Thus *W* is a non-empty set. Let *u, v* be vectors in *W* and $\alpha, \beta \in F$. Then

$$u \in W, v \in W$$

So $\quad u, v \in W$ and $\alpha, \beta \in F$

Therefore $\alpha u + \beta v \in W$ since $\alpha u + \beta v$ is in all W_i's. Thus $W = \underset{a}{\cap} W_a$ is a subspace of *V(F)*.

Definition: Let S_1, S_2, ..., S_n are subsets of a vector space *V*, the set of all sums

$$\alpha_1 + \alpha_2 + ... + \alpha_k$$

of vectors $\alpha_i \in S_i$ is called the sum of the subsets $S_1, S_2, ...S_k$ and is denoted by

$$S_1 + S_2 + ... + S_k \text{ or by } \sum_{\text{i}=1}^{\text{k}} S_i.$$

If $W_1, W_2, W_3 ... W_k$ are subspaces of *V*, then the sum

$$W = W_1 + W_2 + ... + W_k$$

is easily seen to be a subspace of *V* which contains each of the subspaces W_i. From this it follows, that *W* is a subspace spanned by the union of $W_1, W_2, W_3 ... W_k$.

Example 8: Let *F* be a subfield of the field *c* of complex numbers. Suppose

$$\alpha_1 = (1,2,0,3,0)$$

$$\alpha_2 = (0,0,1,4,0)$$

$$\alpha_3 = (0,0,0,0,1)$$

Now a vector α is in the subspace *W* of F^5 spanned by $\alpha_1, \alpha_2, \alpha_3$ if and only if there exist scalars c_1, c_2, c_3 in *F* such that

$$\alpha = c_1\alpha_1 + c_2\alpha_2 + c_3\alpha_3$$

Thus *W* consists of all vectors of the form

$$\alpha = (c_1, 2c_1, c_2, 3c_1 + 4c_2, c_3)$$

Where c_1, c_2, c_3 are arbitrary scalars in *F*. Writing α the set of all 5-tuples

$$\alpha = (x_1, x_2, x_3, x_4, x_5)$$

with $x_i \in F$ such that

$$x_2 = 2x_1$$

$$x_4 = 3x_1 + 4x_2$$

It is clear that the vector $(-3,-6,1,-5,2)$ is in *W*, whereas $(2,4,6,7,8)$ is not in *W*.

Self Assessment

1. Let $V = R^3 = (x,y,z : x,y,z \in R)$ and let *W* be the set of all triples (x,y,z) such that $x - 3y + 4z = 0$

   Show that *W* is a subspace of *V*.

2. Prove that the set *W* of *n*-tuples $(x_1, x_2, \ldots x_{n-1}, 0)$ where all *x*′s belong to *F*, is a subspace of the vector space $V_n(F)$.

3. Show that the set *W* of the elements of the vector space $V_3(R)$, of the form $(x+2y, y, 3y-x), x, y \in R$, is a subspace of $V_3(R)$.

4. Let *V* be the space of all polynomial functions over *F*. Let *S* be the subset of *V* consisting of the polynomial functions $f_0, f_1, f_2, \ldots$ defined by

   $f_n(x) = x^n, n = 0,1,2,\ldots$

   Show that W is the subspace spanned by the set *S*.

5. Show that the vector (3,–1,0,–1) is not in the subspace of R^4 spanned by the vectors $(2,-1,3,2), (-1,1,1,-3)$ and $(1,1,9,-5)$.

2.3 Summary

- If *V* is any vector space, *V* is a subspace of *V*; the subset consisting of the zero vector alone is a space of *V*, and is called the zero subspace.
- Let *F* be a subfield of the field C of complex numbers, and let *V* be the vector space of all 2 × 2 matrices over *F*. Let W_1 be the subset of *V* consisting of all matrices of the form

$$\begin{pmatrix} a & b \\ c & 0 \end{pmatrix}$$

- Consider a matrix *A* an *m* × *n* matrix over *F* and *B* and *C* are *n* × *p* matrices over *F*, then

  $A(a\,B + C) = a\,(AB) + AC$

  for each scalar a in *F*.

- Let S_1, S_2, ..., S_n are subsets of a vector space V, the set of all sums

  $\alpha_1 + \alpha_2 + ... + \alpha_k$

  of vectors $\alpha_i \in S_i$ is called the sum of the subsets $S_1, S_2, ...S_k$ and is denoted by

  $S_1 + S_2 + ... + S_k$ or by $\sum_{i=1}^{k} S_i$.

2.4 Keywords

Symmetric Matrix: An $n \times n$ matrix A over the field F is symmetric if $A_{ij} = A_{ji}$, for each i and j. The symmetric matrices form a subspace of the square of all $n \times n$ matrices over the field F.

Vector Subspace: Let V be a vector space over a field F. Then a non-empty subset W of V is called a vector subspace of V if under the operations of V, W itself, is a vector space of F.

2.5 Review Questions

1. Consider the three sets A,B,C such that

   $A = \{x_1, x_2 ; x_1 \le x_2\}$

   $B = \{x_1, x_2 ; x_1 x_2 \ge 0\}$

   $C = \{x_1, x_2 ; x_1 = x_2\}$

   which of these sets are subspace of $V(2)$? Give reasons.

2. Let $V = R^3 = \left[(x, y, z); x, y, z \in R\right]$ and Let W be the set of all triples (x, y, z) such that

   $2x - 3y + 4z = 0$

   Show that W is a subspace of V.

3. Let V be the vector space of functions from R into R let V_s be the subset of even functions $f(-x) = f(x)$; let V_0 be the subset of odd functions $f(-x) = -f(x)$. Then

   (a) Prove that V_s and V_0 are subspaces of V.

   (b) Prove that $V_s + V_0 = V$

   (c) Prove that $V_s \cap V_0 = \{0\} =$ null vector.

4. Let W_1 and W_2 be subspaces of a vector space V such that $W_1 + W_2 = V$ and $W_1 \cap W_2 = (0)$. Prove that for each α in V there are unique vectors α_1 in W_1 and α_2 in W_2 such that $\alpha = \alpha_1 + \alpha_2$.

Notes

2.6 Further Readings

Books

Kenneth Hoffman and Ray Kunze, *Linear Algebra*

I. N. Herstein, *Topics in Algebra*

Michael Artin, *Algebra*

Unit 3: Bases and Dimensions of Vector Spaces

CONTENTS

Objectives

After studying this unit, you will be able to:

- See that in dealing with a finite dimensional vector space *V* over the *F*, we sometime enquire whether a set of vectors is dependent or independent set.
- Understand that if you find a set of vectors as independent set in a vector space *V* then this set of vectors forms the basis of the space *V* and the number of vectors in the sets defines the dimension of the space *V*.

Introduction

In this unit we explain the concept of linear dependence and linear independence of the set of vectors.

The number of independent set of vectors determines the dimension of the vector space and the set of independent vectors forms the basis of the vector space.

3.1 Linear Dependence and Linear Independence of Vectors

Linear Dependence: Let $V(F)$ be a vector space and let $S = \{u_1, u_2, ...u_n\}$ be a finite subset of V. Then *S* is said to be linearly dependent if there exists scalars $\alpha_1, \alpha_2 ... \alpha_n \in F$, not all zero, such that

$$\alpha_1 u_1 + \alpha_2 u_2 + ... + \alpha_n u_n = 0.$$

Linear Independence: Let $V(F)$ be a vector space and let $S = \{u_1, u_2, ...u_n\}$ be finite subset of *V*. Then *S* is said to be linearly independent if

$$\sum_{i=1}^{n} a_u u_i = 0, \alpha_1 \in F.$$

holds only when $\alpha_i = 0,$ $\quad i = 1, 2, ...n.$

The following are easy consequences of the definition:

1. Any set which contains a linearly dependent set is linearly dependent.
2. Any subset of linearly independent set is linearly independent.
3. Any set which contains 0 vector is linearly dependent.
4. A set *S* of vectors is linearly independent if and only if each finite subset of *S* is linearly independent.

An infinite subset *S* of *V* is said to be linearly independent if every finite subset *S* is linearly independent, otherwise it is linearly dependent.

Illustrative Examples

Example 1: Show that the system of three vectors (1, 3, 2)(1, –7, –8), (2, 1, –1) of $V_3(R)$ is linearly dependent.

Solution: For $\alpha_1, \alpha_2, \alpha_3 \in R$ such that

$$\alpha_1(1,3,2) + \alpha_2(1,-7,-8) + \alpha_3(2,1,-1) = 0$$

$$\Leftrightarrow \quad (\alpha_1 + \alpha_2 + 2\alpha_3, 3\alpha_1 - 7\alpha_2 + \alpha_3, 2\alpha_1 - 8\alpha_2 - \alpha_3) = 0$$

$$\Leftrightarrow \quad \alpha_1 + \alpha_2 + 2\alpha_3 = 0, 3\alpha_1 - 7\alpha_2 + \alpha_3 = 0, 2\alpha_1 - 8\alpha_2 - \alpha_3 = 0$$

$$\Leftrightarrow \quad \alpha_1 = 3, \alpha_2 = 1, \alpha_3 = -2.$$

Therefore, the given system of vectors is linearly dependent.

Example 2: Consider the vector space $R^3(R)$ and the subset $S = \{(1,0,0),(0,1,0),(0,0,1)\}$ of R^3. Prove that *S* is linearly independent.

Solution: For $\alpha_1, \alpha_2, \alpha_3 \in R$,

$$\alpha_1(1,0,0) + \alpha_2(0,1,0) + \alpha_3(0,0,1) = (0,0,0)$$

$$\Leftrightarrow \quad (\alpha_1, \alpha_2, \alpha_3) = (0,0,0)$$

$$\Leftrightarrow \quad \alpha_1 = 0, \alpha_2 = 0, \alpha_3 = 0.$$

This shows that if any linear combination of elements of *S* is zero then the coefficients must be zero. *S* is linearly independent.

Example 3: Show that $\{1, x, 1 + x + x^2\}$ is a linearly independent set of vectors in the vector space of all polynomials over the real number field.

Solution: Let α, β, γ be scalar (real numbers) such that

$$\alpha(1) + \beta(x) + \gamma(1 + x + x^2) = 0 \text{ then}$$

$$\alpha + \beta x + \gamma(1 + x + x^2) = 0$$

$$\Rightarrow \quad \alpha + \gamma + (\beta + \gamma)x + \gamma x^2 = 0$$

$$\Rightarrow \quad \alpha + \gamma = 0, \beta + \gamma = 0, \gamma = 0,$$

$\Rightarrow \qquad \alpha = 0, \beta = 0, \gamma = 0.$

Hence the vectors $1, x, 1+x+x^2$ are linear independent over the field of real numbers.

Example 4: If the set $S = \{\alpha_1, \alpha_2, ...\alpha_n\}$ of vectors of $V(F)$ is linearly independent, then none of the vectors $\alpha_1, \alpha_2, ...\alpha_n$ can be zero vector.

Solution: Let α_r be zero vector where $1 \le r \le n$ then

$$0\alpha_1 + 0\alpha_2 + ... + a\alpha_r + 0\alpha r + ... + 1 + 0\alpha_n = 0$$

for any $a \ne 0$ in F.

Since $a \ne 0$ we notice that S is linearly dependent. This is contrary to what is given.

Hence none of the vectors $\alpha_1, \alpha_2, ...\alpha_n$ can be a zero vector.

3.2 Basis and Dimension of a Vector Space

A subset S of a vector space $V(F)$ is said to be a basis of $V(F)$, if

(i) S consists of linearly independent vectors, and

(ii) S generates $V(F)$ i.e. $\xi(S) \in V$ i.e. each vector in V is a linear combination of the finite number of elements of S.

For example the set (1, 0, 0), (0, 1, 0), (0, 0, 1) is a basis of the vector space $V_3(R)$ over the field of real numbers.

The set $\beta = (v_1, v_2, v_3, \ldots, v_n)$ is a basis of V if every vector w in V can be written in a unique way as a combination $w = x_1v_1 + x_2v_2 + \quad \ldots\ldots\ldots\ldots\ldots\ldots + x_nv_n$.

If every vector can be uniquely written as a combination, of the vectors $v_1, v_2, \ldots v_x$ of β, then β is independent and spans V, so β is a basis.

If V is a finite dimensional vector space, then it contains a finite set $v_1, v_2, \ldots, v_n$ of linearly independent elements that spans V.

If $v_1, v_2, \ldots v_n$ is a basis of V over F and if $w_1, w_2, \ldots w_m$ in V are linearly independent over F, then $m \le n$.

We also see that if V is finite-dimensional over F then any two basis of V has the same number of elements.

Thus for a finite dimensional space V, the basis has a unique number of elements say n. This unique integer, n; in fact, is the number of elements in any basis of V over F.

Definition: The integer n is called the dimension of the vector space over F.

The ***Dimension*** of a finite space V over F is thus the number of elements in any basis of V over F.

A vector space V is finite-dimensional if some finite set of vectors spans V. Otherwise V is infinite dimensional.

The dimension of V will be denoted by dim V.

If W is the subspace of a finite dimensional vector space V, then W is finite dimensional, and dim $W \le$ dim V. Moreover, dim W = dim V if and only if $W = V$.

Illustrative Examples

Example 5: Show that the set

$S = \{(1,2,1)(2,1,0),(1,-1,2)\}$ forms a basis for $V_3(F)$.

Solution: Let $a_1, a_2, a_3 \in F$.

then $a_1(1,2,1) + a_2(2,1,0) + a_3(1,-1,2) = 0$

$\Rightarrow \quad (a_1 + 2a_2 + a_3, 2a_1 + a_2 - a_3, a_1 + 2a_3) = (0,0,0)$

$\Rightarrow \quad a_1 + 2a_2 + a_3 = 0, 2a_1 + a_2 - a_3 = 0, a_1 + 2a_3 = 0$

$\Rightarrow \quad a_1 = a_2 = a_3 = 0.$

Hence the given set is linearly independent.

Now let $(1,0,0) \quad = x(1,2,1) + y(2,1,0) + z(1,-1,2)$

$$= (x + 2y + z, 2x + y - z, x + 2z)$$

so that $x + 2y + z = 1, 2x + y - z = 0, x + 2z = 0$

$\therefore\ x = -2/9, y = 5/9, z = 1/9$

Thus, the unit vector (1,0,0) is a linear combination of the vectors of the given set, i.e.

(1, 0, 0) = –2/9 (1, 2, 1) + 5/9(2, 1, 0) + 1/9 (1, –1, 2)

Similarly,

(0, 1, 0) = 4/9 (1, 2, 1) – 1/9(2, 1, 0) – 2/9 (1, –1, 2) and

(0, 0, 1) = 1/3 (1, 2, 1) – 1/3(2, 1, 0) + 1/3 (1, –1, 2)

Since $V_3(F)$ is generated by the unit vectors (1,0,0), (0,1,0),(0,0,1) we see therefore that ever elements of $V_3(F)$ is a linear combination of the given set *S*. Hence the vectors of this set form a basis of $V_3(F)$.

Example 6: Prove that system *S* consisting *n* vectors

$e_1 = \{1,0,...0\}, e_2 = \{0,1,...,0\}...e_n = \{0,0,...1\}$ is a basis of $V_n(F)$.

Solution: First we shall prove that the given system *S* is linearly independent.

Let $\quad a_1, a_2, \ldots a_n$ be any scalars, then

$a_1e_1 + a_2e_2 + \ldots a_ne_n = 0$

$\Rightarrow \quad a_1(1,0,...,0) + a_2(0,1,...0) + \ldots + a_n(0,0,...,1) = 0$

$\Rightarrow \quad (a_1, a_2, ...a_n) = 0$

$\Rightarrow \qquad a_1 = 0, a_2 = 0, ... a_n = 0$

Therefore, S is linearly independent set.

Further, we must show that $L(S) = V_n(F)$.

Let $v = (v_1, v_2, ... v_n)$ be any vector in $V_n(F)$. We can write

$(v_1, v_2, ... v_n) = v_1(1,0,...,0) + v_2(0,1,...,0) + ... + v_n(0,0,...,1)$

i.e., $v = v_1 e_1 + v_2 e_2 + ... + v_n e_n$.

Hence S is a basis of $V_n(F)$.

Example 7: Prove that the vector space $F(x)$ of polynomials over the field F has a basis S, such that $S = \{1, x, x^2, ...\}$.

Solution: Let a, b, c, ... be scalars such that

$a(1) + b(x) + c(x^2) + ... + = 0$

$\Rightarrow \qquad a = 0, b = 0, c = 0, ...$

$\therefore$ the vectors 1, x, x^2,... are linearly independent.

Let $f(x) = a_0 + a_1 x + a_2 x^2 + ... + a_i x^i$ be a polynomial in the given vector space then

$$f(x) = a_0(1) + a_1(x) + a_2(x^2) + ... a_i(x^i)$$

$\Rightarrow f(x)$ can be expressed as a linear combination of a finite number of elements of $\{1, x, x^2, ...\}$.

Thus $\{1, x, x^2, ...\}$ is a basis.

Self Assessment

1. Find the condition that the vectors (a_1, a_2) and (b_1, b_2) in $V_2(F)$ are linearly dependent.

   [***Ans:*** $a_1 b_2 - a_2 b_1 = 0$]

2. Test the linear dependence or independence of the vectors:

   (i) $\alpha_1 = (0,1,-2), \alpha_2 = (1,-1,1), \alpha_3 = (1,2,1)$ in $V_3(R)$

   (ii) $(1,2,3),(3,-2,1)(2,-6,5)$ in R^3

   (iii) $(1,0,-1),(2,1,3)(-1,0,0)(1,0,1)$ in $V_3(R)$.

   (iv) The set $\{(1,2,1),(3,1,5)(3,4,7)\}$

3. Is the vector $\alpha = (2,-5,-3)$in $V_3(R)$,a linear combination of vectors.

   $\alpha_1 = (1,-3,2)$, $\alpha_2 = (2,-4,-1)$, $\alpha_3 = (1,-5,-7)$?

4. Prove that the number of elements in a basis of a finite dimensional vector space is unique.

5. If $\{e_1, e_2, e_3\}$ is a basis for R^3, then show that

   $\{e_2, e_3 + e_1, e_1 + e_2\}$ is also a basis.

6. Show that the set $S = \{(1,0,0)(1,1,0)(0,1,1),(0,1,0)\}$ spans $V_3(R)$, but does not form a basis.

7. Show that the set $\{(2,-1,0)(3,5,1)(1,1,2)\}$ forms a basis of $V_3(R)$.

3.3 Summary

- Let $V(F)$ be a vector space and let $S = \{u_1, u_2, ...u_n\}$ be a finite subset of V. Then S is said to be linearly dependent if there exists scalars $\alpha_1, \alpha_2 ... \alpha_n \in F$, not all zero, such that

  $\alpha_1 u_1 + \alpha_2 u_2 + ... + \alpha_n u_n = 0.$

- Let $V(F)$ be a vector space and let $S = \{u_1, u_2, ...u_n\}$ be finite subset of V. Then S is said to be linearly independent if

  $$\sum_{i=1}^{n} a_u u_i = 0, \alpha_1 \in F.$$

  holds only when $\alpha_i = 0,\ i = 1, 2, ...n.$

3.4 Keywords

Dimension: The Dimension of a finite space *V* over *F* is thus the number of elements in any basis of *V* over *F*.

Linear Combination: $V_3(F)$ is generated by the unit vectors (1,0,0), (0,1,0), (0,0,1) therefore that elements of $V_3(F)$ is a linear combination of the given set *S*.

3.5 Review Questions

1. Prove that a set of vectors containing null vector is a linearly dependent set.
2. Prove that the three functions $t^2, \cos x$ and e^x are linearly independent.
3. Prove that the set (1,2,0)(2,1,2)(3,1,1) is a basis for R^3.
4. Prove that if two vectors are linearly dependent, one of them is a scalar multiple of the other.

3.6 Further Readings

Books

Kenneth Hoffman and Ray Kunze, *Linear Algebra*

I N Herstein, *Topics in Algebra*

Michael Artin, *Algebra*

Unit 4: Co-ordinates

CONTENTS

Objectives

After studying this unit, you will be able to:

- See that the dimension and basis of a vector space *V* over the field *F* help us in introducing the co-ordinates of a vector.
- Understand how to go from one basis to another basis with the help of an invertible matrix.
- See that the solved examples help you to find the invertible matrix and hence the co-ordinates of the vector in the new basis can be found out.

Introduction

For an abstract vector space *V* over the field *F* can be spanned by a set of independent vectors which form the basis of the vector space V.

There are more than one way of finding the basis and so it is important to know the relation between one basis over the other.

4.1 Co-ordinates

So far we have dealt with basis and dimension in the unit 3. We also showed the linear independence and dependence of vectors. The dimension of a vector space is the number of basis vectors of the vector space *V* over the field. The standard basis for a three dimensional vector space is taken as

$$l_1 = (1,0,0)$$

$$l_2 = (0,1,0)$$

$$l_3 = (0,0,1)$$

and they form an independent set of vectors and span the whole V_3 over the field *R*.

Now we want to introduce co-ordinates in the vector space V analogous to the natural co-ordinates x_{i} of the vector

$$\alpha = (x_1, x_2 ... x_n)$$

in the space F^n. The co-ordinates in the three dimensional space F^3 are x, y, z co-ordinates. So the co-ordinates of a vector α in V relative to the basis β will be the scalars which serve to express α as a linear combination of the vectors in the basis. If the vectors in the basis are $(\varepsilon_1, \varepsilon_2, \varepsilon_3, ... \varepsilon_n)$ then the vector α is expressible in terms of its co-ordinates as well as in terms of the vectors of basis as follows

$$\alpha = (x_1, x_2, x_3, ... x_n) = \sum_{i=1}^{n} x_i \varepsilon_i \qquad ...(1)$$

For another vector β having co-ordinates $y_1, y_2, ... y_n$ we have

$$\beta = (y_1, y_2, y_3, ... y_n) = \sum_{i=1}^{n} y_i \varepsilon_i .$$

writing

$$\alpha = (x_1, x_2, x_3, ... x_n)$$

the vector α has a unique expression as a linear combination of the standard basis vectors (1), and the i^{th} co-ordinates x_{i} of α is the coefficient of ε_i in the expression (1). By this way of 'natural' ordering of the vectors in the standard basis i.e. by writing ε_1 as the first vector, ε_2 as the second vector etc. we define the order of the co-ordinates of the vector α also. So we have the definition:

Definition: If V is a finite-dimensional space, the ordered basis for V is a finite sequence of basis vectors $(\varepsilon_1, \varepsilon_2, \varepsilon_3, ... \varepsilon_n)$ which is a linearly independent set and spans V. So we just say that

$$\beta = (\varepsilon_1, \varepsilon_2, \varepsilon_3, ... \varepsilon_n) \qquad ...(2)$$

is an ordered basis for V. Now suppose V is a finite dimensional vector space over the field F and (2) is an ordered basis for V, there is a unique n-tuple $\alpha = (x_1, x_2, ... x_n)$ of scalars such that:

$$\alpha = \sum_{i=1}^{n} x_i \varepsilon_i .$$

The n-tuple is unique, because if we also have

$$\alpha = \sum_{i=1}^{n} z_i \varepsilon_i$$

then $\quad \sum_{i=1}^{n} (x_i - z_i) \varepsilon_i = 0$

Since ε_i for each i, is an independent set, so

$$x_i - z_i \equiv 0$$

or $\quad x_i = z_i$

We shall call x_i, the ith co-ordinate of α relative to the basis

$$\beta = (\varepsilon_1, \varepsilon_2, \varepsilon_3, ...\varepsilon_n)$$

If γ is an other vector having ordered co-ordinates $(y_1, y_2, ...y_n)$, then

$$\gamma = (y_1, y_2, ...y_n) = \sum_{i=1}^{n} y_i \varepsilon_i,$$

then $$\alpha + \gamma = \sum_{i=1}^{n} x_i \varepsilon_i + \sum_{i=1}^{n} y_i \varepsilon_i$$

$$= \sum_{i=1}^{n} (x_i + y_i) \varepsilon_i \qquad ...(3)$$

So that the i^{th} co-ordinate of $(\alpha + \gamma)$ in this ordered basis is $(x_i + y_i)$. Similarly the i^{th} co-ordinate of $(c\alpha)$ is cx_i. It is clear that every n-tuple $(z_1, z_2, ...z_n)$ is V_n is the n-tuple of co-ordinates of some vector z in V_n namely the vector

$$z = \sum_{i=1}^{n} z_i \varepsilon_i \qquad ...(4)$$

4.2 Change of Basis from One Ordered Basis to Another

In a three dimensional space V_3, we have $\varepsilon_1 = (1,0,0)$, $\varepsilon_2 = (0,1,0)$, $\varepsilon_3 = (0,0,1)$ as three independent set of basis vectors. We also know that by taking a certain combination of these ε_i's we find another set like

$$\varepsilon_1' = (1,1,0), \varepsilon_2' = (1,1,1) \text{ and } \varepsilon_3' = (0,1,1) \qquad ...(4A)$$

which is again independent. The set $\varepsilon_1', \varepsilon_2', \varepsilon_3'$ is related to the set $\varepsilon_1, \varepsilon_2, \varepsilon_3$ by the relations

and $$\left.\begin{aligned} \varepsilon_1' &= \varepsilon_1 + \varepsilon_2 \\ \varepsilon_2' &= \varepsilon_1 + \varepsilon_2 + \varepsilon_3 \\ \varepsilon_3' &= \varepsilon_2 + \varepsilon_3 \end{aligned}\right] \qquad ...(4B)$$

So by taking $\beta' = (\varepsilon_1', \varepsilon_2', \varepsilon_3')$ as a new basis of V_3 the vector α will have new co-ordinate system $\alpha = (x_1', x_2', x_3')$ given by

$$\alpha = \sum_{i=1}^{3} x_i' \varepsilon_i', \qquad ...(5)$$

We can now find a relation between the new co-ordinates (x_1', x_2', x_3') and old co-ordinates $(x_1, x_2, ..., x_n)$ of α in n dimensional space.

Notes

To find the relation, it is more convenient to use the matrix of α relative to the order basis

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \qquad ...(6)$$

rather than the n-tuple $(x_1, x_2, ...x_n)$ of co-ordinates.

This notation will be particularly useful as we now proceed to describe what happens to the co-ordinates of a vector α as we change from one ordered basis to another.

Suppose that we are dealing with a space V which is n dimensional and that the basis β is changed to a new basis β' i.e.

$$\beta = (\varepsilon_1, \varepsilon_2, \varepsilon_3, ...\varepsilon_n), \beta' = (\varepsilon_1^{'}, \varepsilon_2^{'}, \varepsilon_3^{'}, ...\varepsilon_n^{'}) \qquad ...(7)$$

Let there be unique scalars P_{ij} such that

$$\varepsilon_{ij}^{'} = \sum_{i=1}^{n} P_{ij}\varepsilon_i \qquad j = 1, 2, ...n \qquad ...(8)$$

Let $x_1^{'}, x_2^{'}, x_3^{'}, ...x_n^{'}$ be the co-ordinates of a given vector α in the basis β', then

$$\alpha \quad = x_1^{'}\varepsilon_1^{'} + x_2^{'}\varepsilon_2^{'} + ...x_n^{'}\varepsilon_n^{'}$$

$$= \sum_{j=1}^{n} x_j^{'}\varepsilon_j^{'}$$

$$= \sum_{j=1}^{n} x_j^{'} \sum_{i=1}^{n} P_{ij}\varepsilon_i$$

$$\text{or } \alpha \quad = \sum_{j=1}^{n} \varepsilon_i \left(\sum_{j=1}^{n} P_{ij} x_j^{'} \right) \qquad ...(8A)$$

Putting

$$x_i = \sum_{j=1}^{n} P_{ij} x_j^{'} \qquad ...(9)$$

We have

$$\alpha = \sum_{i=1}^{n} x_i \varepsilon_i \qquad ...(10)$$

where now x_i denotes the i^{th} co-ordinate of the vector α in the old β.

In matrix form equation (9) becomes

$$X = PX' \qquad ...(11)$$

where $X = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}, X' = \begin{bmatrix} \varepsilon_1^{'} \\ \varepsilon_2^{'} \\ \vdots \\ \varepsilon_n^{'} \end{bmatrix}$

$$P = \left(P_{ij}\right) \qquad \text{...(12)}$$

where P is an $n \times n$ matrix whose i,j entry is P_{ij} since β and β' basis are independent sets, $X = 0$ is only possible if $X' = 0$ also. Now the transformation matrix P is such that its inverse exists. Hence multiplying (6) by p^{-1} we obtain

$$X' = P^{-1}X \qquad \text{...(13)}$$

So the new set of co-ordinates $\left(x_1^{'}, x_2^{'}, x_3^{'}, ...x_n^{'}\right)$ are related to the old set of co-ordinates $(x_1, x_2, ...x_n)$ of the vector α by the relation (13).

Example 1: From equation (4), P matrix is given by

$$P = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix}$$

let $P = \begin{vmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{vmatrix} = -1$

Thus the new basis $\beta' = \left(\varepsilon_1^{'}, \varepsilon_2^{'}, \varepsilon_3^{'}\right)$ is given in terms of old basis $\beta = (\varepsilon_1, \varepsilon_2, \varepsilon_3)$ by the matrix relation

$$\begin{pmatrix} \varepsilon_1^{'} \\ \varepsilon_2^{'} \\ \varepsilon_3^{'} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix} \qquad \text{...(14)}$$

Now $P^{-1} = \begin{pmatrix} 0 & 1 & -1 \\ 1 & -1 & 1 \\ -1 & 1 & 0 \end{pmatrix}$ ...(15)

If the co-ordinates of $\alpha = (x_1, x_2, x_3)$ in old basis β then in the new basis β' they are given by

$$\begin{pmatrix} x_1^{'} \\ x_2^{'} \\ x_3^{'} \end{pmatrix} = \begin{pmatrix} 0 & 1 & -1 \\ 1 & -1 & 1 \\ -1 & 1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \qquad \text{...(16)}$$

Example 2: Show that the vectors $\varepsilon_1^{'} = (1,1,0,0), \varepsilon_2^{'} = (0,0,1,1), \varepsilon_3^{'} = (1,0,0,4), \varepsilon_4^{'} = (0,0,0,2)$ form a basis for R^4. Find the co-ordinates of each of the standard basis vectors in the ordered basis $\left(\varepsilon_1^{'}, \varepsilon_2^{'}, \varepsilon_3^{'}, \varepsilon_4^{'}\right)$.

Solution: To prove that the set $\varepsilon_1^{'}, \varepsilon_2^{'}, \varepsilon_3^{'}, \varepsilon_4^{'}$ form a basis, we have to show that they are independent.

So let c_1, c_2, c_3, c_4 are scalars not all of them zero such that $\varepsilon'_i s$ are dependent, then

$$c_1\varepsilon_1^{'} + c_2\varepsilon_2^{'} + c_3\varepsilon_3^{'} + c_4\varepsilon_4^{'} = 0$$

or $$c_1 + c_3 = 0$$

$$c_1 = 0$$

$$c_2 = 0$$

$$c_2 + 4c_3 + 2c_4 = 0$$

So we get $c_1 = 0$

$$c_2 = 0$$

$$c_3 = 0$$

$$c_4 = 0$$

Thus the four set of vectors $\varepsilon'_1, \varepsilon'_2, \varepsilon'_3, \varepsilon'_4$ are independent. Let *P* be a matrix such that

$$\begin{pmatrix} \varepsilon'_1 \\ \varepsilon'_2 \\ \varepsilon'_3 \\ \varepsilon'_4 \end{pmatrix} = P \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{pmatrix}$$

where $$\begin{aligned} \varepsilon_1 &= (1,0,0,0) \\ \varepsilon_2 &= (0,1,0,0) \\ \varepsilon_3 &= (0,0,1,0) \\ \varepsilon_4 &= (0,0,0,1) \end{aligned}$$

So $$P = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 4 \\ 0 & 0 & 0 & 2 \end{pmatrix}$$

Let *P* = –2, so, *P* is non-singular and invertible.

$$P^{-1} = \begin{pmatrix} 0 & 0 & 1 & -2 \\ 1 & 0 & -1 & 2 \\ 0 & 1 & 0 & -1/2 \\ 0 & 0 & 0 & 1/2 \end{pmatrix}$$

Thus $\varepsilon_1 = \varepsilon'_3 - 2\varepsilon'_4, \varepsilon_2 = \varepsilon'_1 - \varepsilon'_3 + 2\varepsilon'_4$

$\varepsilon_3 = \varepsilon'_2 - \varepsilon'_{4/2}, \varepsilon_4 = \varepsilon'_{4/2}$ is the answer.

Example 3: Let *V* be the vector space over the complex numbers of all functions from *R* into *C* i.e. the space of all complex-valued functions on the real line. Let $f_1(x) = 1, f_2(x) = e^{ix}, f^3(x) = e^{-1x}$.

(a) Prove that f_1, f_2, f_3 are linearly independent.

(b) Let $g_1(x)=1, g_2(x)=\cos x, g_3(x)=\sin x$. Find an invertible 3×3 matrix P such that

$$g_j=\sum_{i=1}^{3} P_{ij} f_i \qquad \text{for } j=1,2,3$$

Solution: Let $f_1(x), f_2(x), f_3(x)$ be a dependent set then we can find real c_1, c_2, c_3 not all of them zero so that

$$c_1 f_1(x)+c_2 f_2(x)+c_3 f_3(x)=0$$

or $$c_1.1+c_2 e^{ix}+c_3 e^{-ix}=0 \qquad \text{...(1)}$$

Taking real part we have

$$c_1+c_2\cos x+c_3\cos x=0 \qquad \text{...(2)}$$

Taking imaginary part we have

$$c_2-c_3=0 \qquad \text{...(3)}$$

From (2) we have $c_1=0, c_2+c_3=0$ for arbitrary x,

From (3) we have $c_2=c_3$

So we get $c_1=c_2=c_3$.

which contradicts the statement that all c's are not zero. So the set f_1, f_2, f_3 is an independent set.

So find g_1, g_2, g_3 in terms of $f_1(x), f_2(x), f_3(x)$ we see that

$$g_1(x)=f_1(x)=1$$

$$g_2(x)=\cos x=\frac{f_2(x)+f_3(x)}{2}$$

$$g_3(x)=\sin x=\frac{f_2(x)-f_3(x)}{2i}$$

Thus $$\begin{pmatrix} g_1(x)\\ g_2(x)\\ g_3(x)\end{pmatrix}=\begin{pmatrix}1 & 0 & 0\\ 0 & 1/2 & 1/2\\ 0 & 1/2i & -1/2i\end{pmatrix}\begin{pmatrix} f_1(x)\\ f_2(x)\\ f_3(x)\end{pmatrix} \qquad \text{...(4)}$$

Thus $$P=\begin{pmatrix}1 & 0 & 0\\ 0 & 1/2 & 1/2\\ 0 & -i/2 & i/2\end{pmatrix} \qquad \text{...(5)}$$

Also Let $$P=\frac{2i}{4}=\frac{i}{2}\neq 0 \qquad \text{...(6)}$$

So P is invertible 3 × 3 matrix given by (5).

Notes

Self Assessment

1. Find the co-ordinate matrix for the vector $\alpha = (1,0,1)$ in the basis of C^3 consisting of the vectors $(2i,1,0),(2,-1,1),(0,1+i,1-i)$ in that order.

2. Let $\beta' = (\varepsilon_1', \varepsilon_2', \varepsilon_3')$ be the ordered basis for R^3. Consisting of $\varepsilon_1' = (1,0,-1)$, $\varepsilon_2' = (1,1,1)$, $\varepsilon_3' = (1,0,0)$

   What are the co-ordinates of the vector $\alpha = (a,b,c)$ in the above ordered basis β'.

3. Let R be the field of the real numbers and let θ be a fixed real number. Let the new basis $\beta' = (\varepsilon_1', \varepsilon_2')$ be given in terms of the matrix P by the relation

   $$\begin{pmatrix} \varepsilon_1' \\ \varepsilon_2' \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}$$

   Here $\varepsilon_1 = (1,0)$ and $\varepsilon_2 = (0,1)$

   $$P = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

   Find the co-ordinates of the vector $\alpha(x_1, x_2)$ in terms of the new basis β'.

4. Show that the set of vectors $\beta' = (\varepsilon_1', \varepsilon_2', \varepsilon_3')$ given by

   $\varepsilon_1' = (-1,0,0)$

   $\varepsilon_2' = (4,2,0)$

   $\varepsilon_3' = (5,-3,8)$

   form a basis of F^3. Find the co-ordinates of the vector $\alpha = (x_1, x_2, x_3)$ in the basis β'.

4.3 Summary

- The dimension of a vector space is the number of basis vectors of the vector space V over the field. The standard basis for a three dimensional vector space is taken as

  $l_1 = (1,0,0)$

  $l_2 = (0,1,0)$

  $l_3 = (0,0,1)$

  and they form an independent set of vectors and span the whole V_3 over the field R.

- The co-ordinates in the three dimensional space F^3 are x, y, z co-ordinates. So the co-ordinates of a vector α in V relative to the basis β will be the scalars which serve to express α as a linear combination of the vectors in the basis.

- If V is a finite-dimensional space, the ordered basis for V is a finite sequence of basis vectors $(\varepsilon_1, \varepsilon_2, \varepsilon_3, ...\varepsilon_n)$ which is a linearly independent set and spans V. So we just say that

  $$\beta = (\varepsilon_1, \varepsilon_2, \varepsilon_3, ...\varepsilon_n)$$

  is an ordered basis for V.

4.4 Keywords

n-tuple $(z_1, z_2, ...z_n)$: V_n is the n-tuple of co-ordinates of some vector z in V_n namely the vector

$$z = \sum_{i=1}^{n} z_i \varepsilon_i$$

Unique Scalars: P_{ij} are such that

$$\varepsilon'_{ij} = \sum_{i=1}^{n} P_{ij} \varepsilon_i \qquad j = 1, 2, ...n$$

4.5 Review Questions

1. Show that the vectors

   $$\alpha_1 = (1,1,0,0), \alpha_2 = (1,0,0,4),$$
   $$\alpha_3 = (0,0,1,1), \alpha_4 = (0,0,0,2)$$

   form a basis for R^4. Find the co-ordinates of the standard basis vectors in the ordered basis $(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$.

2. Let W be the subspace of C^2 spanned by $\alpha_1 = (1, 0, i)$ and $\alpha_2 = (1+i, 1, -1)$

   (a) Show that α_1 and α_2 form basis for W.

   (b) Show that the vectors $\beta_1 = (1,1,0)$ and $\beta_2 = (1, i, 1+i)$ are in W and form an other basis for W.

   (c) What are the co-ordinates of α_1 and α_2 in the ordered basis (β_1, β_2) for W?

Answers: Self Assessment

1. $\left(-1, \frac{1+2i}{2}, \frac{3+i}{4}\right)$

2. $(b-c, b, a-2b+c)$

3. $x'_1 = \cos\theta x_1 + \sin\theta x_2$
   $x'_2 = -\sin\theta x_1 + \cos\theta x_2$

4. $\left(-x_1 + 2x_2 + \frac{11x_3}{8}, \frac{x_2}{2} + \frac{3x_3}{8}, \frac{x_3}{8}\right)$

4.6 Further Readings

Books

Kenneth Hoffman, Ray Kunze, *Linear Algebra*

Michael Artin, *Algebra*

Unit 5: Summary of Row-Equivalence

Notes

Objectives

After studying this unit, you will be able to:

- Understand the technique of row operations on matrices of $m \times n$ type.
- Know that if B is a matrix obtained from row operations of A then B and A are called row equivalent.
- Understand how to obtain a row reduced echelon matrix.

Introduction

In solving a system of simultaneous equations the method of row operations on $m \times n$ matrix helps in finding the solution.

The idea of row space of a matrix helps in finding the subspace of the row space.

5.1 Matrices and Elementary Row Operations

Suppose F is a field. We consider the problem of finding x-scalars, x_1, x_2, ... x_n which satisfy the conditions

$$\left.\begin{array}{ccccccccccc} A_{11}x_1 & + & A_{12}x_2 & + & A_{13}x_3 & + \cdots + & A_{1n}x_n & = & y_1 \\ A_{21}x_1 & + & A_{22}x_2 & + & A_{23}x_3 & + \cdots + & A_{2n}x_n & = & y_2 \\ \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\ A_{m1}x_1 & + & A_{m2}x_2 & + & A_{m3}x_3 & + \cdots + & A_{mn}x_n & = & y_n \end{array}\right\} \quad \ldots (1)$$

where $y_1,...y_m$ and A_{ij}, $l < i < m$, $i \leq j \leq n$ are given elements of F. We shall now abbreviate the system of equations (1) by

Notes

Where
$$\left.\begin{aligned} &AX = Y \\ &A = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{bmatrix} \\ &X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \end{aligned}\right\} \quad \text{...(2)}$$

In solving the linear system of equations (1) we sometimes use the technique of elimination. We can illustrate this method on the following homogeneous equations:

$2x_1 - x_2 + x_3 = 0$

$x_1 + 3x_2 + 4x_3 = 0$

If we add (–2) times the second equation to the first equation,

we obtain

$-7x_2 - 7x_3 = 0$

or $x_2 = - x_3$

If we add (3) times the first equation to the second equation

we obtain

$7x_1 + 7x_3 = 0$

or $x_1 = -x_3$

So we conclude that if (x_1, x_2, x_3) is the solution then $x_1 = x_2 = -x_3$. Thus the set of solutions consists of all triples $(a, a, -a)$.

For the general system (1), suppose we select m, scalars $c_1, c_2, ... c_m$, multiply the jth equation by c_j and then add, we obtain the equations

$$(C_1A_{11} + ... + C_mA_{m1})x_1 + ... + (C_1A_{1n} + C_2A_{2n} + ... + C_mA_{mn})x_n = \sum_{j=1}^{m} C_j y_j$$

Such an equation is called a linear combination of the equations in (1). Evidently any solution of the entire system of equations (1) will also be the solution of this new equation. This is the fundamental idea of the elimination process. Thus if we have another system of linear equations

$$\left.\begin{matrix} B_{11}x_1 & + & B_{12}x_2 & + & \dots & + & B_{1n}x_n & = & Z_1 \\ \vdots & & \vdots & & & & \vdots & & \vdots \\ B_{K1}x_1 & + & B_{K2}x_2 & & \cdots & & B_{Kn}x_n & = & Z_K \end{matrix}\right\} \quad \text{...(3)}$$

in which each of the K equations is a linear combination of the equations (1), then every solution of (1) is a solution of the new system (2).

Let us say that two systems of linear equations are **equivalent** if each equation in each system is a linear combination of the equations in the other system. We then formally state the following theorem:

Theorem 1: Equivalent systems of linear equations have exactly the same solutions.

Consider now the system (1) as given by the system (2). We call A the matrix of coefficients of the system. We wish now to consider operations on the rows of the matrix A which correspond to forming linear combinations of the equations in the system $AX = Y$. We restrict ourselves to three **elementary row operations** on $m \times n$ matrix A over the field F:

1. Multiplication of one row of A by a non-zero scalar c;

2. Replacement of the r^{th} row of A by row r plus c times row s, c is any scalar and $r \neq s$;

3. interchange of two rows of A.

An elementary row operation is thus a special type of function (rule) e which is associated with each $m \times n$ matrix (A). One can precisely describe e in the three cases as follows:

1. $e(A)_{ij} = A_{ij}$ if $i \neq r$, $e\,(A)_{rj} = cA_{rj}$

2. $e(A)_{ij} = A_{ij}$ if $i \neq r$, $e(A)_{rj} + cA_{rj}$

3. $e(A)_{ij} = A_{ij}$ if i is different from r and s, $e(A)_{rj} = A_{si}$,

$$e(A)_{sj} = A_{rj}$$

A particular e is defined on the class of all m rowed matrices over F. One reason that we restrict ourselves to these simple types of row operations is that having performed such an operation e on a matrix A, we can recapture A by performing a similar operation on $e(A)$.

Definition: If A and B are $m \times n$ matrices over the field F, we say that B is **row-equivalent** to A if B is obtained from A by a finite sequence of elementary row operations. Consider the two systems of equations

$$AX = 0,$$

and $$BX = 0.$$

If matrix B is obtained from A by a finite sequence of elementary row operations we say that B matrix is row equivalent to A. Hence the above two system of equations are equivalent and so they have the same solutions.

Example 1: Consider

$$AX = 0$$

where $$A = \begin{pmatrix} -1 & i \\ -i & 3 \\ 1 & 2 \end{pmatrix}$$

so the system of equations is

$$\begin{aligned} -x_1 + ix_2 &= 0 \\ -ix_1 + 3x_2 &= 0 \\ x_1 + 2x_2 &= 0 \end{aligned}$$

Let us perform row operations on A

$$A = \begin{pmatrix} -1 & i \\ -i & 3 \\ 1 & 2 \end{pmatrix} \xrightarrow{(2)} \begin{pmatrix} 0 & 2+i \\ -i & 3 \\ 1 & 2 \end{pmatrix} \xrightarrow{(2)} \begin{pmatrix} 0 & 2+i \\ 0 & 3+2i \\ 1 & 2 \end{pmatrix} \xrightarrow{(2)} \begin{pmatrix} 0 & 1 \\ 0 & 3+2i \\ 1 & 0 \end{pmatrix} \xrightarrow{(2)} \begin{pmatrix} 0 & 1 \\ 0 & 0 \\ 1 & 0 \end{pmatrix} = B$$

Now $BX = 0$

gives us

$$B = \begin{pmatrix} 0 & 1 \\ 0 & 0 \\ 1 & 0 \end{pmatrix}$$

has only the trivial solution;

$$x_1 = 0$$

$$x_2 = 0$$

Definition: An $m \times n$ matrix B is called **row-reduced** if:

(a) the first non-zero entry in each non-zero row of B is equal to 1;

(b) each column of B which contains the leading non-zero entry of some row has all its other entries 0.

