COURSE GUIDE

MTH 212
LINEAR ALGEBRA

Course Team Prof. Saheed O. Ajibola (Course Developer/Editor) - NOUN
Dr. Christie Y. Ishola (Reviewer) – NOUN

NOUN

NATIONAL OPEN UNIVERSITY OF NIGERIA

National Open University of Nigeria
Headquarters
University Village
Plot 91, Cadastral Zone
Nnamdi Azikiwe Expressway
Jabi, Abuja

Lagos Office
14/16 Ahmadu Bello Way
Victoria Island, Lagos

e-mail : centralinfo@nou.edu.ng
URL: www.nou.edu.ng

First Printed 2012

Reviewed 2022, 2024

ISBN:

CONTENTS **PAGE**

INTRODUCTION

Linear algebra is a branch of mathematics that deals with linear equations and their representations in the vector space using matrices. In other words, the study of linear vectors and functions is what linear algebra is all about. It is one of the most important mathematical issues. The majority of contemporary geometrical ideas are based on linear algebra. Engineering and physics both rely heavily on linear algebra since it makes it easier to simulate a wide range of natural events. The three most critical elements of this topic are vector spaces, matrices, and linear equations. The many concepts related to linear algebra will be covered in more detail in this article.

Students are introduced to the fundamentals of linear algebra in elementary linear algebra. Simple matrix operations, different computations that can be performed on a system of linear equations, and certain characteristics of vectors are all included in this. The following list of key terms related to basic linear algebra includes

Scalars – A scalar is a quantity that only has magnitude and not direction. It is an element that is used to define a vector space. In linear algebra, scalars are usually real numbers.

Vectors – A vector is an element in a vector space. It is a quantity that can describe both the direction and magnitude of an element.

Vector Space – The vector space consists of vectors that may be added together and multiplied by scalars.

Matrix – A matrix is a rectangular array wherein the information is organized in the form of rows and columns. Most linear algebra properties can be expressed in terms of a matrix.

Matrix Operations – These are simple arithmetic operations such as addition, subtraction, and multiplication that can be conducted on matrices.

COURSE COMPETENCIES

The Course

As a 3-credit unit course, 11 study units grouped into 4 modules of 4 units in module 1, 3 units in module 2, 2 units in module 3 and 2 units in module 4.

A quick synopsis of the entire course materials is provided in this course guide. Vector spaces and linear transformations, which are primarily concerned with finite dimensional vector spaces over the set of real numbers, ℝ, or set of complex numbers, ℂ, are the two basic building blocks of linear algebra.

Adding vectors and multiplying them by numbers or scalars gives required linear combinations. The study of certain mappings between two vector spaces, called linear transformations are also initiated in this course.

We shall also demonstrate how a linear transformation can be used to obtain a matrix associated with it, and vice versa. Additionally, by focusing on the corresponding matrix, certain aspects of a linear transformation can be analyzed more clearly. You will find, for instance, that it is frequently simpler to extract a matrix's characteristic roots than a linear transformation.

WORKING THROUGH THIS COURSE

You are required to read the study units, set books and other materials provided by the National Open University to complete the course. You will also need to work through practical and self- assessed exercises and submit assignments for assessment purposes. The course will take you about 60 hours to complete at the end of which you will write a final examination.

This course consists of the following eleven study units:
Having gone the course content, what then is a matrix?

STUDY UNITS

Module 1 Vector Spaces

Unit 1 Vector Spaces
Unit 2 Linear Combinations
Unit 3 Linear Transformation I
Unit 4 Linear Transformation II

Module 2 Matrices

Unit 1 Matrices I
Unit 2 Matrices II
Unit 3 Matrices III

Module 3 Determinants

Unit 1 Determinants I
Unit 2 Determinants II

Module 4 Eigenvalues and Eigenvectors

Unit 1 Eigenvalues and Eigenvectors
Unit 2 Characteristic and Minimal Polynomials

The first four units deal with vector spaces and linear transformation which are the two fundamental elements that form the basis of linear algebra which is an area of mathematics that deals with the common features of algebraic systems made up of sets, as well as a logical concept of a "linear combination" of element in the set. The remaining seven units involve matrix theory which occupies an important position in pure as well as applied mathematics.

Each study unit involves specific objectives, how to study the reading materials, references to set books and other related sources and summaries of vital points and ideas. The units direct you to work on exercises related to the require reading and to carry out solutions to some exercises where appropriate. A number of self-tests are associated with each unit. These tests give you an indication of your progress. The exercises as well as the tutor-marked assignments will help you in achieving the stated learning outcomes of each unit and of the course.

PRESENTATION SCHEDULE

The weekly activities are presented in Table 1 while the required hours of study and the activities are presented in Table 2. This will guide your study time. You may spend more time in completing each module or unit.

Table I: Weekly Activities

Week	**Activity**
1	Orientation and Course Guide
2	Module 1 Unit 1
3	Module 1 Unit 2
4	Module 1 Unit 3
5	Module 1 Unit 4
6	Module 2 Unit 1
7	Module 2 Unit 2
8	Module 2 Unit 3
9	Module 3 Unit 1
10	Module 3 Unit 2

11	Module 4 Unit 1
12	Module 4 Unit 2
13	Response to Exercises
14	Revisions
15	Examination

The activities in Table I include facilitation hours (synchronous and asynchronous), class works and assignments. How do you know the hours to spend on each? A guide is presented in Table 2.

Table 2: Required Minimum Hours of Study

S/N	Activity	Hour per Week	Hour per Semester
1	Synchronous Facilitation (Video Conferencing)	2	26
2	Asynchronous Facilitation (Read and respond to posts including Facilitator's comments, self-study)	4	52
3	Assignments, mini-project, laboratory practical and portfolios	1	13
Total		**7**	**91**

ASSESSMENT

Table 3 presents the mode you will be assessed.

Table 3: Assessment

S/N	**Method of Assessment**	**Score (%)**
3	Computer-Based Tests	30
4	Final Examination	70
	Total	100

ASSIGNMENTS

Take the assignment and click on the submission button to submit. The assignment will be scored, and you will receive feedback.

EXAMINATION

Finally, the examination will help to test the cognitive domain. The test items will be mostly application, and evaluation test items that will lead to creation of new knowledge/idea.

HOW TO GET THE MOST FROM THE COURSE

To get the most in this course, you:

- Need a personal laptop. The use of mobile phone only may not give you the desirable environment to work.
- Need regular and stable internet.
- Need to install the recommended software.
- Must work through the course step by step starting with the programme orientation.
- Must not plagiarize or impersonate. These are serious offences that could terminate your studentship. Plagiarism check will be used to run all your submissions.
- Must do all the assessments following given instructions.
- Must create time daily to attend to your study.

FACILITATION

There will be two forms of facilitation–synchronous and asynchronous. The synchronous will be held through video conferencing according to weekly schedule.

During the synchronous facilitation:

- There will be two hours of online real time contact per week making a total of 26 hours for thirteen weeks of study time.
- At the end of each video conferencing, the video will be uploaded for view at your pace.
- You are to read the course material and do other assignments as may be given before video conferencing time.
- The facilitator will concentrate on main themes.
- The facilitator will take you through the course guide in the first lecture at the start date of facilitation.