Example 2: One example of a row-reduced matrix is the $n \times n$ **identity matrix** I. This is the $n \times n$ matrix defined by

$$I_{ij} = \delta_{ij} = \begin{cases} 1 & if\ i = j \\ 0 & if\ i \neq j \end{cases}$$

Here we have introduced **Kronecker delta** ●.

Example 3: Find a row reduced matrix which is equivalent to

$$A = \begin{bmatrix} 2 & -1 & 3 & 2 \\ 1 & 4 & 0 & -1 \\ 2 & 6 & -1 & 5 \end{bmatrix}$$

Now

$$A = \begin{bmatrix} 2 & -1 & 3 & 2 \\ 1 & 4 & 0 & -1 \\ 2 & 6 & -1 & 5 \end{bmatrix} \xrightarrow{(2)} \begin{bmatrix} 0 & -9 & 3 & 4 \\ 1 & 4 & 0 & -1 \\ 2 & 6 & -1 & 5 \end{bmatrix} \xrightarrow{(2)} \begin{bmatrix} 0 & -9 & 3 & 4 \\ 1 & 4 & 0 & -1 \\ 0 & -2 & -1 & 7 \end{bmatrix} \xrightarrow{(1)} \begin{bmatrix} 0 & -9 & 3 & 4 \\ 1 & 4 & 0 & -1 \\ 0 & 1 & \frac{1}{2} & -\frac{7}{2} \end{bmatrix}$$

$$\xrightarrow{(2)} \begin{bmatrix} 0 & -9 & 3 & 4 \\ 1 & 0 & -2 & 13 \\ 0 & 1 & \frac{1}{2} & -\frac{7}{2} \end{bmatrix} \xrightarrow{(2)} \begin{bmatrix} 0 & 0 & \frac{15}{2} & -\frac{55}{2} \\ 1 & 0 & -2 & 13 \\ 0 & 1 & \frac{1}{2} & -\frac{7}{2} \end{bmatrix} \xrightarrow{(1)} \begin{bmatrix} 0 & 0 & 1 & -\frac{11}{3} \\ 1 & 0 & -2 & 13 \\ 0 & 1 & \frac{1}{2} & -\frac{7}{2} \end{bmatrix}$$

$$\xrightarrow{(2)} \begin{bmatrix} 0 & 0 & 1 & -\frac{11}{3} \\ 1 & 0 & 0 & \frac{17}{3} \\ 0 & 1 & \frac{1}{2} & -\frac{7}{2} \end{bmatrix} \xrightarrow{(2)} \begin{bmatrix} 0 & 0 & 1 & -\frac{11}{3} \\ 1 & 0 & 0 & \frac{17}{3} \\ 0 & 1 & 0 & \frac{-5}{3} \end{bmatrix}$$

The row-equivalence of A with the final matrix in the above sequence tells us in particular that the solutions of

$$AX = 0$$

i.e.,

$$\begin{aligned} 2x_1 - x_2 + 3x_3 + 2x_4 &= 0 \\ x_1 + 4x_2 - x_4 &= 0 \\ 2x_1 + 6x_2 - x_3 + 5x_4 &= 0 \end{aligned}$$

and

$$\begin{aligned} x_3 - \frac{11}{3}x_4 &= 0 \\ x_1 + \frac{17}{3}x_4 &= 0 \\ x_2 - \frac{5}{3}x_4 &= 0 \end{aligned}$$

are exactly the same. In the second system it is apparent that

$$\begin{aligned} x_3 &= \frac{11}{3}x_4 \\ x_1 &= -\frac{17}{3}x_4 \\ x_2 &= \frac{5}{3}x_4 \end{aligned}$$

Thus if x_4 = C then we obtain a solution $\left(-\frac{17}{3}C, \frac{5}{3}C, \frac{11}{3}C, C\right)$ and also that every solution is of this form.

Self Assessment

1. If $A = \begin{bmatrix} 3 & -1 & 2 \\ 2 & 1 & 1 \\ 1 & -3 & 0 \end{bmatrix}$, find all solutions of $AX = 0$ by row-reducing A.

2. Find a row-reduced matrix which is row-equivalent to $A = \begin{bmatrix} i & -(1+i) & 0 \\ 1 & -2 & 1 \\ 1 & 2i & -1 \end{bmatrix}$

5.2 Row-reduced Echelon Matrices

Definition: An $m \times n$ matrix R is called a row-reduced echelon matrix if:

(a) R is row-reduced;

(b) every row of R which has all its entries 0 occurs below every row which has a non-zero entry;

(c) if rows $1,..., r$ are the non-zero rows of R, and if the leading non-zero entry of row i occurs is column k_i, $i = 1,..., r$, then $k_1 < k_2 < ... < k_r$.

One can also describe an $m \times n$ row-reduced echelon matrix R as follows. Either every entry in R is 0, or there exists a positive integer r, $1 \le r \le m$, and r positive integers $k_1 ,..., k_r$ with $1 \le k_i \le n$ and

(a) $R_{ij} = 0$ for $i > r$, and $R_{ij} = 0$ if $j < k_i$.

(b) $R_{iki} = \delta_{ij}$, $1 \le i \le r$, $1 \le j \le r$.

(c) $k_1 < ... < k_r$.

Example 4: Two examples of row-reduced echelon matrices are the $n \times n$ identity matrix, and the $m \times n$ zero matrix $0^{m,\,n}$, in which all entries are 0. The reader should have no difficulty in making other examples, but we should like to give one non-trivial one:

$$\begin{bmatrix} 0 & 1 & -3 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Theorem 2: Every $m \times n$ matrix A is row-equivalent to a row-reduced echelon matrix.

Proof: We know that A is row-equivalent to a row-reduced matrix. All that we need observe is that by performing a finite number of row interchanges on a row-reduced matrix we can bring it to row-reduced echelon form.

In Examples 1 and 3, we saw the significance of row-reduced matrices in solving homogeneous systems of linear equations. Let us now discuss briefly the system $RX = 0$, when R is a row-reduced echelon matrix. Let rows $1,..., r$ be the non-zero rows of R, and suppose that the leading non-zero entry of row i occurs in column k_i. The system $RX = 0$ then consists of r non-trivial equations. Also the unknown x_k, will occur (with non-zero coefficient) only in the i^{th} equation. If we let $u_1,...,u_{n-r}$ denote the $(n - r)$ unknowns which are different from $x_{k1},...,x_{kr}$, then the r non-trivial equations in $RX = 0$ are of the form

$$\left.\begin{array}{ccccc} x_{k1} & + & \sum\limits_{j=1}^{n-r} C_{1j}u_j & = & 0 \\ \vdots & & & & \vdots \\ x_{kr} & + & \sum\limits_{j=1}^{n-r} C_{rj}u_j & = & 0 \end{array}\right\} \quad ...(1)$$

All the solutions to the system of equations $RX = 0$ are obtained by assigning any values whatsoever to $u_1,...,u_{n-r}$ and then computing the corresponding values of $x_{k1},...,x_{kr}$ from (1). For example, if R is the matrix displayed in Example 4, then $r = 2$, $k_1 = 2$, $k_2 = 4$, and the two non-trivial equations in the system $RX = 0$ are

$$x_2 - 3x_3 + \frac{1}{2}x_5 = 0 \text{ or } x_2 = 3x_3 - \frac{1}{2}x_5$$

$$x_4 + 2x_5 = 0 \text{ or } x_4 = -2x_5$$

So we may assign any values to x_1, x_3, and x_5, say $x_1 = a$, $x_3 = b$, $x_5 = c$, and obtain the solution $(a, 3b - \frac{1}{2}c,\ b, -2c, c)$.

Let us observe one thing more in connection with the system of equations $RX = 0$. If the number r of non-zero rows in R is less than n, then the system $RX = 0$ has a non-trivial solution, that is, a solution $(x_1,...,x_n)$ in which not every x_j in 0. For, since $r < n$, we can choose some x_j which is not among the r unknowns $x_{k1},...,x_{kr}$, and we can then construct a solution as above in which this x_j is 1. This observation leads us to one of the most fundamental facts concerning systems of homogeneous linear equations.

Theorem 3: If A is an $m \times n$ matrix and $m < n$, then the homogeneous system of linear equations $AX = 0$ has a non-trivial solution.

Proof: Let R be a row-reduced echelon matrix which is row-equivalent to A. Then the systems $AX = 0$ and $RX = 0$ have the same solutions by Theorem 3. If r is the number of non-zero rows in R, then certainly $r \leq m$, and since $m < n$, we have $r < n$. It follows immediately from our remarks above that $AX = 0$ has a non-trivial solution.

Theorem 4: If A is an $n \times n$ (square) matrix, then A is row-equivalent to the $n \times n$ identity matrix if and only if the system of equations $AX = 0$ has only the trivial solution.

Proof: If A is row-equivalent to I, then $AX = 0$ and $IX = 0$ have the same solutions. Conversely, suppose $AX = 0$ has only the trivial solution $X = 0$. Let R be an $n \times n$ row-reduced echelon matrix which is row-equivalent to A, and let r be the number of non-zero rows of R. Then $RX = 0$ has no non-trivial solution. Thus $r \geq n$. But since R has n rows, certainly $r \leq n$, and we have $r = n$. Since this means that R actually has a leading non-zero entry of 1 in each of its n rows, and since these 1's occur each in a different one of the n columns, R must be the $n \times n$ identity matrix.

Let us now ask what elementary row operations do toward solving a system of linear equations $AX = Y$ which is not homogeneous. At the outset, one must observe one basic difference between this and the homogeneous case, namely, that while the homogeneous system always has the trivial solution $x_1 = \cdots = x_n = 0$, an inhomogeneous system need have no solution at all.

We form the augmented matrix A' of the system $AX = Y$. This is the $m \times (n + 1)$ matrix whose first n columns are the columns of A and whose last column is Y. More precisely,

$$A'_{ij} = A_{ij},\ if\ j \leq n$$
$$A'_{i(n+1)} = y_i$$

Suppose we perform a sequence of elementary row operations on A arriving at a row-reduced echelon matrix R. If we perform this same sequence of row operations on the augmented matrix A', we will arrive at a matrix R' whose first n columns are the columns of R and whose last column contains certain scalars $z_1,...,z_m$. The scalars z_i are the entries of the $m \times 1$ matrix

$$Z = \begin{bmatrix} z_1 \\ \vdots \\ z_m \end{bmatrix}$$

which results from applying the sequence of row operations to the matrix Y. It should be clear to the reader that, just as in the proof of Theorem 3 the systems $AX = Y$ and $RX = Z$ are equivalent and hence have the same solutions. It is very easy to determine whether the system $RX = Z$ has any solutions and to determine all the solutions if any exist. For, if R has r non-zero rows, with the leading non-zero entry of row i occurring in column k_i, $i = 1,...,r$, then the first r equations of

$RX = Z$ effectively express $x_{k1},...,x_{kr}$ in terms of the $(n - r)$ remaining x_j and the scalars $z_1,...,z_r$. The last $(m - r)$ equations are

$$\begin{array}{ccc} 0 & = & z_{r+1} \\ \vdots & & \vdots \\ 0 & = & z_m \end{array}$$

and accordingly the condition for the system to have a solution is $z_\text{i} = 0$ for $i > r$. If this condition is satisfied, all solutions to the system are found just as in the homogeneous case, by assigning arbitrary values the $(n - r)$ of the x_j and then computing x_{ki} from the ith equation.

Example 5: Let F be the field of rational numbers and

$$A = \begin{bmatrix} 1 & -2 & 1 \\ 2 & 1 & 1 \\ 0 & 5 & -1 \end{bmatrix}$$

and suppose that we wish to solve the system $AX = Y$ for some y_1, y_2 and y_3. Let us perform a sequence of row operations on the augmented matrix A' which row-reduces A:

$$\begin{bmatrix} 1 & -2 & 1 & y_1 \\ 2 & 1 & 1 & y_2 \\ 0 & 5 & -1 & y_3 \end{bmatrix} \xrightarrow{(2)} \begin{bmatrix} 1 & -2 & 1 & y_1 \\ 0 & 5 & -1 & (y_2 - 2y_1) \\ 0 & 5 & -1 & y_3 \end{bmatrix} \xrightarrow{(2)}$$

$$\begin{bmatrix} 1 & -2 & 1 & y_1 \\ 0 & 5 & -1 & (y_2 - 2y_1) \\ 0 & 0 & 0 & (y_3 - y_2 + 2y_1) \end{bmatrix} \xrightarrow{(1)} \begin{bmatrix} 1 & -2 & 1 & y_1 \\ 0 & 1 & -\frac{1}{5} & \frac{1}{5}(y_2 - 2y_1) \\ 0 & 0 & 0 & (y_3 - y_2 + 2y_1) \end{bmatrix} \xrightarrow{(2)}$$

$$\begin{bmatrix} 1 & 0 & \frac{3}{5} & \frac{1}{5}(y_1 + 2y_2) \\ 0 & 1 & -\frac{1}{5} & \frac{1}{5}(y_2 - 2y_1) \\ 0 & 0 & 0 & (y_3 - y_2 + 2y_1) \end{bmatrix}$$

The condition that the system $AX = Y$ have a solution is thus

$$2y_1 - y_2 + y_3 = 0$$

and if the given scalars y_i satisfy this condition, all solutions are obtained by assigning a value c to x_3 and then computing

$$x_1 = -\frac{3}{5}c + \frac{1}{5}(y_1 + 2y_2)$$

$$x_2 = \frac{1}{5}c + \frac{1}{5}(y_2 - 2y_1)$$

Let us observe one final thing about the system $AX = Y$. Suppose the entries of the matrix A and the scalars $y_1,...y_m$ happen to lie in a subfield F_1 of the field F. If the system of equations $AX = Y$ has a solution with $x_1,...,x_n$ in F, it has a solution with $x_1,...,x_n$ in F_1. For, over either field, the condition for the system to have a solution is that certain relations hold between $y_1,...,y_m$ in F_1

(the relations $z_i = 0$ for $i > r$, above). For example, if $AX = Y$ is a system of linear equations in which the scalars y_k and A_{ij} are real numbers, and if there is a solution in which $x_1,\ldots,x_n$ are complex numbers, then there is a solution with $x_1,\ldots,x_n$ real numbers.

Self Assessment

3. Find all solutions to the following system of equations by row-reducing the coefficient matrix:

$$\begin{aligned} &\frac{3}{3}x_1 + 2x_2 - 6x_3 = 0 \\ &-4x_1 \qquad + 5x_3 = 0 \\ &-3x_1 + 6x_2 - 13x_3 = 0 \\ &-\frac{7}{3}x_1 + 2x_2 - \frac{8}{3}x_3 = 0 \end{aligned}$$

4. Find a row-reduced echelon matrix which is row-equivalent to

$$A = \begin{bmatrix} 1 & -i \\ 2 & 2 \\ i & 1+i \end{bmatrix}$$

What are the solutions of $AX = 0$?

5.3 Summary of Row-Equivalence

In this section we shall utilize some elementary facts on bases and dimension in finite-dimensional vector spaces to complete our discussion of row-equivalence of matrices. We recall that if A is an $m \times n$ matrix over the field F the row vectors of A are the vectors $\alpha_1,\ldots,\alpha_m$ in F^n defined by

$$\alpha_i = (A_{ij},\ldots, A_{in})$$

and that the row space of A is the subspace of F^n spanned by these vectors. The row rank of A is the dimension of the row space of A.

If P is a $k \times m$ matrix over F, then the product $B = PA$ is a $k \times n$ matrix whose row vectors $\beta_1,\ldots,\beta_k$ are linear combinations

$$\beta_i = P_{i1\alpha1} + \ldots + P_{im}\alpha_m$$

of the row vectors of A. Thus the row space of B is a subspace of the row space of A. If P is an $m \times m$ invertible matrix, then B is row-equivalent to A so that the symmetry of row-equivalence, or the equation $A = P^{-1}B$, implies that the row space of A is also a subspace of the row space of B.

Theorem 5: Row-equivalent matrices have the same row space.

Thus we see that to study the row space of A we may as well study the row space of a row-reduced echelon matrix which is row-equivalent to A. This we proceed to do.

Theorem 6: Let R be a non-zero row-reduced echelon matrix. Then the non-zero row vectors of R form a basis for the row space of R.

Proof: Let $\rho_1,\ldots,\rho_r$ be the non-zero row vectors of R:

$$\rho_i = (R_{i1},\ldots,R_{in})$$

Certainly these vectors span the row space of R; we need only prove they are linearly independent. Since R is a row-reduced echelon matrix, there are positive integers $k_1,\ldots,k_r$ such that, for $i \leq r$

$$\left.\begin{array}{lll}\text{(a)} & R(i, j) = 0 & \text{if } j < k_i \\ \text{(b)} & R(i, k_j) = \delta_{ij} & \\ \text{(c)} & k_1 < ... < k_r & \end{array}\right\} \qquad ...(1)$$

Suppose $\beta = (b_1,...,b_n)$ is a vector in the row space of R:

$$\beta = c_1\rho_1 + ... + c_r\rho_r \qquad ...(2)$$

Then we claim that $c_j = b_{ki}$. For, by

$$b_{kj} = \sum_{i=1}^{r} c_i R(i, k_j)$$

$$= \sum_{i=1}^{r} c_i \delta_{ij} \qquad ...(3)$$

$$= c_j$$

In particular, if $\beta = 0$, i.e., if $c_1\rho_1 + ... + c_r\rho_r = 0$, then c_j must be the k_jth coordinate of the zero vector so that $c_j = 0$, $j = 1,..., r$. Thus $\rho_1,...,\rho_r$ are linearly independent.

Theorem 7: Let m and n be positive integers and let F be a field. Suppose W is a subspace of F^n and dim $W \leq m$. Then there is precisely one $m \times n$ row-reduced echelon matrix over F which has W as its row space.

Proof: There is at least one $m \times n$ row-reduced echelon matrix with row space W. Since dim $W \leq m$, we can select some m vectors $\alpha_1,...,\alpha_m$ in W which span W. Let A be the $m \times n$ matrix with row vectors $\alpha_1,...,\alpha_m$ and let R be a row-reduced echelon matrix which is row-equivalent to A. Then the row space of R is W.

Now let R be any row-reduced echelon matrix which has W as its row space. Let $\rho_1,...,\rho_r$ be the non-zero row vectors of R and suppose that the leading non-zero entry of ρ_i occurs in column k_i, $i = 1,...,r$. The vectors $\rho_1,...,\rho_r$ form a basis for W. In the proof of Theorem, we observed that if $\beta = (b_1,...,b_n)$ is in W, then

$$\beta = c_1\rho_1 + ... + c_r\rho_r,$$

and $c_i = b_{ki}$; in other words, the unique expression for β as a linear combination of $\rho_1,...,\rho_r$ is

$$\beta = \sum_{i=1}^{r} b_{ki}\rho_i \qquad ...(4)$$

Thus any vector β is determined if one knows the coordinates b_{ki}, $i = 1,..., r$. For example, ρ_s is the unique vector in W which has k_sth coordinate 1 and k_ith coordinate 0 for $i \neq$ s.

Suppose β is in W and $\beta \neq 0$. We claim the first non-zero coordinate of β occurs in one of the columns k_s. Since

$$\beta = \sum_{i=1}^{r} b_{ki}\rho_i$$

and $\beta \neq 0$, we can write

$$\beta = \sum_{i=s}^{r} b_{ki}\rho_i, \qquad b_{ks} \neq 0 \qquad ...(5)$$

From the conditions (1) one has $R_{ij} = 0$ if $i > s$ and $j \leq k_s$. Thus

$$\beta = (0,..., 0, \beta_{ks},...,b_n),\ b_{ks} \neq 0$$

and the first non-zero coordinate of β occurs in column k_8. Note also that for each k_8, $S = 1,..., r$, there exists a vector in W which has a non-zero k_sth coordinate, namely ρ_s.

It is now clear that R is uniquely determined by W. The description of R in terms of W is as follows. We consider all vectors $\beta = (b_1,...,b_{\mathrm{n}})$ in W. If $\beta \neq 0$, then the first non-zero coordinate of β must occur in some column t:

$$\beta = (0,...,0, b_t,..., b_n),\ b_t \neq 0$$

Let $k_1,...,k_{\mathrm{r}}$ be those positive integers t such that there is some $\beta \neq 0$ in W, the first non-zero coordinate of which occurs in column t. Arrange $k_1,...,k_r$ in the order $k_1 < k_2 < ... < k_r$. For each of the positive integers k_{s} there will be one and only one vector ρ_{s} in W such that the k_{s}th coordinate of ρ_s is 1 and the k_ith coordinate of ρ_s is 0 for $i \neq s$. Then R is the $m \times n$ matrix which has row vectors $\rho_1,...,\rho_r$, 0, ..., 0.

Corollary. Each $m \times n$ matrix A is row-equivalent to one and only one row-reduced echelon matrix.

Proof: We know that A is row-equivalent to at least one row-reduced echelon matrix R. If A is row-equivalent to another such matrix R', then R is row-equivalent to R'; hence, R and R' have the same row space and must be identical.

Corollary: Let A and B be $m \times n$ matrices over the field F. Then A and B are row-equivalent if and only if they have the same row space.

Proof: We know that if A and B are row-equivalent, then they have the same row space. So suppose that A and B have the same row space. Now A is row-equivalent to a row-reduced echelon matrix R and B is row-equivalent to a row-reduced echelon matrix R'. Since A and B have the same row space, R and R' have the same row space. Thus $R = R'$ and A is row-equivalent to B.

To summarize—if A and B are $m \times n$ matrices over the field F, the following statements are equivalent:

1. A and B are row-equivalent.
2. A and B have the same row space.
3. $B = PA$, where P is an invertible $m \times m$ matrix.

A fourth equivalent statement is that the homogeneous systems $AX = 0$ and $BX = 0$ have the same solutions; however, although we know that the row-equivalence of A and B implies that these systems have the same solutions, it seems best to leave the proof of the converse until later.

5.4 Summary

- Such an equation is called a linear combination of the equations in (1). Evidently any solution of the entire system of equations (1) will also be the solution of this new equation. This is the fundamental idea of the elimination process.
- A particular e is defined on the class of all m rowed matrices over F. One reason that we restrict ourselves to these simple types of row operations is that having performed such an operation e on a matrix A, we can recapture A by performing a similar operation on $e(A)$.
- An $m \times n$ matrix B is called row-reduced if:

  (a) the first non-zero entry in each non-zero row of B is equal to 1;

(b) each column of B which contains the leading non-zero entry of some row has all its other entries 0.

5.5 Keywords

Equivalent: Two systems of linear equations are equivalent if each equation in each system is a linear combination of the equations in the other system.

Row-equivalent: If A and B are $m \times n$ matrices over the field F, we say that B is **row-equivalent** to A if B is obtained from A by a finite sequence of elementary row operations.

5.6 Review Questions

1. Find all solutions to the system of equations

$$(1 - i)x_1 - ix_2 = 0$$
$$2x_1 + (1 - i)x_2 = 0$$

2. Let $A = \begin{bmatrix} 3 & -1 & 2 \\ 2 & 1 & 1 \\ 1 & -3 & 0 \end{bmatrix}$

For which triples (y_1, y_2, y_3) does this system $AX = Y$ have a solution?

Answers: Self Assessment

1. Row-reduced matrix is $\begin{bmatrix} 0 & 1 & 1/4 \\ 1 & 0 & 3/8 \\ 0 & 0 & 3/8 \end{bmatrix}$ the solution is $x_1 = x_2 = x_3 = 0$

3. Row-reduced matrix is $\begin{bmatrix} 0 & 1 & -\frac{64}{24} \\ 1 & 0 & -\frac{5}{4} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ the solution is $x_1 = \frac{67}{24}C, x_2 = \frac{5}{4}C, x_3 = C$ where C is a constant.

4. Row-reduced matrix is $\begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 1 & 0 \end{bmatrix}$. The solution is $x_1 = x_2 = 0$

5.7 Further Readings

Books

Kenneth Hoffman and Ray Kunze *Linear Algebra*

I.N. Herstein *Topics in Algebra*

Unit 6: Computation Concerning Subspaces

CONTENTS

Objectives

After studying this unit, you will be able to:

- See that the units (3), (4) and (5) are quite suitable to find if a set of vectors α_1, α_2,...α_m are linearly independent.
- Determine whether another vector β is a linear combination of α_1,...α_m.
- See that the detailed examples in this unit clarify most ideas covered in the last few units.

Introduction

This unit mostly summarizes the ideas of row-operations in helping to find out the basis of a vector-subspace.

One can understand how a vector α belongs to the vector sub-space spanned by the basis vectors.

6.1 Computation Concerning Subspaces

In this unit we should like to show how elementary row operations helps us in understanding in a concrete way the subspaces of F^n. This discussion applies to any n-dimensional vector space over the field F, if one selects a fixed ordered basis β and describes each vector α in V by the n-tuple (x_1, x_2,...,x_n) which gives the co-ordinates of α in the ordered basis β.

Suppose we are given m vectors α_1,...,α_m in F^n. We consider the following questions.

1. How does one determine if the vectors α_1, α_2,...,α_m are linearly independent? How does one find the dimension of the subspace W spanned by these vectors?
2. Given β in F^n, how does one determine whether β is a linear combination of α_1,...,α_m, i.e., whether β is in the subspace W?
3. How can one give an explicit description of the subspace W?

The third question is a little vague, since it does not specify what is meant by an 'explicit description'; however, we shall clear up this point by giving the sort of description we have in mind. With this description, questions (1) and (2) can be answered immediately.

Let A be the $m \times n$ matrix with row vectors α_i:

$$\alpha_i = (A_{i1},\ldots,A_{in})$$

Perform a sequence of elementary row operations, starting with A and terminating with a row-reduced echelon matrix R. We have previously described how to do this. At this point, the dimension of W (the row space of A) is apparent, since this dimension is simply the number of non-zero row vectors of R. If $\rho_1,\ldots,\rho_r$ are the non-zero row vectors of R, then $\mathcal{B} = \{\rho_1,\ldots,\rho_r\}$ is a basis for W. If the first non-zero coordinate of ρ_i is the k_ith one, then we have for $i \le r$

(a) $R(i, j) = 0$, if $j < k_i$

(b) $R(i, k_j) = \delta_{ij}$

(c) $k_1 < \ldots < k_r$

The subspace W consists of all vectors

$$\beta = c_1\rho_1 + \ldots + c_r\rho_r$$

$$= \sum_{i=1}^{r} c_i (R_{i1},\ldots,R_{in})$$

The coordinates $b_1,\ldots,b_n$ of such a vector β are then

$$b_j = \sum_{i=1}^{r} c_i R_{ij} \qquad \ldots(1)$$

In particular, $b_{ki} = c_j$, and so if $\beta = (b_1,\ldots,b_n)$ is a linear combination of the ρ_i, it must be the particular linear combination.

$$\beta = \sum_{i=1}^{r} b_{ki}\rho_i \qquad \ldots(2)$$

The conditions on β that (2) should hold are

$$b_j = \sum_{i=1}^{r} b_{ki} R_{ij} \qquad j = 1,\ldots,n. \qquad \ldots(3)$$

Now (3) is the explicit description of the subspace W spanned by $\alpha_1,\ldots,\alpha_m$, that is, the subspace consists of all vectors β in F^n whose coordinates satisfy (3). What kind of description is (3)? In the first place it describes W as all solutions $\beta = (b_1,\ldots,b_n)$ of the system of homogeneous linear equations (3). This system of equations is of a very special nature, because it expresses $(n - r)$ of the coordinates as linear combinations of the r distinguished coordinates $b_{k1},\ldots,b_{kr}$. One has complete freedom of choice in the coordinates b_{ki}, that is, if $c_1,\ldots,c_r$ are any r scalars, there is one and only one vector β in W which has c_i as its k_ith coordinate.

The significant point here is this: Given the vectors α_i, row-reduction is a straightforward method of determining the integers $r, k_1,\ldots,k_r$ and the scalars R_{ij} which give the description of the subspace spanned by $\alpha_1,\ldots,\alpha_m$. One should observe that every subspace W of F^n has a description of the type (3). We should also point out some things about question (2). We have already stated how one can find an invertible $m \times m$ matrix P such that $R = PA$. The knowledge of P enables one to find the scalars $x_1,\ldots,x_m$ such that

$$\beta = x_1\alpha_1 + \ldots + x_m\alpha_m$$

when this is possible. For the row vectors of R are given by

$$\rho_{\mathrm{i}} = \sum_{j=1}^{m} P_{ij}\alpha_j$$

so that if β is a linear combination of the α_j, we have

$$\begin{aligned}\beta &= \sum_{i=1}^{r} b_{ki}\rho_i \\ &= \sum_{i=1}^{r} b_{ki}\sum_{j=1}^{m} P_{ij}\alpha_j \\ &= \sum_{j=1}^{m}\sum_{i=1}^{r} b_{ki}P_{ij}\alpha_j\end{aligned}$$

and thus $$x_j = \sum_{i=1}^{r} b_{ki}P_{ij}$$

is one possible choice for the x_j (there may be many).

The question of whether $\beta = (b_1,...,b_n)$ is a linear combination of the α_i, and if so, what the scalars x_i are, can also be looked at by asking whether the system of equations

$$\sum_{i=1}^{m} A_{ij}x_i = b_j, \qquad j = 1,...,n$$

has a solution and what the solutions are. The coefficient matrix of this system of equations is then $n \times m$ matrix B with column vectors $\alpha_1,...,\alpha_m$. In unit 5, we discussed the use of elementary row operations in solving a system of equations $BX = Y$. Let us consider one example in which we adopt both points of view in answering questions about subspaces of F^n.

6.2 Illustrative Examples

Example 1: Let us pose the following problem. Let W be the subspace of R^4 spanned by the vectors

$$\begin{aligned}\alpha_1 &= (1, 2, 2, 1) \\ \alpha_2 &= (0, 2, 0, 1) \\ \alpha_3 &= (-2, 0, -4, 3)\end{aligned}$$

(a) Prove that α_1, α_2, α_3 form a basis for W, i.e., that these vectors are linearly independent.

(b) Let $\beta = (b_1, b_2, b_3, b_4)$ be a vector in W. What are the coordinates of β relative to the ordered basis $\{\alpha_1, \alpha_2, \alpha_3\}$?

(c) Let

$$\begin{aligned}\alpha_1' &= (1, 0, 2, 0) \\ \alpha_2' &= (0, 2, 0, 1) \\ \alpha_3' &= (0, 0, 0, 3)\end{aligned}$$

Show that α_1', α_2', α_3' form a basis for W.

(d) If β is in W, let X denote the coordinate matrix of β relative to the α-basis and X' the coordinate matrix of β relative to the α′-basis. Find the 3 × 3 matrix P such that $X = PX'$ for every such β.

To answer these questions by the first method we form the matrix A with row vectors $\alpha_1, \alpha_2, \alpha_3$, find the row-reduced echelon matrix R which is row-equivalent to A and simultaneously perform the same operations on the identity to obtain the invertible matrix Q such that $R = QA$:

$$\begin{bmatrix} 1 & 2 & 2 & 1 \\ 0 & 2 & 0 & 1 \\ -2 & 0 & -4 & 3 \end{bmatrix} \to R = \begin{bmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \to Q = \frac{1}{6}\begin{bmatrix} 6 & -6 & 0 \\ -2 & 5 & -1 \\ 4 & -4 & 2 \end{bmatrix}$$

(a) Clearly R has rank 3, so α_1, α_2 and α_3 are independent.

(b) Which vectors $\beta = (b_1, b_2, b_3, b_4)$ are in W? We have the basis for W given by ρ_1, ρ_2, ρ_3, the row vectors of R. One can see at a glance that the span of ρ_1, ρ_2, ρ_3 consists of the vectors β for which $b_3 = 2b_1$. For such a β we have

$$\begin{aligned} \beta &= b_1\rho_1 + b_2\rho_2 + b_4\rho_4 \\ &= [b_1, b_2, b_4]R \\ &= [b_1\ b_2\ b_4]QA \\ &= x_1\alpha_1 + x_2\alpha_2 + x_3a_3 \end{aligned}$$

where $x_i = [b_1\ b_2\ b_4]Q_i$:

$$\left.\begin{aligned} x_1 &= b_1 - \frac{1}{3}b_2 + \frac{2}{3}b_4 \\ x_2 &= -b_1 + \frac{5}{6}b_2 - \frac{2}{3}b_4 \\ x_3 &= \quad -\frac{1}{6}b_2 + \frac{1}{3}b_4 \end{aligned}\right\} \qquad \text{... (1)}$$

(c) The vectors $\alpha_1', \alpha_2', \alpha_3'$ are all of the form (y_1, y_2, y_3, y_4) with $y_3 = 2y_1$ and thus they are in W. One can see at a glance that they are independent.

(d) The matrix P has for its columns

$$P_j = \left[\alpha_j^{'}\right]\mathcal{B}$$

where $\beta = \{\alpha_1, \alpha_2, \alpha_3\}$. The equations (1) tell us how to find the coordinate matrices for $\alpha_1', \alpha_2', \alpha_3'$.

For example with $\beta = \alpha_1^{'}$ we have $b_1 = 1$, $b_2 = 0$, $b_3 = 2$, $b_4 = 0$, and

$$x_1 = 1 - \frac{1}{3}(0) + \frac{2}{3}(0) = 1$$

$$x_2 = -1 + \frac{5}{6}(0) + \frac{2}{3}(0) = -1$$

$$x_3 = -\frac{1}{6}(0) + \frac{1}{3}(0) = 0$$

Notes

Thus $\alpha_1^{'} = \alpha_1 - \alpha_2$. Similarly we obtain $\alpha_2^{'} = \alpha_2$ and $\alpha_3^{'} = 2\alpha_1 - 2\alpha_2 + \alpha_3$.

Hence

$$P = \begin{bmatrix} 1 & 0 & 2 \\ -1 & 1 & -2 \\ 0 & 0 & 1 \end{bmatrix}$$

Now let us see how we would answer the questions by the second method which we described. We form the 4 × 3 matrix B with column vectors $\alpha_1, \alpha_2, \alpha_3$:

$$B = \begin{bmatrix} 1 & 0 & -2 \\ 2 & 2 & 0 \\ 2 & 0 & -4 \\ 1 & 1 & 3 \end{bmatrix}$$

We inquire for which y_1, y_2, y_3, y_4 the system $BX = Y$ has a solution.

$$\begin{bmatrix} 1 & 0 & -2 & y_1 \\ 2 & 2 & 0 & y_2 \\ 2 & 0 & -4 & y_3 \\ 1 & 1 & 3 & y_4 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & -2 & y_1 \\ 0 & 2 & 4 & y_2 - 2y_1 \\ 0 & 0 & 0 & y_3 - 2y_1 \\ 0 & 1 & 5 & y_4 - y_1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & -2 & y_1 \\ 0 & 0 & -6 & y_2 - 2y_4 \\ 0 & 1 & 5 & y_4 - y_1 \\ 0 & 0 & 0 & y_3 - 2y_1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & 0 & y_1 - \frac{1}{3}y_2 + \frac{2}{3}y_4 \\ 0 & 0 & 1 & \frac{1}{6}(2y_4 - y_2) \\ 0 & 1 & 0 & -y_1 + \frac{5}{6}y_2 - \frac{2}{3}y_4 \\ 0 & 0 & 0 & y_3 - 2y_1 \end{bmatrix}$$

Thus the condition that the system $BX = Y$ have a solution is $y_3 = 2y_1$. So $\beta = (b_1, b_2, b_3, b_4)$ is in W if and only if $b_3 - 2b_1$. If β is in W, then the coordinates (x_1, x_2, x_3) in the ordered basis $\{\alpha_1, \alpha_2, \alpha_3\}$ can be read off from the last matrix above. We obtain once again the formulas (1) for those coordinates

The questions (*c*) and (*d*) are now answered as before.

Example 2: We consider the 5 × 5 matrix

$$A = \begin{bmatrix} 1 & 2 & 0 & 3 & 0 \\ 1 & 2 & -1 & -1 & 0 \\ 0 & 0 & 1 & 4 & 0 \\ 2 & 4 & 1 & 10 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

and the following problems concerning A

(a) Find an invertible matrix P such that PA is a row-reduced echelon matrix R.

(b) Find a basis for the new row space W of A.

(c) Say which vectors $(b_1, b_2, b_3, b_4, b_5)$ are in W.

(d) Find the coordinate matrix of each vector $(b_1, b_2, b_3, b_4, b_5)$ in W in the ordered basis chosen in (b).

(e) Write each vector $(b_1, b_2, b_3, b_4, b_5)$ in W as a linear combination of the rows of A.

(f) Give an explicit description of the vector space V of all 5 × 1 column matrices X such that $AX = 0$.

(g) Find a basis for V.

(h) For what 5 × 1 column matrices Y does the equation $AX = Y$ have solutions X?

To solve these problems we form the augmented matrix A' of the system $AX = Y$ and apply an appropriate sequence of row operations to A'.

$$\begin{bmatrix} 1 & 2 & 0 & 3 & 0 & y_1 \\ 1 & 2 & -1 & -1 & 0 & y_2 \\ 0 & 0 & 1 & 4 & 0 & y_3 \\ 2 & 4 & 1 & 10 & 1 & y_4 \\ 0 & 0 & 0 & 0 & 1 & y_5 \end{bmatrix} \to \begin{bmatrix} 1 & 2 & 0 & 3 & 0 & y_1 \\ 0 & 0 & -1 & -4 & 0 & -y_1 + y_2 \\ 0 & 0 & 1 & 4 & 0 & y_3 \\ 0 & 0 & 1 & 4 & 1 & -2y_1 + y_4 \\ 0 & 0 & 0 & 0 & 1 & y_5 \end{bmatrix} \to$$

$$\begin{bmatrix} 1 & 2 & 0 & 3 & 0 & y_1 \\ 0 & 0 & 1 & 4 & 0 & y_1 - y_2 \\ 0 & 0 & 0 & 0 & 0 & -y_1 + y_2 + y_3 \\ 0 & 0 & 0 & 0 & 1 & -3y_1 + y_2 + y_4 \\ 0 & 0 & 0 & 0 & 1 & y_5 \end{bmatrix} \to \begin{bmatrix} 1 & 2 & 0 & 3 & 0 & y_1 \\ 0 & 0 & 1 & 4 & 0 & y_1 - y_2 \\ 0 & 0 & 0 & 0 & 1 & y_5 \\ 0 & 0 & 0 & 0 & 0 & -y_1 + y_2 + y_3 \\ 0 & 0 & 0 & 0 & 0 & -3y_1 + y_2 + y_4 - y_5 \end{bmatrix}$$

(a) If

$$PY = \begin{bmatrix} y_1 \\ y_1 - y_2 \\ y_5 \\ -y_1 + y_2 + y_3 \\ -3y_1 + y_2 + y_4 - y_5 \end{bmatrix}$$

for all Y, then

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ -1 & 1 & 1 & 0 & 0 \\ -3 & 1 & 0 & 1 & -1 \end{bmatrix}$$

hence PA is the row-reduced echelon matrix

$$R = \begin{bmatrix} 1 & 2 & 0 & 3 & 0 \\ 0 & 0 & 1 & 4 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

It should be stressed that the matrix P is not unique. There are, in fact, many invertible matrices P (which arise from different choices for the operations used to reduce A') such that $PA = R$.

(b) As a basis for W we may take the non-zero rows

$$\rho_1 = (1\ 2\ 0\ 3\ 0)$$

$$\rho_2 = (0\ 0\ 1\ 4\ 0)$$

$$\rho_3 = (0\ 0\ 0\ 0\ 1)$$

of R

(c) The row-space W consists of all vectors of the form

$$\begin{aligned}\beta &= c_1\rho_1 + c_2\rho_2 + c_3\rho_3 \\ &= (c_1, 2c_1, c_2, 3c_1 + 4c_2, c_3)\end{aligned}$$

where c_1, c_2, c_3 are arbitrary scalars. Thus $(b_1, b_2, b_3, b_4, b_5)$ is in W if and only if

$(b_1, b_2, b_3, b_4, b_5) = b_1\rho_1 + b_3\rho_2 + b_5\rho_3$

which is true if and only if

$$\begin{aligned}b_2 &= 2b_1 \\ b_4 &= 3b_1 + 4b_3.\end{aligned}$$

These equations are instances of the general system (3) in unit 5, and using them we may tell at a glance whether a given vector lies in W. Thus (–5, –10, 1, –11, 20) is a linear combination of the rows of A, but (1, 2, 3, 4, 5) is not.

(d) The coordinate matrix of the vector $(b_1, 2b_1, b_3, 3b_1 + 4b_3, b_5)$ in the basis $\{\rho_1, \rho_2, \rho_3\}$ is evidently

$$\begin{bmatrix} b_1 \\ b_3 \\ b_5 \end{bmatrix}$$

(e) There are many ways to write the vectors in W as linear combinations of the rows of A.

$$\begin{aligned}\beta &= (b_1, 2b_1, b_3, 3b_1 + 4b_3, b_5) \\ &= [b_1, b_3, b_5, 0, 0]\ .\ \mathrm{R} \\ &= [b_1, b_3, b_5, 0, 0]\ .\ \mathrm{PA}\end{aligned}$$

$$= [b_1, b_3, b_5, 0, 0]\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ -1 & 1 & 1 & 0 & 0 \\ -3 & 1 & 0 & 1 & -1 \end{bmatrix} A$$

$$= [b_1 + b_3, -b_3, 0, 0, b_5]\ .\ \mathrm{A}$$

In particular, with β = (–5, –10, 1, –11, 20) we have

$$\beta = (-4, -1, 0, 0, 20)\begin{bmatrix} 1 & 2 & 0 & 3 & 0 \\ 1 & 2 & -1 & -1 & 0 \\ 0 & 0 & 1 & 4 & 0 \\ 2 & 4 & 1 & 10 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Notes

(f) The equations in the system $RX = 0$ are

$$x_1 + 2x_2 + 3x_4 = 0$$

$$x_3 + 4x_4 = 0$$

$$x_5 = 0$$

Thus V consists of all columns of the form

$$X = \begin{bmatrix} -2x_2 - 3x_4 \\ x_2 \\ -4x_4 \\ x_4 \\ 0 \end{bmatrix}$$

where x_2 and x_4 are arbitrary.

(g) The columns

$$\begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} -3 \\ 0 \\ -4 \\ 1 \\ 0 \end{bmatrix}$$

form a basis of V.

(h) The equation $AX = Y$ has solutions X if and only if

$$-y_1 + y_2 + y_3 = 0$$

$$-3y_1 + y_2 + y_4 - y_5 = 0$$

Self Assessment

1. In C^3, let

$$\alpha_1 = (1, 0, -i),\ \alpha_2 = (1 + i, 1 - i, 1),\ \alpha_3 = (i, i, i)$$

Prove that these vectors form a basis for C^3. What are the co-ordinates of the vector (a, b, c) in this basis?

2. Let $\alpha_1 = (1, 1, -2, 1)$, $\alpha_2 = (3, 0, 4, -1)$, $\alpha_3 = (-1, 2, 5, 2)$

Let

$\alpha = (4, -5, 9, -7)$, $\beta = (3, 1, -4, 4)$, $\gamma = (-1, 1, 0, 1)$

Which of the vectors α, β, γ are in the sub-space of R^4 spanned by the α_i?

6.3 Summary

- In this unit it is shown how elementary row operations help us in understanding the basis of the subspace F^n.
- The detailed examples show how to go from one basis vector to another by means of an invertible matrix.

- Given the vectors α_i, row-reduction is a straightforward method of determining the integers $r, k_1,...,k_r$ and the scalars R_{ij} which give the description of the subspace spanned by $\alpha_1,...,\alpha_m$.
- The question of whether $\beta = (b_1,...,b_n)$ is a linear combination of the α_i, and if so, what the scalars x_i are, can also be looked at by asking whether the system of equations

$$\sum_{i=1}^{m} A_{ij}x_i = b_j, \qquad j = 1,...,n$$

  has a solution and what the solutions are.
- The unit helps in finding an invertible matrix P such that the co-ordinates of a vector α in the two system of basis β and β' are related by the relation $X = PX'$ for every basis β.

6.4 Keywords

Basis of the Subspace: The basis of the subspace W is found by the row vectors of R. So one can test whether a vector β belongs to W or not.

Row Reduction of a Matrix: The row reduction of a matrix A helps whether a set of vectors α_1, α_2, α_3 form a basis by forming the matrix A with row vectors and finding its rank.

6.5 Review Questions

1. Let $\beta = (u_1, u_2,...,u_n)$ and $\beta' = (v_1, v_2, v_3,..,v_n)$ be two bases of a vector space V. Show that the base change matrix P is uniquely determined by the two bases β and β' and is an invertible matrix.
2. Solve completely the system of equations $AX = 0$ and $AX = B$, where

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \end{bmatrix} \text{ and } B = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$$

Answers: Self Assessment

1. $\left[\frac{a-b}{5}(1-2i), \frac{a}{5}(1-2i) + \frac{b}{5}(7i-1) - ic, \frac{a}{5}(3-i) - \frac{b}{5}(6-i) - c(1+i)\right]$
2. α, β

6.6 Further Readings

Books

Kenneth Hoffman and Ray Kunze, *Linear Algebra*

I.N. Herstein, *Topics in Algebra*

Michael Artin, *Algebra*

Unit 7: Algebra of Linear Transformation

Objectives

After studying this unit, you will be able to:

- Know that linear transformation on the space is quite important. It helps in understanding the space under various transformations.
- See that the knowledge of the basis and dimension help us that the properties of linear transformation on the basis vector is central to the ideas of matrix mechanics.

Introduction

It will be seen that in the development of the algebra linear transformation plays an important part in understanding the properties of spaces. It is seen that the set of linear transformations also satisfy the properties of vector spaces.

7.1 Homomorphism

Consider two vector spaces V and W over the same field F i.e.

$$V = \{v, F, +, \oplus, \odot\}$$
$$W = \{w, F, +, +, \odot\}$$

The vectors of two different systems might have different names, and the vector operations of two systems might be defined in different ways.

A mapping H of V into W is called a homomorphism provided that all α, $B \in V$ and all $a \in F$,

$$(\alpha \oplus B)\, H = \alpha H + \beta H \qquad \text{...(1)}$$

and $$(\alpha \odot \alpha)\, H = a \,.\, \alpha H \qquad \text{...(2)}$$

If every vector of W is in the range of H, H is said to be *homomorphism* of V onto W.

A one-to-one homomorphism H of V onto W is called an *isomorphism*. If such a mapping exists, V, and W are said to be isomorphic.

We now show that

$$(\alpha \odot \alpha \oplus b \odot \beta)H = (a \cdot \alpha)\, H + (b\,.\,\beta)H \qquad ...(3)$$

clearly equation 1 follows from equation (3) by selecting $a = 1 = b$ and equation (2) follows by choosing $b = 0$.

7.2 Linear Transformation

Condition (3) is the requirement of linearly and since homomorphism is a mapping we call a homomorphism a *linear transformation*.

Thus a linear transformation T from a vector space V to a vector space W, both over the same field is a mapping of V onto W such that for all $\alpha, \beta \in V$ and for all $a, b, \in F$,

$$(a\alpha + b\beta)T = a\,(\alpha T) + b\,(\beta T)$$

Example 1: Identity transformation. If V is any vector space, then the identity transformation I defined by $I\alpha = \alpha$, is linear transformation from V into V.

The zero transformation 0, defined by $0\alpha = 0$, is a linear transformation from V into V.

Example 2: If V be the space of polynomial function f from the field F into F, given by

$$f(x) = C_0 + C_1 x + C_2 x^2 + ... + C_n x^n$$

Let $\quad Df(x) = C_1 + 2C_2 x + 3C_3 x^2 + ... + nC_n x^{n-1}$

Then D is a linear transformation from V into V the differentiation transformation.

Example 3: In two dimension space V_2, the transformation

$(x, y)T = (x \cos\theta - y \sin\theta, x \sin\theta + y \cos\theta)$ is a linear transformation

Example 4: In the space V_2 represented geometrically by the plane the transformation

$(x, y)T = (ax, by)$

Example 5: Let R be the field of real numbers and let V be the space of all functions from into R which are continuous. Define T by

$$(Tf)\,(x) = \int_0^x f(t)dt.$$

Then T is a linear transformation from V into V. The function Tf is not only continuous but has a continuous first derivative. The linearity of integration is one of its fundamental properties.

Example 6: Let A being a fixed $m \times n$ matrix with entries in the field F. The function T defined by $T(X) = AX$ is a linear transformation from $F^{n \times 1}$ into $F^{m \times 1}$. The function U defined by $U(\alpha) = \alpha A$ is a linear transformation from F^m into F^n.

Example 7: Let P be a fixed $m \times m$ matrix with entries in the field F and let Q be a fixed $n \times m$ matrix over F. Define a function T from the space $F^{m \times n}$ into itself by $T = PAQ$.

Notes

Then T is a linear transformation from $F^{m \times n}$ into $F^{m \times n}$,

because

$$\begin{aligned} T(CA + B) &= P(CA + B)Q \\ &= (CPA + PB)Q \\ &= C\,PAQ + PBQ \\ &= CT(A) + T(B) \end{aligned}$$

Example 8: The linear transformation preserves the linear combination; that is, if $\alpha_1, \alpha_2, ...\alpha_n$ are vectors in V and $C_1, C_2, ... C_n$ are scalars, then

$$T(C_1\alpha_1 + C_2\alpha_2 + ... + C_n\alpha_n) = C_1(T\alpha_1) + C_2T(\alpha_2) + ... + C_n(T\alpha_n).$$

This follows readily from the definition. For example

$$T(C_1\alpha_1 + C_2\alpha_2) = C_1(T\alpha_1) + C_2(T\alpha_2)$$

Theorem 1: Let V be a finite dimensional vector space over the field F and let $(\alpha_1, \alpha_2, ...\alpha_n)$ be an ordered basis for V. Let W be a vector space over the same field F and let $\beta_1, \beta_2, ... \beta_n$, be a set of any vectors in W. There is precisely one linear transformation T from V into W such that

$$T\alpha_i = \beta_i, \; i = 1, 2, ... n$$

Proof: To prove that there is some linear transformation T with $T\alpha_i = \beta i$, we proceed as follows, given α in V, there is a unique n-tuple $(x_1, x_2, ... x_n)$ such that

$$\alpha = x_1\,\alpha_i + x_2\,\alpha_2 + ... + x_n\,\alpha_n$$

For this α we define

$$T\alpha = x_1\beta_1 + x_2\beta_2 + ... + x_{\text{n}}\beta_n.$$

Then T is a well define rule for associating with each vector α in V a vector $T\alpha$ in W. From the definition it is clear that $T\alpha_j = \beta_j$ for each j. To see that T is linear, let

$$\beta = y_1\alpha_1 + y_2\alpha_2 + ... + y_n\alpha_n$$

be in V and let C be any scalar. Now

$$C\alpha + \beta = (Cx_1 + y_1)\alpha_1 + (Cx_2 + y_2)\alpha_2 + ... + (Cx_n + y_n)\alpha_n$$

and so by definition

$$T(C\alpha + \beta) = (Cx_1 + y_1)\beta_1 + (Cx_2 + y_2)\beta_2 + ... + (Cx_n + y_n)\beta_n$$

on the other hand

$$\begin{aligned} C(T\alpha) + T\beta &= e\sum_{i=1}^{n} x_i\beta_i + \sum_{i=1}^{n} y_i\beta_i \\ &= \sum_{i=1}^{n} (Cx_i + y_i)\beta_i \end{aligned}$$

and thus

$$T(C\alpha + \beta) = C(T\alpha) + (T\beta)$$

If U is a linear transformation from V into W with $U\alpha_i = \beta_i$, $i = 1, 2, \dots n$, then for the vector $\alpha = \sum_{i=1}^{n} x_i\alpha_i$, we have

$$U\alpha = U\left(\sum_{i=1}^{n} x_i\alpha_i\right)$$

$$= \sum_{i=1}^{n} x_i(U\,\alpha_i)$$

$$= \sum_{i=1}^{n} x_i\beta_i$$

So that U is exactly the rule T which we defined above. This shows that the linear transformation T with $T\alpha_i = \beta_i$ is unique.

Relations and operations of Linear Transformations

1. Two linear transformations T_1 and T_2 from v to w are said to be equal if and only if

$$\alpha T_1 = \alpha T_2 \text{ for all } \alpha \in v.$$

2. The sum $T_1 \oplus T_2$ of linear transformation from v to w are defined, respectively, by

$$\alpha(T_1 \oplus T_2) = \alpha T_1 + \alpha T_2 \text{ for all } \alpha \in v.$$

3. The scalar multiple $C \odot T_1$ of linear transformations from v to w are defined as

$$\alpha(c \odot T_1) = c(\alpha T_1), \text{ for all } \alpha \in v, c \in F.$$

Special Linear Transformation

(a) The zero linear transformation Z is defined from v to w by

$$\alpha Z = \ominus \text{ for every } \alpha \in v$$

(b) Negative transformation $(-T)$ from v to w, is defined by

$$\alpha(-T) = -\alpha T \text{ for every } \alpha \in v$$

(c) Identity linear transformation I from v to v is defined by

$$\alpha I = \alpha \text{ for every } \alpha \in v$$

(d) Product transformation $T_1 \boxdot T_2$.

Let v, w and y be vector spaces over the field F; let T_1 be a linear transformation from v to w and T_2 be a linear transformation from w to y. Then the product transformation $T_1 \boxdot T_2$ is the mapping from v to y defined by

$$\alpha(T_1 \boxdot T_2) = (\alpha T_1)T_2 \text{ for every } \alpha \in v.$$

Thus for every T we have

$$T \oplus Z = T$$

$$T \oplus -T = Z$$

$$T \boxdot I = I \boxdot T + T.$$

Notes

Example 9: In the space V_2 let T_1, T_2 and T_3 be defined by

$(x, y)T_1 = (x, 0)$

$(x, y)T_2 = (0, y)$

$(x, y)T_3 = (y, x)$

All these transformations are linear, now

$(x, y)T_1T_2 = (x, 0)T_2 = (0, 0)$, so $T_1T_2 = Z$

But $T_1 \neq Z$ and $T_2 \neq Z$ Hence a product of non-zero transformation can be the zero transformation. Also

$(x, y)T_2T_3 = (0, y)T_3 = (y, 0)$

But

$(x, y)T_3T_2 = (y, x)T_2 = (0, x)$. Hence

$\mathbf{T}_2T_3 \neq T_3\ T_2$.

So the multiplication of transformation is not commutative. Observe that

$(x, y)T_1T_1 = (x, 0)T_1 = (x, 0) = (x, y)T_1$,

so that $T_1^2 = T_1$. Thus there exist idempotent transformation i.e.

$$T_1^k = T_1$$

for integer k, other than I and Z.

Rank and Nullity of a Linear Transformation

Consider a linear transformation from a space v into a space w. The domain of T is the space v and the range of T is a subset R_T of w, the set of all images αT of the vectors of v:

$R_T = \{\beta \in w \mid \beta = \alpha T \text{ for some } \alpha \in v\}$

Another set associated with any vector space homomorphism T is the Kernel K_T of the homomorphism, which is defined to be the set of all vectors in v which are mapped into θ.

$$K_T = \{\alpha \in v \mid \alpha T = 0\}.$$

To see that K_T is a subspace of v, let $\alpha, \beta \in K_T$, and $C \in F$. Then

$$(\alpha + \beta)T = \alpha T + \beta T = \theta + \theta = \theta,$$

so that $\alpha + \beta \in K_T$, also $(C\ \alpha)T = C(\alpha\ T) = \theta$, so $c\ \alpha \in K_T$,

Thus K_T is a subspace of v.

These two subspaces, R_T and K_T, are called respectively the range space of T and the null space of T.

The range space R_T of a linear transformation T is the set of all images $\alpha T \in w$ as ranges over v. The rank $p\ (T)$ of a linear transformation T is the dimension of its range space.

The nullity $v\ (T)$ of a linear transformation T is the dimension of its null space.

Consider an n dimensional vector space v_n. If T is a linear transformation from v_n to w, then

$$P(T) \ = v(T) = n$$

Theorem 2: Let $\{\alpha_1, \alpha_2, ...\alpha_{v(T)}\}$ be a basis for K_T. Extend this basis to any basis $\{\alpha_1, \alpha_2, ...\alpha_{v(T)}, \alpha_{v(T)+1}, ... \alpha_n\}$ for v_n.

Then $\{\alpha_{v(T)+1} T, ..., \alpha_n T\}$ is a basis for R_T.

Proof: Let $\{\alpha_1, \alpha_2, ...\alpha_v\}$ be any basis for v_n. Any vector of R_T is of the form ξT for some $\xi \in V_n$. Let

$$\xi T = \sum_{i=1}^{n} a_1 \alpha_i ;$$

then

$$\xi T = \left(\sum_{i=1}^{n} a_i \alpha_i\right) T = \sum_{i=1}^{n} a_i (\alpha_i T) = \sum_{i=v(T)+1}^{n} a_i (\alpha_i T),$$

since

$$\alpha_i \, T = \theta \text{ for } i = 1, 2, ... v(T)$$

Hence $(\alpha_{v(T)+1} T, ..., \alpha_n T)$ spans R_T. As the dimension of R_T is not known we have to prove linear independence of the above vectors. Suppose scalars b_i, not all zero, exist such that

$$\theta = \sum_{v(T)+1}^{n} \mathrm{b}_i (\alpha_i T) = \sum_{i=v(T)+1}^{n} \mathrm{b}_i \, \alpha_i T$$

Then $\sum_{v(T)+1}^{n} b_i \alpha_i \in K_T$; but $\{\alpha_1, ...\alpha_{v(T)}\}$ spans K_T, so for suitable scalars c_i

$$\sum_{v(T)+1}^{n} b_i \alpha_i = \sum_{i=1}^{v(t)} c_i \alpha_i$$

This contradicts the linear independence of $\{\alpha_1, ...\alpha_n\}$, so the vectors $\{\alpha_{v(T)+1}, ...\alpha_n\}$ are linearly independent and therefore form a basis of R_T.

Theorem 3: If T is a linear transformation from V_n to w, then

$$p\,(\mathrm{T}) + v(T) = n.$$

Self Assessment

1. In the space of all polynomials $p(x)$ of all degrees define mapping M and D by:

   $$D\,p(x) = \frac{d}{dx} p(x), M\,p(x) = x\,p(x)$$

   Find

   (i) $DM - MD$

   (ii) $M^2 D^2 + MD$

2. Let v be the infinite dimensional space of all real polynomials. Let D and J be the Linear Transformation defined by

   $$D\,p(x) = \frac{d}{dx} p(x)$$

   $$J\,p(x) = \int_0^x p(t) dt$$

   for $p(x) \in v$,

Find

(i) $DJ\,p(x)$

(ii) Is $JD = DJ$?

3. Which of the following functions T from R^2 into R^2 are linear transformations?

(i) $T(x_1, x_2) = (1 + x_1, x_2)$;

(ii) $T(x_1, x_2) = (x_2, x_1)$;

(iii) $T(x_1, x_2) = (x_1^2, x_2)$;

(iv) $T(x_1, x_2) = (x_1 - x_2, 0)$.

7.3 Algebra of Linear Transformation

In the study of linear transformation from v into w it is of fundamental importance that the set of these transformations inherits a natural vector space structure.

Theorem 4: Let v and w be vector spaces over the field F. Let T and U be linear transformations from v into w. The function $(T + U)$ defined by

$$(T + U)\,(\alpha) = T\alpha + U\alpha$$

is a linear transformation from v into w. IF c is any element of F, the function (cT) defined by

$$(CT)\,(\alpha) = C(T\alpha)$$

is a linear transformation from v into w. The set of all linear transformations from v into w, together with the addition and scalar multiplication defined above is a vector space over the field F.

Proof: Suppose T and U are linear transformations from v into w and that we define $(T + U)$ as above. Then

$$\begin{aligned}(T + U)\,(C\alpha\ + \beta) &= T(C\alpha + \beta) + U\,(C\alpha + \beta)\\ &= C(T\alpha) + T\beta + C\,(U\alpha) + U\beta\\ &= C(T\alpha + U\alpha) + (T\beta + U\beta)\\ &= C(T + U)\alpha + (T + U)\beta\end{aligned}$$

which shows that $T + U$ is a linear transformation.

Similarly

$$\begin{aligned}(CT)\,(d\alpha\ + \beta) &= C[T(d\alpha + \beta]\\ &= C[d(T\alpha) + T\beta]\\ &= Cd(T\alpha) + C\,(T\beta)\\ &= d[c(T\alpha)] + c(T\beta)\\ &= d[(CT)\alpha] + C\,(T\beta)\end{aligned}$$

which shows that (CT) is a linear transformation. One must directly check that the vector addition and scalar multiplication are also satisfied along the above set of linear transformations of v into w.

We shall denote the space of linear transformations from v into w by $L(v, w)$. It is to be understood that $L(v, w)$ is defined only when v and w are vector spaces over the same field F.

Theorem 5: Let v be an n-dimensional vector space over the filed F; and let w be an m-dimensional vector space over F. Then the space $L(v, w)$ is finite dimensional and has dimension mn.

Thus let F be field, v and w vector spaces over F and L the set of all linear transformations from v into w. The system

$$£ = \{L, F, +, .; \oplus, \odot\}$$

is a vector space over F.

A special situation arises when we consider the system of all linear transformations of a vector space v into v itself £ is then a vector space in which the "vectors" are linear mappings of v into v. So we can define a product $S \odot T$ of vectors. This vector space over F in which a suitable product of vectors is defined is called an algebra of linear transformations over F.

A linear algebra £ over the field F is a system

$$£ = \{L, F, +, .; \oplus, \odot, \boxdot\}$$

which satisfies postulates:

(a) the system $\{L, F, T, ..., \oplus, \odot\}$ is a vector space over F.

(b) $\boxdot$ is a binary operation on £, which is closed, associative and bilinear

i.e.

closed $T_1, T_2 \in £$

Associative $T_1(T_2T_3) = (T_1T_2)T_3$

Bilinear $T_1(aT_2 + bT_3) = aT_1T_2 + bT_1T_3$

$(aT_2 + bT_3)\, T_1 = aT_2T_1 + bT_3T_1$

Also the dimension of £ is defined to be its dimension as a vector space.

Theorem 6: Let v, w and z be vector spaces over the field F. Let T be a linear transformation from v into w and u a linear transformation from w into z. Then the composed function UT defined by $(UT)(\alpha) = U(T(\alpha))$ is a linear transformation from v into z.

Proof:

$$\begin{aligned}(UT)(C\alpha + \beta) &= U[T(C\alpha + \beta)]\\ &= U(CT\alpha + T\beta)\\ &= C[U(T\alpha)] + U(T\beta)\\ &= C(UT)(\alpha) + (UT)(\beta)\end{aligned}$$

we shall be primarily concerned with linear transformation of a vector space into itself. So we from now on we write 'T is a linear operator on V' instead of writing 'T is a linear transformation from v into V'.

Definition: If v is a vector space over the field, a linear operator on v is a linear transformation from v into v.

Lemma: Let v be a vector space over the field F; let U, T_1 and T_2 be linear operators on v; let c be an element of F.

(a) $IU = UI = U$;

(b) $U(T_1 + T_2) = UT_1 + UT_2$; $(T_1 + T_2)\, U = T_1U + T_2U$;

(c) $C(UT_1) = (eU)\, T_1 = U(eT_1)$.

Proof: (a) This property of the identity function is obvious. We have stated here merely for emphasis.

(b)

$$[U(T_1 + T_2)](\alpha) = U[(T_1 + T_2)(\alpha)]$$
$$= U(T_1\alpha + T_2\alpha)$$
$$= U(T_1\alpha)] + U(T_2\alpha)$$
$$= (UT_1)(\alpha) + (UT_2)(\alpha)$$

so that

$$U(T_1 + T_2) = UT_1 + UT_2$$

Also $$[(T_1 + T_2)U](\alpha) = (T_1 + T_2)(U\alpha)$$
$$= T_1(U\alpha)] + T_2(U\alpha)$$
$$= (T_1U)(\alpha) + (T_2U)(\alpha)$$

so that $(T_1 + T_2)U = T_1U + T_2U$.

(c) It is easy to prove (c) in a simple way.

Non-singular Transformations

A linear transformation T from v and w is said to be non-singular transformation if and only if there exists a mapping T^* from R_T onto v such that $TT^* = I$, where I is the identity mapping on V. Thus $T^* = T^{-1}$. Thus $TT^{-1} = T^{-1}T = I$, T^{-1} is called inverse of T.

The function T from v into w is called invertible if there exists a function U from w into v such that UT is the identity function on v and TU is the identity function on w. If T is invertible, the function U is unique and is denoted by T^{-1}. Further more T is invertible if and only if

1. T is 1:1, that is, $T\alpha = T\beta$ implies $\alpha = \beta$;

2. T is onto, that is, the range of T is w.

Theorem 7: Let v and w be vector spaces over the field F and let T be a linear transformation from v into w. If T is invertible, then the inverse function T^{-1} is a linear transformation from w onto v.

Proof: What we are proving here is that if a linear transformation T is invertible, then the inverse T^{-1} is also linear.

Let β_1 and β_2 be vectors in w and let c be a scalar. We wish to show that

$$T^{-1}(C\beta_1 + \beta_2) = CT^{-1}\beta_1 + T^{-1}\beta_2$$

Let $\alpha_i = T^{-1}\beta_i$, $i = 1, 2$, that is, let α_i be the unique vector in v such that $T\alpha_i = \beta_i$. Since T is linear,

$$T(C\alpha_1 + \alpha_2) = CT\alpha_1 + T\alpha_2$$
$$= C\beta_1 + \beta_2.$$

Thus $C\alpha_1 + \alpha_2$ is the unique vector in v which is sent by T into $C\beta_1 + \beta_2$ and so

$$T^{-1}(C\beta_1 + \beta_2) = C\alpha_1 + \alpha_2$$
$$= CT^{-1}\beta_1 + T^{-1}\beta_2$$

and thus T^{-1} is linear.

Theorem 8: Let T be a linear transformation on v_n to w_n the following statements are equivalent.

1. T is non-singular

2. For all α, $\beta \in v_n$, if $\alpha T = \beta T$, then $\alpha = \beta$.

3. $K_T = [\theta]$

4. $v(T) = 0$

5. T is onto, that is, the range of T is w_n i.e. $p(T) = n$.

6. T maps any basis for v_n onto a basis for w_n.

Proof: Let n = dim v = din w. Now

$$\text{rank } (T) + \text{nullity } (T) = n$$

Since T is non-singular if and only if nullity $(T) = 0$ and rank $(T) = n$. Therefore T is non-singular if and only if $T(v_n) = w_n$. So, if either condition (1) or (2) holds the other is satisfied as well and T is invertible.

The above equations are also equivalent, there is some basis $(\alpha_1, \alpha_2, \alpha_n)$ for v such that $(T\alpha_1, T\alpha_2, ..., T\alpha_n)$ is basis for w.

Example 10: Let F be a field and let T be the linear operator on F^2 defined by

$$T(\text{x}_1, x_2) = (x_1, x_2, x_1)$$

Then T is non-singular.

Proof: If T is singular than $T(x_1, x_2) = 0$, means we have

$$x_1 + x_2 = 0$$

$$x_1 = 0$$

so the solution is $x_1 = 0$, $x_2 = 0$. We also see that T is onto; for let (z_1, z_2) be any vector in F^2. To show that (z_1, z_2) is in the range of T we must find scalars z_1 and z_2 such that

$$x_1 + x_2 = z_1$$

$$x_1 = z_2$$

and the obvious solution is $x_1, = z_2$, $x_2 = z_1 - z_2$. This last result gives us an explicit for T^{-1}, namely

$$T^{-1}(x_1, x_2) = (z_2, z_1 - z_2)$$

Self Assessment

4. If T and U be the linear operator on R^2 defined by

   $T(x_1, x_2) = (x_2, x_1)$ and $U(x_1, x_2) = (x_1, 0)$

   give rules like the ones defining T and U for each of the transformations

   (i) $U + T$

   (ii) UT

   (iii) TU

Notes

5. Let T be the unique linear operator on C^3 for which

   $T\in_1 = (1, 0, i)$, $T\in_2 = (0, 1, 1,)$ $T\in_3 = (i, 1, 0)$.

   Is T invertible?

7.4 Summary

- The properties of linear transformations are important in understanding the properties of the vector space.
- The basis vectors play an important part in the study of linear transformations.
- It is also explained that not all transformations are linear.
- A linear transformation T from a vector space V to a vector space W, both over the same field is a mapping of V onto W such that for all $\alpha, \beta \in V$ and for all $a, b, \in F$,

$$(a\alpha + b\beta)T = a\,(\alpha T) + b\,(\beta T)$$

7.5 Keywords

Homomorphism: If every vector of W is in the range of H, H is said to be homomorphism of V onto W.

Isomorphism: A one-to-one homomorphism H of V onto W is called an isomorphism. If such a mapping exists, V, and W are said to be isomorphic.

Linear Transformation: If T_1 is a linear transformation of v into w and T_2 is the linear transformation of w into z space, then $T_1\ T_2$ is a linear transformation of v into z.

7.6 Review Questions

1. Let T be a linear transformation on R^3 defined by

   $T(x_1, x_2, x_3) = (3x_1, x_1 - x_2, 2x_1 + x_2 + x_3)$

   (a) Is T invertible? If so, find a rule for T^{-1} like the one which defines T.

   (b) Find the value of

   $(T^2 - I)\ (T - 3I)\ (x_1, x_2, x_3)$.

2. Let $C^{2\times 2}$ be the complex vector space of 2 × 2 matrices with complex entries. Let

$$B = \begin{bmatrix} 1 & -1 \\ -4 & 4 \end{bmatrix}$$

   and let T be a linear operator on $C^{2\times 2}$ defined by

$$T(A) = BA - AB$$

   for any $A \in C^{2\times 2}$. What is the rank of T?

3. A transformation T on vector $\vec{V}$ of a vector space w is defined by

$$T(\vec{V}) + \vec{A} \times \vec{V}$$

   where the given vector $\vec{A} \in W$ and 'x' means the vector product. Find

$$(T^2 + A^2T)\,(\vec{V})\,.$$

Answers: Self Assessment

1. (i) I → identity transformation

   (ii) $(MD)^2$

2. (i) $DJ\,p(x) = I\,p(x)$

   (ii) no $JD \neq DJ$

3. (ii), (iv) are linear transformations.

4. (i) $(U + T)\,(x_1, x_2) = (x_1 + x_2, x_1)$

   (ii) $(U + T)\,(x_1, x_2) = (x_2, 0)$

   (iii) $(TU)\,(x_1, x_2) = (0, x_1)$

5. Yes T is invertible as $\in_1, \in_2, \in_3$ are standard basis of C^3 space.

7.7 Further Readings

Books Kenneth Hoffman and Ray Kunze, *Linear Algebra*

I.N. Herstein, *Topics in Algebra*

Michael Artin, *Algebra*.

Unit 8: Isomorphism

CONTENTS

Objectives

After studying this unit, you will be able to:

- Understand the linear transformation T is such that T transforms a subspace S of independent vectors of vector space into an independent subspace $T(S)$ of W.
- See that isomorphism is a homomorphism if the linear transformation T on V onto W is one-one.
- Know that for finite vector space the linear transformation T is non-singular if and only if dim V = dim W and T is isomorphism of V onto W.

Introduction

In dealing with two vector spaces over the same field, a transformation T from V into W can be homomorphism or isomorphism.

After studying this unit one can see that a fine n-dimensional vector space and a space of n-tuple co-ordinate space over the same field are isomorphic and so studying of one space gives all information about the other space.

8.1 Isomorphism

If V and W are vector spaces over the field F, any one-one linear transformation T of V onto W is called an isomorphism of V onto W. If there exists an isomorphism of V onto W, we say that V is isomorphic to W.

Note that V is trivially isomorphic to V, the identity transformation operator being an isomorphism of V onto V. Also, if V is isomorphic to W via an isomorphism T, then W is isomorphic to V, because then T is invertible and so T^{-1} is an isomorphism of W onto V. Thus it is easily verified that if V is isomorphic to W and W is isomorphic to Z, then V is isomorphic to Z. Briefly, isomorphism is an equivalence relation on the class of vector spaces. If there exists an isomorphism of V onto W, we sometimes say that V and W are isomorphic.

Theorem 1: Every n-dimensional vector space V_n over the field F is isomorphic to the space F^n.

Proof: Let V_n be an n-dimensional space over the field F and let $\beta = (\alpha_1, \alpha_2 ... \alpha_n)$ be the ordered basis for V. We defined a function T from V into F^n, as follows:

If α vector is V, let $T\alpha$ be the n-tuple $(x_1, x_2 ... x_n)$ of co-ordinates of α relative to the ordered basis β , i.e. the n-tuple such that

$$\alpha = x_1\alpha_1 + x_2\alpha_2 + ... + x_n\alpha_n.$$

given α in V, there is a unique n-tuple $(x_1, x_2 ... x_n)$ of scalars. Thus n-tuple is unique, because if we also have

$$\alpha = \sum_{i=1}^{n} z_i d_i$$

then $$\sum_{i=1}^{n} (x_i - z_i) d_i = 0$$

and the linear independence of the α , tells us that $x_i - z_i = 0$ for each i. We call the ith co-ordinate of α relative to the ordered basis

$$\beta = \{\alpha_1, \alpha_2, ... \alpha_n\}$$

Let another vector γ be given by

$$\gamma = \sum_{i=1}^{n} y_i \alpha_i$$

then $$\alpha + \gamma = \sum_{i=1}^{n} (x_i + y_i) \alpha_i$$

that the ith co-ordinate of $(\alpha + \gamma)$ in this ordered basis β is $(x_i + y_i)$. Similarly the ith co-ordinate of $(c\alpha)$ is $c\alpha_i$. One should note that every n-tuple $(x_1, x_2, ... x_n)$ in F^n is the n-tuple of co-ordinates of some vector in V. Thus, there is a one-one correspondence between the set of all vectors in V and the set of all n-tuples in F^n.

For many purposes one often regards isomorphic vector spaces as being the same, although the vectors and operations in the spaces may be quite different, that is, one often identifies isomorphic spaces. Let us denote the space of linear transformation from V into W by $L(V,W)$ over the same field F.

A Few Comments and Theorems

Suppose T is an isomorphism of V onto W. If S is a subset of V, then we have the following theorem:

Theorem 2: Let T be a linear transformation from V into W. Then T is non-singular if and only if T carries each linearly independent subset of V onto a linearly independent sub-set of W.

Proof: First suppose that T is non-singular. Let S be a linearly independent subset of V. If $\alpha_1, \alpha_2, ... \alpha_n$ are vectors in S, then the vectors $T\alpha_1, T\alpha_2, ... T\alpha_k$ are linearly independent, for if

$$c_1(T\alpha_1) + c_2(T\alpha_2) + ... + c_k(T\alpha_k) = 0$$

then $\quad T\left(c_1\alpha_1 + c_2\alpha_2 + ... + c_k\alpha_k\right) = 0$

and since T is non-singular

$$c_1\alpha_1 + c_2\alpha_2 + ... + c_k\alpha_k = 0$$

from which it follows that each $c_i = 0$, because S is an independent set. The argument shows that the image of S under T is independent.

Suppose that T carries independent subsets onto independent subsets. Let α be a non-zero vector in V. Then the set S consisting of the one vector α is independent. The image of S is the set consisting of the one vector $T\alpha$, and this set is independent. Therefore $T\alpha \neq 0$, because the set consisting of the zero vector alone is dependent. This shows that the null space of T is zero subspace i.e., T is non-singular.

Thus in deciding whether S is independent it does not matter whether we look at S or $T(S)$. From this one sees that an isomorphism is 'dimension preserving', that is any finite-dimensional subspace of V has the same dimension as the image under T. Here is a very simple illustration of this idea. Suppose A is an $m \times n$ matrix over the field F. We have really given two definitions of the solution space of the matrix A. The first is the set of all n-tuples $\left(x_1, x_2 ... x_n\right)$ in F^n which satisfy each of the equations in the system $AX = 0$. The second is the set of all $n \times 1$ column matrices X such that $AX = 0$. The first solution space is thus a subspace of F^n and the second is a subspace of the space of all $n \times 1$ matrices over F. Now there is a completely obvious isomorphism between F^n and F^{n+1}, namely

$$\left(x_1, x_2, ... x_n\right) \rightarrow \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

Under this isomorphism, the first space of A is carried onto the second solution space. These spaces have the same dimension, and so if we want to prove a theorem about the dimension of the solution space, it is immaterial which space we choose to discuss.

Example 1: $F^{(n)}$ is isomorphic $F^{(m)}$ if and only if $n = m$.

Proof: Here $F^{(n)}$ has, as one basis, the set of n vectors (1, 0, 0, …, 0), (0, 1, …, 0), … (0, 0, …, 1). Likewise $F^{(m)}$ has a basis containing m vectors. An isomorphism maps a basis of $F^{(n)}$ onto a basis of $F^{(m)}$. This is only possible if the dimensions of $F^{(n)}$ and $F^{(m)}$ are the same. Hence $n = m$.

Example 2: Prove that

(a) $F^{(1)}$ is not isomorphic to $F^{(n)}$ for $n > 1$.

(b) $F^{(2)}$ is not isomorphic to $F^{(3)}$.

Example 3: Let $V = C$ be the set of complex numbers, remembering only the addition of two elements as $\alpha + \beta$ and multiplication r α of a complex element α by a real number. Then the linear transformation T mapping $R^2 \rightarrow C$ sending $(a, b) \rightarrow a + b\,i$ is an isomorphism.

Example 4: Let $F^{n\times n}$ denote the set of $n \times n$ matrices with entries in a field F. This set is a vector space over F and it is isomorphic to the space of column vectors of length n^2.

Self Assessment

1. Show that $F^{m\times n}$ is isomorphic to F^{mn}.

2. Let V be the set of complex numbers regarded as a vector space over the field of real numbers. Define a function T from V into the space of 2 × 2 real matrices, as follows. If $z = x + iy$ with x and y real numbers, then

$$T(z) = \begin{bmatrix} x+7y & 5y \\ -10y & x-7y \end{bmatrix}.$$

   (a) Verify that $T(z_1 z_2) = T(z_1)T(z_2)$

   (b) Verify that T is a one-one (real) linear transformation of V into the space of 2×2 real matrices.

8.2 Summary

- A homomorphism is a mapping T of the space V into W over the same field F, preserving all the algebraic structures of the system. If T, in addition is one-to-one we call the mapping an isomorphism.
- Two spaces V and W are isomorphic only if the dim V = dim W.

8.3 Keywords

Isomorphism: T is an isomorphism of V into W over the same field F if T transforms a subset S of independent vectors into $T(S)$ a set of independent vectors of W.

Transformation: A transformation T of the space V into W is isomorphic if T is a non-singular transformation.

8.4 Review Questions

1. Let U and V be finite dimensional vector space over the field F. Prove that U and V are isomorphic if and only if dim U = dim V.

2. Let V and W be vector spaces over the field F and let T_1 be an isomorphism of V onto W. Prove that $T_2 \to T_1 T_2 T_1^{-1}$ is an isomorphism of $L(V,V)$ onto $L(W,W)$.

8.5 Further Readings

Books

Kenneth Hoffman and Ray Kunze, *Linear Algebra*

I.N. Herstein, *Topics in Algebra*

Sachin Kaushal, Lovely Professional University

Unit 9: Representation of Transformations by Matrices

CONTENTS

Objectives

After studying this unit, you will be able to:

- Know that the matrix of the linear transformation depends on the basis vectors of V as well as basis vectors of W where T is a linear transformation from V to W.
- See that the matrix of T depends upon the ordered basis relative to β and β' and the matrix of T relative to ordered basis β is different from the previous matrix.
- See that when T defines a transformation from V to V then the idea of similar matrices does come up.
- Understand how to find the matrix of T with the help of detailed solved examples.

Introduction

With the help of linear transformation one can deduce the rules for addition of matrices and multiplication of two matrices.

One can also understand geometrically the meaning of linear transformation clearly.

9.1 Representation of Transformations by Matrices

Although we have been discussing linear transformations for some time, it has always been in a detached way; to us a linear transformation has been a symbol (very often T) which acts in a certain way on a vector space. When one gets right down to it, outside of the few concrete examples encountered in the problems, we have really never come face to face with specific linear transformations. At the same time it is clear that if one were to pursue the subject further there would often arise the need of making a thorough and detailed study of a given linear transformation. To mention one precise problem, presented with a linear transformation; how does go about, in a "practical" and computable way, finding its characteristic roots?

What we seek first is a simple notation, or, perhaps more accurately, representation, for linear transformations. We shall accomplish this by use of a particular basis of the vector space and by use of the action of a linear transformation on this basis. Once this much is achieved by means of the operations in $A(V)$, we can induce operations for the symbols created, making of them an algebra. This new object, infused with an algebraic life of its own, can be studied as a mathematical entity having an interest by itself. This study is what comprises the subject of matrix theory.

However to ignore the source of these matrices, that is, to investigate the set of symbol independently of what they represent, can be costly. Instead we shall always use the interplay between the abstract, $A(V)$, and the concrete, the matrix algebra, to obtain information one about the other.

Let V be an n-dimensional vector space over the field F and let W be an m-dimensional vector space over F. Let $\mathcal{B} = \{\alpha_1, ..., \alpha_n\}$ be an ordered basis for V and $\mathcal{B}' = \{\beta_1, ..., \beta_m\}$ an ordered basis for W. If T is any linear transformation from V into W, then T is determined by its action on the vectors α_j. Each of the n vectors $T\alpha_j$ is uniquely expressible as a linear combination

$$T\alpha_j = \sum_{i=1}^{m} A_{ij}\beta_i \qquad ...(1)$$

of the β_i the scalars $A_{ij}, ...A_{mj}$ being the coordinates of $T\alpha_j$ in the ordered basis $\mathcal{B}'$. Accordingly, the transformation T is determined by the mn scalars A_{ij} via the formula (1). The $m \times n$ matrix A defined by $A(i, j) = A_{ij}$ is called the matrix of T relative to the pair of ordered basis $\mathcal{B}$ and $\mathcal{B}'$. Our immediate task is to understand explicitly how the matrix A determines the linear transformation T.

If $\alpha = x_1\alpha_1 + ... + x_n\alpha_n$ is a vector in V, then

$$T\alpha = T\left(\sum_{j=1}^{n} x_j\alpha_j\right)$$

$$= \sum_{j=1}^{n} x_j\left(T\alpha_j\right)$$

$$= \sum_{j=1}^{n} x_j \sum_{j=1}^{m} A_{ij}\beta_i$$

$$= \sum_{i=1}^{m}\left(\sum_{j=1}^{n} A_{ij}x_j\right)\beta_i.$$

If X is the coordinate matrix of α in the ordered basis β, then the computation above shows that AX is the coordinate matrix of the vector $T\alpha$ in the ordered basis β' , because the scalar

$$\sum_{j=1}^{n} A_{ij}x_j$$

is the entry in the ith row of the column matrix AX. Let us also observe that if A is any $m \times n$ matrix over the field F, then

$$T\left(\sum_{j=1}^{n} x_j\alpha_j\right) = \sum_{i=1}^{m}\left(\sum_{j=1}^{n} A_{ij}x_j\right)\beta_i \qquad ...(2)$$

defines a linear transformation T from V into W, the matrix of which is A, relative to $\mathcal{B}, \mathcal{B}'$.

Theorem 1: Let V be an n-dimensional vector space over the field F and W an m-dimensional vector space over F. Let $\mathcal{B}$ be an ordered basis for V and $\mathcal{B}'$ an ordered basis for W. For each linear transformation T from V into W, there is an $m\times n$ matrix A with entries in F such that

$$[T\alpha]\mathcal{B}'=A[\alpha]\mathcal{B}$$

for every vector α in V. Furthermore, $T \rightarrow A$ is a one-one correspondence between the set of all linear transformations from V into W and the set of all $m\times n$ matrices over the field F.

The matrix A which is associated with T in Theorem 1 is called the matrix of T relative to the ordered basis $\mathcal{B},\mathcal{B}'$. Note that Equation (1) says that A is the matrix whose columns $A_1,...,A_n$ are given by

$$A_j = \left[T\alpha_j\right]\mathcal{B}', \quad j = 1, ..., n.$$

If U is another linear transformation from V into W and B $\left[B_1,...,B_n\right]$ is the matrix of U relative to the ordered basis $\mathcal{B},\mathcal{B}'$ then $cA+B$ is the matrix of $cT+U$ relative $\mathcal{B},\mathcal{B}'$. That is clear because

$$\begin{aligned} cA_j + B_j &= c\left[T\alpha_j\right]\mathcal{B}'+\left[U\alpha_j\right]\mathcal{B}' \\ &= \left[cT\alpha_j + U\alpha_j\right]\mathcal{B}' \\ &= \left[(cT+U)\alpha_j\right]\mathcal{B}'. \end{aligned}$$

Theorem 2: Let V be an n-dimensional vector space over the field F and let W be an m-dimensional vector space over F. For each pair of ordered bases $\mathcal{B},\mathcal{B}'$ for V and W respectively, the function which assigns to a linear transformation T its matrix relative to $\mathcal{B},\mathcal{B}'$ is an isomorphism between the space $L(V,W)$ onto the set of $m\times n$ matrices over the field F.

Proof: We observed above that the function in question is linear, and as stated in Theorem 1, this function is one-one and maps $L(V, W)$ onto the set of $m\times n$ matrices.

We shall be particularly interested in the representation by matrices of linear transformations of a space into itself, i.e., linear operators on a space V. In this case it is most convenient to use the same ordered basis in each case, that is, to take $\mathcal{B}=\mathcal{B}'$. We shall then call the representing matrix simply the matrix of T relative to the ordered basis $\mathcal{B}$. Since this concept will be so important to us, we shall review its definition. If T is a linear operator on the finite-dimensional vector space V and $\mathcal{B}=\{\alpha_1,...,\alpha_n\}$ is an ordered basis for V, the matrix of T relative to $\mathcal{B}$ (or, the matrix of T in the ordered basis $\mathcal{B}$) is the $n\times n$ matrix A whose entries A_{ij} are defined by the equations

$$T\alpha_j = \sum_{i=1}^{n} A_{ij}\alpha_i, \quad j = 1, ..., n \qquad ...(3)$$

One must always remember that this matrix representing T depends upon the ordered basis $\mathcal{B}$, and that there is a representing matrix for T in each ordered basis for V. (For transformations of one space into another the matrix depends upon two ordered bases, one for V and one for W). In order that we shall not forget this dependence, we shall use the notation

$$[T]_{\mathcal{B}}$$

for the matrix of the linear operator T in the ordered basis $\mathcal{B}$. The manner in which this matrix and the ordered basis describe T is that for each α in V

$$[T_{\alpha}]_{\mathcal{B}} = [T]_{\mathcal{B}}[\alpha]_{\mathcal{B}}.$$

Example 1: Let V be the space of n×*1* column matrices over the field F; let W be the space of $m \times 1$ matrices over F; and let A be a fixed $m \times n$ matrix over F. Let T be the linear transformation of V into W defined by $T(X) = AX$. Let β be the ordered basis for V analogous to the standard basis in F^n, i.e., the i^{th} vector in $\mathcal{B}$ in the $n \times 1$ matrix X_i with a 1 in row i and all other entries 0. Let $\mathcal{B}'$ be the corresponding ordered basis for W, i.e. the jth vector in $\mathcal{B}'$ is the m×1 matrix Y_i with a 1 in row j and all other entries 0. Then the matrix of T relative to the pair $\mathcal{B},\mathcal{B}'$ is the matrix A itself. This is clear because the matrix AX_j is the j^{th} column of A.

Example 2: Let F be a field and let T be the operator of F^2 defined by

$$T(x_1, x_2) = (x_1, 0).$$

It is easy to see that T is a linear operator in F^2. Let $\mathcal{B}$ be the standard ordered basis for $F^2, \mathcal{B} = \{\in_1, \in_2\}$. Now

$$T \in_1 = T(1,0) = (1,0) = 1_{\in_1} + 0_{\in_2}$$

$$T \in_2 = T(0,1) = (0,0) = 0_{\in_1} + 0_{\in_2}$$

so the matrix of T in the ordered basis $\mathcal{B}$ is

$$[T]_{\mathcal{B}} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

Example 3: Let V be the space of all polynomial functions from R into R of the form

$$f(x) = c_0 + c_1 x + c_2 x^2 + c_3 x^3$$

that is, the space of polynomial functions of degree three or less. The differentiation operator D maps V into V, since D is 'degree' decreasing'. Let $\mathcal{B}$ be the ordered basis for V consisting of the four functions f_1, f_2, f_3, f_4 defined by $f_i(x) = x^{i-1}$. Then

$$(Df_1)(x) = 0, \qquad Df_1 = 0f_1 + 0f_2 + 0f_3 + 0f_4$$

$$(Df_2)(x) = 1, \qquad Df_2 = 1f_1 + 0f_2 + 0f_3 + 0f_4$$

$$(Df_3)(x) = 2x, \qquad Df_3 = 0f_1 + 2f_2 + 0f_3 + 0f_4$$

$$(Df_4)(x) = 3x^2, \qquad Df_4 = 0f_1 + 0f_2 + 3f_3 + 0f_4$$

so that the matrix of D in the ordered basis $\mathcal{B}$ is

$$[D]_{\mathcal{B}} = -\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

We have seen what happens to representing matrices when transformations are added, namely, that the matrices add. We should now like to ask what happens when we compose transformations. More specifically, let V, W and Z be vector spaces over the field F of respective dimensions n, m and p. Let T be a linear transformation from V into W and U a linear transformation from W into Z. Suppose we have ordered basis

$$\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}, \quad \mathcal{B}' = \{\beta_1, \ldots, \beta_m\}, \quad \mathcal{B}'' = \{\gamma_1, \ldots, \gamma_p\}$$

for the respective spaces V, W and Z. Let A be the matrix of T relative to the pair $\mathcal{B}', \mathcal{B}'$ and let $\mathcal{B}$ be the matrix of U relative to the pair $\mathcal{B}', \mathcal{B}''$. It is then easy to see that the matrix C of the transformation UT relative to the pair $\mathcal{B}, \mathcal{B}''$ is the product of B and A; for, if α is any vector in V.

$$[T\alpha]_{\mathcal{B}'} = A[\alpha]_{\mathcal{B}}$$

$$[U(T\alpha)]_{\mathcal{B}''} = B[T\alpha]_{\mathcal{B}'}$$

and so $\quad [(UT)(\alpha)]_{\mathcal{B}''} = BA[\alpha]_{\mathcal{B}}$

and hence, by the definition and uniqueness of the representing matrix, we must have $C = BA$. One can also see this by carrying out the computation

$$\begin{aligned}
(UT)(\alpha_j) &= U(T\alpha_j) \\
&= U\left(\sum_{k=1}^{m} A_{kj}\beta_k\right) \\
&= \sum_{k=1}^{m} A_{kj}(U\beta_k) \\
&= \sum_{k=1}^{m} A_{kj} \sum_{i=1}^{p} B_{ik}\gamma_i \\
&= \sum_{i=1}^{p}\left(\sum_{k=1}^{m} B_{ik}A_{kj}\right)\gamma_i
\end{aligned}$$

so that we must have

$$C_{ij} = \sum_{k=1}^{m} B_{ik}A_{kj}. \qquad \text{...(4)}$$

We motivated the definition (4) of matrix multiplication via operations on the rows of a matrix. One sees here that a very strong motivation for the definition is to be found in composing linear transformations. Let us summarize formally.