For the asynchronous facilitation, your facilitator will:

- Present the theme for the week.
- Direct and summarise forum discussions.
- Coordinate activities in the platform.
- Score and grade activities when need be.
- Support you to learn. In this regard personal mails may be sent.
- Send you videos and audio lectures, and podcasts if need be.

Read all the comments and notes of your facilitator especially on your assignments, participate in forum discussions. This will give you opportunity to viiiocialize with others in the course and build your skill for teamwork. You can raise any challenge encountered during your study. To gain the maximum benefit from course facilitation, prepare a list of questions before the synchronous session. You will learn a lot from participating actively in the discussions.

LEARNER SUPPORT

You will receive the following support:

- Technical Support: There will be contact number(s), email addresses and chat bot on the Learning Management System (LMS) where you can chat or send message to get assistance and guidance any time during the course.

- 24/7 communication: You can send personal mail to your facilitator and the study centre at any time of the day. You will receive answer to you mails within 24 hours. There is also opportunity for personal or group chats at any time of the day with those that are online.

- You will receive guidance and feedback on your assessments, academic progress, and receive help to resolve challenges facing your studies.

COURSE INFORMATION

Course Code: MTH212
Course Title: Linear Algebra
Credit Unit: 3 units
Course Status: Compulsory

Course Blub: This is a basic course designed to help students master the contents of a first course in Linear Algebra. Its availability, make it widely used for self-study especially independent student in an online programme. The materials in this course are standard in that the topic covered are vector spaces, linear maps and transformations, matrices, determinants, and eigenvalues and eigenvectors. Another standard is the audience' friendly attribute of the material as well as the numerous examples following each topic
Semester: Second Semester
Course Duration: 13 Weeks
Required Hours for Study: 91 hours

MAIN COURSE

CONTENTS

MODULE 1

This module shall define the mathematical object which experience has shown to be the most useful abstraction of this type of algebraic system. There are two aspects to linear algebra. Abstractly, it is the study of vector spaces over fields and their linear maps and bi-linear forms. Concretely, it is matrix theory, since matrices occur in all parts of mathematics and its applications and the diagonalization of a real symmetric matrix is a skill that is required of everyone working in the mathematical sciences or associated fields. Therefore, it is important to discuss both the abstract and concrete aspects in a course of this kind, even when applications are not covered in great length.

Unit 1 Vector Spaces
Unit 2 Linear Combination
Unit 3 Linear Transformation I
Unit 4 Linear Transformation II

UNIT 1 VECTOR SPACES

Unit Structure

1.1 Introduction
1.2 Learning Outcomes
1.3 Vector Spaces
  1.3.1 Definitions and Examples of Vector Space
  1.3.2 Spaces Associated with Vector Spaces
  1.3.3 Definitions and Examples of Vector Subspace
1.4 Summary
1.5 References/Further Readings

1.1 Introduction

Vector spaces and linear transformations are two fundamental elements that form the basis of linear algebra. The strength of mathematics typically comes from the ability to abstractly formulate a wide range of situations, and that is exactly what shall be done throughout this module. The area of mathematics known as linear algebra deals with the common features of algebraic systems made up of sets, as well as a logical concept of a "linear combination" of element in the set.

Vector spaces, linear maps, and bi-linear forms are dealt with on the theoretical side. Vector spaces over a field F are particularly alluring algebraic objects, since each vector space is completely determined by a single number, its dimension (unlike groups, for example, whose structure is much more complicated). On the practical side,the subject only concerns matrix. A matrix must be used to express any calculation involving a linear map or a bilinear form. As a result, matrices can represent a variety of objects.

1.2 Learning Outcomes

By the end of this unit, you will be able to:

- Define vector space over a field
- Define vector subspace
- Cite examples of vector spaces and subspaces
- State and prove theorems involving vector spaces and subspaces

1.3 Vector Spaces

1.3.1 Definitions and Examples of Vector Space

This unit will mainly be concerned with finite dimensional vector spaces over the set of real numbers, $\mathbb{R}$, or set of complex numbers, $\mathbb{C}$.

Note that the real and complex numbers have the property that any pair of elements can be added, subtracted or multiplied. Division is also allowed by a non-zero element. Such sets in mathematics are called field, thus the sets of rational numbers, Q, real numbers, R, and complex numbers, C, are examples of field and they have infinite number of elements. But, in mathematics, we do have fields that have only finitely many elements. For example, consider the sets $Z_4 = \{0,1,2,3\}$ and $Z_5 = \{0,1,2,3,4\}$.

In Z_4,Z_5, we define addition and multiplication, respectively, as

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

and

$\times$	0	1	2	3
0	0	0	0	0
1	0	1	2	3
2	0	2	0	2
3	0	3	2	1

$+$	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

and

$\times$	0	1	2	3	4
0	0	0	0	0	0
1	0	1	2	3	4
2	0	2	4	1	3
3	0	3	1	4	2
4	0	4	3	2	1

Then, we see that the elements of both $Z_4$and Z_5 can be added, subtracted and multiplied.

Thus, $Z_4$and Z_5 indeed behave like a field. So, in this module, F will represent a field.

Examples of fields are the set of complex numbers, the set of real numbers, the set of rational numbers, and even the finite set of "binary numbers", $\{0,1\}$.

Definition 1.3.1: A **field** is an algebraic system consisting of a non-empty set ***F*** equipped with two binary operations + (addition) and · (multiplication) satisfying the conditions:

(Recall from MTH103 that the vectors in $R^2$and R^3satisfy the conditions)

1) **Vector Addition:** To every pair$u, v \in R^3$, there corresponds a unique element $u + v \in R^3$
(Called the addition of vectors) such that
- a) $u + v = v + u$ (Commutative law)
- b) $(u + v) + w = u + (v + w)$ (Associative Law)
- c) R^3 has a unique element, denoted 0, called the zero vector that satisfies $u + 0 = u$, for every $u \in R^3$ (called the additive identity).
- d) For every$u \in R^3$, there is an element $w \in R^3$ that satisfies$u + w = 0$.

2) **Scalar Multiplication:** For each $u \in R^3$ and$\alpha \in R$, there corresponds a unique element
$\alpha \cdot u \in R^3$ (Called the scalar multiplication) such that
- a) $\alpha \cdot (\beta \cdot u) = (\alpha \cdot \beta) \cdot u$, for every $\alpha, \beta \in R$ and $u \in R^3$
- b) $1 \cdot u = u$ for every $u \in R^3$; where $1 \in IR$.

3) **Distributive Laws: Relating vector addition with scalar multiplication**
For any $\alpha, \beta \in R$ and$u, v \in R^3$; the following distributive laws hold:
- a) $\alpha \cdot (u + v) = (\alpha \cdot u) + (\alpha \cdot v)$
- b) $(\alpha + \beta) \cdot u = (\alpha \cdot u) + (\beta \cdot v)$

For the above properties to hold for any collection of vectors, we have the following definitions:

Definition 1.3.2: A **vector space** V over F, denoted by $V(F)$ or in short V (if the field $F$is clear from the context), is a non-empty set, in which vector addition, scalar multiplication can be defined.

In other words, a vector space is composed of three objects; - a set and two operations.

Further, with these definitions, the properties of vector addition, scalar multiplication and distributive laws (see items 1, 2 and 3 above) are satisfied.

Remarks:

A. The elements of F are called scalars.

B. The elements of V are called vectors.

C. The zero element of $F$is denoted by **0** whereas the zero element of V is also denoted by **0**

D. Note that the condition 1d) above implies that for every$u \in V$, the vector $w \in V$ such that$u + w = 0$holds, is unique.

   i. For if, $w_1, w_2 \in V$with$u + w_i = 0$, for $i = 1,2$.
   then by **commutativity** of vector addition, we see that

$$w_1 = w_1 + 0 = w_1 + (u + w_2) = (w_1 + u) + w_2$$
$$= 0 + w_2 = w_2$$

   ii. Hence, we represent this unique vector by $-u$ and call it the **additive inverse**.

E. If V is a vector space over Rthen $V$is called a real vector space.

F. If V is a vector space over C then $V$is called a complex vector space.

G. In general, a vector space over Ror $C$is called a linear space.

For better understanding of the conditions given above, the following theorem and proof are presented:

Theorem 1.3.1: Let V be a vector space over F. Then,

i. $u + v = u$ implies $v = 0$

ii. $\alpha \cdot u = 0$ if and only if either $u = 0$or $\alpha = 0$

iii. $(-1) \cdot u = -u$, for every $u \in V$.

Proof:

Part 1: By Condition 1d) and Remarks D above, for each $u \in V$there exists $-u \in V$ such that $-u + u = 0$.

Hence $u + v = u$ implies $0 = -u + v = -u + (u + v) = (-u + u) + v = 0 + v = v$

Part 2: As $0 = 0 + 0$, using Condition 3, $\alpha \cdot 0 = \alpha \cdot (0 + 0) = (\alpha \cdot 0) + (\alpha + 0)$
By Part 1, $\alpha \cdot 0 = 0$ for any $\alpha \in F$
Similarly, $0 \cdot u = (0 + 0) \cdot u = (0 \cdot u) + (0 \cdot u)$ implies $0 \cdot u = 0$, for any $u \in V$.
Now, suppose $\alpha \cdot u = 0$.
If $\alpha = 0$ then the proof is over.
Assume that $\alpha \neq 0, \alpha \in F$, then, $(\alpha)^{-1} \in F$ and using
$1 \cdot u = u1 \cdot u = u$ for every vector $u \in V$ (see Condition 2b), we have

$$0 = (\alpha)^{-1} \cdot 0 = (\alpha)^{-1} \cdot (\alpha \cdot u) = ((\alpha)^{-1} \cdot \alpha) \cdot u = 1 \cdot u = u$$

Thus, if $\alpha \neq 0$ and $\alpha \cdot u = 0$ then $u = 0$.

Part 3:
As $0 = 0 \cdot u = \big(1 + (-1)\big) \cdot u = u + (-1) \cdot u$,
Then, $(-1) \cdot u = -u$
The following is one of the two most important definitions in the entire course.

Definition 1.3.3:
Suppose that V is a set upon which we have defined two operations: (i) vector addition, which combines two elements of V and is denoted by $(+)$, and (ii) scalar multiplication, which combines a complex number C with an element of V and is denoted by $(\cdot)$. Then V, along with the two operations, is a **vector space** over C if the following ten properties hold:

i. **Additive Closure**
ii. If $u, v \in V$, then $u + v \in V$
iii. **Scalar Closure**
iv. If $\alpha \in C$ and $u \in V$ then $\alpha \cdot u \in V$
v. **Commutativity**
vi. If $u, v \in V$, then $u + v = v + u$
vii. **Additive Associativity**
viii. If $u, v, w \in V$, then $u + (v + w) = (u + v) + w$
ix. **Zero Vector**
x. There is a vector,0, called the zero vector, such that $u + 0 = u$, for all $u \in V$.
xi. **Additive Inverses**
xii. If $u \in V$, then there exists a vector $-u \in V$ such that $u + (-u) = 0$
xiii. **Scalar Multiplication Associativity:**
xiv. If $\alpha, \beta \in C$ and $u \in V$, then $\alpha \cdot (\beta \cdot u) = (\alpha \cdot \beta) \cdot u$.
xv. **Distributivity across Vector Addition:**
xvi. If $\alpha \in C$ and $u, v \in V$, then $\alpha \cdot (u + v) = \alpha \cdot u + \alpha \cdot v$.
xvii. **Distributivity across Scalar Addition:**
xviii. If $\alpha, \beta \in C$ and $u \in V$, then $(\alpha + \beta) \cdot u = \alpha \cdot u + \beta \cdot v$.
xix. **One:** If $u \in V$, then $1 \cdot u = u$.

The objects in V are called **vectors**, no matter what else they might really be, simply by virtue of being elements of a vector space.

Examples

1. **The Euclidean plane R^2** is a real vector space. In other words, two vectors can be added together as well as multiply a vector by a scalar (a real number).

There are two approaches to explain these definitions.

a) The ***geometric*** definition: Think of a vector as an arrow starting at the origin and ending at a point of the plane, then addition of two vectors is done by the *parallelogram law* (see Figure 1.3.1). The scalar multiple av is the vector whose length is $|a|$ times the length of v, in the same direction if $a > 0$ and in the opposite direction if $a < 0$.

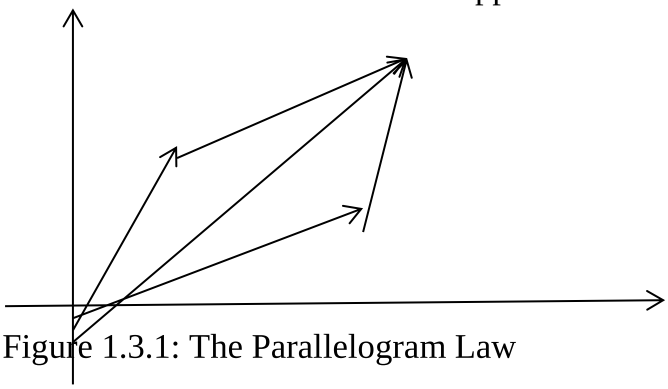

Figure 1.3.1: The Parallelogram Law

b) The ***algebraic*** definition: The points of the plane are represented by Cartesian coordinates (x, y) such that a vector ***v*** is just a pair (x, y) of real numbers. Now we define addition and scalar multiplication by

$$(x_1, y_1) + (x_2, y_2) = (x_1 + x_2, y_1, y_2)$$
$$a(x, y) = (ax, ay)$$

Let's check that the rules for a vector space are really satisfied, for instance, we check the law

$$a(v + w) = av + aw.$$

Let $(x_1, y_1) + (x_2, y_2)$

Then, $a(v + w) = a[(x_1, y_1) + (x_2, y_2)]$
$= a(x_1 + x_2, y_1 + y_2)$
$= (ax_1 + ax_2, ay_1 + ay_2)$
$= (ax_1, ay_1) + (ax_2, ay_2)$
$= av + aw.$

In the algebraic definition, we say that the operations of addition and scalar multiplication are **coordinate-wise**, that is, two vectors can be added coordinate by coordinate, and similarly for scalar multiplication. The generalized form of this example using coordinates is given by the next example:

2. **The *n-tuple*, F^n:**
Let n be any positive integer and F be any field, and let $V = F^n$, be the set of all *n*-tuples of elements of F. Then V is a vector space over F where the operations are defined coordinate-wise:
Let $\alpha = (x_1 + x_2, \dots, x_n)$ of scalars x_i and $\beta = (y_1 + y_2, \dots, y_n)$ with $y_i \in F$.
The sum of $\alpha$and β is defined by

$$\alpha + \beta = (x_1 + x_2, \dots, x_n) + (y_1 + y_2, \dots, y_n)$$
$$= (x_1 + y_1, x_2 + y_2, \dots, x_n + y_n)$$

The scalar multiplication condition would be $c\alpha = (cx_1, cx_2, \dots, cx_n)$
The product of a scalar c and vector α is defined by $c\alpha = (cx_1, cx_2, \dots, cx_n)$

The fact that the vector addition and scalar multiplication satisfy conditions (1) and (2), it is easy to verify, using the similar properties of addition and multiplication of elements of F.