Theorem 3: Let V, W, and Z be finite-dimensional vector spaces over the field F; let T be a linear transformation from V into W and U a linear transformation from W into Z. If $\mathcal{B},\mathcal{B}'$ and $\mathcal{B}''$ are ordered basis for the spaces V, W and Z respectively, if A is the matrix of T relative to the pair $\mathcal{B},\mathcal{B}'$ and $\mathcal{B}''$ is the matrix of U relative to the pair $\mathcal{B}',\mathcal{B}''$, then the matrix of the composition UT relative to the pair $\mathcal{B}$, $\mathcal{B}''$ is the product matrix $C = BA$.

We remark that Theorem 3 gives a proof that matrix multiplication is associative – a proof which requires no calculations.

It is important to note that if T and U are linear operators on a space V and we are representing by a single ordered basis $\mathcal{B}$, then Theorem 3 assumes the simple form $[UT]_{\mathcal{B}} = [U]_{\mathcal{B}}[T]_{\mathcal{B}}$. Thus in this case, the correspondence which $\mathcal{B}$ determines between operators and matrices is not only a vector space isomorphism but also preserve products. A simple consequence of this is that the linear operator T is invertible if and only if $[T]_{\mathcal{B}}$ is an invertible matrix. For, the identity operator I is represented by the identity matrix in any order basis, and thus

$$UT = TU = I$$

is equivalent to

$$[U]_{\mathcal{B}}[T]_{\mathcal{B}} = [T]_{\mathcal{B}}[U]_{\mathcal{B}} = I.$$

Of course, when T is invertible

$$\left[T^{-1}\right]_{\mathcal{B}} = [T]_{\mathcal{B}}^{-1}.$$

Now we should like to inquire what happens to representing matrices when the ordered basis is changed. For the sake of simplicity, we shall consider this question only for linear operators on a space V, so that we can use a single ordered basis. The specific question is this. Let T be a linear operator on the finite-dimensional space V, and let

$$\mathcal{B}=\{\alpha_1,\ldots,\alpha_n\} \text{and } \mathcal{B}'=\{\alpha_1^{'},\ldots,\alpha_n^{'}\}$$

be two ordered basis for V. How are the matrices $[T]_{\mathcal{B}}$ and $[T]_{\mathcal{B}'}^{'}$ related? There is a unique (invertible) $n\times n$ matrix P such that

$$[\alpha]_{\mathcal{B}} = P[\alpha]_{\mathcal{B}'} \qquad \text{...(5)}$$

for every vector α in V. It is the matrix $P = [P_1,\ldots,P_n]$ where $Pj = \left[\alpha_j^{'}\right]_{\mathcal{B}}$. By definition

$$[T\alpha]_{\mathcal{B}} = [T]_{\mathcal{B}}[\alpha]_{\mathcal{B}}. \qquad \text{...(6)}$$

Applying (5) to the vector $T\alpha$, we have

$$[T\alpha]_{\mathcal{B}} = P[T\alpha]_{\mathcal{B}'}. \qquad \text{...(7)}$$

Combining (5), (6) and (7), we obtain

$$[T]_{\mathcal{B}}P[\alpha]_{\mathcal{B}'} = P[T\alpha]_{\mathcal{B}'}$$

$$P^{-1}[T]_{\mathcal{B}}P[\alpha]_{\mathcal{B}'} = [T\alpha]_{\mathcal{B}'}$$

Notes

and so it must be that

$$[T]_{\mathcal{B}'} = P^{-1}[T]_{\mathcal{B}}P. \qquad ...(8)$$

This answers our questions.

Before stating this result formally, let us observe the following. There is a unique linear operator U which carries $\mathcal{B}$ onto $\mathcal{B}'$, defined by

$$U\alpha_j = \alpha'_j, \quad j = 1,...,\text{n}$$

This operator U is invertible since it carries a basis for V onto a basis for V. The matrix P (above) is precisely the matrix of the operator U in the ordered basis $\mathcal{B}$. For, P is defined by

$$\alpha'_j = \sum_{i-1}^{n} P_{ij}\alpha_i$$

and since $U\alpha_j = \alpha'_j$, this equation can be written as

$$U\alpha_j = \sum_{i=1}^{n} P_{ij}\alpha_i.$$

So P = $[U]_{\mathcal{B}}$, by definition.

Theorem 4: Let V be a finite-dimensional vector space over the field F, and let

$$B = \{\alpha_1,...,\alpha_\text{n}\} \text{ and } \mathcal{B}' = \{\alpha'_1,...,\alpha'_\text{n}\}$$

be ordered basis for V. Suppose T is linear operator on V. If $P = \{P_1,...,P_\text{n}\}$ is the $n\times n$ matrix with columns $P_j = P^{-1}[T]_{\mathcal{B}}$, then

$$[T]_{\mathcal{B}'} = P^{-1}[T]_{\mathcal{B}}P.$$

Alternatively, if U is the invertible operator on V defined by $U\alpha_j = \alpha'_j, j = 1,...,n$ then

$$[T]_{\mathcal{B}'} = [U]_{\mathcal{B}'}^{-1}[T]_{\mathcal{B}}[U]_{\mathcal{B}}.$$

Self Assessment

1. Let T be the linear transformation $T: R^3 \rightarrow R^3$, defined by

   $T(x,y,z) = (2y+z, x-4y, 3z)$

   find the matrix T, with respect to the basis

   $E_1 = (1,1,1), E_2 = (1,1,0)$ and $E_3 = (1,0,0)$

2. A transformation T is defined by

$$T(x,y) = \frac{1}{\sqrt{2}}\{(x-y), x+y\}$$

(i) Show that T is linear

(ii) Find the matrix M represented by T w.r.t. basis (1,0) and (0,1)

9.2 Illustrative Examples

Example 4: Let T be the linear transformation defined by

$$T(x_1, x_2) = (x, 0).$$

The matrix of T in the standard basis $\varepsilon_1 = (1,0), \varepsilon_2 = (0,1)$

is $$[T]_\beta = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

Let β' be the ordered basis for R^2 given by $\varepsilon'_1 = (1,1), \varepsilon'_2 = (2,1)$.

Then $\varepsilon'_1 = \varepsilon_1 + \varepsilon_2, \varepsilon'_2 = 2\varepsilon_1 + \varepsilon_2$, so that P matrix is

$$P = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix} \text{ and } P^{-1} = \begin{bmatrix} -1 & 2 \\ 1 & -1 \end{bmatrix}$$

Thus $$[T]_{\beta'} = P^{-1} T_\beta P$$

$$= \begin{bmatrix} -1 & 2 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} -1 & 2 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} -1 & -2 \\ 1 & 2 \end{bmatrix}$$

We can easily check that this is correct because

$$T \in'_1 = (1,0) = - \in'_1 + \in'_2$$

$$T \in'_2 = (2,0) = -2 \in'_1 + 2 \in'_2 .$$

Example 5: Let V be the space of polynomial functions from R into R which have 'degree' less than or equal to 3. As in Example 3, let D be the differentiation operator on V, and let

$$\mathcal{B} = \{f_1, f_2, f_3, f_4\}$$

be the ordered basis for V defined by $f_1(x) = x^{i-1}$. Let t be a real number and define $g_1(x) = (x+t)^{i-1}$, that is

$$g_1 = f_1$$

Notes

$$g_2 = tf_1 + f_2$$

$$g_3 = t^2 f_1 + 2tf_2 + f_3$$

$$g_4 = t^3 f_1 + 3t^2 f_2 + 3tf_3 + f_4.$$

Since the matrix

$$P = \begin{bmatrix} 1 & t & t^2 & t^3 \\ 0 & 1 & 2t & 3t^2 \\ 0 & 0 & 1 & 3t \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

is easily seen to be invertible with

$$P^{-1} = \begin{bmatrix} 1 & -t & t^2 & -t^3 \\ 0 & 1 & -2t & 3t^2 \\ 0 & 0 & 1 & -3t \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

it follows that $\mathcal{B}' = \{g_1, g_2, g_3, g_4\}$ is an ordered basis for V. In Example 3, we found that the matrix of D in the ordered basis $\mathcal{B}$ is

$$[D]_{\mathcal{B}} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

The matrix of D in the ordered basis $\mathcal{B}'$ is thus

$$P^{-1}[D]_{\mathcal{B}}P = \begin{bmatrix} 0 & -t & t^2 & t^3 \\ 0 & 1 & -2t & 3t^2 \\ 0 & 0 & 1 & -3t \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & t & t^2 & t^3 \\ 0 & 1 & 2t & 3t^2 \\ 0 & 0 & 1 & 3t \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & -t & t^2 & t^3 \\ 0 & 1 & -2t & 3t^2 \\ 0 & 0 & 1 & -3t \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 2t & 3t^2 \\ 0 & 0 & 2 & 6t \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Thus D is represented by the same matrix in the ordered basis $\mathcal{B}$ and $\mathcal{B}'$. Of course, one can see this somewhat more directly since

$$Dg_1 = 0$$

$$Dg_2 = g_1$$

$$Dg_3 = 2g_2$$

$$Dg_4 = 3g_3.$$

This example illustrates a good point. If one knows the matrix of a linear operator in some ordered basis $\mathcal{B}$ and wishes to find the matrix in another ordered basis $\mathcal{B}'$, it is often most convenient to perform the coordinate change using the invertible matrix P; however, it may be a much simpler task to find the representing matrix by a direct appeal to its definition.

Definition: Let A and B be $n\times n$ (square) matrices over the field F. We say that B is similar to A over F if there is an invertible $n\times n$ matrix P over F such that $\beta = P^{-1}AP$.

According to Theorem 4, we have the following: If V is an n-dimensional vector space over F and $\mathcal{B}$ and $\mathcal{B}'$ are two ordered bases for V_i then for each linear operator T on V the matrix $B = [T]_{\mathcal{B}'}$ is similar to the matrix $A = [T]_{\mathcal{B}'}$. The argument also goes in the other direction. Suppose A and B are $n\times n$ matrices and that B is similar to A. Let V be any n-dimensional space over F and let $\mathcal{B}$ be an ordered basis for V. Let T be the linear operator on V which is represented in the basis $\mathcal{B}$ by A. If $\beta = P^{-1}AP$, let $\mathcal{B}'$ be the ordered basis for V obtained from $\mathcal{B}$ by P, i.e.

$$\alpha_j' = \sum_{i=1}^{n} P_{ij}\alpha_i.$$

Then the matrix of T in the ordered basis $\mathcal{B}'$ will be B.

Thus the statement that B is similar to A means that on each n-dimensional space over F the matrices A and B represent the same linear transformation in two (possibly) different ordered basis.

Note that each $n\times n$ matrix A is similar to itself, using $P = I$; if B is similar to A, then A is similar to B, for $B = P^{-1}AP$ implies that $A = \left(P^{-1}\right)^{-1} BP^{-1}$; if B is similar to A and C is similar to B, then C is similar to A, for $B = P^{-1}AP$ and $C = Q^{-1}BQ$ imply that $C = (PQ)^{-1}A(PQ)$. Thus, similarity is an equivalence relation on the set of $n\times n$ matrices over the field F. Also note that the only matrix similar to the identity matrix I is I itself, and that the only matrix similar to the zero matrix is the zero matrix itself.

Self Assessment

3. Let T be the linear transformation on R^3 defined by

   $$T(x_1, x_2, x_3) = (3x_1 + x_3, -2x_1 + x_2, -x_1 + 2x_2 + 4x_3)$$

   (i) What is the matrix of T in the standard ordered basis for R^3?

   (ii) What is the matrix of T in the ordered basis $(\alpha_1, \alpha_2, \alpha_3)$ where

   $$\alpha_1 = (1,0,1), \alpha_2 = (-1,2,1) \text{and } \alpha_3 = (2,1,1)$$

4. Let T be the linear transformation from R^3 into R^2 defined by

   $$T(x_1, x_2, x_3) = (x_1 + x_2, 2x_3 - x_1)$$

   If B is the standard ordered basis for R^3 and β' is the standard ordered basis for R^2, what is the matrix of T relative to the pair β, β'?

9.3 Summary

- One can identify the effect of linear transformation on the space and study its effects by means of algebra of matrices.
- This way one has insight of the meaning of similar matrices.
- The linear transformation T for R^3 to R^2.

9.4 Keywords

Degree Decreasing: The differentiation operator D maps V into V, since D is 'degree' decreasing.

Linear Transformation: The statement that B is similar to A means that on each n-dimensional space over F the matrices A and B represent the same linear transformation in two (possibly) different ordered basis.

Unique Linear Operator: A unique linear operator U which carries $\mathcal{B}$ onto $\mathcal{B}'$, defined by $U\alpha_j = \alpha'_j, \; j = 1,...,n$

9.5 Review Questions

1. Let T be the linear transformation on R^2 defined by

   $T(x_1, x_2) = (-x_2, x_1)$

   (a) What is the matrix of T in the standard basis for R^2?

   (b) What is the matrix of T in the standard basis $\beta(\alpha_1, \alpha_2)$ where $\alpha_1 = (1,2)$ and $\alpha_2 = (1,-1)$?

2. Let $(\alpha_1, \alpha_2, \alpha_3)$ be the basis for V_3 and let $\beta_1 = \alpha_1 - 2\alpha_2, \beta_2 = \alpha_1 + \alpha_2 + \alpha_3, \beta_3 = \alpha_2 - \alpha_3$

   (a) Prove $(\beta_1, \beta_2, \beta_3)$ is a basis and express $\alpha_1, \alpha_2, \alpha_3$ as a linear combination of β_1, β_2 and β_3.

   (b) If T is defined by $T\alpha_i = \beta_i, \quad i = 1,2,3,...$

   find a matrix A which represents T relative to α basis.

9.6 Further Readings

Books

Kenneth Hoffman and Ray Kunze, *Linear Algebra*

I.N. Herstein, *Topics in Algebra*

Michael Artin, *Algebra*

Unit 10: Linear Functionals

CONTENTS

Objectives

After studying this unit, you will be able to:

- Understand in a better way the discussion of subspaces, linear equations and co-ordinates.
- See that a few examples of linear functional cited in this unit.
- Know the concept of dual basic vectors for the dual vector space V^*.
- See that how to obtain the basis of the dual spaces which is done by examples.

Introduction

The concept of linear function is important in the study of finite dimensional spaces because the linear functional method helps to organize and clarify the discussion of subspaces.

The method is illustrated by means of a few theorems and a few solved examples.

10.1 Linear Functionals

If V is a vector space over the field F, a linear transformation f from V into the scalar field F is also called a **linear functional** on V. If we start from scratch, this means that f is a function from V into F such that

$$f.(c\alpha + \beta) = cf(\alpha) + f(\beta)$$

for all vectors α and β in V and all scalars c in F. The concept of linear functional is important in the study of finite-dimensional spaces because it helps to organize and clarify the discussion of subspaces, linear equations, and coordinates.

Example 1: Let F be a field and let $a_1, ..., a_n$ be scalars in F. Define a function f on F^n by

$$f(x_1,...,x_n) = a_1x_1 + ... + a_nx_n$$

Then f is a linear functional on F^n. It is the linear functional which is represented by the matrix $[a_1 ... a_n]$ relative to the standard ordered basis for F^n and the basis {1} for F:

$$a_j = f(\varepsilon_j), \qquad j = 1,...,n.$$

Every linear functional on F^n is of this form, for some scalars a_1, ..., a_n. That is immediate from the definition of linear functional because we define $a_j = f(\varepsilon_j)$ and use the linearity

$$\begin{aligned} f(x_1, ..., x_n) &= f\left(\sum_j x_j \varepsilon_j\right) \\ &= \sum_j x_j f(\varepsilon_j) \\ &= \sum_j a_j x_j \end{aligned}$$

Example 2: Here is an important example of a linear functional. Let n be a positive integer and F is field. If A is an $n \times n$ matrix with entries in F, the **trace** of A is the scalar

$$tr\, A = A_{11} + A_{22} + ... + A_{nn}.$$

The trace function is a linear functional on the matrix space $F^{n \times n}$ because

$$\begin{aligned} tr(cA + B) &= \sum_{i=1}^{n} (cA_{ii} + B_{ii}) \\ &= c\sum_{i=1}^{n} cA_{ii} + \sum_{i=1}^{n} B_{ii} \\ &= c\ tr\ A + tr\ B. \end{aligned}$$

Example 3: Let V be the space of all polynomial functions from the field F into itself. Let t be an element of F. If we define

$$L_t(p) = p(t)$$

then L_t is a linear functional on V. One usually describes this by saying that, for each t, 'evaluation at t' is a linear functional on the space of polynomial functions. Perhaps we should remark that the fact that the functions are polynomials plays no role in this example. Evaluation at t is a linear functional on the space of all functions from F into F.

Example 4: This may be the most important linear functional in mathematics. Let $[a, b]$ be a closed interval on the real line and let $C([a, b])$ be the space of continuous real-valued functions on $[a, b]$. Then

$$L(g) = \int_a^b g(t)\ dt$$

Theorem 1: Let V be an n-dimensional vector space over the field F, and let W be an m-dimensional vector space over F. Then the space $L(V, W)$ is finite-dimensional and has dimension mn.

Proof: Let

$$\beta = \{\alpha_1, \alpha_2, ...\alpha_n\} \text{ and } \beta' = \{\beta_1, \beta_2, ...\beta_n\}$$

be ordered basis for V and W, respectively. For each pair of integers (p, q) with $1 \le p \le m$ and $1 \le q \le n$, we define a linear transformation $E^{p,q}$ from V into W by

$$E^{p,q}(\alpha_i) = \begin{cases} 0 & \text{if} \quad i \neq q \\ \beta_p & \text{if} \quad i = q \end{cases}$$

$$= \delta_{iq}\,\beta_b$$

According to the theorem 1 of unit 7, there is a unique linear transformation from V into W satisfying these conditions. The claim is that the mn transformations $E^{p,q}$ from a basis for $L(V, W)$.

Let T be a linear transformation from V into W. For each j, $i \le j \le n$, let $A_{1j}, A_{2j}, \ldots A_{mj}$ be the co-ordinates of the vector $T\alpha_j$ in the ordered basis β', i.e.,

$$T\alpha_j = \sum_{p=1}^{m} A_{pj}\,\beta_p \qquad \ldots(1)$$

we wish to show that

$$T = \sum_{p=1}^{m}\sum_{q=1}^{n} A_{pq}\,E^{p,q} \qquad \ldots(2)$$

Let U be the linear transformation in the right hand member of (2). Then for each j

$$U\alpha_j = \sum_{p}\sum_{q} A_{pq}\,E^{p,q}(\alpha_j)$$

$$= \sum_{p}\sum_{q} A_{pq}\,\delta_{jq}\,\beta_p$$

$$= \sum_{p=1}^{m} A_{pj}\,\beta_p$$

$$= T\alpha_{\mathrm{j}}$$

and consequently $U = T$. Now (2) shows that the $E^{p,q}$ span $L(V, W)$; we must prove that they are independent. But this is clear from what we did above; for, if the transformation

$$U = \sum_{p}\sum_{q} A_{pq}\,E^{p,q}$$

is the zero transformation, then $U\alpha_j = 0$ for each j, so

$$\sum_{p=1}^{m} Ap_j\,\beta_j = 0$$

and the independence of the β_{p} implies that $A_{pj} = 0$ for every p and j.

If V is finite-dimensional vector space, the collection of linear functionals of V forms a vector space in a natural way. It is the space $L(V, F)$. We denote this space by V^*. From the above theorem we know the following about the space V^* that

$$\dim V^* = \dim V. \qquad \ldots(3)$$

Let $\beta = (\alpha_1, \alpha_2, ...\alpha_n)$ be a basis for V. According to theorem 1 of unit 7, there is (for each i) a unique linear functional f_i on V such that

$$f_i\,(\alpha_i) = \delta_{ij} \qquad ...(4)$$

In this way we obtain from β a set of n distinct linear functionals $f_1, f_2, ... f_n$ on V. These functionals are also linearly independent. For, suppose

$$f = \sum_{i=1}^{n} c_i f_i \qquad ...(5)$$

Then

$$\begin{aligned} f(\alpha_j) &= \sum_{i=1}^{n} c_i f_i(\alpha_j) \\ &= \sum_{i=1}^{n} c_i \delta_{ij} \\ &= c_j. \end{aligned}$$

In particular, if f is the zero functional $f(\alpha_j) = 0$ for each j and hence the scalars c_j are all 0. Now $f_1, ...f_n$ are n linearly independent functionals, and since we know that V^* has dimension n, it must be that $\mathcal{B}^* = \{f_1, ..., f_n\}$ is a basis for V^*. This basis is called the dual basis of $\mathcal{B}$.

Theorem 2: Let V be a finite-dimensional vector space over the field F, and let $\mathcal{B} = \{\alpha_1, ..., \alpha_n\}$ be a basis for V. Then there is a unique dual basis $\mathcal{B}^* = \{f_1, ..., f_n\}$ for V^* such that $f_i(\alpha_j) = \delta_{ij}$. For each linear functional f on V we have

$$f = \sum_{i=1}^{n} f(\alpha_i) f_i \qquad ...(6)$$

and for each vector α in V we have

$$\alpha = \sum_{i=1}^{n} f_1(\alpha)\, \alpha_i. \qquad ...(7)$$

Proof: We have shown above that there is a unique basis which is 'dual' to $\mathcal{B}$. If f is a linear functional on V, then f is some linear combination (5) of the f_i, and as we observed after (5) the scalars c_j must be given by $c_j = f(\alpha_j)$. Similarly, if

$$\alpha = \sum_{i=1}^{n} x_i\, \alpha_i$$

is a vector in V, then

$$\begin{aligned} f_j(\alpha) &= \sum_{i=1}^{n} x_i f_i(\alpha_i) \\ &= \sum_{i=1}^{n} x_i \delta_{ij} \\ &= x_j \end{aligned}$$

so that the unique expression for α as a linear combination of the α_i, is

$$\alpha = \sum_{i=1}^{n} f_i(\alpha)\alpha_i.$$

Equation (7) provides us with a nice way of describing what the dual basis is. It says if $\mathcal{B} = \{\alpha_1, \alpha_2 ..., \alpha_n\}$ is an ordered basis for V and $\mathcal{B}^* = \{f_1, ..., f_n\}$ is the dual basis, then f_i is precisely the function which assigns to each vector α in V the ith coordinate of α relative to the ordered basis $\mathcal{B}$. Thus we may also call the f_i the coordinate functions for $\mathcal{B}$. The formula (6), when combined with tells us the following:

If f is in V^*, and we let $f(\alpha_i) = a_i$, then when

$$\alpha = x_1\alpha_1 + ... + x_n\alpha_n$$

we have

$$f(x) = a_1x_1 + ... + a_nx_n. \qquad ...(8)$$

In other words, if we choose an ordered basis $\mathcal{B}$ for V and describe each vector in V by its n-tuple of coordinates $(x_1,....x_n)$ relative to $\mathcal{B}$, then every linear functional on V has the form. This is the natural generalization of Example 1, which is the special case $V = F^n$ and $\mathcal{B} = \{\varepsilon_1, ..., \varepsilon_n\}$.

Example 5: Let V be the vector space of all polynomial functions from R into R which have degree less than or equal to 2. Let t_1, t_2 and t_3 be any three distinct real numbers, and let

$$L_i(p) = p(t_i)$$

Then L_1, L_2 and L_3 are linear functionals on V. These functionals are linearly independent; for, suppose

$$L = c_1L_1 + c_2L_2 + c_3L_3$$

If $L = 0$, i.e., if $L(p) = 0$ for each p in V, then applying L to the particular polynomial 'functions' 1, x, x^2, we obtain

$$c_1 + c_2 + c_3 = 0$$

$$t_1c_1 + t_2c_2 + t_3c_3 = 0$$

$$t_1^2c_1 + t_2^2c_2 + t_3^2c_3 = 0$$

From this it follows that $c_1 = c_2 = c_3 = 0$, because (as a short computation shows) the matrix

$$\begin{bmatrix} 1 & 1 & 1 \\ t_1 & t_2 & t_3 \\ t_1^2 & t_2^2 & t_3^2 \end{bmatrix}$$

is invertible when t_1, t_2 and t_3 are distinct. Now the L_i are independent and since V has dimension 3, these functional from a basis for V^*. What is the basis for V, of which this is the dual? Such a basis $\{p_1, p_2, p_3\}$ for V must satisfy

$$L_i(p_i) = \delta_{ij}$$

or

$$p_j(t_i) = \delta_{ij}.$$

These polynomial functions are easily shown to be

$$p_1(x) = \frac{(x-t_2)(x-t_3)}{(t_1-t_2)(t_1-t_3)}$$

$$p_2(x) = \frac{(x-t_1)(x-t_3)}{(t_2-t_1)(t_2-t_3)}$$

$$p_3(x) = \frac{(x-t_1)(x-t_2)}{(t_3-t_1)(t_3-t_2)}$$

The basis $\{p_1, p_2, p_3\}$ for V is interesting, because according to (7) we have to each p in V.

$$p = p(t_1)p_1 + p(t_2)p_2 + p(t_3)p_3.$$

Thus, if c_1, c_2 and c_3 are any real numbers, there is exactly one polynomial function p over R which has degree at most 2 and satisfies $p(t_j) = -c_j, j = 1,2,3.$ This polynomial function is $p = c_1p_1 + c_2p_2 + c_3p_3$.

Now let us discuss the relationship between linear functionals and subspaces. If f is a non-zero linear functional, then the rank of f is 1 because the range of f is a non-zero subspace of the scalar field and must (therefore) be the scalar field. If the underlying space V is finite-dimensional, the rank plus nullity theorem tells us that the null space N_f has dimension

$$\dim N_f = \dim V - 1.$$

In a vector space of dimension n, a subspace of dimension $n - 1$ is called a **hyperspace**. Such spaces are sometimes called hyperplanes or subspaces of co-dimension 1. Is every hyperspace the null space of a linear functional? The answer is easily seem to be yes. It is not much more difficult to show that each d-dimensional subspace of an n-dimensional space is the intersection of the null spaces of $(n - d)$ linear functionals (Theorem below).

Definition: If V is a vector space over the field F and S is a subset of V, the **annihilator** of S is the set $S°$ of linear functionals on V such that $f(\alpha) = 0$ for every α in S.

It should be clear that $S°$ is a subspace of V^*, whether S is a subspace of V or not. If S is the set consisting of the zero vector alone, then $S° = V^*$. If $S = V$, then $S°$ is the zero subspace of V^*. (This is easy to see when V is finite-dimensional.)

Theorem 3. Let V be a finite-dimensional vector space over the field F, and let W be a subspace of V. Then

$$\text{din } W + \dim W° = \dim V.$$

Proof: Let k be the dimension of W and $\{\alpha_1, ..., \alpha_k\}$ a basis for W. Choose vector $\alpha_{k+1}, ..., \alpha_n$, in V such that $\{\alpha_1,..., \alpha_n\}$ is a basis for V. Let $\{f_1, ..., f_n\}$ be the basis for V^* which is dual to this basis for V.

This claim is that $\{f_{k+1},...f_n\}$ is a basis for the annihilator $W°$. Certainly f_i belongs to $W°$ for $i \geq k + 1$, because

$$f_i(\alpha_i) = \delta_{ij}$$

and $\delta_{ij} = 0$ if $i \geq k + 1$ and $j \leq k$; from this it follows that, for $i \geq k + 1$, $f_i(\alpha) = 0$ whenever α is a linear combination of $\alpha_1,...,\alpha_k$. The functionals $f_{k+1},...,f_n$ are independent, so all we must show is that they span $W°$. Suppose f is in V^*.

Now,

$$f = \sum_{i=1}^{n} f(\alpha_i) f_i$$

so that if f is in W° we have $f(\alpha_i) = 0$ for $i \le k$ and

$$f = \sum_{i=k+1}^{n} f(\alpha_i) f_i.$$

We have shown that if dim $W = k$ and dim $V = n$ then dim $W^\circ = n - k$.

Corollary: If W is a k-dimensional subspace of an n-dimensional vector space V, then W is the intersection of $(n - k)$ hyperspaces in V.

Proof: This is a corollary of the proof of Theorem 3 rather than its statement. In the notation of the proof, W is exactly the set of vectors α such that $f_i(\alpha) - 0$, $i = k+1, ..., n$. In case $k = n - 1$, W is the null space of f_n.

Corollary: If W_1 and W_2 are subspaces of a finite-dimensional vector space, then $W_1 = W_2$ if and only if $W_1^0 = W_2^0$.

Proof: If $W_1 = W_2$, then of course $W_1^0 = W_2^0$. If $W_1 \neq W_2$, then one of then two subspaces contains a vector which is not in the other. Suppose there is a vector α which is in W_2 but not in W_1. By the previous corollaries (or the proof of Theorem 3) there is a linear functional f such that $f(\beta) = 0$ for all β in W, but $f(\alpha) \neq 0$. Then f is in W_1^0 but not in W_2^0 and $W_1^0 \neq W_2^0$.

10.2 System of Linear Equations

The first corollary says that, if we select some ordered basis for the space, each k-dimensional subspace can be described by specifying $(n - k)$ homogeneous linear conditions on the coordinates relative to that basis.

Let us look briefly at system of homogeneous linear equations from the point of view of linear functionals. Suppose we have a system of linear equations,

$$\begin{aligned} A_{11}x_1 + \cdots + A_{1n}x_n &= 0 \\ \vdots \qquad\quad \vdots \\ A_{m1}x_1 + \cdots + A_{mn}x_n &= 0 \end{aligned}$$

for which we wish to find the solutions. If we let f_i, $i = 1, ..., m$, be the linear functional on F^n defined by

$$f_i(x_i, ..., x_n) = A_{ix}x_i + ... + A_{in}x_n$$

then we are seeking the subspace of F^n of all α such that

$$f_i(\alpha) = 0, \qquad i = 1, ..., m.$$

In other words, we are seeking the subspace annihilated by $f_1, ..., f_m$. Row-reduction of the coefficient matrix provides us with a systematic method of finding this subspace. The n-tuple $(A_{i1}, ...A_{in})$ gives the coordinates of the linear functional f_i relatives to the basis which is dual to the standard basis for F^n. The row space of the coefficient matrix may thus be regarded as the space of linear functionals spanned by $f_1, ..., f_m$. The solution space is the subspace annihilated by this space of functionals.

Notes

Now one may look at the system of equations from the 'dual' point of view. That is, suppose that we are given m vectors in F^n.

$$\alpha_1 = (A_{i1},\ldots,A_{in})$$

and we wish to find the annihilator of the subspace spanned by these vectors. Since a typical linear functional on F^n has the form

$$f(x_1,\ldots x_n) = c_1x_1 + \ldots + c_nx_n$$

the condition that f be in this annihilator is that

$$\sum_{j=1}^{n} A_{ij}c_j = 0, \qquad i = 1,\ldots,m$$

that is, that $(c_1,\ldots., c_n)$ be a solution of the system $AX = 0$. From this point of view, row-reduction gives us a systematic method of finding the annihilator of the subspace spanned by a given finite set of vectors in F^n.

Example 6: Here are three linear functionals on R^4:

$$\begin{aligned} f_1(x_1,x_2,x_3,x_4) &= x_1 + 2x_2 + 2x_3 + x_4 \\ f_2(x_1,x_2,x_3,x_4) &= 2x_2 + x_4 \\ f_3(x_1,x_2,x_3,x_4) &= -2x_1 - 4x_2 + 3x_4. \end{aligned}$$

The subspace which they annihilate may be found explicitly by finding the row-reduced echelon form of the matrix

$$A = \begin{bmatrix} 1 & 2 & 2 & 1 \\ 0 & 2 & 0 & 1 \\ -2 & 0 & -4 & 3 \end{bmatrix}$$

A short calculation, shows that A goes over $2R$ as

$$R = \begin{bmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Therefore, the linear functionals

$$\begin{aligned} g_1(x_1,x_2,x_3,x_4) &= x_1 + 2x_3 \\ g_2(x_1,x_2,x_3,x_4) &= x_2 \\ g_3(x_1,x_2,x_3,x_4) &= x_4 \end{aligned}$$

span the same subspace of $(R^4)^*$ and annihilate the same subspace of R^4 as do f_1, f_2, f_3. The subspace annihilated consists of the vectors with

$$\begin{aligned} x_1 &= -2x_3 \\ x_2 &= x_4 = 0 \end{aligned}$$

Notes

Example 7: Let W be the subspace of R^4 which is spanned by the vectors

$$\alpha_1 = (2,\ -2, 3, 4, -1) \qquad \alpha_3 = (0, 0, -1, -2, 3)$$

$$\alpha_2 = (-1, 1, 2, 5, 2) \qquad \alpha_4 = (1, -1, 2, 3, 0).$$

How does one describe W^0, the annihilator of W? Let us form the 4 × 5 matrix A with row vectors α_1, α_2, α_3, α_4, and find the row-reduced echelon matrix R which is row-equivalent of A:

$$A = \begin{bmatrix} 2 & -2 & 3 & 4 & -1 \\ -1 & 1 & 2 & 5 & 2 \\ 0 & 0 & -1 & -2 & 3 \\ 1 & -1 & 2 & 3 & 0 \end{bmatrix} \rightarrow R = \begin{bmatrix} 1 & -1 & 0 & -1 & 0 \\ 0 & 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

If f is a linear functional on R^5:

$$f(x_1,....,x_5) = \sum_{j=1}^{5} c_j x_j$$

then f is in W° if and only if $f(\alpha_i) = 0, i = 1, 2, 3, 4,$ i.e., if and only if

$$\sum_{j=1}^{5} A_{ij} c_j = 0, \qquad 1 \le i \le 4.$$

This is equivalent to

$$\sum_{j=1}^{5} R_{ij} c_j = 0, \qquad 1 \le i \le 3$$

or

$$c_1 - c_2 - c_4 = 0$$

$$c_3 + 2c_4 = 0$$

$$c_5 = 0$$

We obtain all such linear functionals f by assigning arbitrary values to c_2 and c_4, say $c_2 = a$ and $c_4 = b$, and then finding the corresponding $c_1 = a + b$, $c_3 = -2b$, $c_5 = 0$. So W° consists of all linear functionals f of the form

$$f(x_1, x_2, x_3, x_4, x_5) = (a+b)x_1 + ax_2 - 2bx_3 + bx_4$$

The dimension of W^* is 2 the basis (f_1, f_2) for W^* can be found by first taking $a = 1$, $b = 0$ and then $a = 0$ and $b = 1$:

$$f_1(x_1, x_2, x_3, x_4, x_5) = x_1 + x_2$$

$$f_2(x_1, x_2, x_3, x_4, x_5) = x_1 - 2x_3 + x_4$$

The above general f in W^* is $f = a\ f_1 + b\ f_2$.

Self Assessment

1. Let W be the subspace of R^5 which is spanned by the vectors

   $\alpha_1 = \varepsilon_1 + 2\varepsilon_2 + \varepsilon_3,\ \alpha_2 = \varepsilon_2 + 3\varepsilon_3 + 3\varepsilon_4 + \varepsilon_5$
   $\alpha_3 = \varepsilon_1 + 4\varepsilon_2 + 6\varepsilon_3 + 4\varepsilon_4 + \varepsilon_5$

   Find a basis for W^*.

2. Let W be the subspace spanned by R^5, which is spanned by the vectors

   $\alpha_1 = (1, 2, 0, 3, 0),\ \alpha_2 = (1, 2, -1, -1, 0)$
   $\alpha_3 = (0, 0, 1, 4, 0),\ \alpha_4 = (2, 4, 1, 10, 1)$
   $\alpha_5 = (0, 0, 0, 0, 1)$

   How does one describe W^*, the annihilator of W.

10.3 Summary

- The concept of linear functional helps us to clarify the discussion of subspaces, linear equations and co-ordinates.
- In this unit the idea of dual basis for V^* is obtained i.e. if $B = (\alpha_1, \alpha_2, ...\alpha_n)$ be the basis of V then there is a unique dual basis $\beta^* = (f_1, ...f_n)$ for V^*.
- The concept of linear functional is important in the study of finite-dimensional spaces because it helps to organize and clarify the discussion of subspaces, linear equations, and coordinates.
- Let V be the space of all polynomial functions from the field F into itself. Let t be an element of F. If we define

  $$L_t(p) = p(t)$$

  then L_t is a linear functional on V. One usually describes this by saying that, for each t, 'evaluation at t' is a linear functional on the space of polynomial functions.

10.4 Keywords

Dual Basis: In particular, if f is the zero functional $f(\alpha_j) = 0$ for each j and hence the scalars c_j are all 0. Now $f_1, ...f_n$ are n linearly independent functionals, and since we know that V^* has dimension n, it must be that $\mathcal{B}^* = \{f_1, ..., f_n\}$ is a basis for V^*. This basis is called the dual basis of $\mathcal{B}$.

Linear Functional: If V is a vector space over the field F, a linear transformation f from V into the scalar field F is also called a linear functional on V.

Trace: If A is an $n \times n$ matrix with entries in F, the trace of A is the scalar $tr\, A = A_{11} + A_{22} + ... + A_{nn}$.

10.5 Review Questions

1. In R^3, $\alpha_1 = (1, 0, 1)\ \alpha_2 = (0, 1, -2),\ \alpha_3 = (-1, -1, 0)$

   If f is a linear functional on R^3 such that

Notes

$f(\alpha_1) = 1, f(\alpha_2) = -1, f(\alpha_3) = 3$

and if $\alpha = (a, b, c)$, find $f(\alpha)$.

2. Let $\beta = (\alpha_1, \alpha_2, \alpha_3)$ be the basis for C^3 defined by

$\alpha_1 = (1, 0, -1), \alpha_2 = (1, 1, 1), \alpha_3 = (2, 2, 0).$

Find the dual basis of β.

Answers: Self Assessment

1. The dimension of *W** is 2 and the basis (f_1, f_2) for *W** is given by

$$f_1(x_1, x_2, ...x_5) = -4x_1 - 3x_2 + 2x_3 + x_4$$
$$f_2(x_1, x_2, ...x_5) = -5x_1 + 2x_2 + x_3 + x_5$$

2. The dimension of *W** is 2 and the basis (f_1, f_2) for *W** is given by

$$f_1(x_1, x_2, x_3, x_4, x_5) = -2x_1 + x_2$$
$$f_2(x_1, x_2, x_3, x_4, x_5) = -3x_1 + 4x_3 + x_4$$

10.6 Further Readings

Books

Ervin Kreyszig, *Introductory Functional Analysis with Applications*

Kenneth Hoffman and Ray Kunze, *Linear Algebra*

Richa Nandra, Lovely Professional University

Unit 11: The Double Dual

CONTENTS

Objectives

After studying this unit, you will be able to:

- Understand the meanings of $V*$ and $V**$ and their corresponding basis $\beta*$ and $\beta**$.
- Know that the mapping $\alpha \to L_\alpha$ is an isomorphism of V onto $V**$.
- See that if S is any subset of a finite dimensional vector space then $\left(S^0\right)^0$ is the subspace spanned by S.
- Understand that the T', the transpose of the linear transformation T is often called the adjoint of T; however in this unit we use only the word transpose.
- See that if A be the matrix of T relative to basis β, β' and β be the matrix of T', relative to dual basis $\beta'*$ and β^α then $B_{ij} = A_{ji}$.

Introduction

In this unit the idea of dual and double dual finite dimensional spaces and their basis vectors are explained.

Also the transpose T' of the linear transformation is introduced. The alternate name of the transpose transformation is word adjoint transformation.

11.1 The Double Dual

One question about dual bases which we did not answer in the last section was whether every basis for $V*$ is the dual of some basis for V. One way to answer that question is to consider $V**$, the dual space of $V*$.

If α is a vector in V, then α includes a linear functional L_α on V^* defined by

$$L_\alpha(f) = f(\alpha), \quad f \text{ in } V^*. \qquad \text{...(1)}$$

The fact that L_α is linear is just a reformulation of the definition of linear operations in V^*:

$$\begin{aligned} L_\alpha(cf+g) &= (cf+g)(\alpha) \\ &= (cf)(\alpha) + g(\alpha) \\ &= cf(\alpha) + g(\alpha) \\ &= cL_\alpha(f) + L_\alpha(g). \qquad \text{...(2)} \end{aligned}$$

If V is finite-dimensional and $\alpha \neq 0$, then $L_\alpha \neq 0$; in other words, there exists a linear functional f such that $f(\alpha) \neq 0$. The proof is very simple. Choose an ordered basis $\mathcal{B}=\{\alpha_1,...,\alpha_n\}$ for V such that $\alpha_1=\alpha$ and let f be the linear functional which assigns to each vector in V its first coordinate in the ordered basis $\mathcal{B}$.

Theorem 1: Let V be a finite-dimensional vector space over the field F. For each vector α in V define

$$L_\alpha(f) = f(\alpha), \quad f \text{ in } V^*.$$

The mapping $\alpha \to L_\alpha$ is then an isomorphism of V onto V^{**}.

Proof: We showed that for each α the function L_α is linear. Suppose α and β are in V and c is in F, and let $\gamma = c\alpha + \beta$. Then for each f in V^*.

$$\begin{aligned} L_\gamma(f) &= f(\gamma) \\ &= f(c\alpha+\beta) \\ &= cf(\alpha) + f(\beta) \\ &= cL_\alpha(f) + L_\beta(f) \end{aligned}$$

and so

$$L_\gamma = CL_\alpha + L_\beta$$

This shows that the mapping $\alpha \to L_\alpha$ is a linear transformation from V into V^{**}. This transformation is non-singular; for, according to the remarks above $L_\alpha = 0$ if and only if $\alpha = 0$. Now $\alpha \to L_\alpha$ is a non-singular linear transformation from V into V^{**}, and since

$$\dim V^{**} = \dim V^* = \dim V \qquad \text{...(3)}$$

Therefore this transformation is invertible, and is therefore an isomorphism of V onto V^{**}.

Corollary: Let V be a finite-dimensional vector space over the field F. If L is a linear functional on the dual space V^* of V, then there is a unique vector α in V such that

$$L(f) = f(\alpha) \qquad \text{...(4)}$$

for every f in V^*.

Corollary: Let V be a finite-dimensional vector space over the field F. Each basis for V^* is the dual of some basis for V.

Proof: Let $\mathcal{B}^*=\{f_1,...,f_n\}$ be a basis for V^*. By Theorem 2 of unit 10 there is a basis $\{L_1,...,L_n\}$ for V^{**} such that

$$L_i(f_i)=\delta_{ij}. \qquad ...(5)$$

Using the corollary above, for each i there is a vector α, in V such that

$$L_i(f) = f(\alpha_i)$$

for every f in V^*, i.e., such that $L_i = L_{\alpha i}$. It follows immediately that $\{\alpha_1,...,\alpha_n\}$ is a basis for V and that $\mathcal{B}^*$ is the dual of this basis.

In view of Theorem 1, we usually identify α with L_α and say that V 'is' the dual space of V^* or that the spaces V, V^* are naturally in duality with one another. Each is the dual space of the other. In the last corollary we have an illustration of how that can be useful. Here is a further illustration.

If E is a subset of V^*, then the annihilator E^0 is (technically) a subset of V^{**}. If we choose to identify V and V^{**} as in Theorem (1), then E^0 is a subspace of V, namely, the set of all α in V such that $f(\alpha) = 0$ for all f in E. In a corollary of Theorem 3 of unit 10 we noted that each subspace W is determined by its annihilator W°. How is it determined? The answer is that W is the subspace annihilated by all f in W°, that is, the intersection of the null spaces of all $f's$ in W°. In our present notation for annihilators, the answer may be phrased very simply: $W = (W^\circ)^\circ$.

Theorem 2: If S is any subset of a finite-dimensional vector space V, then $(S^\circ)^\circ$ is the subspace spanned by S.

Proof: Let W be the subspace spanned by S. Clearly $W^\circ = S^\circ$. Therefore, what we are to prove is that $W = W^{\circ\circ}$. We have given one proof. Here is another. By Theorem 3 of unit 10.

$$\left.\begin{aligned}\dim W + \dim W^\circ &= \dim V\\ \dim W^\circ + \dim W^{\circ\circ} &= \dim V^*\end{aligned}\right\} \qquad ...(6)$$

and since $\dim V = \dim V^*$ we have

$$\dim W = \dim W^{\circ\circ}.$$

Since W is a subspace of $W^{\circ\circ}$, we see that $W = W^{\circ\circ}$.

The results of this section hold for arbitrary vector spaces; however the proofs require the use of the so-called Axiom of Choice. Here we shall not tackle annihilators for general vector spaces. But, there are two results about linear functionals on arbitrary vector spaces which are so fundamental that we should include them.

Let V be a vector space. We want to define hyperspaces in V. Unless V is finite-dimensional, we cannot do that with the dimension of the hyperspace. But, we can express the idea that a space N falls just one dimension short of filling out V, in the following way:

1. N is a proper subspace of V;
2. If W is a subspace of V which contains N, then either $W = N$ or $W = V$.

Conditions (1) and (2) together say that N is a proper subspace and there is no larger proper subspace, in short, N is a maximal proper subspace.

Notes

Definition: If V is a vector space, a *hyperspace* in V is a maximal proper subspace of V.

Theorem 3. If f is a non-zero linear functional on the vector space V, then the null space of f is a hyperspace in V. Conversely, every hyperspace in V is the null space of a (not unique) non-zero linear functional on V.