3. The space of $(m \times n)$ matrices $F^{m\times n}$
Let F be any field and let m and n be positive integers. Let $F^{m\times n}$ be the set of all $(m \times n)$
matrices over the field F.

i. The **sum** of two vectors A and B in $F^{m\times n}$ is defined by $(A + B)_{ij} = A_{ij} + B_{ij}$
ii. The **product** of a scalar c and the matrix $(A + B)_{ij}$ is defined by

$$c(A + B)_{ij} = cA_{ij} + cB_{ij}$$

Self-Assessment: From what you just read, can you recount the definition of vector space over C and some of its properties?

1.3.2 Spaces Associated with Vector Spaces

A) The space of functions from a set to a field:
Definition 1.3.4: Let F be any field and S be any non-empty set. Let V be the set of all functions from the set S into F, then,

i. The **sum** of two vectors $f$and g in V is the vector $f + g$, i.e., the function from S into F is defined by $(f + g)(s) = f(s) + g(s)$
ii. The **product** of the scalar c and the function f is the function cfdefined by

$$(cf)(s) = cf(s)$$

B) The space of polynomial functions over a field *F*

Definition 1.3.5: Let F be a field and let V be the set of all functions f from F into F which have a rule of the form: $f(x) = c_0 + c_1x + c_2x^2 + \cdots + c_nx^n$, where $c_0, c_1, c_2, \dots c_n$ are fixed scalars in F(independent of x). A function of this type is called a polynomial function on F.

N.B: Let addition and scalar multiplication be defined as in definition 3.2.1.
One must observe here that if f and g are polynomial functions and $c \in F$, then $f + g$ and cf are again polynomial functions.

Definition 1.3.6: Let $R^n = \{(a_1, \dots, a_n)^T : a_i \in R, 1 \leq i \leq n\}$;$u = (a_1, \dots, a_n)^T$, $v = (b_1, \dots, b_n)^T \in V$ and $\alpha \in R$, we define $u + v = (a_1 + b_1, \dots, a_n + b_n)^T$ and $\alpha \cdot u = (\alpha a_1, \dots, \alpha a_n)^T$ (called component-wise operations), then, V is a **real vector space**.
The vector space R^n is called **the real vector space of *n-tuples*.**

Definition 1.3.7: Let $m, n \in N$ and $M_{m+n}(C) = \{A_{m\times n} = [a_{ij}] \in C\}$, then, with the usual addition and scalar multiplication of matrices, $M_{m+n}(C)$ is **a complex vector space**.
If $m = n$, the vector space $M_{m+n}(C)$ is denoted by $M_n(C)$.

Definition 1.3.8: Let S be a non-empty set and let $R^S = \{f\}$ such that f is a function from S to R. For $f, g \in R^S$ and $\alpha \in R$, define $(f + \alpha g)(x) = f(x) + \alpha g(x)$ for all $x \in S$, then, R^S **is a real vector space.**
In particular, for $S = N$, observe that R^N consists of all real sequences and forms a real vector space.

Let $C(R, R) = \{f: R \to R\}$, such that f is continuous.
Then $C(R, R)$ is a real vector space, where $(f + \alpha g)(x) = f(x) + \alpha g(x) \forall x \in R$.

1.3.3 Definition and Examples of Vector Subspace

Definition 1.3.9: A **subspace** is a vector space that is contained within another vector space. In other words, every subspace is a vector space in its own right, but it is also defined relative to some other (larger) vector space.

The principal definition of subspace is presented below:

Definition 1.3.10: Suppose that *V* and *W* are two vector spaces that have identical definitions of vector addition and scalar multiplication, and that *W* is a subset of *V*, that is, $W \subseteq V$, then *W* is a **subspace** of *V*.

Let us look at an example of a vector space inside another vector space.

Vector Subspace

Definition 1.3.12: Let V be a vector space over F. Then, a non-empty subset W of V is called a subspace of V if W is also a vector space with vector addition and scalar multiplication in W coming from that in V (compute the vector addition and scalar multiplication in V and then the computed vector should be an element of W).

Theorem 1.3.2: Let $V(F)$ be a vector space and $W \subseteq V, W \neq \emptyset$. Then, W is a subspace of V if and only if $\alpha u + \beta v \in W$ whenever $\alpha, \beta \in F$ and $u, v \in W$.

Proof: Let W be a subspace of V and let $u, v \in W$. As $u, v \in W$ is a subspace, the scalar $\alpha, \beta \in F$ and $\alpha u + \beta v \in W$.
Now, we assume that $\alpha u + \beta v \in W$, whenever $\alpha, \beta \in F$ and $u, v \in W$.
To show, W is a subspace of V:

i. Taking $\alpha = 0$ and $\beta = 0 \Rightarrow 0 \in W$. So, W is non-empty.
ii. Taking $\alpha = 1$ and $\beta = 1$, we see that $u + v \in W$, for every $u, v \in W$.
iii. Taking $\beta = 0$, we see that $\alpha u \in W$, for every $\alpha \in F$ and $u \in W$.
iv. Hence, using Theorem 1.3.1(iii) above, $-u = (-1)u \in W$.
v. The commutative and associative laws of vector addition hold as they hold in V.
vi. The conditions related with scalar multiplication and the distributive laws also hold as they hold in V.

Theorem 1.3.3: A non-empty subset W of V is a subspace of V if and only if for each pair of vectors $\alpha, \beta \in W$ and each scalar $c \in F$ the vector $c\alpha + \beta \in W$.

Proof:

Suppose that W is a non-empty subset of V such that $c\alpha + \beta \in W$ for all vector $\alpha, \beta \in W$ and all scalars $c \in F$.
Since W is non-empty, there is a vector $\phi \in W$, which implies that $(-1)\phi + \phi = 0$ is in W.
If α is any vector in W, and c, any scalar, the vector $c\alpha = c\alpha + 0$ is in W.
In particular, $(-1)\alpha = -\alpha$ is in W.
Also, if $\alpha, \beta \in W$, then $\alpha + \beta = 1\alpha + \beta$ is in W.
Thus, W is a subspace of V.
Conversely, if W is a subspace of V, $\alpha, \beta \in W$ and c is a scalar, certainly $c\alpha + \beta \in W$.

Examples:

1. If V is any vector space and V is a subspace of V; the subset consisting of the zero vector alone is a subspace of V, called the zero subspace of V.
2. In F^n, the set of *n-tuples* $x_1, x_2, \dots, x_n$with $x_i = 0$ is a subspace; however, the set of *n-tuples* with $x_1 = 1 + x_2$ is not a subspace $(n \geq 2)$.
3. The space of polynomial functions over the field $F$is a subspace of the space of all functions from F into F.
4. An $n \times n$ (square) matrix Aover the field $F$is symmetric if $A_{ij} = A_{ji}$, for each i and j. The symmetric matrices form a subspace of the space of all $n \times n$ matrices over F.
5. An $n \times n$ (square) matrix $\boldsymbol{A}$ over the field $\boldsymbol{C}$ of complex numbers is Hermitian (or self-adjoint) if $A_{jk} = \bar{A}_{kj}$for each j, k, the bar denoting complex conjugation.

Definition 1.3.13: A 2×2 matrix is **Hermitian** if and only if it has the form

$\begin{bmatrix} z & x+iy \\ x-iy & w \end{bmatrix}$, where x, y, z, w, are real numbers.

The set of all Hermitian matrices is not a subspace of the space of all $n \times n$matrices over $\boldsymbol{C}$. If A is Hermitian, its diagonal entries $A_{11}, A_{22}, \dots, A_{nn}$ are all real numbers, but the diagonal entries of $A_{ij}$are in general not real. On the other hand, it is easily verified that the set of $n \times n$complex Hermitian matrices is a vector space over the field ***R*** of real numbers (with the usual operations).

Theorem 1.3.4: The solution space of a system of homogeneous linear equations:

Let A be an $m \times n$ matrix over *F*, then the set of all $n \times 1$ (column) matrices X over F 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 $A(cX + Y) = 0$ when $AX = 0, AY = 0$, and *c* is an arbitrary scalar in *F*.

This follows immediately from the following general fact.

Lemma: If A is an $m \times n$ matrix over *F* and *B*, *C* are $n \times p$ matrices over *F* then

$A(dB + C) = d(AB) + AC$ for each scalar $d \in F$

Proof:

$$[A(dB + C)]_{ij} = \sum_k A_{ik}(dB + C)_{kj}$$

$$= \sum_k \left(dA_{ik}B_{kj} + A_{ik}C_{kj}\right)$$
$$= d\sum_k A_{ik}B_{kj} + \sum_k A_{ik}C_{kj}$$
$$= d(AB)_{ij} + (AC)_{ij}$$
$$= [d(AB) + AC]_{ij}$$

Similarly, one can show that $(dB + C)A = d(BA) + CA$, if the matrix sums and products are defined.

Theorem 1.3.5: 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:
Let $\{W_a\}$ be a collection of subspaces of V, and let $W = \cap_a W_a$ be their intersection.
Recall that W is defined as the set of all elements belonging to every W_a.
Since each $W_a$is a subspace, each contains the zero vector. Thus, the zero vector is in the intersection W, and W is non-empty.
Let α and β be vectors in W and let c be a scalar.

By definition of W, both $\alpha$and βbelong to each W_a, and because each W_a is a subspace, the vector $c\alpha + \beta$ is in every W_a. Thus$(c\alpha + \beta)$ is again in W.
Thus, by Theorem 1.3.1, W is a subspace of V.
From Theorem 1.3.2, it follows that if S is any collection of vectors in V, then there is a smallest subspace of V which contains S, that is, a subspace which contains S and which is contained in every other subspace containing S.

SELF- ASSESSMENT EXERCISE

i. By Definition 1.3.3, show that the ten properties hold using the two operations on a **vector space** over C.
ii. Enumerate any three (3) examples of vector space.

1.4 Summary

In this unit we have covered the following points:

- A vector space is composed of three objects, a set and two operations which satisfy some properties

- The Euclidean plane R^2, the *n-tuple*, F^n and the space of $(m \times n)$ matrices $F^{m \times n}$ are examples of vector spaces
- A subspace is a vector space that is contained within another vector space.

A **vector space** V over the field F, denoted by $V(F)$, is a non-empty set, in which vector addition, scalar multiplication can be defined.

The vector space R^n is called **the real vector space of *n-tuples*.**
A non-empty subset *W* of a vector space *V* over *F* is called a subspace of *V* if *W* is also a vector space with vector addition and scalar multiplication in *W* coming from that in *V*.
The intersection of any collection of subspaces of *V* is a subspace of *V*.

1.5 References/Further Readings

Robert A. Beezer (2014). A First Course in Linear Algebra Congruent Press, Gig Harbor, Washington, USA 3(40).

Arbind K Lal Sukant Pati (2018). Linear Algebra through Matrices.

Peter J. Cameron (2008). Notes on Linear Algebra.

UNIT 2 LINEAR COMBINATIONS

Unit Structure

2.1 Introduction

The heart of linear algebra is in two operations, both with vectors. We add vectors to obtain $v + w$ and multiply them by numbers or scalars c and d to get cv and dw.Combining those two operations (adding cv to dw) gives the **linear combination** $cv + dw$.

Linear combinations are all-important in this subject! Sometimes, one particular combination is required, the specific choice $c = 2$ and $d = 1$ that produces $cv + dw = (4,5)$.

Other times, we require all the combinations of v and w (coming from all c and d), the vectors cv lie along a line. When w is not on that line, the combinations $cv + dw$fill the whole two-dimensional plane ("two-dimensional" because linear algebra allows higher-dimensional planes).
For four vectors u, v, w, z in four-dimensional space and their combinations:

$cu + dv + ew + jz$ are likely to fill the space but not always.
The vectors and their combinations could even lie on one line.

2.2 Learning Outcomes

By the end of this unit, you will be able to:

- Define Linear Combination of Column Vectors
- Form Linear Combinations given different scalars
- Define Finite dimension of a Vector Space
- Define Consistency of a System
- Obtain the Solution(s) of a system containing a linear combination of the columns
- Define Linear Span of a Collection of Vectors

2.3 Linear Combinations

2.3.1 Linear Combinations of Column Vectors

Definition 2.3.1: Given n vectors $u_1, u_2, u_3, \ldots, u_n$ from the column vector C^m and n-scalars $\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_n$, their linear combination is the vector $\alpha_1 u_1 + \alpha_2 u_2 + \alpha_3 u_3 + \cdots + \alpha_n u_n$.

The definition above combines an equal number of scalars and vectors using the two operations (scalar multiplication and vector addition), thus forming a single vector of the same size as the original vectors.

Example 1: Linear combinations in C^5
Let $\alpha_1 = -1\alpha_2 = -3\alpha_3 = 4\alpha_4 = 2$ and

$$u_1 = \begin{bmatrix} 1 \\ -2 \\ 0 \\ 1 \\ 5 \end{bmatrix} u_2 = \begin{bmatrix} 3 \\ -2 \\ 4 \\ -3 \\ 1 \end{bmatrix} u_3 = \begin{bmatrix} 2 \\ 1 \\ -2 \\ 0 \\ 4 \end{bmatrix} u_4 = \begin{bmatrix} 3 \\ 6 \\ -4 \\ -1 \\ 1 \end{bmatrix}$$