Proof: Let f be a non-zero linear functional on V and N_f its null space. Let α be a vector in V which is not in N_f, i.e., a vector such that $f(\alpha) \neq 0$. We shall show that every vector in V is in the subspace spanned by N_f and α. That subspace consists of all vectors

$$\gamma + c\alpha, \qquad \gamma \text{ in } N_f, c \text{ in } F.$$

Let β be in V. Define

$$c = \frac{f(\beta)}{f(\alpha)}$$

which makes sense because $f(\alpha) \neq 0$. Then the vector $\gamma = \beta - c\alpha$ is in N_f since

$$\begin{aligned} f(\gamma) &= f(\beta - c\alpha) \\ &= f(\beta) - cf(\alpha) \\ &= 0. \end{aligned} \qquad ...(7)$$

So β is in the subspace spanned by N_f and α.

Now let N be a hyperspace in V. Fix some vector α which is not in N. Since N is a maximal proper subspace, the subspace spanned by N and α is the entire space V. Therefore each vector β in V has the form

$$\beta = \gamma + c\alpha, \qquad \gamma \text{ in } N, c \text{ in } F.$$

The vector γ and the scalar c are uniquely determined by β. If we have also

$$\beta = \gamma' + c'\alpha, \qquad \gamma' \text{ in } N, c' \text{ in } F. \qquad ...(8)$$

then $\quad (c' - c)\alpha = \gamma - \gamma'$

If $c' - c \neq 0$, then α would be in N; hence, $c' = c$ and $\gamma' = \gamma$. Another way to phrase our conclusion is this: If β is in V, there is a unique scalar c such that $\beta - c\alpha$ is in N. Call that scalar $g(\beta)$. It is easy to see that g is a linear functional on V and that N is the null space of g.

Lemma: If f and g are linear functionals on a vector space V, then g is a scalar multiple of f if and only if the null space of g contains the null space of f, that is, if and only if $f(\alpha) = 0$ implies $g(\alpha) = 0$.

Proof: If $f = 0$ then $g = 0$ as well and g is trivially a scalar multiple of f. Suppose $f \neq 0$ so that the null space N_f is a hyperspace in V. Choose some vector α in V with $f(\alpha) \neq 0$ and let

$$c = \frac{g(\alpha)}{f(\alpha)}. \qquad ...(9)$$

The linear functional $h = g - cf$ is 0 on N_f since both f and g are 0 there, and $h(\alpha) = g(\alpha) - cf(\alpha) = 0$.

Thus h is 0 on the subspace spanned by N_f and α – and that subspace is V. We conclude that $h = 0$, i.e. that $g = cf$.

Theorem 4: Let $g, f_1, ..., f_r$ be linear functionals on a vector space V with respective null space $N, N_1, ..., N_r$. Then g is a linear combination of $f_1, ..., f_r$ if and only if N contains the intersection $N_1 \cap ... \cap N_r$.

Proof: If $g = c_1 f_1 + ... + c_r f_r$ and $f_i(\alpha) = 0$ for each i, then clearly $g(\alpha) = 0$. Therefore, N contains $N_1 \cap ... \cap N_r$.

We shall prove the converse (the 'if' half of the theorem) by induction on the number r. The preceding lemma handles the case $r = 1$. Suppose we know the result for $r = k - 1$, and let $f_1, ..., f_r$ be linear functionals with null spaces $N_1, ..., N_k$ such that $N_1 \cap ... \cap N_k$ is contained in N, the pull space of g. Let $g', f_1', ..., f_{k-1}'$ be the restrictions of $g, f_1, ..., f_{k-1}$ to the subspace N_k. Then $g', f_1', ..., f_{k-1}'$ are linear functionals on the vector space N_k. Furthermore, if α is a vector in N_k and $f_i'(\alpha) = 0$, $i = 1, ..., k-1$, then α is in $N_1 \cap ... \cap N_k$ and so $g'(\alpha) = 0$. By the induction hypothesis (the case $r = k - 1$), there are scalars c_i such that

$$g' = c_1 f_1' + ... + c_{k-1} f_{k-1}'$$

Now let

$$h = g - \sum_{i=1}^{k-1} c_i f_i. \qquad ...(10)$$

Then h is a linear functional on V and (10) tells us that $h(\alpha) = 0$ for every α in N_k. By the preceding leema, h is a scalar multiple of f_k. If $h = c_k f_k$, then

$$g = \sum_{i=1}^{k} c_i f_i.$$

Self Assessment

1. Let n be a positive integer and F a field. Let W be the set of all vectors $(x_1, ..., x_n)$ in F^n such that $x_1 + ... + x_n = 0$.

   (a) Prove that W^0 consists of all linear functionals f of the form

   $$f(x_1, ..., x_n) = c \sum_{j=1}^{n} x_j.$$

(b) Show that the dual space W^* of W can be 'naturally' identified with the linear functionals

$$f(x_1,...,x_n) = c_1x_1 + ... + c_nx_n$$

on F^n which satisfy $c_1 + ... + c_n = 0$.

2. Use Theorem 4 to prove the following. If W is a subspace of a finite-dimensional vector space V and if $\{g_1,...,g_r\}$ is any basis for W°, then

$$W = \bigcap_{i=1}^{r} N_g.$$

11.2 The Transpose of a Linear Transformation

Suppose that we have two vector spaces over the field F, V and W, and a linear transformation T from V into W. Then T induces a linear transformation from W^* into V^*, as follows. Suppose g is a linear functional on W, and let

$$f(\alpha) = g(T\alpha) \qquad ...(11)$$

for each α in V. Then (11) defines a function f from V into F, namely the composition of T, a function from V into W, with g, a function from W into F. Since both T and g are linear, Theorem 5 of unit 7 tells us that f is also linear, i.e., f is a linear functional on V. Thus T provides us with a rule T^t which associates with each linear functional g on W a linear functional $f = T^tg$ on V, defined by (11). Note also that T^t is actually a linear transformation from W^* into V^*; for, if g_1 and g_2 are in W^* and c is a scalar

$$\begin{aligned}\left[T^t(cg_1 + g_2)\right](\alpha) &= (cg_1 + g_2)(T\alpha)\\ &= cg_1(T\alpha) + g_2(T\alpha)\\ &= c(T^tg_1)(\alpha) + (T^tg_2)(\alpha) \qquad ...(12)\end{aligned}$$

so that $T^t(cg_1 + g_2) = cT^tg_1 + T^tg_2$. Let us summarize.

Theorem 5: Let V and W be vector spaces over the field F. For each linear transformation T from V into W, there is a unique linear transformation T^t from W^* into V^* such that

$$(T^tg)(\alpha) = g(T\alpha) \qquad ...(13)$$

for every g in W^* and α in V.

We shall call T^t the transpose of T. This transformation T^t is often called the adjoint of T; however, we shall not use this terminology.

Theorem 6: Let V and W be vector spaces over the field F, and let T be a linear transformation from V into W. The null space of T^t is the annihilator of the range of T. If V and W are finite-dimensional, then

(i) rank T^t = rank (T)

(ii) the range of T^t is the annihilator of the null space T. ...(14)

Proof: If g is in W^*, then by definition

$$\left(T^t g\right)(\alpha) = g(T\alpha)$$

for each α in V. The statement that g is in the null space of T^t means that $g(T\alpha) = 0$ for every α in V. Thus the null space of T^t is precisely the annihilator of the range of T.

Suppose that V and W are finite-dimensional, say dim $V = n$ and dim $W = m$. For (i): Let r be the rank of T, i.e., the dimension of the range of T. By Theorem 3 of unit 10, the annihilator of the range of T then has dimension $(m - r)$. By the first statement of this theorem, the nullity of T^t must be $(m - r)$. But then since T^t is a linear transformation on an m-dimensional space, the rank of T^t is $m - (m - r) = r$, and so T and T^t have the same rank. For (ii): Let N be the null space of T. Every functional in the range of T^t is in the annihilator of N; for suppose $f = T^t g$ for some g in W^*; then, if α is in N

$$f(\alpha) = \left(T^t g\right)(\alpha) = g(T\alpha) = g(0) = 0.$$

Now the range of T^t is a subspace of the space N^0, and

$$\dim N^0 = n - \dim N = \text{rank}(\text{T}) = \text{rank}\ (T^t) \qquad \text{...(15)}$$

so that the range of T^t must be exactly N^0.

Theorem 7: Let V and W be finite-dimensional vector spaces over the field F. Let $\mathcal{B}$ be an ordered basis for V with dual basis $\mathcal{B}^*$, and let $\mathcal{B}'$ be an ordered basis for W with dual basis $\mathcal{B}'^*$. Let T be a linear transformation from V into W; let A be the matrix of T relative to $\mathcal{B}, \mathcal{B}'$ and let B be the matrix of T^t relative to $\mathcal{B}'^*, \mathcal{B}^*$. Then $B_{ij} = A_{ji}$.

Proof: Let

$$\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}, \mathcal{B}' = \{\beta_1, \ldots, \beta_m\},$$
$$\mathcal{B}^* = \{f_1, \ldots, f_n\}, \mathcal{B}'^* = \{g_1, \ldots, g_m\}.$$

By definition,

$$\left.\begin{aligned} T\alpha_j &= \sum_{i=1}^{m} A_{ij}\beta_i i \quad j = 1, \ldots, n \\ T^t g_j &= \sum_{i=1}^{n} B_{ij} f_i j \quad j = 1, \ldots, m \end{aligned}\right] \qquad \text{...(16)}$$

On the other hand,

$$\begin{aligned} \left(T^t g_j\right)(\alpha_i) &= g_j(T\alpha_i) \\ &= g_j\left(\sum_{k=1}^{m} A_{ki}\beta_k\right) \end{aligned}$$

$$= \sum_{k=1}^{m} A_{ki} g_j(\beta_k)$$

$$= \sum_{k=1}^{m} A_{ki} \delta_{jk}$$

$$= A_{ji}.$$

For any linear functional *f* on *V*

$$f = \sum_{i=1}^{m} f(\alpha_i) f_i. \qquad ...(17)$$

If we apply this formula to the functional $f = T^t g_j$ and use the fact that $\left(T^t g_j\right)(\alpha_i) = A_{ji}$, we have

$$T^t g_j = \sum_{i=1}^{n} A_{ji} f_i. \qquad ...(18)$$

from which it immediately follow that $B_{ij} = A_{ji}$.

Definition: If *A* is an *m* × *n* matrix over the field *F*, the transpose of *A* is *n* × *m* matrix A^t defined by $A^t_{ij} = A_{ji}$.

Theorem 8: Thus states that if *T* is a linear transformation from *V* into *W*, the matrix of which in some pair of bases is *A*, then the transpose transformation T^t is represented in the dual pair of bases by the transpose matrix A^t.

Theorem 9: Let *A* be any *m* × *n* matrix over the field *F*. Then the row rank of *A* is equal to the column rank of *A*.

Proof: Let $\mathcal{B}$ be the standard ordered basis for F^n and $\mathcal{B}'$ the standard ordered basis for F^m. Let *T* be the linear transformation from F^n into F^m such that the matrix of *T* relative to the pair $\mathcal{B}, \mathcal{B}'$ is A, i.e.,

$$\left.\begin{aligned} . \qquad & T(x_1, ..., x_n) = (y_1, ..., y_m) \\ \text{where} \qquad & y_i = \sum_{j=1}^{n} A_{ij} x_j. \end{aligned}\right] \qquad ...(19)$$

The column rank of *A* is the rank of transformation *T*, because the range of *T* consists of all *m*-tuples which are linear combinations of the column vectors of *A*.

Relative to the dual bases $\mathcal{B}'^*$ and $\mathcal{B}^*$, the transpose mapping T^t is represented by the matrix A^t. Since the columns of A^t are the rows of *A*, we see by the same reasoning that the row rank of *A* (the column rank of A^t) is equal to the rank of T^t. By Theorem 7, *T* and T^t have the same rank, and hence the row rank of *A* is equal to the column rank of *A*.

Now we see that if A is an $m \times n$ matrix over F and T is the linear transformation from F^n into F^m defined above, then

$$\text{rank } (T) = \text{row rank } (A) = \text{column rank } (A) \qquad \text{...(20)}$$

and we shall call this number simply the rank of A.

Example 1: This example will be of a general nature – more discussion than example. Let V be an n-dimensional vector space over the field F, and let T be a linear operator on V. Suppose $\mathcal{B}=\{\alpha_1,...,\alpha_n\}$ is an ordered basis for V. The matrix of T in the ordered basis $\mathcal{B}$ is defined to be the $n \times n$ matrix A such that

$$T\alpha_j = \sum_{j=1}^{n} A_{ij}\alpha_i \qquad \text{...(21)}$$

in other words, A_{ij} is the ith coordinate of the vector $T\alpha_j$ in the ordered basis $\mathcal{B}$. If $\{f_1,...,f_n\}$ is the dual basis of $\mathcal{B}$, this can be stated simply

$$A_{ij} = f_i\left(T\alpha_j\right) \qquad \text{...(22)}$$

Let us see what happens when we change basis. Suppose

$$\mathcal{B}'=\{\alpha_1',...,\alpha_n'\}$$

is another ordered basis for V, with dual basis $\{f_1',...,f_n'\}$. If B is the matrix of T in the ordered basis $\mathcal{B}'$, then

$$B_{ij} = f_i'\left(T\alpha_j'\right). \qquad \text{...(23)}$$

Let U be the invertible linear operator such that $U\alpha_j = \alpha_j'$. Then the transpose of U is given by $U^t f_t' = f_i$. It is easy to verify that since U is invertible, so is U^t and $(U^t)^{-1} = (U^{-1})^t$. Thus $f_t' = \left(U^{-1}\right)^t f_i, i = 1,...,n$. Therefore,

$$\begin{aligned} B_{ij} &= \left[\left(U^{-1}\right)^t f_i\right]\left(T\alpha_j'\right) \\ &= f_i\left(U^{-1}T\alpha_j'\right) \\ &= f_i\left(U^{-1}TU\alpha_j\right). \qquad \text{...(24)} \end{aligned}$$

Now what does this say? Well, $f_i\left(U^{-1}TU\alpha_j\right)$ is the *i, j* entry of the matrix of $U^{-1}TU$ in the ordered basis $\mathcal{B}$. Our computation above shows that this scalar is also the *i, j* entry of the matrix of T in the ordered basis $\mathcal{B}'$. In other words

$$\left.\begin{aligned} [T]_{\mathcal{B}'} &= \left[\mathrm{U}^{-1}TU\right]_{\mathcal{B}} \\ &= \left[\mathrm{U}^{-1}\right]_{\mathcal{B}}[T]_{\mathcal{B}}[U]_{\mathcal{B}} \\ &= [\mathrm{U}]_{\mathcal{B}}^{-1}[T]_{\mathcal{B}}[U]_{\mathcal{B}} \end{aligned}\right] \quad \text{...(25)}$$

and this is precisely the change-of-basis formula which we derived earlier.

Self Assessment

3. Let V be a finite dimensional vector space over the field F and let T be a linear operator on V. Let C be a scalar and suppose there is a non-zero vector α in V such that $T\alpha = c\alpha$. Prove that there is a non-zero linear functional F on V such that $T'f = cf$.

4. For all A, B matrices in F^m, prove that–

   (a) $(A')' = A$

   (b) $(A+B)' = A' + B'$

   (c) $(AB)' = B'A'$

11.3 Summary

- A vector α induces a linear functional α_α in V^* and the mapping $\alpha \to L_\alpha$ is an isomorphism of V and V^{**}.
- If T is the linear transformation from V into W then it also induces a transformation from W^* into V^* through its transpose.
- The alternate name of the transpose transformation is word adjoint transformation.

11.4 Keywords

Adjoint: T^t is the transpose of T. This transformation T^t is often called the adjoint of T.

Transpose: If A is an $m \times n$ matrix over the field F, the transpose of A is $n \times m$ matrix A^t defined by $A^t_{ij} = A_{ji}$.

11.5 Review Questions

1. Let S be a set, F a field and $V(S,F)$ the space of all functions from S into F:

   $$\begin{aligned}(f+g)(x) &= f(x)+g(x)\\ (cf)(x) &= cf(x).\end{aligned}$$

   Let W be any n-dimensional space of $V(S,F)$. Show that there exists points $x_1,\ldots,x_n$ in S and functions $f_1, f_2, \ldots, f_n$ in W such that $f_i\left(x_j\right) = S_{ij}$.

2. Let F be a field and let f be the linear functional on F^2 defined by $f(x_1, x_2) = ax_1 + bx_2$. For each of the following operations T, let $g = T'f$ and find $g(x_1, x_2)$

(a) $T(x_1, x_2) = (-x_2, x_1)$;

(b) $T(x_1, x_2) = (x_1 - x_2, x_1 + x_2)$.

11.6 Further Readings

Books

Kenneth Hoffman and Ray Kunze, *Linear Algebra*

I N Herstein, *Topics in Algebra.*

Unit 12: Introduction and Characteristic Values of Elementary Canonical Forms

CONTENTS

Objectives

After studying this unit, you will be able to:

- Know that when the matrix of the linear transformation is in the diagonal form for some ordered basis the properties of the transformation can be seen at a glance.
- See that a matrix *A* of a linear operator *T* can be cast into a diagonal form under similarity transformations.
- See that a matrix *A* and $P^{-1}AP$ where *P* is an invertible have the same characteristic values.

Introduction

In this unit it is shown how a matrix has a diagonal form.

For this purpose the characteristic values and characteristic vectors are worked out and an invertible matrix is worked out of the characteristic vectors that can diagonalize the given matrix.

12.1 Overview

One of our primary aim in these units is to study linear transformation on finite dimensional vector spaces. So far we have studied many specific properties of linear transformations. In terms of ordered basis vectors we have represented such types of matrices by matrices. In terms of matrices we see lots of insight of the linear transformation. We also explored the linear algebra $L(V, V)$ consisting of the linear transformations of a space into itself.

In the next few units we shall concentrate ourselves with linear operators on a finite dimensional vector space. If we consider the ordered basis $\beta = (\alpha_1, \alpha_2, ... \alpha_n)$ then the effect of T on α_i is

$$T\alpha_i = \sum_{j=1}^{n} A_{ji}\beta_j \qquad i = 1, 2, ... n$$

where the new ordered basis is $\beta' = (\beta_1, \beta_2, ... \beta_n)$. If we now choose the basis $\beta = (\alpha_1, \alpha_2, ... \alpha_n)$ in such a way that

$$T\alpha_i = c_i\alpha_i \qquad \ldots (1)$$

for i = 1 to n then the matrix of the linear transformation is given by

$$D = \begin{bmatrix} C_1 & 0 & 0 & 0 & 0 & \ldots & 0 \\ 0 & C_2 & 0 & 0 & 0 & \ldots & 0 \\ 0 & 0 & C_3 & 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & \ldots & \ldots & C_n \end{bmatrix} \qquad \ldots(2)$$

with the help of equation (2) we would gain considerable information about T. Simple numbers associated with T, such as the rank of T or the determinant of T, would be determined with little more than a glance. The range of T would be the subspace spanned by those α_i's for which $c_i \neq 0$, and the null space would be spanned by the remaining α_i's. Indeed, it seems fair to say that, if we knew a basis β and a diagonal matrix D such that $[T] = D$, we could answer readily any question about T which might arise.

In the following we are interested in finding out if a linear operator can be represented by a diagonal matrix. How can we find the basis for such type of linear operator and what are the values of c_i's.

12.2 Characteristic Values

Guided by the equation (1) we should study vectors which on application of linear operator T transformed into the scalar multiples of themselves.

Let V be a vector space over the field F and T be a linear operator on V. A characteristic value of T is a scalar C in F such that there is a non-zero vector α in V with $T\alpha = c\alpha$. If c is a characteristic value of T, then

(a) Any α such that $T\alpha = c\alpha$, is called characteristic vector of T.

(b) The collection of all α such that $T\alpha = c\alpha$, is called the characteristic space associated with c.

If T is any linear operator and c is any scalar, the set of vectors α, such that $T\alpha = c\alpha$ is a sub-space of V. It is null space of linear transformation $(T - cI)$. We call c a characteristic value of T if this subspace is different from the zero subspace, i.e., if $(T - cI)$ fails to be 1:1. If the underlying space V is finite-dimensional, $(T - cI)$ fails to be 1:1 precisely when its determinant is different from 0.

Theorem 1: Let T be a linear operator on a finite-dimensional space V and let c be a scalar. The following are equivalent:

(i) c is a characteristic value of T.

(ii) The operator $(T - cI)$ is singular (not invertible)

(iii) det $(T - cI) = 0$.

The determinant criterion (iii) is very important because it tells us where to look for the characteristic values of T. Since det $(T - cI)$ is a polynomial of degree n in the variable c, we will find the characteristic values as the roots of that polynomial.

If $\mathcal{B}$ is any ordered basis of V and $A = [T]_\beta$, then $(T - cI)$ is invertible if and only if the matrix $(A - cI)$ is invertible. Accordingly, we make the following definition.

Definition: If A is an $n \times n$ matrix over the field F, a characteristic value of A in F is a scalar c in F such that the matrix $(A - cI)$ is singular (not invertible).

Notes

Since c is a characteristic value of A if and only if det $(A - cI) = 0$, we form the matrix $(xI - A)$ with polynomial entries, and consider the polynomial f = det $(xI - A)$. Clearly the characteristic values of A in F are just the scalars c in F such that $f(c) = 0$. For this reason f is called the characteristic polynomial of A. It is important to note that f is a monic polynomial which has degree exactly n. This is easily seen from the formula for the determinant of a matrix in terms of its entries.

Lemma: Similar matrices have the same characteristic polynomial.

Proof: If $B = P^{-1} AP$, then

$$\begin{aligned} \det(xI - B) &= \det(xI - P^{-1}PA) \\ &= \det(P^{-1}(xI - A)P) \\ &= \det P^{-1}.\ \det(xI - A)\ .\ \det P \\ &= \det(xI - A) \end{aligned}$$

This lemma enables us to define sensibly the characteristic polynomial of the operator T as the characteristic polynomial of any $n \times n$ matrix which represents T in some ordered basis for V. Just as for matrices, the characteristic values of T will be the roots of the characteristic polynomial for T. In particular, this shows us that T cannot have more than n distinct characteristic values. It is important to point out that T may not have any characteristic values.

Example 1: Let T be the linear operator on R^2 which is represented in the standard ordered basis by the matrix

$$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

The characteristic polynomial for T (or for A) is

$$\det(xI - A) = \begin{bmatrix} x & 1 \\ -1 & x \end{bmatrix} = x^2 + 1.$$

Since this polynomial has no real roots, T has no characteristic values. If U is the linear operator on C^2 which is represented by A in the standard ordered basis, then U has two characteristic value, i and $-i$. Here we see a subtle point. In discussing the characteristic values of a matrix A, we must be careful to stipulate the field involved. The matrix A above has no characteristic values in R, but has the two characteristic values, i and $-i$ in C.

Example 2: Let A be the (real) 3×3 matrix

$$\begin{bmatrix} 3 & 1 & -1 \\ 2 & 2 & -1 \\ 2 & 2 & 0 \end{bmatrix}$$

Then the characteristic polynomial for A is

$$\begin{bmatrix} x-3 & -1 & 1 \\ -2 & x-2 & 1 \\ -2 & -2 & x \end{bmatrix} = x^3 - 5x^2 + 8x - 4 = (x-1)(x-2)^2.$$

Thus the characteristic values of A are 1 and 2.

Suppose that T is the linear operator on R^3 which is represented by A in the standard basis. Let us find the characteristic vectors of T associated with the characteristic values, 1 and 2. Now

$$A - I = \begin{bmatrix} 2 & 1 & -1 \\ 2 & 1 & -1 \\ 2 & 2 & -1 \end{bmatrix}$$

It is obvious at a glance that $A–I$ has rank equal to 2 (and hence $T - I$ has nullity equal to 1). So the space of characteristic vectors associated with the characteristic value 1 is one-dimensional. The vector $\alpha_1 = (1, 0, 2)$ spans the null space of $T - I$. Thus $T\alpha = \alpha$ if and only if α is a scalar multiple of α_1. Now consider

$$A - 2I = \begin{bmatrix} 1 & 1 & -1 \\ 2 & 0 & -1 \\ 2 & 2 & -2 \end{bmatrix}$$

Evidently $A - 2I$ also has rank 2, so that the space of characteristic vectors associated with value 2 has dimension 1. $T\alpha = 2\alpha$ is possible if α is a scalar multiple of $\alpha_2 = (1, 1, 2)$.

Example 3: Find the characteristic values and associated characteristic vector for the matrix

$$A = \begin{bmatrix} 8 & -6 & 2 \\ -6 & 7 & -4 \\ 2 & -4 & 3 \end{bmatrix}$$

Solution: We know that the characteristic equation is $|A - \lambda I| = 0$, *i.e.,*

$$\begin{bmatrix} 8-\lambda & -6 & 2 \\ -6 & 7-\lambda & -4 \\ 2 & -4 & 3-\lambda \end{bmatrix} = 0$$

or $\quad \{(8 - \lambda)\} (7 - \lambda) (3 - \lambda) - 16\} + 6\{3 - \lambda\} (-6) + 8\} + 2 \{24 - 2(7 - \lambda)\} = 0$

or $\quad -\lambda^3 + 18\lambda^2 - 45\lambda = 0$

or $\quad \lambda(\lambda^2 + 18\lambda + 45) = 0$

or $\quad \lambda(\lambda - 3) (\lambda - 15) = 0$

$\quad \therefore \lambda = 0, 3, 15.$

Hence the characteristic roots are $\lambda_1 = 0, \lambda_2 = 3, \lambda_3 = 15$. The characteristic vector associated with is $\lambda_1 = 0$ is given by

$$\begin{bmatrix} 8 & -6 & 2 \\ -6 & 7 & -4 \\ 2 & -4 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

This gives $8x_1 - 6x_2 + 2x_3 = 0$

$-6x_1 + 7x_2 - 4x_3 = 0$

$2x_1 - 4x_2 + 3x_3 = 0$

On solving these equations, we get

$$\frac{x_1}{1} = \frac{x_2}{2} = \frac{x_3}{2} = k_1 \text{ (say)}$$

Hence the required characteristic vector corresponding to the characteristic root λ_1, = 0, is

$$X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} k_1 \\ 2k_1 \\ 2k_1 \end{bmatrix}$$

The characteristic vector corresponding to the root $\lambda_2 = 3$ is given by

$$\begin{bmatrix} 8-3 & -6 & 2 \\ -6 & 7-3 & -4 \\ 2 & -4 & 3-3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

or $$\begin{bmatrix} 5 & -6 & 2 \\ -6 & 4 & -4 \\ 2 & -4 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

This gives $5x_1 - 6x_2 + 2x_3 = 0$

$6x_1 + 4x_2 - 4x_3 = 0$

$2x_1 - 4x_2 = 0$

On solving these equations, we get

$$\frac{x_1}{2} = \frac{x_2}{1} = \frac{x_3}{-2} = k_2 \text{ (say) } k_2 \neq 0$$

Thus $x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2k_2 \\ k_2 \\ -2k_2 \end{bmatrix}$ is the required characteristic vector for $\lambda = 3$.

Similarly, for $\lambda = 15$, the characteristic vector will be

$$\begin{bmatrix} 8-15 & -6 & 2 \\ -6 & 7-15 & -4 \\ 2 & -43 & 3-15 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

or $$\begin{bmatrix} -7 & -6 & 2 \\ -6 & -8 & -4 \\ 2 & -4 & -12 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

which give $7x_1 + 6x_2 - 2x_3 = 0$

$3x_1 + 4x_2 + 2x_3 = 0$

$x_1 - 2x_2 - 6x_3 = 0$

On solving these, we get

$$\frac{x_1}{2} = \frac{x_2}{-2} = \frac{x_3}{1} = k_3 \text{ (say) } k_3 \neq 0$$

Hence, latent vector corresponding to the latent root, $\lambda_3 = 15$ will be

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2k_3 \\ -2k_3 \\ k_3 \end{bmatrix}$$

Example 4: If $a + b + c = 0$, find the characteristic values of the matrix

$$\begin{bmatrix} a & c & b \\ c & b & a \\ b & a & c \end{bmatrix}$$

Solution: We have the characteristic equation of A

$$|A - \lambda I| = 0$$

or
$$\begin{bmatrix} a-\lambda & c & b \\ c & b-\lambda & a \\ b & a & c-\lambda \end{bmatrix} = \begin{bmatrix} a+b+c-\lambda & c & b \\ a+b+c-\lambda & b-\lambda & a \\ a+b+c-\lambda & a & c-\lambda \end{bmatrix}$$

On replacing C_1 by $C_1 + C_2 + C_3$.

$$= \begin{bmatrix} -\lambda & c & b \\ -\lambda & b-\lambda & a \\ -\lambda & a & c-\lambda \end{bmatrix} \qquad [\because\ a + b + c = 0]$$

$$= \begin{bmatrix} -\lambda & c & b \\ 0 & b-\lambda-c & c-b \\ 0 & a-c & c-\lambda-b \end{bmatrix}$$

On operating $R_2 - R_1$ and $R_3 - R_1$

$$= \lambda\ [(a^2 + b^2 + c^2 - ab - bc - ca) - \lambda^2]$$

But $\quad a + b + c = 0$, *i.e.*, $(a + b + c)^2 = 0$

or $\quad a^2 + b^2 + c^2 + 2ab + 2bc + 2ca = 0$

or $\quad -(ab + bc + ca) = \frac{1}{2}(a^2 + b^2 + c^2)$

$\because$ $\quad$ Characteristic equation becomes

$$\lambda\left[(a^2 + b^2 + c^2 + \frac{1}{2}(a^2 + b^2 + c^2) - \lambda^2\right] = 0$$

or
$$\lambda\left[\frac{3}{2}(a^2 + b^2 + c^2) - \lambda^2\right] = 0$$

which gives $\lambda = 0$ or $\lambda = \pm\left[\frac{3}{2}(a^2 + b^2 + c^2)\right]^{1/2}$

Example 5: If A be a square matrix, show that the characteristic values of the matrix A are the same as those of its transpose A'.

Solution: The characteristic equation of the square matrix A is given by

$$|A - \lambda I| = 0$$

Similarly the characteristic equation of the matrix

$$A' \text{ is } (A' - \lambda I| = 0$$

Now, we have to prove that the characteristic roots of $|A - \lambda I| = 0$ and $|A' - \lambda I| = 0$ identical.

Since interchange of row and column does not change the value of the determinant, hence we have

$$|A - \lambda I| = |A' - \lambda I|$$

Hence the roots of the equations $|A - \lambda I| = 0$ and $|A' - \lambda I| = 0$ are same.

Lemma: If $\lambda \in F$ is a characteristic value of T, then for any polynomial $q(x) \in F(x)$, $q(\lambda)$ is a characteristic value of $q(T)$.

Proof: Suppose $\lambda \in F$ and $T\alpha = \lambda\alpha$ for non-zero vector α in V. Now $T^2\lambda = T(T\alpha) = T(\lambda\alpha) = \lambda T\alpha = \lambda^2\alpha$, continuing in this way we obtain $T^3\alpha = \lambda^3\alpha$, $T^4\alpha$, $T^4\alpha = \lambda^4\alpha$, ... $T^k\alpha = \lambda^k\alpha$, for all positive integers k. If

$$q(x) = a_0 x^m = a_1 x^{m-1} + \ldots + a_m \in F, \text{ then}$$

$$q(T) = a_0 T^m = a_1 T^{m-1} + \ldots + a_m,$$

hence $\quad q(T)\alpha = a_0\lambda^m\alpha + a_1\lambda^{m-1}\alpha + \ldots + a_m\alpha$

$$= q(\lambda)d.$$

Thus $\quad [q(T) - q(\lambda)I]\alpha = 0$, since $\alpha \neq 0$ so $q(\lambda)$ is characteristic value of $q(T)$.

Definition: Let T be a linear operator on the finite dimensional space V. We say that T is diagonalizable if there is a basis for V each vector of which is a characteristic vector of T.

The reason for the name should be apparent; for, if there is an ordered basis $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ for V in which each α_i is a characteristic vector of T, then the matrix of T in the ordered basis $\mathcal{B}$ is diagonal. If $T\alpha_i = c_i\alpha_i$, then

$$[T]_{\mathcal{B}} = \begin{bmatrix} c_1 & 0 & \ldots & 0 \\ 0 & c_2 & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & c_n \end{bmatrix}$$

We certainly do not require that the scalars $c_1, \ldots c_n$ be distinct; indeed, they may all be the same scalar (when T is a scalar multiple of the identity operator).

One could also define T to be diagonalizable when the characteristic vectors of T span V. This is only superficially different from our definition, since we can select a basis out of any spanning set of vectors.

For Examples 1 and 2 we purposely chose linear operators T on R^n which are not diagonalizable. In Example 1, we have a linear operator on R^2 which is not diagonalizable, because it has no characteristic values. In Example 2, the operator T has characteristic values; in fact, the characteristic polynomial for T factors completely over the real number field: $f = (x - 1)(x - 2)^2$. Nevertheless T fails to be diagonalizable. There is only a one-dimensional space of characteristic vectors associated with each of the two characteristic values of T. Hence, we cannot possibly form a basis for R^3 which consists of characteristic vectors of T.

Suppose that T is a diagonalizable linear operator. Let $c_1, \ldots c_k$ be the distinct characteristic values of T. Then there is an ordered basis $\mathcal{B}$ in which T is represented by a diagonal matrix which has for its diagonal entries the scalars c_i, each repeated a certain number of times. If c_i is repeated d_i times, then (we may arrange that) the matrix has the block form

$$[T]_{\mathcal{B}} = \begin{bmatrix} c_1I_1 & 0 & \dots & 0 \\ 0 & c_2I_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & c_kI_k \end{bmatrix}$$

where I_j is the $d_j \times d_j$ identity matrix. From that matrix we see two things. First, the characteristic polynomial for T is the product of (possibly repeated) linear factors:

$$f = (x - c_1)^{d_1} \dots (x - c_k)^{d_k}$$

If the scalar field F is algebraically closed, e.g., the field of complex numbers, every polynomial over F can be so factored; however, if F is not algebraically closed, we are citing a special property of T when we say that its characteristic polynomial has such a factorization. The second thing we see that d_i, the number of times which c_i is repeated as root of f, is equal to the dimension of the space of characteristic vectors associated with the characteristic value c_i. That is because the nullity of a diagonal matrix is equal to the number of zeros which it has on its main diagonal, and the matrix $[T - c_iI]_{\mathcal{B}}$ has d_i zeros on its main diagonal. This relation between the dimension of the characteristic space and the multiplicity of the characteristic value as a root of f does not seem exciting at first; however, it will provide us with a simpler way of determining whether a given operator is diagonalizable.

Lemma: Let T be a linear operator on the finite dimensional space V. Let $c_1, \dots, c_k$ be the distinct characteristic values of T and let W_i be the space of characteristic vectors associated with the characteristic value c_i. If $W = W_i + \dots + W_k$, then

$$\dim W = \dim W_1 + \dots + \dim W_k.$$

In fact if B_i is an ordered basis for W_i, then $\mathcal{B} = (\mathcal{B}_1, \dots, \mathcal{B}_k)$ is an ordered basis for W.

Proof: The space $W = W_i + \dots + W_k$ is the subspace spanned by all of the characteristic vectors of T. Usually when one forms the sum W of subspaces W_i, one expects that $\dim W < \dim W_i + \dots + \dim W_k$ because of linear relations which may exist between vectors in the various spaces. This lemma states that the characteristic spaces associated with different characteristic values are independent of one another.

Suppose that (for each i) we have a vector β_i in W_i, and assume that $\beta_i + \dots + \beta_k = 0$. We shall show that $\beta_i = 0$ for each i. Let f be any polynomial. Since $T\beta_i = c_i\beta_i$, the preceding lemma tells us that

$$\begin{aligned} 0 = f(T)0 &= f(T)\beta_1 + \dots + f(T)\beta_k \\ &= f(c_1)\beta_1 + \dots + f(c_k)\beta_k \end{aligned}$$

Choose polynomial $f_1, \dots, f_k$ such that

$$f_i(c_j) = \delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j. \end{cases}$$

Then

$$\begin{aligned} 0 = f_i(T) &= \sum_j \delta_{ij}\beta_j \\ &= \beta_i. \end{aligned}$$

Now, let β_i be an ordered basis for W_i, and let β be the sequence $\beta = (\beta_1, \dots, \beta_k)$. Then β spans the subspace $W = W_1 + \dots + W_k$. Also, β is a linearly independent sequence of vectors, for the following reason. Any linear relation between the vectors in β will have the form $\beta_1 + \dots + \beta_k = 0$, where β_i

is some linear combination of the vectors in β_i. From what we just did, we know that $\beta_i = 0$ for each i. Since each β_i is linearly independent, we see that we have only the trivial linear relation between the vectors in β.

Theorem 2: Let T be a linear operator on a finite-dimensional space V. Let $c_1, ..., c_k$ be the distinct characteristic values of T and let W_i be the null space of $(T - c_iI)$. The following are equivalent:

(i) T is diagonalizable

(ii) The characteristic polynomial for T is

$$f = (x - c_i)^{di} \ldots (x - c_k)^{dk}$$

and dim $W_i = di, i = 1, \ldots k$.

(iii) dim W_i + ... + dim W_k = dim V.

Proof: We have observed that (i) implies (ii). If the characteristic polynomial f is the product of linear factors, as in (ii), then $d_i + .. + d_k$ = dim V. For, the sum of the d_i's is the degree of the characteristic polynomial, and that degree is dim V. Therefore (ii) implies (iii). Suppose (iii) holds. By the lemma, we must have $V = W_1 + \ldots + W_k$, i.e., the characteristic vectors of T span V.

The matrix analogue of Theorem 2 may be formulated as follows. Let A be an $n \times n$ matrix with entries in a field F, and let $c_1, \ldots c_k$ be the distinct characteristic values of A in F. For each i, let W_i be the space of column matrices X (with entries in F) such that

$$(A - c_iI)X = 0,$$

and let β_i be an ordered basis for W_i. The bases $\beta_1, \ldots, \beta_k$ collectively string together to form the sequence of columns of a matrix P:

$$P = [P_1, P_2, \ldots] = (\beta_1, \ldots, \beta_k)$$

The matrix A is similar over F to a diagonal matrix if and only if P is a square matrix. When P is square, P is invertible and $P^{-1} AP$ is diagonal.

Example 6: Let T be the linear operator on R^3 which is represented in the standard ordered basis by the matrix.

$$A = \begin{bmatrix} 5 & -6 & -6 \\ -1 & 4 & 2 \\ 3 & -6 & -4 \end{bmatrix}$$

Let us indicate how one might compute the characteristic polynomial, using various row and column operations:

$$\begin{bmatrix} x-5 & 6 & 6 \\ 1 & x-4 & -2 \\ -3 & 6 & x+4 \end{bmatrix} = \begin{bmatrix} x-5 & 0 & 6 \\ 1 & x-2 & -2 \\ -3 & 2-x & x+4 \end{bmatrix}$$

$$= (x-2)\begin{bmatrix} x-5 & 0 & 6 \\ 1 & 1 & -2 \\ -3 & -1 & x+4 \end{bmatrix}$$

$$= (x-2)\begin{bmatrix} x-5 & 0 & 6 \\ 1 & 1 & -2 \\ -2 & 0 & x+2 \end{bmatrix}$$

$$= \quad (x-2)\begin{bmatrix} x-5 & 6 \\ -2 & x+2 \end{bmatrix}$$

$$= \quad (x-2)\,(x^2-3x+2)$$

$$= \quad (x-2)^2\,(x-1).$$

What are the dimensions of the spaces of characteristic vectors associated with the two characteristic values? We have

$$A - I \;=\; \begin{bmatrix} 4 & -6 & -6 \\ -1 & 3 & 2 \\ 3 & -6 & -5 \end{bmatrix}$$

$$A - 2I \;=\; \begin{bmatrix} 3 & -6 & -6 \\ -1 & 2 & 2 \\ 3 & -6 & -6 \end{bmatrix}$$

We know that $A - I$ is singular and obviously rank $(A - I) \geq 2$. Therefore, rank $(A - I) = 2$. It is evident that rank $(A - 2I) = 1$.

Let W_1, W_2 be the spaces of characteristic vectors associated with the characteristic values 1, 2. We know that dim $W_1 = 1$ and dim $W_2 = 2$. By Theorem 2, T is diagonalizable. It is easy to exhibit a basis for R^3 in which T is represented by a diagonal matrix. The null space of $(T - I)$ is spanned by the vector $\alpha_1 = (3, -1, 3)$ and so $\{\alpha_1\}$ is a basis for W_1. The null space of $T - 2I$ (i.e., the space W_2) consists of the vectors (x_1, x_2, x_3) with $x_1 = 2x_2 + 2x_3$. Thus, one example of a basis for W_2 is

$$\alpha_2 \;=\; (2, 1, 0)$$

$$\alpha_3 \;=\; (2, 0, 1).$$

If $\beta = (\alpha_1, \alpha_2, \alpha_3)$, then $[T]_\beta$ is the diagonal matrix

$$D \;=\; \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

The fact that T is diagonalizable means that the original matrix A is similar (over R) to the diagonal matrix D. The matrix P which enables us to change coordinates from the basis β to the standard basis is (of course) the matrix which has the transposes of α_1, α_2, α_3 as its column vectors:

$$P \;=\; \begin{bmatrix} 3 & 2 & 2 \\ -1 & 1 & 0 \\ 3 & 0 & 1 \end{bmatrix}$$

Furthermore, $AP = PD$, so that

$$P^{-1}AP = D.$$

Self Assessment

1. In each of the following cases, let T be the linear operator on R^2 which is represented by the matrix A in the standard ordered basis for R^2, and let U be the linear operator on C^2 represented by A in the standard ordered basis. Find the characteristic polynomial for T

and that for U, find the characteristic values of each operator, and for each such characteristic value c find a basis for the corresponding space of characteristic vectors.

$$A = \begin{bmatrix} 1 & 4 \\ 2 & 3 \end{bmatrix}, \quad A = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

2. Let T be the linear operator on R^3 which is represented in the standard ordered basis by the matrix

$$\begin{bmatrix} 4 & 2 & -2 \\ -5 & 3 & 2 \\ -2 & 4 & 1 \end{bmatrix}$$

Prove that T is diagonalizable by exhibiting a basis for R^3, each vector of which is a characteristic vector of T.

3. Let $$A = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & -5 \\ 0 & 1 & -2 \end{bmatrix}$$

Is A similar over the field R to a diagonal matrix? Is A similar over the field C to a diagonal matrix?

12.3 Summary

- When a matrix of a linear operator under a certain ordered basis is in the diagonal form then some properties of the linear operator can be real at a glance on this matrix.
- In this unit the characteristic values and the corresponding characteristic vectors of a matrix are found which help us in answering the question whether the given matrix is diagonalizable over the F or not.

12.4 Keywords

Invertible Matrix: The invertible matrix P formed out of the characteristic vectors of a vector A shows that A and PAP^{-1} are similar and also PAP^{-1} is in the diagonal form.

Null Space: If T is any linear operator and c is any scalar, the set of vectors α, such that $T\alpha = c\alpha$ is a sub-space of V. It is null space of linear transformation ($T- cI$).

12.5 Review Questions

1. If T be the linear operator on C^3 which is represented in the ordered basis by the matrix

$$A = \begin{bmatrix} 1 & i & 1 \\ -i & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

Prove that T is diagonalizable by exhibiting a basis for C^3, each vector of which is a characteristic vector of T.

2. If T be the linear operator on R^3 which is represented in the standard basis by the matrix

$$A = \begin{bmatrix} 5 & -6 & -6 \\ -1 & 4 & 2 \\ 3 & -6 & -4 \end{bmatrix}$$

Prove that T is diagonalizable. Find the diagonalizable matrix P that PAP^{-1} is diagonal.

Answers: Self Assessment

1. For $A = \begin{bmatrix} 1 & 4 \\ 2 & 3 \end{bmatrix}$, characteristic polynomial for T is $T^2 - 4t - 5I = 0$

$\lambda_1 = 5, \alpha_1 = (1, 1) \qquad \lambda_1 = -1, \alpha_2 = (2, -1)$

For $A = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$, the characteristic polynomial for T is $T^2 - 2T = 0$ the characteristic roots are

$\lambda_1 = 0, \quad \alpha_1 = (1, 1)$

$\lambda_2 = 2, \quad \alpha_2 = (1, -1)$

2. In the matrix is diagonalizable has the characteristic values 1, 2, 5 with the characteristic vectors (2, 1, 4), (1, 1, 0), (0, 1, 1) respectively. The diagonalizing matrix is

$$\begin{bmatrix} 2 & 1 & 0 \\ 1 & 1 & 1 \\ 4 & 2 & 1 \end{bmatrix}$$

3. A is not similar over the real field F to a diagonal matrix. But A is similar over the field C to a diagonal matrix

$$\begin{bmatrix} 3 & 0 & 0 \\ 0 & i & 0 \\ 0 & 0 & -i \end{bmatrix}$$

12.6 Further Readings

Books Kenneth Hoffman and Ray Kunze, *Linear Algebra*

I.N. Herstein, *Topics in Algebra*

Unit 13: Annihilating Polynomials

CONTENTS

Objectives

After studying this unit, you will be able to:

- Know about the polynomials over the field *F*, the degree of polynomial, monic polynomial, annihilating polynomials as well as minimal polynomials.
- Understand that the linear operator is annihilated by its characteristic polynomial.
- Understand that we consider all monic polynomials with coefficients in *F* and the degree of the minimal polynomial is the least positive integer such that a linear relation is obtained annihilated.

Introduction

In this unit we investigate more properties of a linear transformation.