Their linear combination is

$$\alpha_1 u_1 + \alpha_2 u_2 + \alpha_3 u_3 + \alpha_4 u_4$$

$$= (-1)\begin{bmatrix} 1 \\ -2 \\ 0 \\ 1 \\ 5 \end{bmatrix} + (-3)\begin{bmatrix} 3 \\ -2 \\ 4 \\ -3 \\ 1 \end{bmatrix} + (4)\begin{bmatrix} 2 \\ 1 \\ -2 \\ 0 \\ 4 \end{bmatrix} + (2)\begin{bmatrix} 3 \\ 6 \\ -4 \\ -1 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} -1 \\ 2 \\ 0 \\ -1 \\ -5 \end{bmatrix} + \begin{bmatrix} -9 \\ 6 \\ -12 \\ 9 \\ -3 \end{bmatrix} + \begin{bmatrix} 8 \\ 4 \\ -8 \\ 0 \\ 16 \end{bmatrix} + \begin{bmatrix} 6 \\ 12 \\ -8 \\ -2 \\ 2 \end{bmatrix} = \begin{bmatrix} 4 \\ 24 \\ -28 \\ 6 \\ 10 \end{bmatrix}$$

Other different linear combinations can be formed given different scalars, for instant, given $\beta_1 = 2; \beta_2 = -1; \beta_3 = -3$; $\beta_4 = 0$
We can form a linear combination given by

$$\beta_1 u_1 + \beta_2 u_2 + \beta_3 u_3 + \beta_4 u_4$$

$$= (2)\begin{bmatrix} 1 \\ -2 \\ 0 \\ 1 \\ 5 \end{bmatrix} + (-1)\begin{bmatrix} 3 \\ -2 \\ 4 \\ -3 \\ 1 \end{bmatrix} + (-3)\begin{bmatrix} 2 \\ 1 \\ -2 \\ 0 \\ 4 \end{bmatrix} + (0)\begin{bmatrix} 3 \\ 6 \\ -4 \\ -1 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} 4 \\ -4 \\ 0 \\ 2 \\ 10 \end{bmatrix} + \begin{bmatrix} -3 \\ 2 \\ -4 \\ 3 \\ -1 \end{bmatrix} + \begin{bmatrix} -6 \\ -3 \\ 6 \\ 0 \\ -12 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} -5 \\ -5 \\ 2 \\ 5 \\ -3 \end{bmatrix}$$

2.3.2 Linear Combination and Consistency of a System

Definition 2.3.2: A system $Ax = B$ is **consistent** if it has a solution and **inconsistent** if it has no solution. The consistency of the system $Ax = B$ leads to the idea that the vector *B* is a **linear combination** of the columns of *A*.

Example 2: Let $A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}$ and $B = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}$. Then, $\begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$

This implies that $\begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}$ is a linear combination of the vectors in $P = \left\{\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}\right\}$

Similarly, the vector $\begin{bmatrix} 8 \\ 13 \\ 18 \end{bmatrix}$ is a linear combination of the vectors in P as

$$\begin{bmatrix} 8 \\ 13 \\ 18 \end{bmatrix} = 3\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + 5\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = A\begin{bmatrix} 3 \\ 5 \end{bmatrix}$$

Thus, a formal definition of linear combination is given below:

Definition 2.3.3: Let *V* be a vector space over *F* and let $P = \{u_1, \dots, u_n\} \subseteq V$. Then, a vector $u \in V$ is called a **linear combination** of elements of P

if we can find $\alpha_1, \ldots, \alpha_n$, such that $u = \alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_n u_n = \sum_{i-1}^{n} \alpha_i u_i$.
Or equivalently, any vector of the form $\sum_{i-1}^{n} \alpha_i u_i$where $\alpha_1, ... + \alpha_n \in F$ is said to be a **linear combination** of the elements of P.
Thus, the system $Ax = B$ has a solution, meaning that B is a linear combination of the columns of A.
Or equivalently, B is a linear combination means the system $Ax = B$ has a solution.

So, recall that when we were solving a system of linear equations, we looked at the point of intersections of lines or plane etc. But here it leads us to the study of whether a given vector is a linear combination of a given set P or not? Or in the language of matrices, is B a linear combination of columns of the matrix A or not?

Examples 3:
a) Is (4, 5, 5) a linear combination of (1, 0, 0), (2, 1, 0) and (3, 3, 1)?

Solution:
Let $A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 4 \\ 5 \\ 5 \end{bmatrix}$

$$9\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + (-10)\begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} + 5\begin{bmatrix} 3 \\ 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \\ 5 \end{bmatrix}$$

Hence (4, 5, 5) a linear combination of (1, 0, 0), (2, 1, 0) and (3, 3, 1) and

$$= \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 9 \\ -10 \\ 5 \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \\ 5 \end{bmatrix}$$

$A\begin{bmatrix} 9 \\ -10 \\ 5 \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \\ 5 \end{bmatrix}$ of the form $Ax = B$

Thus, $x = [9 \quad -10 \quad 5]^T$ is a solution.

b) Find condition(s) on $x, y, z \in R$ such that

i. (x, y, z) is a linear combination of $(1, 2, 3), (-1, 1, 4)$ and $(3, 3, 2)$.
ii. (x, y, z) is a linear combination of $(1, 2, 1), (1, 0, -1)$ and $(1, 1, 0)$.
iii. (x, y, z) is a linear combination of $(1, 1, 1), (1, 1, 0)$ and $(1, 1, 0)$.

Solution:

i. $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + \begin{bmatrix} -1 \\ 1 \\ 4 \end{bmatrix} + \begin{bmatrix} 3 \\ 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ 6 \\ 9 \end{bmatrix}$

ii. $\begin{bmatrix}1\\2\\1\end{bmatrix}+\begin{bmatrix}1\\0\\-1\end{bmatrix}+\begin{bmatrix}1\\1\\0\end{bmatrix}=\begin{bmatrix}3\\3\\0\end{bmatrix}$

iii. $\begin{bmatrix}1\\1\\1\end{bmatrix}+\begin{bmatrix}1\\1\\0\end{bmatrix}+\begin{bmatrix}1\\1\\0\end{bmatrix}=\begin{bmatrix}3\\3\\1\end{bmatrix}$

Self-Assessment: Now, in simple sentences, give a simple description of linear combination of column vectors

2.3.3 Linear Span

Definition 2.3.4: Let *V* be a vector space over *F* and *S* a subset of *V*. We now look at **'linear span'** of a collection of vectors. So, here we ask;"What is the largest collection of vectors that can be obtained as linear combination of vectors from *S*" Or equivalently, what is the smallest subspace of *V* that contains *S*?

Example 4

Let$S=\{(1,0,0),(1,2,0)\}\subseteq R^3$. Let us find the largest possible subspace of R^3 which contains vectors of the form $\alpha(1,0,0),\beta(1,2,0)$ and $(1,0,0),(1,2,0)$ for all possible choices of$\alpha,\beta\in R$. Note that

a. $l_1=\{\alpha(1,0,0):\alpha\in R\}$ gives the *x*-axis.

b. $l_2=\{\beta(1,2,0):\beta\in R\}$ gives the line passing through (0, 0, 0) and (1, 2, 0).

So, we want the largest subspace of R^3 that contains vectors which are formed as sum of any two points on the two lines $l_1$and l_2 , or the smallest subspace of R^3 that contains *S*?

2.3.4 Finite and Infinite Dimensional

Definition 2.3.5: Let *V* be a vector space over *F* and $S\subseteq V$, then

1. the linear span of *S*, denoted by ***Ls(S),*** is defined as $Ls(S)=\{\alpha_1u_1+\cdots+\alpha_nu_n \mid \alpha_1\in F,u_i\in S \text{ for } 1\le i\le n\}$. This implies that *Ls(S)* is the set of all possible linear combinations of finitely many vectors of *S*. If *S* is an empty set, we define $Ls(S)=\{0\}$.
2. *V* is said to be **finite dimensional** if there exists a finite set *S* such that *V = L(S).*
3. If there does not exist any finite subset *S* of *V* such that *V = L(S)* then *V* is called **infinite dimensional**.