We define certain terms like monic polynomial, minimal polynomial as well as annihilating polynomial and characteristic polynomial.

It is seen that the theorem of Cayley-Hamilton in this unit helps us in narrowing down the reach for the minimal polynomials of various operators.

13.1 Overview

Polynomial Over F. Let $F(x)$ be the subspace of F^n spanned by vectors 1, x, x^2..... An element of $F(x)$ is called a polynomial over F.

Degree of a Polynomial: $F(x)$ consists of all (finite) linear combinations of x and its powers. If f is a non-zero polynomial of the form

$$f = f_0x^0 + f_1x + f_2x^2 + \cdots + f_nx^n$$

such that $f_n \neq 0$ and $n \geq 0$ and $f_n = 0$ for all integers $k > n$; this integer is obviously unique and is called the *degree* of *f*.

The scalars $f_0, f_1, f_2, \cdots, f_n$ are sometimes called the coefficients of f in the field F.

Monic Polynomial: A polynomial *f*(*x*) over a field F is called monic polynomial if the coefficient of highest degree term in it is unity i.e $f_n = 0$

Annihilating Polynomials: Let *A* be $n \times n$ matrix over a field *F* and *f*(*x*) be a polynomial over *F*. Then if *f*(*A*) = 0. Then we say that the polynomial *f*(*x*) annihilates the matrix *A*.

13.2 Annihilating Polynomials

It is important to know the class of polynomials that Annihilate *T*.

Suppose *T* is a linear operator on *V*, a vector space over the field *F*. If *p* is a polynomial over *F*, then *p*(*T*) is again a linear operator on *V*. If *q* is another polynomial over *F*, then

$$(p+q)(T) = p(T)+q(T)$$

$$(pq)(T) = p(T)\,q(T)$$

Therefore, the collection of polynomials *p* which annihilate *T*, in the sense that

$$p(T) = 0,$$

is an ideal in the polynomial algebra *F*[*x*]. It may be the zero ideal, i.e., it may be that *T* is not annihilated by any non-zero polynomial. But, that cannot happen if the space *V* is finite-dimensional.

Suppose *T* is a linear operator on the *n*-dimensional space *V*. Look at the first ($n^2 + 1$) powers of *T*:

$$I, T, T^2 \cdots, T^{n2}.$$

This is a sequence of $n^2 + 1$ operators in *L*(*V*, *V*), the space of linear operators on *V*. The space *L*(*V*, *V*,) has dimension n^2. Therefore, that sequence of $n^2 + 1$ operators must be linearly dependent. i.e., we have

$$c_0 I + c_1 T + \cdots + c_{n2} T^{n2} = 0$$

for some scalars c_i not all zero. So, the ideal of polynomials which annihilate *T* contains a non-zero polynomial of degree n^2 or less.

Definition. Let *T* be a linear operator on a finite-dimensional vector space *V* over the field *F*. The minimal polynomial for *T* is the (unique) monic generator of the ideal of polynomials over *F* which annihilate *T*.

The name '*minimal polynomial*' stems from the fact that generator of a polynomial ideal is characterized by being the monic polynomial of minimum degree in the ideal. That means that the minimal polynomial *p* for the linear operator *T* is uniquely determined by these three properties:

1. *p* is a monic polynomial over the scalar field *F*.
2. *p*(*T*) = 0
3. No polynomial over *F* which annihilates *T* has smaller degree than *p* has.

If A an $n \times n$ matrix over *F*, we define the **minimal polynomial** for *A* in an analogous way, as the unique monic generator of the ideal of all polynomials over *F* which annihilate *A*. If the operator *T* is represented in some ordered basis by the matrix *A*, then *T* and *A* have the same minimal polynomial. That is because *f*(*T*) is represented in the basis by the matrix *f*(*A*) so that *f*(*T*) = 0 if and only if *f*(*A*) = 0.

From the last remark about operators and matrices it follows that similar matrices have the same minimal polynomial. That fact is also clear from the definitions because

$$f(P^{-1}AP) = P^{-1}f(A)P$$

for every polynomial *f*.

There is another basic remark which we should make about minimal polynomials of matrices. Suppose that *A* is an $n \times n$ matrix with entries in the field *F*. Suppose that F_1 is a field which contains *F* as a subfield. (For example, *A* might be a matrix with rational entries, while F_1 is the field of real numbers. Or, *A* might be a matrix with real entries, while F_1 is the field of complex numbers.) We may regard *A* either as an $n \times n$ matrix over *F* or as an $n \times n$ matrix over F_1. On the surface, it might appear that we obtain two different minimal polynomials for *A*. Fortunately that is not the case; and we must see why. What is the definition of the minimal polynomial for *A*, regarded as an $n \times n$ matrix over the field *F*? We consider all monic polynomials with coefficients in *F* which annihilate *A*, and we choose the one of least degree. If *f* is a monic polynomial over *F*:

$$f = x^k + \sum_{j=0}^{k-1} a_j x^i \qquad \dots (1)$$

then *f*(*A*) = 0 merely says that we have a linear relation between the powers of *A*:

$$A^k + a_{k-1}A^{k-1} + \cdots + a_1 A + a_0 I = 0 \qquad \dots (2)$$

The degree of the minimal polynomial is the least positive integer *k* such that there is a linear relation of the form (2) between the powers *I*, $A, \cdots, A^k$. Furthermore, by the uniqueness of the minimal polynomial, there is for that *k* one and only one relation of the form (2); i.e., once the minimal *k* is determined, there are unique scalars $a_0, \cdots, a_{k-1}$ in *F* such that (2) holds. They are the coefficients of minimal polynomial.

Now (for each *k*) we have in (2) a system of n^2 linear equations for the 'unknowns' $a_0, \cdots, a_{k-1}$. Since the entries of *A* lie in *F*, the coefficients of the system of equations (2) are in *F*. Therefore, if the system has a solution with $a_0, \cdots, a_{k-1}$ in F_1 it has a solution with $a_0, \cdots, a_{k-1}$ in *F*. It should now be clear that the two minimal polynomials are the same.

What do we know thus far about the minimal polynomial for a linear operator on an *n*-dimensional space? Only that its degree does not exceed n^2. That turns out to be a rather poor estimate, since the degree cannot exceed *n*. We shall prove shortly that the operator is annihilated by its characteristic polynomial. First, let us observe a more elementary fact.

Theorem 1: Let *T* be a linear operator on an *n*-dimensional vector space *V* [or, let *A* be an $n \times n$ matrix]. The characteristic and minimal polynomials for *T* [for *A*] have the same roots, except for multiplicities.

Proof. Let *p* be the minimal polynomial for *T*. Let *c* be a scalar. What we want to show is that *p*(*c*) = 0 if and only if *c* is a characteristic value of *T*.

First, suppose *p*(*c*) = 0. Then

$$p = (x - c)q$$

where *q* is a polynomial. Since deg *q* < deg *p*, the definition of the minimal polynomial *p* tells us that $q(T) \neq 0$. Choose a vector β such that $q(T)\ \beta \neq 0$. Let $\alpha = q(T)\beta$. Then

$$\begin{aligned} 0 &= p(T)\beta \\ &= (T - cI)q(T)\beta \end{aligned}$$

$$= (T - cI)\alpha$$

and thus, c is a characteristic value of T.

Now, suppose that c is a characteristic value of T, say $T\alpha = c\alpha$ with $\alpha \neq 0.$ So

$$p(T)\alpha \ = p(c)\alpha.$$

Since $p(T) = 0$ and $\alpha \neq 0$, we have $p(c)$ 0.

Let T be a diagonalizable linear operator and let $c_1, \cdots, c_k$ be the distinct characteristic values of T. Then it is easy to see that the minimal polynomial for T is the polynomial.

$$p = (x - c_1) \ \cdots \ (x - c_k).$$

If α is a characteristic vector, then one of the operators $T - c_1I, \ \cdots \ , T - c_kI$ sends α into 0. Therefore

$$(T - c_1I)\cdots(T - c_kI)\,\alpha = 0$$

for every characteristic vector α. There is a basis for the underlying space which consists of characteristic vectors of T; hence

$$p(T) = (T - c_1I)\cdots(T - c_kI) = 0.$$

What we have concluded is this. If T is a diagonalizable linear operator, then the minimal polynomial for T is a product of distinct linear factors. As we shall soon see, that property characterizes diagonalizable operators.

Example 1: Let's try to find the minimal polynomials for the operators in Example 1, 2, and 6 in unit 12. We shall discuss them in reverse order. The operator in Example 6 was found to be diagonalizable with characteristic polynomial.

From the preceding paragraph we know that the minimal polynomial for T is.

$$p = (x - 1)(x - x).$$

The reader might find it reassuring to verify directly that

$$(A - I)(A - 2I) = 0.$$

In Example 2, the operator T also had the characteristic polynomial $f = (x - 1)(x - 2)^2$. But, this T is not diagonalizable, so we don't know that the minimal polynomial is $(x - 1)\ (x - 2)$. What do be know about the minimal polynomial in this case? We know that its roots are 1 and 2, with some multiplicities allowed. Thus we search for p among polynomials of the form $(x - 1)^k$ $(x - 2)^l, k \geq 1, l \geq 1.$ Try $(x - 1)\ (x - 2)$:

$$(A - I)\ (A - 2I) \ = \begin{bmatrix} 2 & 1 & -1 \\ 2 & 1 & -1 \\ 2 & 2 & -1 \end{bmatrix}\begin{bmatrix} 1 & 1 & -1 \\ 2 & 0 & -1 \\ 2 & 2 & -2 \end{bmatrix}$$

$$= \begin{bmatrix} 2 & 0 & -1 \\ 2 & 0 & -1 \\ 4 & 0 & -2 \end{bmatrix}$$

Thus, the minimal polynomial has degree at least 3. So, next we should try either $(x - 1)^2\ (x - 2)$ or $(x - 1)\ (x - 2)^2$. The second being the characteristic polynomial, would seem a less random choice. One can readily compute that $(A - I)\ (A - 2I)^2 = 0$. Thus the minimal polynomial for T is its characteristic polynomial.

Notes

In Example 1 in unit 12 we discussed the linear operator T on R^2 which is represented in the standard basis by the matrix.

$$\text{A} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

The characteristic polynomial is $x^2 + 1$, which has no real roots. To determine the minimal polynomial, forget about T and concentrate on A. As a complex 2×2 matrix, A has the characteristic values i and $-i$. Both roots must appear in the minimal polynomial. Thus the minimal polynomial is divisible by $x^2 + 1$. It is trivial to verify that $A^2 + I = 0$. So the minimal polynomial is $x^2 + 1$.

Theorem 2 (Cayley-Hamilton): Let T be a linear operator on a finite dimensional vector space V. If f is the characteristic polynomial for T, then $f(T) = 0$; in other words, the minimal polynomial divides the characteristic polynomial for T.

Proof: The proof, although short, may be difficult to understand. Aside from brevity, it has the virtue of providing an illuminating and far from trivial application of the general theory of determinants.

Let K be the commutative ring with identity consisting of all polynomials in T. Of course, K is actually a commutative algebra with identity over the scalar field. Choose an ordered basis $\{\alpha_1, \cdots, \alpha_n\}$ for V, and let A be the matrix which represents T in the given basis. Then

$$T\alpha_i = \sum_{i=j}^{n} A_{ji}\alpha_j, \qquad 1 \le j \le n$$

These equations may be written in the equivalent form

$$\sum_{j=1}^{n} (\delta_{ij}T - A_{ji}I)\alpha_j = 0, \qquad 1 \le i \le n.$$

Let B denote the element of $K^{n\times n}$ with entries

$$B_{ij} = \delta_{ij}T - A_{ji}I.$$

When $n = 2$

$$B = \begin{bmatrix} T - A_{11}I & -A_{21}I \\ -A_{12}I & T - A_{22}I \end{bmatrix}$$

and

$$\begin{aligned} \det B &= (T - A_{11}I)(T - A_{22}I) - A_{12}A_{21}I \\ &= T^2 - (A_{11} + A_{22})T + (A_{11}A_{12} - A_{12}A_{21})I \\ &= f(T) \end{aligned}$$

where f is the characteristic polynomial:

$$f = x^2 - (\text{trace } A)x + \det A.$$

For the case $n > 2$, it is also clear that

$$\det B = f(T)$$

since f is the determinant of the matrix $xI - A$ whose entries are the polynomials

$$(xI - A)_{ij} = \delta_{ij}x - A_{ji}.$$

We wish to show that $f(T) = 0$. In order that $f(T)$ be the zero operator, it is necessary and sufficient that $(\det B)_{\alpha k} = 0$ for $k = 1, \cdots, n$. By the definition of B, the vectors $\alpha_1, \cdots \alpha_n$ satisfy the equations

$$\sum_{j=1}^{n} B_{ij}\alpha_j = 0, \quad 1 \le i \le n. \qquad \dots (3)$$

When $n = 2$, it is suggestive to write (3) in the form

$$\begin{bmatrix} T - A_{11}I & -A_{21}I \\ -A_{12}I & T - A_{22}I \end{bmatrix}\begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

In this case, the classical adjoint, *adj B* is the matrix

$$\tilde{B} = \begin{bmatrix} T - A_{22}I & A_{21}I \\ A_{12}I & T - A_{11}I \end{bmatrix}$$

and

$$\tilde{B}B = \begin{bmatrix} \det B & 0 \\ 0 & \det B \end{bmatrix}.$$

Hence, we have

$$\begin{aligned} (\det B)\begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} &= (\tilde{B}B)\begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} \\ &= \tilde{B}\left(B\begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}\right) \\ &= \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \end{aligned}$$

In the general case, let $\tilde{B} = adj\ B$. Then by (3)

$$\sum_{j=1}^{n} \tilde{B}_{ki}B_{ij}\alpha_j = 0$$

for each pair k, i, and summing on i, we have

$$\begin{aligned} 0 &= \sum_{i=1}^{n}\sum_{j=1}^{n} \tilde{B}_{ki}B_{ij}\alpha_i \\ &= \sum_{j=1}^{n}\left(\sum_{i=1}^{n} \tilde{B}_{ki}B_{ij}\right)\alpha_j. \end{aligned}$$

Now $\tilde{B}\,B$ = (det B)I, so that

$$\sum_{i=1}^{n} \tilde{B}_{ki}B_{ij} = \delta_{kj} \det B.$$

Therefore

$$0 = \sum_{j=1}^{n} \delta_{ki}\left(\det B\right)\alpha_j$$

$$= (\det B)_{\alpha k,} \qquad 1 \le k \le n.$$

The Cayley-Hamilton theorem is useful to us at this point primarily because it narrows down the search for the minimal polynomials of various operators. If we know the matrix A which represents T in some ordered basis, then we can compute the characteristic polynomial f. We know that the minimal polynomial p divides f and that the two polynomials have the same roots. There is no method for computing precisely the roots of a polynomial (unless its degree is small); however, if f factors

$$f = (x - c_1)^{d1} \cdots (x - c_k)^{dk}, C_{11} \cdots, C_k, \text{distinct}, d_i \ge 1 \qquad \text{.... (4)}$$

then

$$p = (x - c_1)^{r1} \cdots (x - c_k)^{rk}, \qquad 1 \le r_j \le d_j \qquad \text{.... (5)}$$

That is all we can say in general. If f is the polynomial (4) and has degree n, then for every polynomial p as in (5) we can find an $n \times n$ matrix which has f as its characteristic polynomial and p as its minimal polynomial. We shall not prove this now. But, we want to emphasize the fact that the knowledge that the characteristic polynomial has the form (4) tells us that the minimal polynomial has the form (5), and it tells us nothing else about p.

Example 2: Let A be the 4×4 (rational) matrix

$$A = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}$$

The powers of A are easy to compute

$$A^2 = \begin{bmatrix} 2 & 0 & 2 & 0 \\ 0 & 2 & 0 & 2 \\ 2 & 0 & 2 & 0 \\ 0 & 2 & 0 & 2 \end{bmatrix}$$

$$A^3 = \begin{bmatrix} 0 & 4 & 0 & 4 \\ 4 & 0 & 4 & 0 \\ 0 & 4 & 0 & 4 \\ 4 & 0 & 4 & 0 \end{bmatrix}$$

Thus $A^3 = 4A$, i.e., if $p = x^3 - 4x = x(x + 2)(x - 2)$, then $p(A) = 0$. The minimal polynomial for A must divide p. That minimal polynomial is obviously not of degree 1, since that would mean that A was a scalar multiple of the identity. Hence, the candidates for the minimal polynomial are:

Notes

p, $x(x + 2)$, $x(x - 2)$, $x^2 - 4$. The three quadratic polynomials can be eliminated because it is obvious at a glance that $A^2 \neq -2A$, $A^2 \neq 2A$, $A^2 \neq 4I$. Therefore p is the minimal polynomial for A. In particular 0, 2, and – 2 are the characteristic values of A. One of the factors x, $x - 2$, $x + 2$ must be repeated twice in the characteristic polynomial. Evidently, rank (A) = 2. Consequently there is a two-dimensional space of characteristic vectors associated with the characteristic value 0. From Theorem 2, it should now be clear that the characteristic polynomial is $x^2 (x^2 - 4)$ and that A is similar over the field of rational numbers to the matrix

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & -2 \end{bmatrix}.$$

Example 3: Verify Cayley-Hamilton's theorem for the linear transformation T represented by the matrix A.

$$\text{A} = \begin{bmatrix} 0 & 0 & 1 \\ 3 & 1 & 0 \\ -2 & 1 & 4 \end{bmatrix}$$

Solution: The characteristic polynomial is given by

$$|A - x\,I| = \begin{bmatrix} 0-x & 0 & 1 \\ 3 & 1-x & 0 \\ -2 & 1 & 4-x \end{bmatrix}$$

$$= -x\left[(1-x)(4-x)\right] + (3+2-2x)$$

$$= -x\left(4-5x+x^2\right)+5-2x$$

$$= -x^3+5x^2-6x+5=0$$

or

$$f(x) = x^3-5x^2+6x-5=0$$

Now

$$A^2 = \begin{bmatrix} 0 & 0 & 1 \\ 3 & 1 & 0 \\ -2 & 1 & 4 \end{bmatrix}\begin{bmatrix} 0 & 0 & 1 \\ 3 & 1 & 0 \\ -2 & 1 & 4 \end{bmatrix} = \begin{bmatrix} -2 & 1 & 4 \\ 3 & 1 & 3 \\ -5 & 5 & 14 \end{bmatrix}$$

$$A^3 = \begin{bmatrix} -2 & 1 & 4 \\ 3 & 1 & 3 \\ -5 & 5 & 14 \end{bmatrix}\begin{bmatrix} 0 & 0 & 1 \\ 3 & 1 & 0 \\ -2 & 1 & 4 \end{bmatrix} = \begin{bmatrix} -5 & 5 & 14 \\ -3 & 4 & 15 \\ -13 & 19 & 51 \end{bmatrix}$$

So

$$f(A) = A^3 - 5A^2 + 6A - 5I$$

$$= \begin{bmatrix} -5 & 5 & 14 \\ -3 & 4 & 15 \\ -13 & 19 & 51 \end{bmatrix} - \begin{bmatrix} -10 & 5 & 20 \\ 15 & 5 & 15 \\ -25 & 25 & 14 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 6 \\ 18 & 6 & 0 \\ -12 & 6 & 24 \end{bmatrix} - \begin{bmatrix} 5 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 5 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = 0$$

where 0 being null matrix. So $f(A) = 0$

Self Assessment

1. Let A be the following 3×3 matrix over F;

$$A = \begin{bmatrix} 2 & -1 & 1 \\ -2 & 2 & -1 \\ 1 & -1 & 2 \end{bmatrix}$$

Find the characteristic polynomial for A and also the minimal polynomial for A.

2. Let A be the following 3×3 matrix over F;

$$A = \begin{bmatrix} 1 & 3 & 7 \\ 4 & 2 & 3 \\ 1 & 2 & 1 \end{bmatrix}$$

Find the characteristic polynomial for A and also find the minimal polynomial for A.

13.3 Summary

- In this unit certain terms related to linear operator T are defined, i.e., the monic polynomial, annihilating polynomials, minimal polynomials as well as characteristic polynomials.
- With the help of Cayley-Hamilton theorem it becomes easier to search for the minimal polynomials of various operators.

13.4 Keywords

Annihilating Polynomial: Annihilating polynomial $f(x)$ over the field F is such that for a matrix A of $n \times n$ matrix over the field $f(A) = 0$, then we say that the polynomial annihilates the matrix.

If a linear operator T is represented by the matrix then $f(T) = 0$ gives us the annihilating polynomial for the linear operator T.

Monic Polynomial: The monic polynomial is a polynomial $f(x)$ whose coefficient of the highest degree in it is unity.

13.5 Review Questions

1. Let A be the following 3×3 matrix over F;

$$A = \begin{bmatrix} 2 & 4 & 3 \\ 0 & -1 & 1 \\ 2 & 2 & -1 \end{bmatrix}$$

Find the characteristic polynomial and minimal polynomial for A.

2. Let A be the following 3×3 matrix over F;

$$A = \begin{bmatrix} 1 & 2 & 0 \\ 2 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

Find the characteristic polynomial and minimal polynomial for A.

Notes

Answers: Self Assessment

1. The characteristic polynomial is given by

   $f(x) = x^3 - 6x^2 + 9x - 4 = 0$

   and that this is also the minimal polynomial for A.

2. The characteristic polynomial for A is

   $f(x) = x^3 - 4x^2 - 20x - 35 = 0$,

   and that this is also the minimal polynomial for A.

13.6 Further Readings

Books Kenneth Hoffman and Ray Kunze *Linear Algebra*

I.N Herstein *Topics in Algebra*

Unit 14: Invariant Subspaces

Notes

CONTENTS

Objectives

After studying this unit, you will be able to:

- Know about few concepts which are useful in analysing further properties of the linear operator *T*.
- Understand concepts like invariant subspace, the restriction operator *Tw*, the *T*-conductor of a vector α into subspace *W*.
- See that all these concepts help us in understanding the structure of minimal polynomial of linear operator.
- Understand the restriction operator *Tw* helps in writing the matrix *A* of the linear operator in a block form and so the characteristic polynomial for *Tw* divides the characteristic polynomial for *T*.

Introduction

In this unit we are still studying the properties of a linear operator on the vector space *V*. The concept of invariant subspace, the restriction operator *Tw* help us in finding the characteristic polynomial of *T* as well as its annihilator and so it helps in diagonalization of the matrix *A* of the linear operator *T*.

14.1 Invariant Subspaces: Definitions

In this unit, we shall introduce a few concepts which are useful in analysing further the properties of a linear operator. We shall use these concepts to obtain characterizations of diagonalizable (and triangulable) operators in terms of their minimal polynomials.

Invariant Subspace

A subspace *W* of the vector space *V* is invariant of more precisely *T*-invariant if for each vector α in *W* the vector $T\alpha$ is in *W*, i.e., $T(w)$ is contained in *W*. When this is so *T* induces a linear operator on *W*, called restriction to *W*. We often denote the restriction by *Tw*. The linear operator *Tw* is defined by $Tw(\alpha) = T(\alpha)$, for α in *W*, but *Tw* is quite a different object from *T* since its domain is *W* and not *V*.

The T-conductor of ● into W

Let W be an invariant subspace for T and let α be a vector in V. The T-conductor of α into W is the set $S_T(\alpha; W)$, which consists of all polynomials g (over the scalar field), such that $g(T)\alpha$ is in W. Some authors call that collection of polynomials the 'stuffer'. In the special case $W = \{0\}$, the conductor is called T-annihilator of α.

14.2 Theorems and Examples

Example 1: If T is any linear operator on V, then V is invariant under T, as is the zero subspace. The range of T and the null space of T are also invariant under T.

Example 2: Let F be a field and let D be the differentiation operator on the space $F[x]$ of polynomials over F. Let n be a positive integer and let W be the subspace of polynomials of degree not greater than n. Then W is invariant under D. This is just another way of saying that D is 'degree decreasing'.

Example 3: Here is a very useful generalization of Example 1. Let T be a linear operator on V. Let U be any linear operator on V which commutes with T, i.e., $TU = UT$. Let W be the range of U and let N be the null space of U. Both W and N are invariant under T. If α is in the range of U, say $\alpha = U\beta$, then $T\alpha = T(U\beta) = U(T\beta)$ so that $T\alpha$ is in the range of U. If α is in N, then $U(T\alpha) = T(U\alpha) = T(0) = 0$; hence $T\alpha$ is in N.

A particular type of operator which commutes with T is an operator $U = g(T)$, where g is a polynomial. For instance, we might have $U = T - cI$, where c is a characteristic value of T. The null space of U is familiar to us. We see that this example includes the (obvious) fact that the space of characteristic vectors of T associated with the characteristic value c is invariant under T.

Example 4: Let T be the linear operator on R^2 which is represented in the standard ordered basis by the matrix

$$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

Then the only subspaces of R^2 which are invariant under T are R^2 and the zero subspace. Any other invariant subspace would necessarily have dimension 1. But, if W is the subspace spanned by some non-zero vector α, the fact that W is invariant under T means that α is a characteristic vector, but A has no real characteristic values.

When V is finite-dimensional, the invariance of W under T has a simple matrix interpretation, and perhaps we should mention it at this point. Suppose we choose an ordered basis $\mathcal{B} = \{\alpha_1,...,\alpha_n\}$ for V such that $\mathcal{B}' = \{\alpha_1,...,\alpha_r\}$ is an ordered basis for W (r = dim W). Let $A = [T]_{\mathcal{B}}$ so that

$$T\alpha_j = \sum_{i=1}^{n} A_{ij}\alpha_i \qquad ...(1)$$

Since W is invariant under T, the vector $T\alpha_j$ belongs to W for $j \leq r$. This means that

$$T\alpha_j = \sum_{i=1}^{r} A_{ij}\alpha_i, \qquad j \leq r \qquad ...(2)$$

Notes

In other words, $A_{ij} = 0$ if $j \leq r$ and $i \geq r$.

Schematically, A has the block form

$$A = \begin{bmatrix} B & C \\ 0 & D \end{bmatrix} \qquad ...(3)$$

where B is an $r \times r$ matrix, C is an $r \times (n - r)$ matrix, and D is an $(n - r) \times (n - r)$ matrix. The reader should note that according to (2) the matrix B is precisely the matrix of the induced operator Tw in the ordered basis $\mathcal{B}'$.

Most often, we shall carry out arguments about T and Tw without making use of the block form of the matrix A in (3). But we should note how certain relations between Tw and T are apparent from that block form.

Lemma: Let W be an invariant subspace for T, the characteristic polynomial for the restriction operator Tw divides the characteristic polynomial for T. The minimal polynomial for Tw divides the minimal polynomial for T.

Proof: We have

$$A = \begin{bmatrix} B & C \\ 0 & D \end{bmatrix}$$

where $A = [T]_{\mathcal{B}}$ and $B = [T_w]_{\mathcal{B}'}$. Because of the block form of the matrix

$$\det(xI - A) = \det(xI - B) \det(xI - D)$$

That proves the statement about characteristic polynomials. Notice that we used I to represent identity matrices of three different sizes.

The k^{th} power of the matrix A has the block form

$$A^k = \begin{bmatrix} B^k & C_k \\ 0 & D^k \end{bmatrix}$$

where C_k is some $r \times (n - r)$ matrix. Therefore, any polynomial which annihilates A also annihilates B (and D too). So, the minimal polynomial for B divides the minimal polynomial for A.

Example 5: Let T be any linear operator on a finite-dimensional space V. Let W be the subspace spanned by all of the characteristic vectors of T. Let $c_1,...,c_k$ be the distinct characteristic values of T. For each i, let W_i be the space of characteristic vectors associated with the characteristic value c_i, and let $\mathcal{B}_i$ be an ordered basis for W_i. The lemma before Theorem 2 of unit 12 tells us that $\mathcal{B}' = (\mathcal{B}_1,...,\mathcal{B}_k)$ is an ordered basis for W. In particular,

$$\dim W = \dim W_1 + ... + \dim W_k.$$

Let $\mathcal{B}' = \{\alpha_1,...,\alpha_r\}$ so that the first few α's form the basis $\mathcal{B}_1$, the next few $\mathcal{B}_2$, and so on. Then

$$T\alpha_i = t_i\alpha_i, \qquad i = 1,...,r$$

where $(t_1,..., t_r) = (c_1, c_1,..., c_1,..., c_k, c_k,..., c_k)$ with c_i repeated $\dim W_i$ times.

Now W is invariant under T, since for each α in W we have

$$\alpha = x_1\alpha_1 + ... + x_r\alpha_r$$

$$T_\alpha = t_1x_1\alpha_1 + ... + t_rx_r\alpha_r$$

Choose any other vectors $\alpha_{r+1},..., \alpha_n$ in V such that $\mathcal{B} = \{\alpha_1,...,\alpha_n\}$ is a basis for V. The matrix of T relative to $\mathcal{B}$ has the block form (3), and the matrix of the restriction operator Tw relative to the basis $\mathcal{B}'$ is

$$B = \begin{bmatrix} t_1 & 0 & \cdots & 0 \\ 0 & t_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & t_r \end{bmatrix}$$

The characteristic polynomial of B (i.e., of Tw) is

$$g = (x - c_1)^{e_1} ... (x - c_k)^{e_k}$$

where e_i = dim W_i. Furthermore, g divides f, the characteristic polynomial for T. Therefore, the multiplicity of c_i as a root of f is at least dim W_i.

All of this should make Theorem 2 of unit 12 transparent. It merely says that T is diagonalizable if and only if $r = n$, if and only if $e_1 + ... + e_k = n$. It does not help us too much with the non-diagonalizable case, since we don't know the matrices C and D of (3).

Lemma: If W is an invariant subspace for T, then W is invariant under every polynomial in T. Thus, for each α in V, the conductor $S_T(\alpha; W)$ is an ideal in the polynomial algebra $F[x]$.

Proof: If β is in W, then $T\beta$ is in W. Consequently, $T(T\beta) = T^2\beta$ is in W. By induction, $T^k\beta$ is in W for each k. Take linear combinations to see that $f(T)\beta$ is in W for every polynomial f.

The definition of $S_T(\alpha; W)$ makes sense if W is any subset of V. If W is a subspace, then $S_T(\alpha; W)$ is a subspace of $F[x]$, because

$$(cf + g)(T) = cf(T) + g(T)$$

If W is also invariant under T, let g be a polynomial in $S_T(\alpha; W)$, i.e., let $g(T)\alpha$ be in W. If f is any polynomial, then $f(T)[g(T)\alpha]$ will be in W. Since

$$(fg)(T) = f(T)g(T)$$

fg is in $S_T(\alpha; W)$. Thus the conductor absorbs multiplication by any polynomial.

The unique monic generator of the ideal $S_T(\alpha; W)$ is also called the T-conductor of α into W (the T-annihilator in case $W = \{0\}$). The T-conductor of α into W is the monic polynomial g of least degree such that $g(T)\alpha$ is in W. A polynomial f is in $S_T(\alpha; W)$ if and only if g divides f. Note that the conductor $S_T(\alpha; W)$ always contains the minimal polynomial for T; hence, every T-conductor divides the minimal polynomial for T.

As the first illustration of how to use the conductor $S_T(\alpha; W)$, we shall characterize triangulable operators. The linear operator T is called triangulable if there is an ordered basis in which T is represented by a triangular matrix.

Lemma. Let V be a finite-dimensional vector space over the field F. Let T be a linear operator on V such that the minimal polynomial for T is a product of linear factors

$$p = (x - c_1)^{r_1} ... (x - c_k)^{r_k}, \quad c_i \text{ in } F$$

Let W be a proper ($W \neq V$) subspace of V which is invariant under T. There exists a vector α in V such that

(a) α is not in W;

(b) $(T - cI)\ \alpha$ is in W, for some characteristic value c of the operator T.

Proof: What (a) and (b) say is that the T-conductor of α into W is a linear polynomial. Let β be any vector in V which is not in W. Let g be the T-conductor of β into W. Then g divides p, the minimal polynomial for T. Since β is not in W, the polynomial g is not constant. Therefore,

$$g = (x - c_1)^{e_1} \dots (x - c_k)^{e_k}$$

where at least one of the integers e_i is positive. Choose j so that $e_j > 0$.

Then $(x - c_j)$ divides g:

$$g = (x - c_j)h$$

By the definition of g, the vector $\alpha = h(T)\beta$ cannot be in W. But

$$\begin{aligned}(T - c_jI)\alpha &= (T - c_jI)h(T)\beta \\ &= g(T)\beta\end{aligned}$$

is in W.

Theorem 1: Let V be a finite-dimensional vector space over the field F and let T be a linear operator on V. Then T is triangulable if and only if the minimal polynomial for T is a product of linear polynomials over F.

Proof: Suppose that the minimal polynomial factors

$$p = (x - c_1)^{r_1} \dots (x - c_k)^{r_k}$$

By repeated application of the lemma above, we shall arrive at an ordered basis $\mathcal{B} = \{\alpha_1,\dots,\alpha_n\}$ in which the matrix representing T is upper triangular:

$$[T]_{\mathcal{B}} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ 0 & a_{22} & a_{23} & \cdots & a_{2n} \\ 0 & 0 & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & a_{nn} \end{bmatrix} \qquad \dots(4)$$

Now (4) merely says that

$$T\alpha_j = \alpha_{1j}\alpha_1 + \dots + \alpha_{jj}\alpha_j, \qquad 1 \le j \le n \qquad \dots(5)$$

that is, $T\alpha_j$ is in the subspace spanned by $\alpha_1,\dots,\alpha_j$. To find $\alpha_1,\dots,\alpha_n$, we start by applying the lemma to the subspace $W = \{0\}$, to obtain the vector α_1. Then apply the lemma to W_1, the space spanned by α_1, and we obtain α_2. Next apply the lemma to W_2, the space spanned by α_1 and α_2. Continue in that way. One point deserves comment. After $\alpha_1,\dots, \alpha_i$ have been found, it is the triangular-type relations (5) for $j = 1,\dots, i$ which ensure that the subspace spanned by $\alpha_1,\dots, \alpha_i$ is invariant under T.

If T is triangulable, it is evident that the characteristic polynomial for T has the form

$$f = (x - c_1)^{d_1} \dots (x - c_k)^{d_k}, \qquad c_i \text{ in } F$$

Just look at the triangular matrix (4). The diagonal entries $a_{11},\dots, a_{1n}$ are the characteristic values, with c_i repeated d_i times. But, if f can be so factored, so can the minimal polynomial p, because it divides f.

Corollary: Let F be an algebraically closed field, e.g., the complex number field. Every $n \times n$ matrix over F is similar over F to a triangular matrix.

Theorem 2: Let V be a finite-dimensional vector space over the field F and let T be a linear operator on V. Then T is diagonalizable if and only if the minimal polynomial for T has the form

Notes

$$p = (x - c_1) \dots (x - c_k)$$

where $c_1, \dots, c_k$ are distinct elements of F.

Proof: We have noted earlier that, if T is diagonalizable, its minimal polynomial is a product of distinct linear factors. To prove the converse, let W be the subspace spanned by all of the characteristic vectors of T, and suppose $W \neq V$. By the lemma used in the proof of Theorem 1, there is a vector α not in W and a characteristic value c_j of T such that the vector

$$\beta = (T - c_j I)\alpha$$

lies in W. Since β is in W,

$$\beta = \beta_1 + \dots + \beta_k$$

where $T\beta_i = c_i\beta_i$, $1 \leq i \leq k$, and therefore the vector

$$h(T)\beta = h(c_1)\beta_1 + \dots + h(c_k)\beta_k$$

is in W, for every polynomial h.

Now $p = (x - c_j)q$, for some polynomial q. Also

$$q - q(c_j) = (x - c_j)h$$

We have

$$q(T)\alpha - q(c_j)\alpha = h(T)(T - c_j I)\alpha = h(T')\beta$$

But $h(T)\beta$ is in W and, since

$$0 = p(T)\alpha = (T - c_j I)q(T)\alpha$$

the vector $q(T)\alpha$ is in W. Therefore, $q(c_j)\alpha$ is in W. Since α is not in W, we have $q(c_j) = 0$. That contradicts the fact that p has distinct roots.

In addition to being an elegant result, Theorem 2 is useful in a computational way. Suppose we have a linear operator T, represented by the matrix A in some ordered basis, and we wish to know if T is diagonalizable. We compute the characteristic polynomial f. If we can factor f:

$$f = (x - c_1)^{d_1} \cdots (x - c_k)^{d_k}$$

we have two different methods for determining whether or not T is diagonalizable. One method is to see whether (for each i) we can find d_i independent characteristic vectors associated with the characteristic value c_i. The other method is to check whether or not $(T - c_1 I) \cdots (T - c_k I)$ is the zero operator.

Theorem 1 provides a different proof of the Cayley-Hamilton theorem. That theorem is easy for a triangular matrix. Hence, via Theorem 1, we obtain the result for any matrix over an algebraically closed field. Any field is a subfield of an algebraically closed field. If one knows that result, one obtains a proof of the Cayley-Hamilton theorem for matrices over any field. If we at least admit into our discussion the Fundamental Theorem of Algebra (the complex number field is algebraically closed), then Theorem 1 provides a proof of the Cayley-Hamilton theorem for complex matrices, and that proof is independent of the one which we gave earlier.

Self Assessment

1. Let T be the linear operator on R^2, the matrix of which in the standard ordered basis is

$$A = \begin{bmatrix} 1 & -1 \\ 2 & 2 \end{bmatrix}$$

(a) Prove that the only subspaces of R^2 invariant under T are R^2 and the zero subspace.

(b) If U is the linear operator on C^2, the matrix of which in the standard ordered basis is A, show that U has 1-dimensional invariant subspaces.

2. Let W be an invariant subspace for T. Prove that the minimal polynomial for the restriction operator T_W divides the minimal polynomial for T, without referring to matrices.

14.3 Summary

- In this unit the idea of invariant subspace of a linear operator T on the n dimension space helps in introducing a restriction operator Tw as well as a conductor of a vector $\alpha \in V$ into the subspace W.
- These concepts generally help us in the diagonalizing of the matrix of the linear operator T.
- These concepts also lead to triangular form of the matrix A of the linear operator T if A is diagonalizable.

14.4 Keywords

Invariant: If T is any linear operator on V, then V is invariant under T, as is the zero subspace. The range of T and the null space of T are also invariant under T.

Restriction Operator: By introducing the concepts of the restriction operator T_w and the conductor of a vector into the invariant sub-space the characteristic polynomial of the linear operator is cast into a form where the matrix of T can be seen to be diagonalizable or not.

Restriction: T induces a linear operator on W, called restriction to W.

14.5 Review Questions

1. Show that for the matrix A

$$A = \begin{bmatrix} 2 & -2 & -4 \\ -1 & 3 & 4 \\ 1 & -2 & -3 \end{bmatrix}$$

$A^2 = A$.

Find the characteristic values of A.

2. Show that every matrix A such that $A^2 = A$ is similar to a diagonal matrix.

14.6 Further Readings

Books

Kenneth Hoffman and Ray Kunze, *Linear Algebra*

Michael Artin, *Algebra*

Richa Nandra, Lovely Professional University

Notes

Unit 15: Simultaneous Triangulation and Simultaneous Diagonalization

CONTENTS

Objectives

After studying this unit, you will be able to:

- Know the structure of the triangular form of a matrix of a linear operator T on a space V over the field F.
- Understand that we can diagonalize two or more commuting matrices simultaneously.
- Know that the matrix of a linear operator T commutes with that of a polynomial of a linear operator T.

Introduction

In this unit we are again exploring the properties of a linear operator on the space$V$ over the field F.

In an upper triangular or lower triangular matrix the elements in the diagonal are the characteristic values.

15.1 Simultaneous Triangulation and Simultaneous Diagonalization

Let V be a finite-dimensional space and let $\mathcal{F}$ be a family of linear operators on V. We ask when we can simultaneously triangulate or diagonalize the operators in $\mathcal{F}$, i.e., find one basis $\mathcal{B}$ such that all of the matrices $[T]\mathcal{B}$, T in $\mathcal{F}$, are triangular (or diagonal). In the case of diagonalization, it is necessary that $\mathcal{F}$ be a commuting family of operators: $UT = TU$ for all T, U in $\mathcal{F}$. That follows from the fact that all diagonal matrices commute. Of course, it is also necessary that each operator in $\mathcal{F}$ be a diagonalizable operator. In order to simultaneously triangulate, each operator in $\mathcal{F}$ must be triangulable. It is not necessary that $\mathcal{F}$ be a commuting family; however that condition is sufficient for simultaneous triangulation (if each T can be individually triangulated). These results follow from minor variations of the proofs of Theorems 1 and 2 of unit 14.

The subspace W is invariant under (the family of operators) $\mathcal{F}$ if W is invariant under each operator in $\mathcal{F}$.

Lemma: Let $\mathcal{F}$ be a commuting family of triangulable linear operator on V. Let W be a proper subspace of V which is invariant under $\mathcal{F}$. There exists a vector α in V such that

(a) α is not in W;

(b) for each T in $\mathcal{F}$, the vector $T\alpha$ is in the subspace spanned by α and W.

Proof: It is no loss of generality to assume that $\mathcal{F}$ contains only a finite number of operators, because of this observation. Let $\{T_1,...,T_n\}$ be a maximal linearly independent subset of $\mathcal{F}$, i.e., a basis for the subspace spanned by $\mathcal{F}$. If α is a vector such that (b) holds for each T_i, then (b) will hold for every operator which is a linear combination of $T_1,..., T_r$.

By the lemma before Theorem 1 of unit 14 (this lemma for a single operator), we can find a vector β_1 (not in W) and a scalar c_1 such that $(T_1 - c_1I)\beta_1$ is in W. Let V_1 be the collection of all vectors β in V such that $(T_1 - c_1I)\beta$ is in W. Then V_1 is a subspace of V which is properly larger than W. Furthermore, V_1 is invariant under $\mathcal{F}$, for this reason. If T commutes with T_1, then

$$(T_1 - c_1I)(T\beta) = T(T_1 - c_1I)\beta$$

If β is in V_1, then $(T_1 - c_1I)\beta$ is in W. Since W is invariant under each T in $\mathcal{F}$, we have $T(T_1 - c_1I)\beta$ in W, i.e., $T\beta$ in V_1, for all β in V_1 and all T in $\mathcal{F}$.

Now W is a proper subspace of V_1. Let U_2 be the linear operator on V_1 obtained by restricting T_2 to the subspace V_1. The minimal polynomial for U_2 divides the minimal polynomial for T_2. Therefore, we may apply the lemma before Theorem 1 of unit 14 to that operator and the invariant subspace W. We obtain a vector β_2 in V_1 (not in W) and a scalar c_2 such that $(T_2 - c_2I)\ \beta_2$ is in W. Note that

(a) β_2 is not in W;

(b) $(T_1 - c_1I)\beta_2$ is in W;

(c) $(T_2 - c_2I)\beta_2$ is in W.

Let V_2 be the set of all vectors β in V_1 such that $(T_2 - c_2I)\beta$ is in W. Then V_2 is invariant under $\mathcal{F}$. Apply the lemma before Theorem 1 of unit 14 to U_3, the restriction of T_3 to V_2. If we continue in this way, we shall reach a vector $\alpha = \beta_r$ (not in W) such that $(T_j - c_jI)\alpha$ is in W, $j = 1,..., r$.

Theorem 1: Let V be a finite-dimensional vector space over the field F. Let $\mathcal{F}$ be a commuting family of triangulable linear operators on V. There exists an ordered basis for V such that every operator in $\mathcal{F}$ is represented by a triangular matrix in that basis.

Proof: Given the lemma which we just proved, this theorem has the same proof as does Theorem 1 of unit 14, if one replaces T by $\mathcal{F}$.

Corollary: Let $\mathcal{F}$ be a commuting family of $n \times n$ matrices over an algebraically closed field F. There exists a non-singular $n \times n$ matrix P with entries in F such that $P^{-1}AP$ is upper-triangular, for every matrix A in $\mathcal{F}$.

Theorem 2: Let F be a commuting family of diagonalizable linear operators on the finite-dimensional vector space V. There exists an ordered basis for V such that every operator in $\mathcal{F}$ is represented in that basis by a diagonal matrix.

Proof: We could prove this theorem by adapting the lemma before Theorem 1 to the diagonalizable case, just as we adapted the lemma before Theorem 1 of unit 14 to the diagonalizable case in order to prove Theorem 2 of unit 14. However, at this point it is easier to proceed by induction on the dimension of V.

If dim $V = 1$, there is nothing to prove. Assume the theorem for vector spaces of dimension less than n, and let V be an n-dimensional space. Choose any T in $\mathcal{F}$ which is not a scalar multiple of the identity. Let $c_1,..., c_k$ be the distinct characteristic values of T, and (for each i) let W_i be the null space of $T - c_iI$. Fix an index i. Then W_i is invariant under every operator which commutes with T. Let $\mathcal{F}_i$ be the family of linear operators on W_i obtained by restricting the operators in $\mathcal{F}$ to the (invariant) subspace W_i. Each operator in $\mathcal{F}_i$ is diagonalizable, because its minimal polynomial divides the minimal polynomial for the corresponding operator in $\mathcal{F}$. Since dim $W_i <$ dim V, the operators in $\mathcal{F}_i$ can be simultaneously diagonalized. In other words, W_i has a basis $\mathcal{B}_i$ which consists of vectors which are simultaneously characteristic vectors for every operator in $\mathcal{F}_i$.

Since T is diagonalizable, the lemma before Theorem 2 of unit 12 tells us that $\mathcal{B} = (\mathcal{B}_1, ..., \mathcal{B}_k)$ is a basis for V. That is the basis we seek.

If we consider finite dimensional vector space V over a complex field F, then there is a basis such that the matrix of the linear operator T is diagonal. This is due to the key fact that every complex polynomial of positive degree has a root. This tells us that every linear operator has at least one eigenvector.

From the theorem above we now have that every complex $n \times n$ matrix A is similar to an upper triangular matrix i.e. there is a matrix P, such that $P^{-1} AP$ is upper triangular.

Equally we also state that for a linear operator T on a finite dimensional complex vector space V, there is a basis β of V such that the matrix of T with respect to that basis is upper triangular.

Let V contain an eigenvector of A, call it v_1. Let λ be its eigenvalue. We extend (v_1) to a Basis $\beta = (v_1, v_2, \ldots, v_n)$ for V. There will be a matrix P for which the new matrix $A' = P^{-1} A\, P$ has the block form

$$A' = \left[\begin{array}{c|c} \lambda & * \\ \hline O & D \end{array}\right]$$

where D is an $(n - 1) \times (n - 1)$ matrix, λ is a 1 × 1 matrix of the restriction of T to $W(v_1)$. Here O denotes $n - 1$ zeros below λ in the first column. By induction on n, we may assume that there exists a matrix Q such that $Q^{-1} D\, Q$ is upper triangular. If we denote Q_1 by the relation

$$Q_1 = \left[\begin{array}{c|c} 1 & O \\ \hline O & Q \end{array}\right]$$

then

$$A'' = Q_1^{-1} A' Q_1 = \left[\begin{array}{c|c} \lambda & * \\ \hline O & Q^{-1}DQ \end{array}\right]$$

is the upper triangular and thus

$$A'' = (P\, Q_1)^{-1} A\, (P\, Q_1).$$

Knowing one vector v corresponding to the characteristic value λ we can find a linear operator P and then Q_1 to find A''.

Self Assessment

1. Find an invertible real matrix P such that $P^{-1}AP$ and $P^{-1}BP$ are both diagonal, where A and B are the real matrices

   (a) $A = \begin{bmatrix} 1 & 2 \\ 0 & 2 \end{bmatrix}$, $\quad B = \begin{bmatrix} 3 & -8 \\ 0 & -1 \end{bmatrix}$

   (b) $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$, $\quad B = \begin{bmatrix} 1 & a \\ a & 1 \end{bmatrix}$

2. Let $\mathcal{F}$ be a commuting family of 3 × 3 complex matrices. How many linearly independent matrices can $\mathcal{F}$ contain? What about the $n \times n$ case?

15.2 Summary

- In this unit we are dealing with matrices that commute with each other.
- In a triangular matrix the main diagonal has the entries of the characteristic values and it has not zero entries in the upper part of the diagonal only or non-zero entries in the lower of the main diagonal.
- If two or more matrices commute then we can diagonalize them simultaneously.

15.3 Keywords

Diagonalizable: Each operator in $\mathcal{F}_i$ is diagonalizable, because its minimal polynomial divides the minimal polynomial for the corresponding operator in $\mathcal{F}$.

Ordered Basis: There exists an ordered basis for V such that every operator in $\mathcal{F}$ is represented by a triangular matrix in that basis.

15.4 Review Question

1. Let T be a linear operator on a n-dimension space and suppose that T has n distinct characteristic values. Prove that any linear operator which commutes with T is a polynomial in T.

Answers: Self Assessment

1. (a) $P = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$, (b) $P = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$

2. 3, n

15.5 Further Readings

Books Kenneth Hoffman and Ray Kunze, *Linear Algebra*

I.N. Herstein, *Topics in Algebra*

Sachin Kaushal, Lovely Professional University

Unit 16: Direct Sum Decompositions of Elementary Canonical Forms

CONTENTS

Objectives

After studying this unit, you will be able to:

- Understand the meanings of invariant subspaces as well as decomposition of a vector space into the invariant direct sums of the independent subspaces.
- Know the projection operators and their properties
- See that there is less emphasis is on matrices and more attention is given to subspaces.

Introduction

This unit and the next units are slightly more complicated than the other previous units. The ideas of invariant subspaces and their relations with the vector space V is obtained.

The ideas of projection operators and their properties are introduced. These ideas will help in expressing the given linear operator T in terms of the direct sums of the operators T_{1j} T_K as seen in the next unit.

16.1 Overview

We are again going to analyse a single linear operator on a finite dimensional space V over the field F. In the next three units we shall stress less in terms of matrices and stress more on the subspaces, in order to find an ordered basis in which the matrix of T assumes an especially a simple form. Our aim in three units will be as follows: To decompose the underlying space V into a sum of invariant subspaces for T such that the restriction operators on these subspaces are simple. These subspaces will be taken as independent subspaces of the vector space V and after finding the independent basis of each independent subspace the ordered basis of the whole space will be constructed. Given such a decomposition of the vector space we then see that T induces a linear operator T_i on each subspace W_i, by restriction. We shall describe this situation by saying that the linear operator is the invariant direct sum of the operators $T_1, T_2, \ldots, T_k$. Once the space is decomposed in terms of invariant subspaces, we shall introduce the concepts of projection operators on V.

16.2 Direct-sum Decompositions

Definition: Let $W_1,..., W_k$ be subspaces of the vector space V. We say that $W_1,..., W_k$ are independent if

$$\alpha_1 + ... + \alpha_k = 0, \qquad \alpha_i \text{ in } W_i$$

implies that each α_i is 0.

For $k = 2$, the meaning of independence is {0} intersection, i.e., W_1 and W_2 are independent if and only if $W_1 \cap W_2 = \{0\}$. If $k > 2$, the independence of $W_1,..., W_k$ says much more than $W_1 \cap ... \cap W_k = \{0\}$. It says that each W_j intersects the sum of the other subspaces W_i only in the zero vector.

The significance of independence is this. Let $W = W_1 + ... + W_k$ be the subspace spanned by $W_1,..., W_k$. Each vector α in W can be expressed as a sum

$$\alpha = \alpha_1 + ... + \alpha_k, \qquad \alpha_i \text{ in } W_i.$$

If $W_1,..., W_k$ are independent, then that expression for α is unique; for if

$$\alpha = \beta_1 + ... + \beta_k, \qquad \beta_i \text{ in } W_i$$

then $0 = (\alpha_1 - \beta_1) + ... + (\alpha_k - \beta_k)$, hence $\alpha_i - \beta_i = 0$, $i = 1,..., k$. Thus, when $W_1,..., W_k$ are independent, we can operate with the vectors in W as k-tuples $(\alpha_1,..., \alpha_k)$, α_i in W_i, in the same way as we operate with vectors in R^k as k-tuples of numbers.

Lemma: Let V be a finite-dimensional vector space. Let $W_1,..., W_k$ be subspaces of V and let $W = W_1 + ... + W_k$. The following are equivalent.

(a) $W_1,..., W_k$ are independent.

(b) For each j, $2 \leq j \leq k$, we have

$$W_j \cap (W_1 + ... + W_{j-1}) = \{0\}$$

(c) If $\mathcal{B}_i$ is an ordered basis for W_i, $1 \leq i \leq k$, then the sequence $\mathcal{B} = (\mathcal{B}_1,..., \mathcal{B}_k)$ is an ordered basis for W.

Proof: Assume (a). Let α be a vector in the intersection $W_j \cap (W_1 + ... + W_{j-1})$. Then there are vectors $\alpha_1,..., \alpha_{j-1}$ with α_i in W_i such that $\alpha = \alpha_1 + ... + \alpha_{j-1}$. Since

$$\alpha_1 + ... + \alpha_{j-1} + (-\alpha) + 0 + ... + 0 = 0$$

and since $W_1, ..., W_k$ are independent, it must be that $\alpha_1 = \alpha_2 = ... = \alpha_{j-1} = \alpha = 0$.

Now, let us observe that (b) implies (a). Suppose

$$0 = \alpha_1 + ... + \alpha_k, \qquad \alpha_i \text{ in } W_i$$

Let j be the largest integer i such that $\alpha_i \neq 0$. Then

$$0 = \alpha_1 + ... + \alpha_j, \qquad \alpha_j \neq 0.$$

Thus $\alpha_j = -\alpha_1 - ... - \alpha_{j-1}$ is a non-zero vector in $W_j \cap (W_1 + ... + W_{j-1})$.

Now that we know (a) and (b) are the same, let us see why (a) is equivalent to (c). Assume (a). Let $\mathcal{B}_i$ be basis for W_i, $1 \leq i \leq k$, and let $\mathcal{B} = (\mathcal{B}_1,..., \mathcal{B}_k)$. Any linear relation between the vectors in $\mathcal{B}$ will have the form

$$\beta_1 + ... + \beta_k = 0$$

where β_i is some linear combination of the vectors in $\mathcal{B}_i$. Since $W_1,.., W_k$ are independent, each β_i is 0. Since each $\mathcal{B}_i$ is independent, the relation we have between the vectors in $\mathcal{B}$ is the trivial relation.

Notes

If any (and hence all) of the conditions of the last lemma hold, we say that the sum $W = W_1 + ... + W_k$ is direct or that W is the direct sum of $W_1,..., W_k$ and we write

$$W = W_1 \oplus \cdots \oplus W_k$$

In the literature, the reader may find this direct sum referred to as an independent sum or the interior direct sum of $W_1,..., W_k$.

Example 1: Let V be a finite-dimensional vector space over the field F and let $\{\alpha_1,..., \alpha_n\}$ be any basis for V. If W_i is the one-dimensional subspace spanned by α_i, then $V = W_1 \oplus \cdots \oplus W_n$.

Example 2: Let n be a positive integer and F a subfield of the complex numbers, and let V be the space of all $n \times n$ matrices over F. Let W_1 be the subspace of all symmetric matrices, i.e., matrices A such that $A^t = A$. Let W_2 be the subspace of all skew-symmetric matrices, i.e., matrices A such that $A^t = -A$. Then $V = W_1 \oplus W_2$. If A is any matrix in V, the unique expression for A as a sum of matrices, one in W_1 and the other in W_2, is

$$A = A_1 + A_2$$

$$A_1 = \frac{1}{2}(A + A^t)$$

$$A_2 = \frac{1}{2}(A - A^t)$$

Example 3: Let T be any linear operator on a finite-dimensional space V. Let $c_1,.., c_k$ be the distinct characteristic values of T, and let W_i be the space of characteristic vectors associated with the characteristic value c_i. Then $W_1,..., W_k$ are independent. In particular, if T is diagonalizable, then $V = W_1 \oplus \cdots \oplus W_k$.

Definition: If V is a vector space, a projection of V is a linear operator E on V such that $E^2 = E$.

Suppose that E is a projection. Let R be the range of E and let N be the null space of E.

1. The vector β is in the range R if and only if $E\beta = \beta$. If $\beta = E\alpha$, then $E\beta = E^2\alpha = E\alpha = \beta$. Conversely, if $\beta = E\beta$, then (of course) β is in the range of E.
2. $V = R \oplus N$.
3. The unique expression for α as a sum of vectors in R and N is $\alpha = E\alpha + (\alpha - E\alpha)$.

From (1), (2), (3) it is easy to see the following. If R and N are subspaces of V such that $V = R \oplus N$, there is one and only one projection operator E which has range R and null space N. That operator is called the projection on R along N.

Any projection E is (trivially) diagonalizable. If $\{\alpha_1,..., \alpha_r\}$ is a basis for R and $\{\alpha_{r+1},..., \alpha_n\}$ a basis for N, then the basis $\mathcal{B} = \{\alpha_1,..., \alpha_n\}$ diagonalizes E.

$$[E]_{\mathcal{B}} = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}$$

where I is the $r \times r$ identity matrix. That should help explain some of the terminology connected with projections. The reader should look at various cases in the plane R^2 (or 3-space, R^3), to convince himself that the projection on R along N sends each vector into R by projecting it parallel to N.

Projections can be used to describe direct-sum decompositions of the space V. For, suppose $V = W_1 \oplus \cdots \oplus W_k$. For each j we shall define an operator E_j on V. Let α be in V, say $\alpha = \alpha_1 + \cdots + \alpha_k$ with α_i in W_i. Define $E_j\alpha = \alpha_j$. Then E_j is a well-defined rule. It is easy to see that E_j is linear, that the range of E_j is W_j, and that $E_j^2 = E_j$. The null space of E_j is the subspace

$$(W_1 + \cdots + W_{j-1} + W_{j+1} + \cdots + W_k)$$

for, the statement that $E_j\alpha = 0$ simply means $\alpha_j = 0$, i.e., that α is actually a sum of vectors from the spaces W_i with $i \neq j$. In terms of the projection E_j we have

$$\alpha = E_1\alpha + \cdots + E_k\alpha$$

for each α in V. What (1) says is that

$$I = E_1 + \cdots + E_k$$

Note also that if $i \neq j$, then $E_iE_j = 0$, because the range of E_j is the subspace W_j which is contained in the null space of E_i. We shall now summarize our findings and state and prove a converse.

Theorem 1: If $V = W_1 \oplus \cdots \oplus W_k$, then there exist k linear operators $E_1,..., E_k$ on V such that

(i) each E_i is a projection $(E_1^2 = E_i)$;

(ii) $E_iE_j = 0$, if $i \neq j$;

(iii) $I = E_1 + \cdots + E_k$;

(iv) the range of E_i is W_i.

Conversely, if $E_1,..., E_k$ are k linear operators on V which satisfy conditions (i), (ii) and (iii), and if we let W_i be the range of E_i, then $V = W_i \oplus \cdots \oplus W_k$.

Proof: We have only to prove the converse statement. Suppose $E_1,..., E_k$ are linear operators on V which satisfy the first three conditions, and let W_i be the range of E_i. Then certainly

$$V = W_1 + \cdots + W_k;$$

for, by condition (iii) we have

$$\alpha = E_1\alpha + \cdots + E_k\alpha$$

for each α in V, and $E_i\alpha$ is in W_i. This expression for α is unique, because if

$$\alpha = \alpha_1 + \cdots + \alpha_k$$

with α_i in W_i, say $\alpha_i = E_i\beta_i$, then using (i) and (ii) we have

$$\begin{aligned} E_j\alpha &= \sum_{i=1}^{k} E_j\alpha_i \\ &= \sum_{i=1}^{k} E_jE_i\beta_i \\ &= E_j^2\beta_j \\ &= E_j\beta_j \\ &= \alpha_j \end{aligned}$$

This shows that V is the direct sum of the W_i.

Self Assessment

1. Let V be a finite dimensional vector space and W_1 is any subspace of V. Prove that there is a subspace W_2 of V such that $V = W_1 \oplus W_2$.

2. True or false? If a diagonalizable operator has only the characteristic values 0 and 1, it is a projection.

3. Let $E_1, E_2, \ldots E_K$ be linear operators on the space V such that $E_1 + E_2 + \ldots + E_K = I$. Prove that if $E_iE_j = 0$ for $i \neq j$, then $E_i^2 = E_i$ for each i.

4. Let V be a finite dimensional vector space and let $W_1, \ldots W_K$ be subspaces of V such that

$$V = W_1 + W_2 + \ldots + W_K \text{ and } \dim V = \dim W_1 + \ldots + W_K$$

Prove that $V = W_1 \oplus W_2 \oplus \ldots \oplus W_K$.

16.3 Summary

- In this unit the importance is given to the ideas of invariant subspaces of a vector space V for a linear operator T.
- The vector space V is decomposed into a set of linear invariant subspaces.
- The sum of the bases vectors of the invariant subspaces defines the basis vectors of the vector space V.

16.4 Keywords

Skew-symmetric Matrices: Skew-symmetric matrices, i.e., matrices A such that $A^t = -A$.

Subspaces: These subspaces will be taken as independent subspaces of the vector space V and after finding the independent basis of each independent subspace the ordered basis of the whole space will be constructed.

16.5 Review Questions

1. If E_1, E_2 are projections onto independent subspaces, then $E_1 + E_2$ is a projection. True or false?

2. Let E_1, E_2 be linear operators on the space V such that $E_1 + E_2 = I$, and $E_1^2 = E_1$ and $E_2^2 = E_2$, then prove that $E_1E_2 = 0$.

Answer: Self Assessment

2. Yes, true

16.6 Further Readings

Books

Kenneth Hoffman and Ray Kunze, *Linear Algebra*

I.N. Herstein, *Topics in Algebra*

Unit 17: Invariant Direct Sums

CONTENTS

Objectives

After studying this unit, you will be able to:

- See that the vector space *V* is decomposed as a direct sum of the invariant subspaces under some linear operator *T*.
- Understand that the linear operator induces a linear operator T_i on each invariant subspaces W_i by restriction.
- Know that if α_i is the vector in the invariant subspace W_i then the vector α in the finite vector space *V* is obtained as a linear combinations of its projections α_i in the subspace W_i.

Introduction

In this unit we again consider a linear transformation *T* on the finite vector space. Here the vector space is decomposed as the direct sum of the invariant subspaces W_i. The linear operator induces a linear operator T_i for each invariant subspaces W_i.

The method of finding the projection operators and their properties is discussed.

17.1 Overview

In this unit we are primarily interested in the direct sum decomposition $V = W_1 \oplus W_2 \oplus + ... + W_K$, where each of the subspaces W_i is invariant under some linear operator *T*. Given such a decomposition of *V*, *T* induces a linear operator T_i on each W_i by restriction. If α_i is the vector in W_i then the vector α in *V* can be given as a linear combinations of its projection α_i in the invariant subspace W_i. Thus the action of *T* is then understood as follows:

If α is a vector in *V*, we have unique vectors $\alpha_1,..., \alpha_k$ with α_i in W_i such that

$$\alpha = \alpha_1 + ... + \alpha_k$$

and then

$$T\alpha = T_1\alpha_1 + ... + T_k\alpha_k$$

We shall describe this situation by saying that *T* is the direct sum of the operators $T_1,..., T_k$. It must be remembered in using this terminology that the T_i are not linear operators on the space *V* but on the various subspaces W_i. The fact that $V = W_1 \oplus ... \oplus W_k$ enables us to associate with each α in

V a unique k-tuple $(\alpha_1,..., \alpha_k)$ of vectors α_i in W_i (by $\alpha = \alpha_1 + ... + \alpha_k$) in such a way that we can carry out the linear operations in V by working in the individual subspaces W_i. The fact that each W_i is invariant under T enables us to view the action of T as the independent action of the operators T_i on the subspaces W_i. Our purpose is to study T by finding invariant direct-sum decompositions in which the T_i are operators of an elementary nature.

Before looking at an example, let us note the matrix analogue of this situation. Suppose we select an ordered basis $\mathcal{B}_i$ for each W_i, and let it be the ordered basis for V consisting of the union of the $\mathcal{B}_i$ arranged in the order $\mathcal{B}_1,..., \mathcal{B}_k$, so that $\mathcal{B}$ is a basis for V. From our discussion concerning the matrix analogue for a single invariant subspace, it is easy to see that if $A = [T]_{\mathcal{B}}$ and $A_i = [T_i]_{\mathcal{B}'}$, then A has the block form

$$A = \begin{bmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & A_k \end{bmatrix} \qquad ...(1)$$

In (1), A_i is a $d_i \times d_i$ matrix ($d_i = \dim W_i$), and the 0′s are symbols for rectangular blocks of scalar 0′s of various sizes. It also seems appropriate to describe (1) by saying that A is the direct sum of the matrices $A_1,..., A_k$.

Most often, we shall describe the subspace W_i by means of the associated projections E_i (Theorem 1 of unit 16). Therefore, we need to be able to phrase the invariance of the subspaces W_i in terms of the E_i.

17.2 Some Theorems

Theorem 1: Let T be a linear operator on the space V, and $W_1,..., W_k$ and $E_1,..., E_k$ be as in Theorem 1 of unit 16. Then a necessary and sufficient condition that each subspace W_i be invariant under T is that T commutes with each of the projections E_i, i.e.,

$$TE_i = E_iT, \qquad i = 1,..., k$$

Proof: Suppose T commutes with each E_i. Let α be in W_j. Then $E_j\alpha = \alpha$, and

$$\begin{aligned} T\alpha &= T(E_j\alpha) \\ &= E_j(T\alpha) \end{aligned}$$

which shows that $T\alpha$ is in the range of E_j, i.e., that W_j is invariant under T.

Assume now that each W_i is invariant under T. We shall show that $TE_j = E_jT$. Let α be any vector in V. Then

$$\begin{aligned} \alpha &= E_1\alpha + ... + E_k\alpha \\ T\alpha &= TE_1\alpha + ... + TE_k\alpha \end{aligned}$$

Since $E_i\alpha$ is in W_i, which is invariant under T, we must have $T(E_i\alpha) = E_i\beta_i$ for some vector β_i. Then

$$\begin{aligned} E_jTE_i\alpha &= E_jE_i\beta_i \\ &= \begin{cases} 0, & \text{if } i \neq j \\ E_j\beta_j, & \text{if } i = j \end{cases} \end{aligned}$$

Thus

$$\begin{aligned} E_jT\alpha &= E_jTE_1\alpha + ... + E_jTE_k\alpha \\ &= E_j\beta_j \end{aligned}$$

$$= TE_j\alpha$$

This holds for each α in V, so $E_jT = TE_j$.

We shall now describe a diagonalizable operator T in the language of invariant direct sum decompositions (projections which commute with T). This will be a great help to us in understanding some deeper decomposition theorems later. The description which we are about to give is rather complicated, in comparison to the matrix formulation or to the simple statement that the characteristic vectors of T span the underlying space. But, we should bear in mind that this is our first glimpse at a very effective method, by means of which various problems concerned with subspaces, bases, matrices, and the like can be reduced to algebraic calculations with linear operators. With a little experience, the efficiency and elegance of this method of reasoning should become apparent.

Theorem 2: Let T be a linear operator on a finite-dimensional space V. If T is diagonalizable and if $c_1,..., c_k$ are the distinct characteristic values of T, then there exist linear operators $E_1,..., E_k$ on V such that

(i) $T = c_1E_1 + ... + c_kE_k$;

(ii) $I = 'E_1 + ... + E_k$;

(iii) $E_iE_j = 0, i \neq j$;

(iv) $E_1^2 = E_i$ (E_i is a projection);

(v) the range of E_i is the characteristic space for T associated with c_i.

Conversely, if there exist k distinct scalars $c_1,..., c_k$ and k non-zero linear operators $E_1,..., E_k$ which satisfy conditions (i), (ii), and (iii), then T is diagonalizable, $c_1,..., c_k$ are the distinct characteristic values of T, and conditions (iv) and (v) are satisfied also.

Proof: Suppose that T is diagonalizable, with distinct characteristic values $c_1,..., c_k$. Let W_i be the space of characteristic vectors associated with the characteristic value c_i. As we have seen,

$$V = W_1 \oplus ... \oplus W_k$$

Let $E_1,...,E_k$ be the projections associated with this decomposition, as in Theorem 1 of unit 16. Then (ii), (iii), (iv) and (v) are satisfied. To verify (i), proceed as follows. For each α in V,

$$\alpha = E_1\alpha + ... + E_k\alpha$$

and so

$$T\alpha = TE_1\alpha + ... + TE_k\alpha$$

$$= c_1E_1\alpha + ... + c_kE_k\alpha$$

In other words, $T = c_1E_1 + ... + c_kE_k$.

Now suppose that we are given a linear operator T along with distinct scalars c_i and non-zero operators E_i which satisfy (i), (ii) and (iii). Since $E_iE_j = 0$ when $i \neq j$, we multiply both sides of $I = E_1 + ... + E_k$ by E_i and obtain immediately $E_i^2 = E_i$. Multiplying $T = c_1E_1 + ... + c_kE_k$ by E_i, we then have $TE_i = c_iE_i$, which shows that any vector in the range of E_i is in the null space of $(T - c_iI)$. Since we have assumed that $E_i \neq 0$, this proves that there is a non-zero vector in the null space of $(T - c_iI)$, i.e., that c_i is a characteristic value of T. Furthermore, the c_i are all of the characteristic values of T; for, if c is any scalar, then

$$T - cI = (c_1 - c)E_1 + ... + (c_k - c)E_k$$

so if $(T - cI)\alpha = 0$, we must have $(c_i - c)E_i\alpha = 0$. If α is not the zero vector, then $E_i\alpha \neq 0$ for some i, so that for this i we have $c_i - c = 0$.

Certainly T is diagonalizable, since we have shown that every non-zero vector in the range of E_i is a characteristic vector of T, and the fact that $I = E_1 + ... + E_k$ shows that these characteristic vectors span V. All that remains to be demonstrated is that the null space of $(T - c_iI)$ is exactly the range of E_i. But this is clear, because if $T\alpha = c_i\alpha$, then

$$\sum_{j=1}^{k}(c_j - c_i)E_j\alpha = 0$$

hence

$$(c_j - c_i)E_j\alpha = 0 \qquad \text{for each } j$$

and then

$$E_j\alpha = 0 \qquad j \neq i$$

Since $\alpha = E_1\alpha + ... + E_k\alpha$, and $E_j\alpha = 0$ for $j \neq i$, we have $\alpha = E_i\alpha$, which proves that α is in the range of E_i.

One part of Theorem 1 of unit 16 says that for a diagonalizable operator T, the scalars $c_1,..., c_k$ and the operators $E_1,..., E_k$ are uniquely determined by conditions (i), (ii), (iii), the fact that the c_i are distinct, and the fact that the E_i are non-zero. One of the pleasant features of the decomposition $T = c_1E_1 + ... + c_kE_k$ is that if g is any polynomial over the field F, then

$$g(T) = g(c_1)E_1 + ... + g(c_k)E_k.$$

To see how it is proved one need only compute T^r for each positive integer r. For example,

$$\begin{aligned}\mathrm{T}^2 &= \sum_{i=1}^{k}c_iE_i\sum_{j=1}^{k}c_jE_j\\ &= \sum_{i=1}^{k}\sum_{j=1}^{k}c_ic_jE_iE_j\\ &= \sum_{i=1}^{k}c_i^2E_i^2\\ &= \sum_{i=1}^{k}c_i^2E_i\end{aligned}$$

The reader should compare this with $g(A)$ where A is a diagonal matrix; for then $g(A)$ is simply the diagonal matrix with diagonal entries $g(A_{11})$, ..., $g(A_{nn})$.

We should like in particular to note what happens when one applies the Lagrange polynomials corresponding to the scalars $c_1,..., c_k$:

$$p_j = \prod_{i\neq j}\frac{(x - c_i)}{(c_j - c_i)}$$

We have $p_j(c_i) = \delta_{ij}$, which means that

$$\begin{aligned}p_j(T) &= \sum_{i=1}^{k}\delta_{ij}E_i\\ &= E_j\end{aligned}$$

Thus the projections E_j not only commute with T but are polynomials in T.

Such calculations with polynomials in T can be used to give an alternative proof of Theorem 2 of unit 14, which characterized diagonalizable operators in terms of their minimal polynomials. The proof is entirely independent of our earlier proof.

If T is diagonalizable, $T = c_1E_1 + ... + c_kE_k$, then

$$g(T) = g(c_1)E_1 + ... + g(c_k)E_k$$

for every polynomial g. Thus $g(T) = 0$ if and only if $g(c_i) = 0$ for each i. In particular, the minimal polynomial for T is

$$p = (x - c_1) ... (x - c_k)$$

Now suppose T is a linear operator with minimal polynomial $p = (x - c_1) ... (x - c_k)$, where $c_1,..., c_k$ are distinct elements of the scalar field. We form the Lagrange polynomials

$$p_j = \prod_{i \neq j} \frac{(x - c_i)}{(c_j - c_i)}$$

So that $p_j(c_i) = \delta_{ij}$ and for any polynomial g of degree less than or equal to $(k - 1)$ we have

$$g = g(c_1)p_1 + ... + g(c_k)p_k$$

Taking g to be the scalar polynomial 1 and then the polynomial x, we have

$$\left.\begin{aligned} 1 &= p_1 + \cdots + p_k \\ x &= c_1p_1 + \cdots + c_kp_k \end{aligned}\right\} \qquad ...(2)$$

You will note that the application to x may not be valid because k may be 1. But if $k = 1$, T is a scalar multiple of the identity and hence diagonalizable). Now let $E_j = p_j(T)$. From (2) we have

$$\left.\begin{aligned} I &= E_1 + \cdots + E_k \\ T &= c_1E_1 + \cdots + c_kE_k \end{aligned}\right\} \qquad ...(3)$$

Observe that if $i \neq j$, then p_ip_j is divisible by the minimal polynomial p, because p_ip_j contains every $(x - c_r)$ as a factor. Thus

$$E_iE_j = 0, \qquad i \neq j \qquad ...(4)$$

We must note one further thing, namely, that $E_i \neq 0$ for each i. This is because p is the minimal polynomial for T and so we cannot have $p_i(T) = 0$ since p_i has degree less than the degree of p. This last comment, together with (3), (4), and the fact that the c_i are distinct enables us to apply Theorem 2 to conclude that T is diagonalizable.

Self Assessment

1. Let T be the diagonalizable linear operator on R^3 which is represented by the matrix

$$A = \begin{bmatrix} 5 & -6 & -6 \\ -1 & 4 & 2 \\ 3 & -6 & -4 \end{bmatrix}$$

use the Lagrange polynomials to write the representing matrix A in the form $A = E_1 + 2E_2$, $E_1 + E_2 = I$, $E_1E_2 = 0$. Where I is a unit matrix and 0 is zero matrix.

2. Let T be the linear operator on R^4 which is represented by the 4 × 4 matrix

$$A = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}$$

Find the matrices E_1, E_2, E_3 such that

$A = C_1E_1 + C_2E_2 + C_3E_3$, $E_1 + E_2 + E_3 = I$ and $E_iE_j = 0$ for $i \neq j$

17.3 Summary

- In this unit the finite dimensional vector space is decomposed into a direct sum of the invariant subspaces.
- The linear operator induces a linear operator T_i on each invariant subspace W_i by restriction.
- The projection operators can be obtained from the Lagrange polynomials once we know the characteristic values.

17.4 Keywords

Projection Operator: The projection operator E has the property that $E^2 = E$ so its characteristic values can be equal to 0 and unit.

Restriction: When the finite space V is decomposed into the direct sum of the invariant subspaces the linear operator induces a linear operator by the process known as restriction.

The Lagrange Polynomials: Help us to find the projection operators for any linear operator T in terms of the matrix representing T and its characteristic values.

17.5 Review Questions

1. Let T be a linear operator on V. Suppose $V = W_1 \oplus ... \oplus W_k$, where each W_i is invariant under T. Let T_i be the induced (restriction) operator on W_i. Prove that the characteristic polynomial for f is the product of the characteristic polynomials $f_1, f_2,..., f_k$.
2. Let T be a linear operator on three dimensional space which is represented by the matrix

$$A = \begin{bmatrix} 4 & 2 & -2 \\ -5 & 3 & 2 \\ -2 & 4 & 1 \end{bmatrix},$$

Find the matrices E_1, E_2, E_3 such that $A = C_1E_1 + C_2E_2 + C_3E_3$

$E_1 + E_2 + E_3 = I$, $E_iE_j = 0$ for $i \neq j$

Answers: Self Assessment

1. $E_1 = 2I - A$, $E_2 = A - I$, Here $E_1 + E_2 = I$, $A = E_1 + 2E_2$ and $E_1E_2 = 0$
2. Here $c_1 = 0$, $c_2 = -2$, $c_3 = 2$

$E_1 = I - A^2/4$

$$E^2 = \frac{1}{8}(A - 2I)A$$

$$E^3 = \frac{A}{8}(A - 2I)$$

17.6 Further Readings

Books

Kenneth Hoffman and Ray Kunze, *Linear Algebra*

I.N. Herstein, *Topics in Algebra*

Michael Artin, *Algebra*

Unit 18: The Primary Decomposition Theorem

CONTENTS

Objectives

After studying this unit, you will be able to:

- See that in considering a linear operator *T* on a finite dimensional space the minimal polynomial for the linear operator is a product of a number of irreducible monic polynomials $p_i^{r_i}$ over the field *F* where r_i are positive integers.
- Know that this structure of the minimal polynomial helps in decomposing the space *V* as the direct sum of the invariant subspaces W_i.
- Understand that the general linear operator *T* induces a linear operator T_i on W_i by restriction and the minimal polynomial for T_i is the irreducible $p_i^{r_i}$.

Introduction

In this unit the idea of the direct sum decomposition of the vector space *V* for a linear operator *T* in terms of invariant subspaces.

The general linear operator *T* induces a linear operator T_i on the invariant subspace, the minimal polynomial of T_i is the $p_i^{r_i}$.

This structure of the induced linear operator helps in introducing the projection operators E_i.

These projections associated with the primary decomposition of *T*, then are polynomials in *T*, and they commute each will an operator that commutes with *T*.

18.1 Overview

We continue our study of a linear operator *T* on the finite dimension space. In this unit we are interested in decomposing *T* into a direct sum of operators which are in some sense elementary. We had already found the characteristic values of the operator and also studied invariant subspaces. The vector space *V* was shown to be direct sum of the invariant subspaces. We can decompose *T* into a direct sum of operators through the characteristic values and vectors of *T* in certain special cases i.e., when the minimal polynomial for *T* factors over the scalar field *F* into a product of distinct monic polynomials of degree 1. In dealing with the general *T* we come across with two problems. First, T may not have a single characteristic value due to the limitation of the scalar field. Second, even if the characteristic polynomial factors completely over F into a product of polynomials of degree 1, there may not be enough characteristic vectors for T to span the space V; which is clearly a deficiency in T. The second situation is illustrated by the operator T on F^3 (F any field) represented in the standard basis by

$$A = \begin{bmatrix} 2 & 0 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

The characteristic polynomial for A is $(x - 2)^2 (x + 1)$ and this is plainly also the minimal polynomial for A (or for T). Thus T is not diagonalizable. One sees that this happens because the null space of $(T - 2I)$ has dimension 1 only. On the other hand, the null space of $(T + \mathrm{I})$ and the null space of $(T - 2I)^2$ together span V, the former being the subspace spanned by $\in_3$ and the latter the subspace spanned by $\in_1$ and $\in_2$.

This will be more or less our general method for the second problem. If (remember this is an assumption) the minimal polynomial for T decomposes

$$p = (x - c_1)^{r1} \dots (x - c_k)^{r2}$$

where $c_1, \dots, c_k$ are distinct elements of F, then we shall show that the space V is the direct sum of the null spaces of $(T - c_iI)^{ri}$, $i = 1, \dots, k$. The hypothesis about p is equivalent to the fact that T is triangulable (Theorem 1 of unit 14); however, that knowledge will not help us.

The theorem which we prove is more general than what we have described, since it works with the primary decomposition of the minimal polynomial, whether or not the primes which enter are all of first degree. The reader will find it helpful to think of the special case when the primes are of degree 1, and even more particularly, to think of the projection-type proof of Theorem 2 of unit 14, a special case of this theorem.

18.2 Primary Decomposition Theorem

Theorem 1 (Primary Decomposition Theorem): Let T be a linear operator on the finite-dimensional vector space V over the field F. Let p be the minimal polynomial for T,

$$p = p_1^{r_1} \cdots p_k^{r}$$

where the p_i are distinct irreducible monic polynomials over F and the r_i are positive integers. Let W_i be the null space of $p_i(T)^{r_i}$, $i = 1, \dots, k$. Then

(i) $V = W_1 \oplus \cdots \oplus W_k$;

(ii) each W_i is invariant under T;

(iii) if T_i is the operator induced on W_i by T, then the minimal polynomial for T_i is $p_1^{r_1}$.

Proof: The idea of the proof is this. If the direct-sum decomposition (i) is valid, how can we get hold of the projections $E_1, \dots, E_k$ associated with the decomposition? The projection E_i will be the identity on W_i and zero on the other W_j. We shall find a polynomial h_i such that $h_i(T)$ is the identity on W_i and is zero on the other W_j, and so that $h_1(T) + \cdots + h_k(T) = I$, etc.

For each i, let

$$f_i = \frac{p}{p_i^{r_i}} = \prod_{j \neq i} p_j^{r_i}.$$

Since $p_1,..., p_k$ are distinct prime polynomials, the polynomials $f_1,..., f_k$ are relatively prime. Thus there are polynomials $g_1,..., g_k$ such that

$$\sum_{i=1}^{n} f_i g_i = 1$$

Note also that if $i \neq j$, then $f_i f_j$ is divisible by the polynomial p, because $f_i f_j$ contains each $p_m^{r_m}$ as a factor. We shall show that the polynomials $h_i = f_i g_i$ behave in the manner described in the first paragraph of the proof.

Let $E_i = h_i(T) = f_i(T)g_i(T)$. Since $h_1 + \cdots + h_k = 1$ and p divides $f_i f_j$ for $i \neq j$, we have

$$E_1 + \cdots + E_k = I$$

$$E_iE_j = 0, \quad \text{if } i \neq j$$

Thus the E_i are projections which correspond to some direct sum decomposition of the space V. We wish to show that the range of E_i is exactly the subspace W_i. It is clear that each vector in the range of E_i is in W_i, for if α is in the range of E_i, then $\alpha = E_i\alpha$ and so

$$\begin{aligned} p_i(T)^{r_i}\alpha &= p_i(T)^{r_i}E_i\alpha \\ &= p_i(T)^{r_i}f_i(T)g_i(T)\alpha \\ &= 0 \end{aligned}$$

because $p^{r_i}f_i g_i$ is divisible by the minimal polynomial p. Conversely, suppose that α is in the null space of $p_i(T)^{r_i}$. If $j \neq i$, then $f_j g_j$ is divisible by $p_i^{r_i}$ and so $f_j(T)g_j(T)\alpha = 0$, i.e., $E_j\alpha = 0$ for $j \neq$ i. But then it is immediate that $E_i\alpha = \alpha$, i.e., that α is in the range of E_i. This completes the proof of statement (i).

It is certainly clear that the subspaces W_i are invariant under T. If T_i is the operator induced on W_i by T, then evidently $p_i(T_i)^{r_i} = 0$, because by definition $p_i(T)^{r_i}$ is 0 on the subspace W_i. This shows that the minimal polynomial for T_i divides $p_i^{r_i}$. Conversely, let g be any polynomial such that $g(T_i) = 0$. Then $g(T)f_i(T) = 0$. Thus gf_i is divisible by the minimal polynomial p of T, i.e., $p_i^{r_i} f_i$ divides gf_i. It is easily seen that $p_i^{r_i}$ divides g. Hence the minimal polynomial for T_i is $p_i^{r_i}$.

Corollary: If $E_1,..., E_k$ are the projections associated with the primary decomposition of T, then each E_i is a polynomial in T, and accordingly if a linear operator U commutes with T then U commutes with each of the E_i, i.e., each subspace W_i is invariant under U.

In the notation of the proof of Theorem 1, let us take a look at the special case in which the minimal polynomial for T is a product of first degree polynomials, i.e., the case in which each p_i is of the form $p_i = x - c_i$. Now the range of E_i is the null space W_i of $(T - c_iI)^{r_i}$. Let us put $D = c_1E_1 + \cdots + c_kE_k$. By Theorem 2 of unit 17, D is a diagonalizable operator which we shall call the diagonalizable part of T. Let us look at the operator $N = T - D$. Now

$$T = TE_1 + ... + TE_k$$

$$D = c_1E_1 + ... + c_kE_k$$

so

$$N = (T - c_1I)E_1 + ... + (T - c_kI)E_k$$

The reader should be familiar enough with projections by now so that he sees that

$$N^2 = (T - c_1I)^2E_1 + ... + (T - c_kI)^2E_k$$

and in general that

$$N^r = (T - c_1I)^rE_1 + ... + (T - c_kI)^rE_k.$$

When $r \geq r_i$ for each i, we shall have $N^r = 0$, because the operator $(T - c_iI)^r$ will then be 0 on the range of E_i.

Definition: Let N be a linear operator on the vector space V. We say that N is nilpotent if there is some positive integer r such that $N^r = 0$.

Theorem 2: Let T be a linear operator on the finite-dimensional vector space V over the field F. Suppose that the minimal polynomial for T decomposes over F into a product of linear polynomials. Then there is a diagonalizable operator D on V and a nilpotent operator N on V such that

(i) $T = D + N$,

(ii) $DN = ND$

The diagonalizable operator D and the nilpotent operator N are uniquely determined by (i) and (ii) and each of them is a polynomial in T.

Proof: We have just observed that we can write $T = D + N$ where D is diagonalizable and N is nilpotent, and where D and N not only commute but are polynomials in T. Now suppose that we also have $T = D' + N'$ where D' is diagonalizable, N' is nilpotent, and $D'N' = N'D'$. We shall prove that $D = D'$ and $N = N'$.

Since D' and N' commute with one another and $T = D' + N'$, we see that D' and N' commute with T. Thus D' and N' commute with any polynomial in T; hence they commute with D and with N. Now we have

$$D + N = D' + N'$$

or

$$D - D' = N' - N$$

and all four of these operators commute with one another. Since D and D′ are both diagonalizable and they commute, they are simultaneously diagonalizable, and $D - D'$ is diagonalizable. Since N and N' are both nilpotent and they commute, the operator $(N' - N)$ is nilpotent; for, using the fact that N and N' commute

$$(N'-N)^r = \sum_{j=0}^{r} \binom{r}{j} (N')^{r-j}(-N)^j$$

and so when r is sufficiently large every term in this expression for $(N' - N)^r$ will be 0. (Actually, a nilpotent operator on an n-dimensional space must have its nth power 0; if we take $r = 2n$ above, that will be large enough. It then follows that $r = n$ is large enough, but this is not obvious from the above expression.) Now $D - D'$ is a diagonalizable operator which is also nilpotent. Such an operator is obviously the zero operator; for since it is nilpotent, the minimal polynomial for this operator is of the form x^r for some $r \leq m$; but then since the operator is diagonalizable, the minimal polynomial cannot have a repeated root; hence $r = 1$ and the minimal polynomial is simply x, which says the operator is 0. Thus we see that $D = D'$ and $N = N'$.

Corollary: Let V be a finite-dimensional vector space over an algebraically closed field F, e.g., the field of complex numbers. Then every linear operator T on V can be written as the sum of a diagonalizable operator D and a nilpotent operator N which commute. These operators D and N' are unique and each is a polynomial in T.

From these results, one sees that the study of linear operators on vector spaces over an algebraically closed field is essentially reduced to the study of nilpotent operators. For vector spaces over non-algebraically closed fields, we still need to find some substitute for characteristic values and vectors. It is a very interesting fact that these two problems can be handled simultaneously and this is what we shall do in the next units.

In concluding this section, we should like to give examples, which illustrate some of the ideas of the primary decomposition theorem. We have chosen to give it at the end of the section since it deals with differential equations and thus is not purely linear algebra.

Example 1: In the primary decomposition theorem, it is not necessary that the vector space V be finite dimensional, nor is it necessary for parts (i) and (ii) that p be the minimal polynomial for T. If T is a linear operator on an arbitrary vector space and if there is a monic polynomial p such that $p(T) = 0$, then parts (i) and (ii) of Theorem 1 are valid for T with the proof which we gave.

Let n be a positive integer and let V be the space of all n times continuously differentiable functions f on the real line which satisfy the differential equation.

$$\frac{d^n f}{dt^n} + a_{n-1}\frac{d^{n-1} f}{dt^{n-1}} + \cdots + a_1 \frac{d_j}{dt} + a_0 f = 0 \qquad \text{...(1)}$$

where $a_0, \ldots, a_{n-1}$ are some fixed constants. If C_n denotes the space of n times continuously differentiable functions, then the space V of solutions of this differential equation is a subspace of C_n. If D denotes the differentiation operator and p is the polynomial

$$p = x^n + a_{n-1}x^{n-1} + \ldots + a_1 x + a_0$$

then V is the null space of the operator $p(D)$, because (1) simply says $p(D)f = 0$. Therefore, V is invariant under D. Let us now regard D as a linear operator on the subspace V. Then $p(D) = 0$.

If we are discussing differentiable complex-valued functions, then C_n and V are complex vector spaces, and $a_0, \ldots, a_{n-1}$ may be any complex numbers. We now write

$$p = (x - c_1)^{r_1} \ldots (x - c_k)^{r_k}$$

where $c_1, \ldots, c_k$ are distinct complex numbers. If W_j is the null space of $(*D - c_j I)^{r_i}$, then Theorem 1 says that

$$V = W_1 \oplus \ldots \oplus W_k$$

In other words, if f satisfies the differential equation (1), then f is uniquely expressible in the form

$$f = f_1 + \ldots + f_k$$

where f_j satisfies the differential equation $(D - c_j I)^{r_j} f_j = 0$. Thus, the study of the solutions to the equation (1) is reduced to the study of the space of solutions of a differential equation of the form

$$(D - cI)^r f = 0 \qquad \text{...(2)}$$

This reduction has been accomplished by the general methods of linear algebra, i.e., by the primary decomposition theorem.

To describe the space of solutions to (2), one must know something about differential equations, that is, one must know something about D other than the fact that it is a linear operator. However, one does not need to know very much. It is very easy to establish by induction on r that if f is in C_r then

$$(D - cI)^r f = e^{ct} D^r (e^{-ct} f)$$

that is,

$$\frac{df}{dt} - cf(t) = e^{ct}\frac{d}{dt}(e^{-ct}f), \quad \text{etc.}$$

Thus $(D - cI)^r f = 0$ if and only if $D^r(e^{-ct}f) = 0$. A function g such that $D^r g = 0$, i.e., $d^r g/dt^r = 0$, must be a polynomial function of degree $(r - 1)$ or less:

$$g(t) = b_0 + b_1 t + \cdots + b_{r-1}t^{r-1}$$

Thus f satisfies (2) if and only if f has the form

$$f(t) = e^{ct}(b_0 + b_1 t + \ldots + b_{r-1}t^{r-1})$$

Accordingly, 'the functions' e^{ct}, te^{ct}, ..., $t^{r-1}e^{ct}$ span the space of solutions of (2). Since 1, t,..., t^{r-1} are linearly independent functions and the exponential function has no zeros, these r functions $t^j e^{ct}$, $0 \le j \le r - 1$, form a basis for the space of solutions.