Example 5:
For each set *S* given below, determine *Ls(S).*
a) $S = \{(1,0)^T, (0,1)^T\} \subseteq R^2$.

Solution: $Ls(S) = \{a(1,0)^T + b(1,0)^T | a, b \in R\} = \{(a, b)^T | a, b \in R\} = R^2$.
Thus, $R^2$is finite dimensional.
b) $S = \{(1,1,1)^T, (2,1,3)^T\}$. What does *L(S)* represent inR^3?

Solution:
$Ls(S) = \{a(1,1,1)^T + b(2,1,3)^T | a, b \in R\} = \{(a + 2b, a + b, a + 3b)^T | a, b \in R\}$.
Note that *L(S)* represents a plane passing through the points $(0,0,0)^T, (1,1,1)^T and (2,1,3)^T$.
To obtain the equation of the plane, we proceed as follows:
Find conditions on *x*, *y* and *z* such that $\{(a + 2b, a + b, a + 3b) = (x, y, z)\}$.
Or equivalently, find conditions on *x*, *y* and *z* such that $a + 2b = x$; $a + b = y$ and
$a + 3b = z$ has a solution for all $a, b \in R$.
The Row-Reduced Echelon Form(RREF) of the augmented matrix equals

$$\begin{bmatrix} 1 & 0 & 2y - x \\ 0 & 1 & x - y \\ 0 & 0 & z + y - 2x \end{bmatrix}$$

Thus, the required condition on *x*, *y* and *z* is given by $z + y - 2x = 0$.
Hence, $Ls(S) = \{a(1,1,1)^T + b(2,1,3)^T | a, b \in R\} = \{(x, y, z)^T \in R^3 | 2x - y - z = 0\}$.
Verify that if $T = S \cup \{(1,1,0)^T\}$, then $Ls(T) = R^3$.
Hence, R^3 is finite dimensional.
In general, for every fixed $n \in N, R^n$is finite dimensional as $R^n = Ls(\{e_1, ... e_n\})$.

c) $S = \{1 + 2x + 3x^2, 1 + x + 2x^2, 1 + 2x + x^3\}$

Solution:
To understand *Ls(S)*, we need to find condition(s) on $\alpha, \beta, \gamma, \delta$such that the linear system:
$a(1 + 2x + 3x^2) + b(1 + x + 2x^2) + c(1 + 2x + x^3) = \alpha + \beta x + \gamma x^2 + \delta x^3$
in the unknowns *a*, *b*, *c* is always consistent.
An application of Gauss-Jordan Elimination (GJE) method gives $\alpha + \beta - \gamma - 3\delta = 0$ as the required condition.
Thus,
$Ls(S) = \{\alpha + \beta x + \gamma x^2 + \delta x^3 \in R[x]: \alpha + \beta - \gamma - 3\delta = 0\}$.

Note that, for every fixed $n \in N, R[x;n]$is finite dimensional as $R[x;n] = Ls(\{1, x, \dots x^n\})$.

d) $S = \left\{ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 1 \\ 1 & 1 & 2 \\ 1 & 2 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 2 \\ 1 & 0 & 2 \\ 2 & 2 & 4 \end{bmatrix} \right\} \subseteq M_3(R)$.

Solution: To get the equation, we need to find conditions on $a_{ij}'s$ such that the system

$$\begin{bmatrix} \alpha & \beta + \gamma & \beta + 2\gamma \\ \beta + \gamma & \alpha + \beta & 2\beta + 2\gamma \\ \beta + 2\gamma & 2\beta + 2\gamma & \alpha + 2\gamma \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

in the unknowns α, β, γ is always consistent.

Now, verify that the required condition equals

$$Ls(S) = \left\{ A = [a_{ij}] \in M_3(R): A = A^T, a_{11} = \frac{a_{22} + a_{33} - a_{13}}{2}, a_{12} = \frac{a_{22} - a_{33} + 3a_{13}}{4}, a_{23} = \frac{a_{22} - a_{33} + 3a_{13}}{2} \right\}$$

In general, for each fixed $m, n \in N$, the vector space $M_{m,n}(R)$ is finite dimensional

$M_{m,n}(R) = Ls(\{e_{ij}: 1 \le i \le m, 1 \le j \le n\})$.

The vector space *R* over *Q* is infinite dimensional.

Definition 2.3.6: Let *S* be a set of vectors in a vector space *V*. The **subspace *spanned*** by *S* is defined to be the intersection *W* of all subspaces of *V* which contain *S*.

When *S* is a finite set of vectors, $S = \{\alpha_1, \alpha_2, \dots, \alpha_n\}$, we shall simply call *W* the *subspace spanned* by the vectors $\alpha_1, \alpha_2, \dots, \alpha_n$.

Theorem 2.3.1: The subspace spanned by a non-empty subset *S* of a vector space *V* is the set of all linear combinations of vectors in *S*.

Proof: Let *W* be the subspace spanned by *S*. Then each linear combination $\alpha = x_1\alpha_1 + x_2\alpha_2 + \cdots + x_m\alpha_m$ of vectors $\alpha_1, \alpha_2, \dots, \alpha_m$in *S* is clearly in *W*.

Thus, *W* contains the set *L* of all linear combinations of vectors in *S*.

The set *L*, on the other hand, contains *S* and is non-empty.

If α, β belong to *L* then $\alpha$is a linear combination, $\alpha = x_1\alpha_1 + x_2\alpha_2 + \cdots + x_m\alpha_m$ of vectors $\alpha_i$in S, and $\beta$is a linear combination $\beta = y_1\beta_1 + y_2\beta_2 + \cdots + y_n\beta_n$ of vectors β_j in *S*

For each scalar *c*, $c\alpha + \beta = \sum_{i=1}^{m}(cx_i)\alpha_i + \sum_{j=1}^{n} y_j\beta_j$

Hence $c\alpha + \beta$belongs to *L*.

Thus, *L* is a subspace of *V*.

Now we have shown that *L* is a subspace of V which contains *S*, and also that any subspace which contains *S* contains *L*. It follows that *L* is the intersection of all subspaces containing *S*, that is, that *L* is the subspace spanned by the set *S*.

Definition 2.3.7: If $S_1, S_2, \ldots, S_k$ are subsets of a vector space *V*, the set of all sums $\alpha_1 + \alpha_2 + \cdots + \alpha_k$ of vectors α_i in S_i is called the **sum of the subsets** $S_1, S_2, \ldots, S_k$ and is denoted by $S_1 + S_2 + \cdots + S_k$ or $\sum_{i=1}^{k} S_i$
If $W_1, W_2, \ldots, W_k$ are subspaces of V, then the sum is easily seen to be a subspace of V which contains each of the subspaces W_i. From this it follows, as in the proof of Theorem 3, that *W* is the subspace spanned by the union of $W_1, W_2, \ldots, W_k$.