Returning to the differential equation (1), which is

$$p(D)f = 0$$

$$p = (x - c_1)^{r_1} \ldots (x - c_k)^{r_k}$$

we see that the n functions $t^m e^{c_j t}$, $0 \le m \le r_j - 1$, $1 \le j \le k$, form a basis for the space of solutions to (1). In particular, the space of solutions is finite-dimensional and has dimension equal to the degree of the polynomial p.

Example 2: Prove that the matrix A

$$A = \begin{bmatrix} 1 & 1 & 1 \\ -1 & -1 & -1 \\ 1 & 1 & 0 \end{bmatrix}$$

is nilpotent. Find its index of nilpotency.

Proof:

$$A^2 = \begin{bmatrix} 1 & 1 & 1 \\ -1 & -1 & -1 \\ 1 & 1 & 0 \end{bmatrix}\begin{bmatrix} 1 & 1 & 1 \\ -1 & -1 & -1 \\ 1 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ -1 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$A^3 = \begin{bmatrix} 1 & 1 & 0 \\ -1 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} 1 & 1 & 1 \\ -1 & -1 & -1 \\ 1 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

So $A^3 = 0$. Hence A is nilpotent of the index of nilpotence 3. Notice that $A^2 \neq 0$. (matrix)

Also the characteristic polynomial of A is $p(x) = x^3$.

Self Assessment

1. If V is the space of all polynomials of degree less than or equal to n over a field F, prove that the differentiation operator on V is nilpotent. Show that its characteristic polynomial is x^n.

2. If N is a nilpotent operator on an n-dimensional vector space V, show that the characteristic polynomial for N is x^n.

18.3 Summary

- The primary decomposition theorem is based on the fact that the minimal polynomial of the linear operator is the product of the irreducible.
- This helps in finding the projection operates which are polynomials in T.
- The direct decomposition of the vector space V in terms of the invariant subspaces helps in inducing linear operators T_i on these subspaces W_i.
- The induced operator T_i on W_i by T has the minimal polynomial as well as due to the factorisation of the minimal polynomial of T.

18.4 Keywords

Invariant Sub-spaces: If a vector α in V is such that α and $T\alpha$ are in the subspace W of V then W is invariant subspace of V over the field F.

Nilpotent Transformation: A nilpotent transformation N on the vector space V represented by a matrix A is such that $A^K = 0$ for some integer K and $A^{K-1} \neq 0$. Here K is the index of nilpotency.

Projection Operators: The projection operator E_i acting on the vector α_i gives $E\alpha_i = \alpha_i$ for the subspace W_i and gives zero for other. Also $E_i^2 = E_i$ and $E_iE_j = 0$ for $i \neq j$

18.5 Review Questions

1. Let T be the linear operator on R^3 which is represented by the matrix

$$A = \begin{bmatrix} 3 & 1 & -1 \\ 2 & 2 & -1 \\ 2 & 2 & 0 \end{bmatrix}$$

in the standard ordered basis. Show that $T = D + N$ where D is a diagonalizable operator and N a nilpotent vector.

2. Show that the linear operator T on R^3 represented by the matrix

$$A = \begin{bmatrix} 1 & -1 & 1 \\ 2 & -2 & 2 \\ -1 & 1 & -1 \end{bmatrix}$$

is nilpotent.

18.6 Further Readings

Books

Kenneth Hoffman and Ray Kunze, *Linear Algebra*

I.N. Herstein, *Topics in Algebra*

Sachin Kaushal, Lovely Professional University

Unit 19: Cyclic Subspaces and Annihilators

CONTENTS

Objectives

After studying this unit, you will be able to:

- Understand clearly the meaning of cyclic vector, cyclic-vector subspace and *T*-annihilator of α.
- See that in the case of a nilpotent linear operator one finds out the basis of the vectors α, $T\alpha$, $T^2\alpha$,... as the basis that spans the space of the linear transformation *T*.
- Know that closely related to the idea of cyclic vector α one understands the *T*-annihilator of α i.e., finds a polynomial *g* in *F* such that $g(T)\alpha = 0$
- See that with the help of these ideas one can understand the rational forms as well as the Jordan forms.

Introduction

The cyclic subspaces, the cyclic vector α and the *T* annihilators of α help us in the factoring of a linear operator *T* on the finite dimensional space to give a simple and elementary form.

In this unit the nilpotent transformation helps us in finding the basis vectors α, $T\alpha$, $T^2\alpha$,... that spans the space and this will help us in introducing the rational and the Jordan forms.

19.1 Cyclic Subspaces

We are considering an arbitrary but fixed linear operator on *V*, a finite-dimension vector space over the field *F*. If α is any vector in *V*, there is a smallest subspace of *V* which is invariant under *T* and contains α. This subspace can be defined as the intersection of all, *T*-invariant subspaces which contains α, if *W* is any subspace of *V* which is invariant under *T* and contains α, then *W* must also contain the vector $T\alpha$; hence *W* must contain $T^2\alpha$, $T^3\alpha$, etc. In other words, *W* must contain $g(T)\alpha$ for every polynomial *g* over *F*. The set of all vectors of the form $g(T)\alpha$, with *g* in $F(x)$, is clearly invariant under *T*, and is thus the smallest *T*-invariant subspace which contains α.

T-cyclic Subspace

If α is any vector in V, the T-cyclic subspace generated by α is the subspace $Z(\alpha; T)$ of all vectors of the form $g(T)\alpha$, g in $F[x]$. If $Z(\alpha; T) = V$, then α is called a cyclic vector for T.

In other words, the subspace $Z(\alpha; T)$ is the subspace spanned by the vectors $T^K\alpha$, $K \geq 0$ and thus α is a cyclic vector for T if and only if these vectors span V.

Example 1: For any T, the T-cyclic subspace generated by the zero vector is the zero space.

If the vector α is a characteristic vector for T the space $Z(\alpha; T)$ is one dimensional.

For the identity operator, every non-zero vector generates a one dimension cyclic subspace, thus, if dim $V > 1$, the identity operator has no cyclic vector.

Example 2: Consider the linear operator T on F^2 which is represented in the standard basis by the matrix

$$A = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$$

Here the cyclic vector is $\in_1 = (1, 0)$; for

$$\mathrm{A}\in_1 = \in_2$$

So that for any vector β given by

$$\beta(a, b)$$

We have $\beta(a, b) = a\in_1 + b\in_2,$

so $= a\in_1 + b\mathrm{A}\in_1$

$= (a + bA)\in_1$

Thus the polynomial g in $F^2(x)$ can be taken as

$$g = a + bx$$

For the same operator T, the cyclic subspace generated by $\in_2 = (0, 1)$ is the one dimensional space spanned by $\in_2$, because $\in_2$ is a characteristic vector of T.

Example 3: Consider the linear operator T on F^3, which is represented by A;

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

Here $A^3 = 0$, but $A^2 \neq 0$. So if α is a vector such that $A^2\nu \neq 0$ i.e., $\nu = \in_1 = (1,0,0)$, then the basic vectors will be $(\nu, T\nu, T^2\nu)$ and space generated by $\in_1$ is a cyclic subspace.

19.2 Annihilators

For any T and α, we shall be interested in linear relations

$$c_0\alpha + c_1T\alpha + \ldots + c_kT^k\alpha = 0$$

between the vectors $T^j\alpha$, that is, we shall be interested in the polynomials $g = c_0 + c_1x + ... + c_kx^k$ which have the property that $g(T)\alpha = 0$. The set of all g in $F[x]$ such that $g(T)\alpha = 0$ is clearly an ideal in $F[x]$. It is also a non-zero ideal, because it contains the minimal polynomial p of the operator $T(p)(T)\alpha = 0$ for every α in V).

Definition: If α is any vector in V, the T-annihilator of α is the ideal $M(\alpha; T)$ in $F[x]$ consisting of all polynomials g over F such that $g(T)\alpha = 0$. The unique monic polynomial p_α which generates this ideal will also be called the T-annihilator of α.

As we pointed out above, the T-annihilator p_α divides the minimal polynomial of the operator T. Please note that $\deg(p_\alpha) > 0$ unless α is the zero vector.

Theorem 1: Let α be any non-zero vector in V and let p_α be the T-annihilator of α.

(i) The degree of p_α is equal to the dimension of the cyclic subspace $Z(\alpha; T)$.

(ii) If the degree of p_α is k, then the vectors $\alpha, T\alpha, T^2\alpha, ..., T^{k-1}\alpha$ form a basis for $Z(\alpha; T)$.

(iii) If U is the linear operator on $Z(\alpha; T)$ induced by T, then the minimal polynomial for U is p_α.

Proof: Let g be any polynomial over the field F. Write

$$g = p_\alpha q + r$$

where either $r = 0$ or $\deg(r) < \deg(p_\alpha) = k$. The polynomial $p_\alpha q$ is in the T-annihilator of α, and so

$$g(T)\alpha = r(T)\alpha$$

Since $r = 0$ or $\deg(r) < k$, the vector $r(T)\alpha$ is a linear combination of the vectors $\alpha, T\alpha, ..., T^{k-1}\alpha$, and since $g(T)\alpha$ is a typical vector in $Z(\alpha; T)$, this shows that these k vectors span $Z(\alpha; T)$. These vectors are certainly linearly independent, because any non-trivial linear relation between them would give us a non-zero polynomial g such that $g(T)\alpha = 0$ and $\deg(g) < \deg(p_\alpha)$, which is absurd. This proves (i) and (ii).

Let U be the linear operator on $Z(\alpha; T)$ obtained by restricting T to that subspace. If g is any polynomial over F, then

$$\begin{aligned} p_\alpha(U)g(T)\alpha &= p_\alpha(U)g(T)\alpha \\ &= g(T)p_\alpha(U)\alpha \\ &= g(T)0 \\ &= 0 \end{aligned}$$

Thus the operator $p_\alpha(U)$ sends every vector in $Z(\alpha; T)$ into 0 and is the zero operator on $Z(\alpha, T)$. Furthermore, if h is a polynomial of degree less than k, we cannot have $h(U) = 0$, for then $h(U)\alpha = h(T)\alpha = 0$, contradicting the definition of p_α. This shows that p_α is the minimal polynomial for U.

A particular consequence of this theorem is the following: If α happens to be a cyclic vector for T, then the minimal polynomial for T must have degree equal to the dimension of the space V; hence, the Cayley-Hamilton theorem tells us that the minimal polynomial for T is the characteristic polynomial for T. We shall prove later that for any T there is a vector α in V which has the minimal polynomial of T for its annihilator. It will then follow that T has a cyclic vector if and only if the minimal and characteristic polynomials for T are identical. But it will take a little work for us to see this.

Our plan is to study the general T by using operators which have a cyclic vector. So, let us take a look at a linear operator U on a space W of dimension k which has a cyclic vector α. By Theorem 1, the vectors $\alpha, ..., U^{k-1}\alpha$ form a basis for the space W, and the annihilator p_α of α is the minimal polynomial for U (and hence also the characteristic polynomial for U). If we let $\alpha_i = U^{i-1}\alpha$, $i = 1,..., k$, then the action of U on the ordered basis $\mathcal{B} = \{\alpha_1,..., \alpha_k\}$ is

$$\left.\begin{aligned} U\alpha_i &= \alpha_{i+1}, \qquad i = 1,...,k-1 \\ U\alpha_k &= -c_0\alpha_1 - c_1\alpha_2 - ... - c_{k-1}\alpha_k \end{aligned}\right\} \quad ... (1)$$

where $p_\alpha = c_0 + c_1x + \cdots + c_{k-1}x^{k-1} + x^k$. The expression for $U\alpha_k$ follows from the fact that $p_\alpha(U)\alpha = 0$, i.e.,

$$U^k\alpha + c_{k-1}U^{k-1}\alpha + \cdots + c_1U\alpha + c_0\alpha = 0$$

This says that the matrix of U in the ordered basis $\mathcal{B}$ is

$$\begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & -c_0 \\ 1 & 0 & 0 & \cdots & 0 & -c_1 \\ 0 & 1 & 0 & \cdots & 0 & -c_2 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -c_{k-1} \end{bmatrix}$$

The matrix (2) is called the companion matrix of the monic polynomial p_α.

Theorem 2: If U is a linear operator on the finite-dimensional space W, then U has a cyclic vector if and only if there is some ordered basis for W in which U is represented by the companion matrix of the minimal polynomial for U.

Proof: We have just observed that if U has a cyclic vector, then there is such an ordered basis for W. Conversely, if we have some ordered basis $\{\alpha_1,..., \alpha_k\}$ for W in which U is represented by the companion matrix of its minimal polynomial, it is obvious that α_1 is a cyclic vector for U.

Corollary: If A is the companion matrix of a monic polynomial p, then p is both the minimal and the characteristic polynomial of A.

Proof: One way to see this is to let U be the linear operator on F^k which is represented by A in the standard ordered basis, and to apply Theorem 1 together with the Cayley-Hamilton theorem. Another method is to use Theorem 1 to see that p is the minimal polynomial for A and to verify by a direct calculation that p is the characteristic polynomial for A.

One last comment—if T is any linear operator on the space V and α is any vector in V, then the operator U which T induces on the cyclic subspace $Z(\alpha; T)$ has a cyclic vector, namely, α. Thus $Z(\alpha; T)$ has an ordered basis in which U is represented by the companion matrix of p_α, the T-annihilator of α.

Self Assessment

1. Consider the linear operator T represented by the matrix

$$A = \begin{bmatrix} 1 & 1 & 1 \\ -1 & -1 & -1 \\ 1 & 1 & 0 \end{bmatrix}$$

Show that A is nilpotent. Find the basis vectors that will span the space of the linear operator T.

2. Let T be a linear operator on the finite dimensional vector space V. Suppose T has a cyclic vector. Prove that if U is any linear operator which commutes with T, then U is a polynomial in T.

19.3 Summary

- Know what is a cyclic vector, cyclic subspaces of a linear operator T acting on a finite dimension vector space.
- See that if the cyclic vector is found then the basis vectors of the sub-space of the linear transformation can be found that span the space of the linear operator.
- Understand how to find the T-annihilator of a vector and also find out that the monic polynomial which generates it has a degree equal to the dimension of the cyclic subspace.

19.4 Keywords

A Cyclic Vector: If the T-cyclic subspace generated by the vector α spans the whole finite dimensional space V then α is called a cyclic vector for the linear T.

Cyclic Subspace: If α is a vector in a finite dimensional space V of a linear operator T, then the invariant subspace W which contains all $g(T)\alpha$ for every polynomial g over F is called T-cyclic subspace generated by α.

T-annihilator of a Vector: α consisting of all polynomials g over F such that $g(T)\alpha = 0$ is called T-annihilator of α. The unique monic polynomial which generates this set will also be called the T-annihilator of α.

19.5 Review Questions

1. Let T be a linear operator on the finite dimensional space V_n. Suppose that T is diagonalizable. If T has a cyclic vector, then show that T has n-distinct characteristic values.
2. If S and T are nilpotent linear transformation which commute, prove that ST and $S + T$ are nilpotent linear transformations.

19.6 Further Readings

Books

Kenneth Hoffman and Ray Kunze, *Linear Algebra*

I.N. Herstein, *Topics in Algebra*

Michael Artin, *Algebra*

Unit 20: Cyclic Decomposition and the Rational Form

Notes

CONTENTS

Objectives

After studying this unit, you will be able to:

- Understand that if T is any linear operator on a finite dimensional space *V*, then there exists vectors $\alpha_1, \alpha_2, \ldots \alpha_n$ in V such that the space *V* is a direct sum of the T-cyclic subspaces $Z(\alpha_i ; T)$ for $i = 1, 2, \ldots n$.
- See that if *W* is any subspace of V, then there exists a subspace *W′*, called *complementary* to *W*, such that $V = W \oplus W'$.
- Know that if *W* is T-invariant and *W′* complementary to *W* is also *T*-invariant then *W* is also *T*-admissible.
- Understand that the Cyclic decomposition theorem says that there exist non-zero vectors $\alpha_1, \alpha_2, \ldots \alpha_n$ in *V* with respective *T*-annihilators $p_1, p_2, \ldots p_r$ such that, *V* is a direct sum of *T*-invariant subspaces along with a proper *T*-admissible subspace *W*.

Introduction

In this unit certain concepts like invariant cyclic subspaces, complimentary subspaces and *T*-admissible proper subspaces are introduced.

The Cyclic decomposition theorem helps us in decomposing the n-dimensional vector space as a direct sum of *T*-invariant cyclic subspaces.

The matrix analogue of the Cyclic Decomposition theorem is that for the cyclic ordered basis $(\alpha, T\alpha, T^2\alpha, \ldots T^{k-1}\alpha)$ the matrix of induced operator T_i is the companion matrix *Ai* if the polynomial p_i, the matrix

$$A = \begin{bmatrix} A_1 & 0 & \ldots & 0 \\ 0 & A_2 & \ldots & 0 \\ \vdots & \vdots & \ldots & 0 \\ 0 & 0 & \ldots & A_n \end{bmatrix}$$

having a *rational form.*

20.1 Overview

In this unit we are interested in dealing with a linear operator T on a finite-dimensional space V, and dealing with the cyclic subspaces $Z\ (\alpha_1; T)$ where the vectors $\alpha_1, \alpha_2, \ldots \alpha_r$ in V. In this case the finite dimensional space V can be decomposed as the direct sum, i.e.,

$$V = Z\ (\alpha_1; T) \oplus \ldots. \oplus Z\ (\alpha_r; T)$$

This will also show that T is the direct sum of a finite number of linear operators each of which has a cyclic vector. The effect of this will be to reduce many questions about the general linear operator.

Let us consider T-invariant subspaces W and W' such that $V = W \oplus W'$. Here for any invariant subspace W, W' is complementary to W. We are interested in those W' which are also T-invariant.

T-admissible invariant subspace W

Let T be a linear operator on a vector space V and let W be a subspace of V. We say that W is *T-admissible* if W is invariant under T and if $f(T)\ \beta$ is in W where f is a polynomial, and there exists a vector Y in W such that

$$f(T)\ \beta \ = \ f(T)\ \gamma$$

our method for arriving at a decomposition

$$V \ = \ Z\ (\alpha_i; T) \oplus \ldots \oplus Z\ (\alpha_r; T)$$

will be to inductively select the vectors $\alpha_1, \alpha_2, \ldots \alpha_r$. Suppose that by some process or another we have selected $\alpha_1, \alpha_2, \ldots. \alpha_j$ and the sub-space

$$W_j \ = \ Z\ (\alpha_i; T) \oplus Z\ (\alpha_2; T) + \ldots + Z\ (\alpha_j; T)$$

is proper. We would like to find a non-zero vector α_{j+1} such that

$$W_j \cap Z\ (\alpha_{j+1}; T) \ = \ \{0\}.$$

Thus W will be a *proper T-invariant* subspace if there is a non-zero vector α such that

$$W \cap Z\ (\alpha\ ; T) \ = \ \{0\} \qquad \ldots(1)$$

Thus the subspace $Z\ (\alpha; T)$ and W are independent if (1) is satisfied and the polynomial f is the T-annihilator of α i.e. $f(T)\ \alpha = 0$.

20.2 Cyclic Decomposition

With the above definition we arrive at the following theorem for the cyclic decomposition of the finite vector space.

Theorem 1 (Cyclic Decomposition Theorem). Let T be a linear operator on a finite-dimensional vector space V and let W_0 be a proper T-admissible subspace of V. There exist non-zero vectors $\alpha_1, \ldots., \alpha_{\underline{r}}$ in V with respective T- annihilators $p_1, \ldots., p_r$ such that

(i) $V = W_0 \oplus Z\ (\alpha_1; T) \oplus \ldots \oplus Z\ (\alpha_r; T)$;

(ii) p_k divides p_{k-1}, $k = 2, \ldots, r$.

Furthermore, the integer r and the annihilators $p_1, \ldots., p_r$ are uniquely determined by (i), (ii), and the fact that no α_k is 0.

Proof: The proof is rather long; hence, we shall divide it into four steps. For the first reading it may seem easier to take $W_0 = \{0\}$, although it does not produce any substantial simplification. Throughout the proof, we shall abbreviate $f(T)\ \beta$ to f^{β}.

Notes

Step 1: There exist non-zero vectors $\beta_1, ..., \beta_r$ in V such that

(a) $V = W_0 + Z(\beta_1; T) + ... + Z(\beta_r; T)$;

(b) if $1 \le k \le r$ and

$$W_k = W_0 + Z(\beta_1; T) + ... + Z(\beta_k; T)$$

then the conductor $p_k = s(\beta_k; W_{k-1})$ has maximum degree among all T-conductors into the subspace W_{k-1}, i.e., for every k

$$\deg p_k = \max_{\alpha \text{ in } V} \deg s(\alpha; W_{k-1})$$

This step depends only upon the fact that W_0 is an invariant subspace. If W is a proper T-invariant subspace, then

$$0 < \max_{\alpha} \deg s(\alpha; W) \le \dim V$$

and we can choose a vector β so that deg s $(\beta; W)$ attains that maximum. The subspace $W + Z(\beta; T)$ is then T-invariant and has dimension larger than dim W. Apply this process to $W = W_0$ to obtain β_1. If $W_1 = W_0 + Z(\beta_1; T)$ is still proper, then apply the process to W_1 to obtain β_2. Continue is that manner. Since dim $W_k >$ dim W_{k-1}, we must reach $W_r = V$ is not more than dim V steps.

Step 2: Let $\beta_1, ... \beta_r$ be non-zero vectors which satisfy conditions (a) and (b) of Step 1. Fix k, $1 \le k \le r$. Let β be any vector in V and let $f = s(\beta; W_{k-1})$. If

$$f\beta = \beta_0 + \sum_{1 \le i < k} g_i\beta_i, \qquad \beta_i \text{ in } W_i$$

then f divides each polynomial g_i and $\beta_0 = f\gamma_0$, where γ_0 is in W_0.

If $k = 1$, this is just the statement that W_0 is T-admissible. In order to prove the assertion for $k > 1$, apply the division algorithm:

$$g_i = fh_i + r_i, \qquad r_i = 0 \qquad \text{or} \qquad \deg r_i < \deg f. \qquad ...(2)$$

We wish to show that $r_i = 0$ for each i, Let

$$\gamma = \beta - \sum_{1}^{k-1} h_i\beta_i. \qquad ...(3)$$

Since $\gamma - \beta$ is in W_{k-1},

$$s(\gamma; W_{k-1}) = s(\beta; W_{k-1}) = f.$$

Furthermore

$$f\gamma = \beta_0 + \sum_{1}^{k-1} r_i\,\beta_i. \qquad ...(4)$$

Suppose that some r_i is different from 0. We shall deduce a contradiction. Let j be the largest index i for which $r_i \ne 0$. Then

$$f\gamma = \beta_0 + \sum_{1}^{j} r_i\,\beta_i, \qquad r_j \ne 0 \qquad \text{and} \qquad \deg r_j < \deg f. \qquad ...(5)$$

Let $p = s(\gamma; W_{j-1})$. Since W_{k-1} contains W_{j-1}, the conductor $f = s(\gamma; W_{k-1})$ must divide p:

$$p = fg.$$

Apply $g(T)$ to both sides of (5):

$$p\gamma = gf\gamma = gr_j\beta_j + g\beta_0 + \sum_{1 \le i < j} gr_i\, \beta_i \qquad ...(6)$$

By definition $p\gamma$ is i W_{j-1}, and the last two terms on the right side of (6) are in W_{j-1}. Therefore, $gr_j\beta_j$ is in W_{j-1}. Now we use condition (b) on Step 1:

$$\begin{aligned} \deg\,(gr_j) &\ge \deg s(\beta_j\,;\, W_{j-1}) \\ &= \deg p_j \\ &= \deg s(\gamma; W_{j-1}) \\ &= \deg p \\ &= \deg\,(fg). \end{aligned}$$

Thus deg $r_j \ge \deg f$, and that contradicts the choice of j. We now know that f divides each g_i and hence that $\beta_0 = f\gamma$. Since W_0 is T-admissible, $\beta_0 = f\gamma_0$ where γ_0 is in W_0. We remark in passing that Step 2 is a strengthened form of the assertion that each of the subspaces $W_1, W_2, ...W_r$ is T-admissible.

Step 3: There exist non-zero vectors $\alpha_1,....\alpha_r$ in V which satisfy conditions (i) and (ii) of Theorem 1.

Start with vectors $\beta_1,.....,\beta_r$ as in Step 1. Fix $k, 1 \le k \le r.$ We apply Step 2 to the vector $\beta = \beta_k$ and the T-conductor $f = p_k$. We obtain

$$p_k\beta_k = p_k\gamma_0 + \sum_{1 \le i < k} p_k h_i \beta_i \qquad ...(7)$$

where γ_0 is in W_0 and $h_i,....,h_{k-1}$ are polynomials. Let

$$\alpha_k = \beta_k - \gamma_0 - \sum_{1 \le i < k} h_i\,\beta_i. \qquad ...(8)$$

Since $\beta_k - \alpha_k$ is in W_{k-1},

$$s(\alpha_k\,;\, W_{k-1}) = s(\beta_k\,;\, W_{k-1}) = p_k \qquad ...(9)$$

and since $p_k\alpha_k = 0,$ we have

$$W_{k-1} \cap Z(\alpha_k\,;\, T) = \{0\}. \qquad ...(10)$$

Because each α_k satisfies (9) and (10), it follows that

$$W_k = W_0 \oplus Z(\alpha_1\,;\, T) \oplus ... \oplus Z\,(\alpha_k\,;\, T)$$

and that p_k is the T-annihilator of α_k. In other words, the vectors α_1, ..., α_r define the same sequence of subspaces W_1, W_2, ... as do the vectors β_1, ..., β_r and the T-conductors $p_k = s\ (\alpha_k, \mathrm{W}_{k-1})$ have the same maximality properties (condition (b) of Step 1). The vectors α_1, ..., α_r have the additional property that the subspaces W_0, $Z\ (\alpha_1; T)$, $Z(\alpha_2; T)$, ... are independent. It is therefore easy to verify condition (ii) in Theorem 1. Since $p_i\alpha_i = 0$ for each i, we have the trivial relation

$$p_k\alpha_k = 0 + p_1\alpha_1 + ... + p_{k-1}\,\alpha_{k-1}.$$

Apply Step 2 with $\beta_1,....,\beta_k$ replaced by $\alpha_1,.....,\alpha_k$ and with $\beta = \alpha_k$.

Conclusion: p_k divides each p_i with $i < k$.

Step 4: The number r and the polynomials $p_1, ... , p_r$ are uniquely determined by the conditions of Theorem 1.

Suppose that in addition to the vectors $\alpha_1,....,\alpha_r$ in Theorem 1 we have non-zero vectors $\gamma_1,....,\gamma_r$ with respective T-annihilators $g_1,....,g_r$ such that

$$V = W_0 \oplus Z(\gamma_i; T) \oplus ... \oplus Z(\gamma_i; T) \qquad ...(11)$$

$$g_k \text{ divides } g_{k-1}, \quad k = 2, ... , s.$$

We shall show that $r = s$ and $p_i = g_i$ for each i.

It is very easy to see that $p_1 = g_1$. The polynomial g_1 is determined from (11) as the T-conductor of V into W_0. Let $S(V; W_0)$ be the collection of polynomials f such that $f\beta$ is in W_0 for each β in V, i.e., polynomials f such that the range of $f(T)$ is contained in W_0. Then $S(V; W_0)$ is a non-zero ideal in the polynomial algebra. The polynomial g_1 is the monic generator of that ideal, for this reason. Each β in V has the form

$$\beta = \beta_0 + f_1\gamma_1 + ... + f_s\gamma_s$$

and so

$$g_1\beta = g_1\beta_0 + \sum_1^s g_1 f_i \gamma_i.$$

Since each g_i divides g_1, we have $g_1\gamma_i = 0$ for all i and $g_1\beta = g_1\beta_0$ is in W_0. Thus g_1 is in $S(V; W_0)$. Since g_1 is the monic polynomial of least degree which sends γ_1 into W_0 we see that g_1 is the monic polynomial of least degree in the ideal $S(V; W_0)$. By the same argument, p_1 is the generator of that ideal, so $p_1 = g_1$.

If f is a polynomial and W is a subspace of V, we shall employ the shorthand fW for the set of all vectors $f\alpha$ with α in W. We have left to the exercises the proofs of the following three facts.

1. $fZ(\alpha; T) = Z(f\alpha; T)$.
2. If $V = V_1 \oplus ... \oplus V_k$, where each V_i is invariant under T, then $fV = fV_1 \oplus ... \oplus fV_k$.
3. If α and γ have the same T-annihilator, then $f\alpha$ and $f\gamma$ have the same T-annihilator and (therefore)

$$\dim Z(f\alpha; T) = \dim Z(f\gamma; T).$$

Now, we proceed by induction to show that $r = s$ and $p_i = g_i$ for $i = 2, ..., r$. The argument consists of counting dimensions in the right way. We shall give the proof that if $r \geq 2$ the $p_2 = g_2$, and from that the induction should be clear. Suppose that $r \geq 2$. Then

$$\dim W_0 + \dim Z(\alpha_1;\, T) < \dim V.$$

Since we know that $p_1 = g_1$, we know that $Z(\alpha_1; T)$ and $Z(\gamma_1; T)$ have the same dimension. Therefore,

$$\dim W_0 + \dim Z(\gamma_1;\, T) < \dim V.$$

which shows that $s \geq 2$. Now it makes sense to ask whether or not $p_2 = g_2$. From the two decompositions of V, we obtain two decompositions of the subspace p_2V:

$$\left.\begin{aligned} p_2V &= p_2W_0 \oplus Z(p_2\alpha_1;T) \\ p_2V &= p_2W_0 \oplus Z(p_2\gamma_1;T) \oplus \ldots \oplus Z(p_2\gamma_s;T). \end{aligned}\right\} \qquad \text{...(12)}$$

We have made use of facts (1) and (2) above and we have used the fact that $p_2\alpha_i = 0, i \geq 2$. Since we know that $p_1 = g_1$, fact (3) above tells us that $Z(p_2\alpha_1;T)$ and $Z(p_2\gamma_1;T)$ have the same dimension. Hence, it is apparent from (12) that

$$\dim Z(p_2\gamma_i;T) = 0, \qquad i \geq 2.$$

We conclude that $p_2\gamma_2 = 0$ and g_2 divides p_2. The argument can be reversed to show that p_2 divides g_2. Therefore $p_2 = g_2$.

Corollary: If T is a linear operator on a finite-dimensional vector space, then every T-admissible subspace has a complementary subspace which is also invariant under T.

Proof: Let W_0 be an admissible subspace of V. If $W_0 = V$, the complement we seek is {0}. If W_0 is proper, apply Theorem 1 and let

$$W_0' = Z(\alpha_1;T) \oplus \ldots \oplus Z(\alpha_r;T).$$

Then W_0' is invariant under T and $V = W_0 \oplus W_0'$.

Corollary: Let T be a linear operator on a finite-dimensional vector space V.

(a) There exists a vector α in V such that the T-annihilator of α is the minimal polynomial for T.

(b) T has a cyclic vector if and only if the characteristic and minimal polynomials for T are identical.

Proof: If $V = \{0\}$, the results are trivially true. If $V \neq \{0\}$, let

$$V = Z(\alpha_1;T) \oplus \ldots \oplus Z(\alpha_r;T) \qquad \text{...(13)}$$

where the T-annihilators $p_1, \ldots., p_r$ are such that p_{k+1} divides p_k, $1 \leq k \leq r-1$. As we noted in the proof of Theorem 1, it follows easily that p_1 is the minimal polynomial for T, i.e., the T-conductor of V into {0}. We have proved (a).

We saw in unit 19 that, if T has a cyclic vector, the minimal polynomial for T coincides with the characteristic polynomial. The content of (b) is in the converse. Choose any α as in (a). If the degree of the minimal polynomial is dim V, then $V = Z(\alpha; T)$.

Theorem 2 (Generalized Cayley-Hamilton Theorem): Let T be a linear operator on a finite-dimensional vector space V. Let p and f be the minimal and characteristic polynomials for T, respectively.

1. p divides f.

2. p and f have the same prime factors, except for multiplication.

3. If

$$p = f_1^{T_1} \ldots. f_1^{T_k} \qquad \text{...(14)}$$

is the prime factorization of p, then

$$f = f_1^{d_1} \ldots. f_k^{d_k} \qquad \text{...(15)}$$

where d_i is the nullity of $f_i(T)^n$ divided by the degree of f_i.

Proof: We disregard the trivial case $V = \{0\}$. To prove (i) and (ii), consider a cyclic decomposition (13) of V obtained from Theorem 1. As we noted in the proof of the second corollary, $p_1 = p$. Let U_i be the restriction of T to $Z(\alpha_i; T)$. Then U_i has a cyclic vector and so p_i is both the minimal polynomial and the characteristic polynomial for U_i. Therefore, the characteristic polynomial f is the product $f = p_1 \ldots p_r$. That is evident from the block form (1) of unit 17 which the matrix of T assumes in a suitable basis. Clearly $p_1 = p$ divides f, and this proves (i). Obviously any prime divisor of p is a prime divisor of f. Conversely, a prime divisor of $f = p_1 \ldots p_r$ must divide one of the factors p_i, which in turn divides p_1.

Let (14) be the prime factorization of p. We employ the primary decomposition theorem (Theorem 1 of unit 18). It tells us that, if V_1 is the null space of $f_i(T)^{r_i}$, then

$$V = V_1 \oplus \ldots \oplus V_k \qquad \ldots(16)$$

and f_1^{ri} is the minimal polynomial of the operator T_i, obtained by restricting T to the (invariant) subspace V_i. Apply part (ii) of the present theorem to the operator T_i. Since its minimal polynomial is a power of the prime f_i, the characteristic polynomial for T_i has the form $f_i^{d_i}$, where $d_i \geq r_i$. Obviously

$$d_i = \frac{\dim V_i}{\deg f_i}$$

and (almost by definition) dim V_i = nullity $f_i(T)^{r_i}$. Since T is the direct sum of the operators $T_1, \ldots, T_k$, the characteristic polynomial f is the product

$$f = f_1^{d_1} \ldots\ldots f_k^{d_k}.$$

Corollary: If T is a nilpotent linear operator on a vector space of dimension n, then the characteristic polynomial for T is x^n.

20.3 The Rational Form

Now let us look at the matrix analogue of the cyclic decomposition theorem. If we have the operator T and the direct-sum decomposition of Theorem 1, let β_i be the 'cyclic ordered basis'

$$\{\alpha_i, T\alpha_i, \ldots\ldots, T^{k_i - 1}\alpha_i\}$$

for $Z(\alpha_i; T)$. Hence k_i denotes the dimension of $Z(\alpha_i; T)$, that is, the degree of the annihilator p_i. The matrix of the induced operator T_i in the ordered basis β_i is the companion matrix of the polynomial p_i. Thus, if we let $\mathcal{B}$ be the ordered basis for V which is the union of the β_i arranged in the order $\beta_1, \ldots\ldots, \beta_r$, then the matrix of T in the ordered basis β will be

$$A = \begin{bmatrix} A_1 & 0 & \ldots & 0 \\ 0 & A_2 & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & A_r \end{bmatrix} \qquad \ldots(17)$$

where A_i is the $k_i \times k_i$ companion matrix of p_i. An $n \times n$ matrix A, which is the direct sum (17) of companion matrices of non-scalar monic polynomials $p_1, \ldots\ldots, p_r$ such that p_{i+1} divides p_i for $i = 1, \ldots, r - 1$, will be said to be in **rational form**. The cyclic decomposition theorem tells us the following concerning matrices.

Theorem 3: Let F be a field and let B be an $n \times n$ matrix over F. The B is similar over the field F to one and only one matrix which is in rational form.

Proof: Let T be the linear operator on F^n which is represented by B in the standard ordered basis. As we have just observed, there is some ordered basis for F^n in which T is represented by a matrix A in rational form. Then B is similar to this matrix A. Suppose B is similar over F to another matrix C which is in rational form. This means simply that there is some ordered basis for F^n in which the operator T is represented by the matrix C. If C is the direct sum of companion matrices C_i of monic polynomials $g_1, ..., g_s$ such that g_{i+1} divides g_i for i = 1, ..., $s - 1$, then it is apparent that we shall have non-zero vectors $\beta_1, ..., \beta_s$ in V with T-annihilators $g_1, ..., g_s$ such that

$$V = Z(\beta_1; T) \oplus ... \oplus Z(\beta_s; T).$$

But then by the uniqueness statement in the cyclic decomposition theorem, the polynomials g_i, are identical with the polynomials p_i which define the matrix A. Thus $C = A$.

The polynomials $p_1, ..., p_r$ are called the **invariant factors** for the matrix B. We shall describe an algorithm for calculating the invariant factors of a given matrix B. The fact that it is possible to compute these polynomials by means of a finite number of rational operations on the entries of B is what gives the rational form its name.

Example 1: Suppose that V is a two-dimensional vector space over the field F and T is a linear operator on V. The possibilities for the cyclic subspace decomposition for T are very limited. For, if the minimal polynomial for T has degree 2, it is equal to the characteristic polynomial for T and T has a cyclic vector. Thus there is some ordered basis for V in which T is represented by the companion matrix of its characteristic polynomial. If, on the other hand, the minimal polynomial for T has degree 1, then T is a scalar multiple of the identity operator. If $T = cI$, then for any two linear independent vectors α_1 and α_2 in V we have

$$V = Z(\alpha_1; T) \oplus Z(\alpha_2; T)$$

$$p_1 = p_2 = x - c.$$

For matrices, this analysis says that every 2 × 2 matrix over the field F is similar over F to exactly one matrix of the types

$$\begin{bmatrix} c & 0 \\ 0 & c \end{bmatrix}, \begin{bmatrix} 0 & -c_0 \\ 1 & -c_1 \end{bmatrix}$$

Example 2: Let T be the linear operator on R^3 which is represented by the matrix.

$$A = \begin{bmatrix} 5 & -6 & -6 \\ -1 & 4 & 2 \\ 3 & -6 & -4 \end{bmatrix}$$

in the standard ordered basis. We have computed earlier that the characteristic polynomial for T is $f = (x-1)(x-2)^2$ and minimal polynomial for T is $p = (x-1)(x-2)$. Thus we know that in the cyclic decomposition for T the first vector α_1 will have p as its T-annihilator.

Since we are operating in a three-dimensional space, there can be only one further vector, α_2. It must generate a cyclic subspace of dimension I, i.e., it must be a characteristic vector for T. T-annihilator p_2 must be $(x - 2)$, because we must have $pp_2 = f$. Notice that this tells us immediately that the matrix A is similar to the matrix

$$B = \begin{bmatrix} 0 & -2 & 0 \\ 1 & 3 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

that is, that T is represented by B in some ordered basis. How can we find suitable vectors α_1 and α_2? Well, we know that any vector which generates a T-cyclic subspace of dimension 2 is a suitable α_1. So let's just try ε_1. We have

$$T\varepsilon_1 = (5, -1, 3)$$

which is not a scalar multiple of ε_1; hence $Z(\varepsilon_1; T)$ has dimension 2. This space consists of all vectors $a\varepsilon_1$ + b $(T\varepsilon_1)$:

$$a(1,0,0)+b(5,-1,3)=(a+5b,-b,3b)$$

or, all vectors (x_1, x_2, x_3) satisfying $x_3 = -3x_2$. Now what we want is a vector α_2 such that $T\alpha_2 = 2\alpha_2$ and $Z(\alpha_2; T)$ is disjoint from $Z(\varepsilon_1; T)$. Since α_2 is to be a characteristic vector for T, the space $Z(\alpha_2; T)$ will simply by the one-dimensional space spanned by α_2, and so what we require is that α_2 not be in $Z(\varepsilon_1; T)$. If $\alpha = (x_1, x_2, x_3)$, one can easily compute that $T\alpha = 2\alpha$ if and only if $x_1 = 2x_2 + 2x_3$. Thus $\alpha_2 = (2, 1, 0)$ satisfies $T\alpha_2 = 2\alpha_2$ and generates a T-cyclic subspace disjoint from $Z(\varepsilon_1; T)$. The reader should verify directly that the matrix of T is the ordered basis.

$$\{(1, 0, 0), (5, -1, 3), (2, 1, 0)\}$$

is the matrix B above.

Example 3: Suppose that T is a diagonalizable linear operator on V. It is interesting to relate a cyclic decomposition for T to a basis which diagonalizes the matrix of T. Let $c_1, \ldots c_k$ be the distinct characteristic values of T and let V_i be the space of characteristic vectors associated with the characteristic value c_i. Then

$$V = V_i \oplus \ldots \oplus V_k$$

and if $d_i = \dim V_i$ then

$$f = (x-c_1)^{d_1} \ldots. (x-c_k)^{d_k}$$

is the characteristic polynomial for T. If α is a vector in V, it is easy to relate the cyclic subspace $Z(\alpha; T)$ to the subspaces $V_1, \ldots, V_k$. There are unique vectors $\beta_1, \ldots, \beta_k$ such that β_i is in V_i and

$$\alpha = \beta_1 + \ldots. + \beta_k.$$

Since $T\beta_1 = c_i\, \beta_i$, we have

$$f(T)\alpha = f(c_1)\,\beta_1 + \ldots + f(c_k)\beta_k \qquad \ldots(18)$$

for every polynomial f. Given any scalars $t_1, \ldots., t_k$ there exists a polynomial f such that $f(c_i) = t_i$, $1 \le i \le k$. Therefore $Z(\alpha; T)$ is just the subspace spanned by the vectors $\beta_1, \ldots., \beta_k$. What is the annihilator of α? According to (18), we have $f(T)\,\alpha = 0$ if and only if $f(c_i)\,\beta_i = 0$ for each i. In other words, $f(T)\alpha = 0$ provided $f(c_i) = 0$ for each i such that $\beta_i \neq 0$. Accordingly, the annihilator of α is the product

$$\prod_{\beta_i \neq 0} (x - c_i). \qquad \ldots(19)$$

Now, let $\beta_i = \{\beta_1^t, \ldots.., \beta_{d_i}^t\}$ be an ordered basis for V_i. Let

$$r = \max_u d_i.$$

We define vectors $\alpha_1, \ldots, \alpha_r$ by

$$\alpha_j = \sum_{d_i \ge j} \beta_j^t, \qquad 1 \le j \le r. \qquad \ldots(20)$$

The cyclic subspace $Z(\alpha_j; T)$ is the subspace spanned by the vectors β_j^t, as i runs over those indices for which $d_i \geq j$. The T-annihilator of α_j is

$$p_j = \prod_{d_i \geq j} (x - c_i). \qquad \text{...(21)}$$

We have

$$V = Z(\alpha_1; T) \oplus \ldots \oplus Z(\alpha_r; T)$$

because each β_j^t belongs to one and only one of the subspaces $Z(\alpha_1; T)$; ...., $Z(\alpha_r; T)$ and $\beta = (\beta_1, ..., \beta_k)$ is a basis for V. By (21) p_{j+1} divides p_j.

Self Assessment

1. Let T be the linear operator on R^3 which is represented in the standard ordered basis by

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

Find the characteristic polynomial for T. What is the minimal polynomial?

2. Show that if T is a diagonalizable linear operator then every T-invariant subspace has a complementary T-invariant subspace.

20.4 Summary

- In this unit the theorem 1 (derived) helps us in finding non-zero vectors α_1, ..., α_r in V with respect to T-annihilators $p_1, p_2, ...p_r$ such that the vector space is a direct sum of T-invariant subspaces along with a proper T-admissible subspace W_0.
- Certain concepts like complementary subspace T-admissible subspace, proper T-invariant subspaces are explained.
- If T is a nilpotent linear operator on a vector space of dimension n, then the characteristic polynomial for T is x^n.
- It is shown that if there is direct sum decomposition theorem 1 the cyclic ordered basis $(\alpha_i, Td_i, ...T^{k-1}\alpha_i)$ for $Z(\alpha_i; T)$ then with the help of companion matrices, A representing T can be put in Jordan Form.

20.5 Keywords

Complementary Subspace: T-invariant subspace W has the property that there exists a subspace W', such that $V = W \oplus W'$, where W' is called a complementary subspace of W. W' can also be T-invariant.

Rational Form: An $n \times n$ matrix A

$$A = \begin{bmatrix} A_1 & 0 & 0 & \ldots & 0 \\ 0 & A_2 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & & 0 \\ \vdots & \vdots & & & 0 \\ 0 & 0 & \ldots & & A_n \end{bmatrix}$$

which is direct sum of companion matrices A_i has a rational form.

***T*-admissible Subspace:** An invariant subspace W with another T-invariant subspace W', such that

$$V = W \oplus W$$

is called T-admissible subspace.

20.6 Review Questions

1. Let T be the linear operator on R^3 which is represented in the standard ordered basis by

$$\begin{bmatrix} 3 & -4 & -4 \\ -1 & 3 & 3 \\ 2 & -4 & -3 \end{bmatrix}$$

Find non-zero α_1, α_2, α_3 satisfying the conditions of theorem 1.

2. Find the minimal polynomial and the rational forms of the following real matrices

$$\begin{bmatrix} 0 & -1 & -1 \\ 1 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix}, \begin{bmatrix} C & 0 & -1 \\ 0 & C & 1 \\ -1 & 1 & C \end{bmatrix}$$

Answer: Self Assessment

1. The characteristic polynomial is

$$f = (x-1)(x+2)^2$$

The minimal polynomial is

$$p = (x-1)(x+2)$$

20.7 Further Readings

Books

Kenneth Hoffman and Ray Kunze, *Linear Algebra*

Michael Artin, *Algebra*