Example 6: Let *F* be a subfield of the field *C* of complex numbers. Suppose

$$\alpha_1 = (1,2,0,3,0), \alpha_2 = (0,0,1,4,0), \alpha_3 = (0,0,0,0,1).$$

By Theorem 2.3.1, 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 $\alpha = c_1, 2c_1, c_2, 3c_1 + 4c_2, \alpha_3$, where c_1, c_2, c_3 are arbitrary scalars in *F*.
Alternatively, W can be described as the set of all 5-tuples $\alpha = (x_1, x_2, x_3, x_4, x_5)$ with x_i in *F* such that $x_2 = 2x_1$ and $x_4 = 3x_1 + 4x_3$
Thus $(-3, -6, 1, 5, 2)$ is in *W*, whereas $(2, 4, 6, 7, 8)$ is not.

2.4 Linear Independence

2.4.1 Linearly Independent Vectors

Definition 2.4.1:
(a) Let *V* be a vector space over the field *K*, and let $S = \{v_1, \ldots, v_n\}$ be a non-empty subset of containing vectors in *V*. The vectors $v_1, v_2 \ldots, v_n$ are **linearly independent** if, whenever there exists scalars $c_1, \ldots, c_n$ satisfying $c_1v_1 + c_2v_2 \ldots + c_nv_n = 0$, then necessarily $c_1 = c_2 = \cdots = c_n = 0_n$.
(b) The vectors $v_1, v_2 \ldots, v_n$ are *spanning* if, for every vector $v \in V$, we can find scalars $c_1, c_2 \ldots, c_n \in K$ such that $v = c_1v_1 + c_2v_2 \ldots + c_nv_n$.
In this case, we write $V = (v_1, v_2 \ldots, v_n)$.
(c) The vectors $v_1, v_2 \ldots, v_n$ form a *basis* for V if they are linearly independent and spanning.

Remarks:
- Linear independence is a property of a *list* of vectors.
- A list containing the zero vector is never linearly independent.
- Also, a list in which the same vector occurs more than once is never linearly independent.

Definition 2.4.2: Let *V* be a vector space over the field *K*, then *V* is ***finite dimensional*** if vectors $v_1, v_2 \dots, v_n \in V$ can be found to form a basis for *V*.

Proposition: The following three conditions are equivalent for the vectors $v_1, v_2 \dots, v_n$ of the vector space *V* over *K*:

a) $v_1, v_2 \dots, v_n$ is a basis;

b) $v_1, v_2 \dots, v_n$ is a maximal linearly independent set (that is, if we add any vector to the list, then the result is no longer linearly independent);

c) $v_1, v_2 \dots, v_n$ is a minimal spanning set (that is, if we remove any vector fromthe list, then the result is no longer spanning).

2.4.2 Properties of Linear Independence:

Theorem 2.4.1 (The Exchange Lemma) Let *V* be a vector space over *K*. Suppose that the vectors $v_1, v_2 \dots, v_n$ are linearly independent, and that the vectors $w_1, w_2 \dots, w_m$ are linearly independent, where $m > n$. Then we can find a number i with $1 \le i \le m$ such that the vectors $v_1, \dots, v_n, w_i$ are linearly independent.

A lemma about systems of equations would be used to prove this theorem.
Lemma: Given a system $(*)$

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1m}x_m &= 0 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2m}x_m &= 0 \\ &\vdots \\ &\vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nm}x_m &= 0 \end{aligned} \qquad \dots\dots\dots\dots (*)$$

of homogeneous linear equations, where the number *n* of equations is strictly less than the number *m* of variables, there exists a non-zero solution $(x_1, \dots, x_m)$ (that is,$x_1, \dots, x_m$ are not all zero).

Proof:
This is proved by induction on the number of variables. If the coefficients $a_{11}, a_{21} \dots, a_{n1}$ of x_1 are all zero, then putting $x_1 = 1$and the other variables zero gives a solution.

If one of these coefficients is non-zero, then we can use the corresponding equation to express x_1 in terms of the other variables, obtaining $(n-1)$ equations in $(m-1)$ variables.
By hypothesis, $n-1 < m-1$.
So, by the induction hypothesis, these new equations have a non-zero solution.

Computing the value of x_1 gives a solution to the original equations.
Now we turn to the proof of the Exchange Lemma.

Let us argue for a contradiction by assuming that the result is false; that is, assume that none of the vectors w_i can be added to the list $(v_1, v_2 \dots, v_n)$ to produce a larger linearly independent list, this means that, for all *j*, the list $(v_1, \dots, v_n, w_i)$ is linearly dependent.
So, there are coefficients $c_1, \dots, c_n, d$, not all zero, such that $c_1v_1 + \cdots + c_nv_n + dw_i = 0$.

We cannot have $d = 0$; for this would mean that we had a linear combination of $v_1, v_2 \dots, v_n$ equal to zero, contrary to the hypothesis that these vectors are linearly independent.

So, we can divide the equation through by *d*, and take w_i to the other side, to obtain (changing notation slightly)

$$w_i = a_{1i}v_1 + a_{2i}v_2 + \cdots + a_{ni}v_n = \sum_{j=1}^{n} a_{ji}v_j \qquad (i)$$

We do this for each value of $i = 1, \dots, m$.
Now take a non-zero solution to the set of equations (i) above: that is,

$$\sum_{i=1}^{m} a_{ji}x_i = 0; \quad j = 1, \dots, n \qquad (ii)$$

Multiplying the formula for w_i by x_i and adding, we obtain

$$x_1w_1 + \cdots + x_mw_m = \sum_{j=1}^{n}\left(\sum_{i=1}^{m} a_{ji}x_i\right)v_j = 0 \qquad (iii)$$

But the coefficients are not all zero, so this means that the vectors $(w_1, w_2 \dots, w_m)$ are not linearly dependent, contrary to hypothesis.
So, the assumption that no w_i can be added to $(v_1, v_2 \dots, v_n)$ in order to obtain a linearly independent set must be wrong, and the proof is complete.

Definition 2.4.3: Let *B* be a subset of a set *A*. Then, *B* is said to be a **maximal subset** of *A* having property *P* if

1. *B* has property *P*
2. No proper superset of *B* in *A* has property *P*.

Example 7: Let T = {2, 3, 4, 7, 8, 10, 12, 13, 14, 15}. Then, a maximal subset of *T* of consecutive integers is *S* = {2, 3, 4}.
Other maximal subsets are {7, 8}, {10} and {12, 13, 14, 15}. Note that {12, 13} is not maximal. Why?

Definition 2.4.4: Let V be a vector space over F. Then, S is called a **maximal linearly independent subset** of V if

1. S is linearly independent and
2. no proper superset of $S \in V$ is linearly independent.

Example 8:

a. In R^3, the set $S = \{e_1, e_2\}$ is linearly independent but not maximal as $S \cup \{(1,0,0)^T\}$ is a linearly independent set containing S.

b. In R^3, set $S = \{(1,0,0)^T, (1,1,0)^T, (1,1,-1)^T\}$ is a maximal linearly independent set as S is linearly independent and any collection of four or more vectors from R^3 is linearly dependent.

c. Is the set $\{1, x, x^2, \ldots\}$ a maximal linearly independent subset of $C[x]$ over C ?

d. Is the set $\{1 \leq i \leq m, 1 \leq j \leq n\}$ a maximal linearly independent subset of $M_{m,n}(C)$ over C?

Theorem 2.4.2:
Let V be a vector space over F and S a linearly independent set in V. Then, S is maximal linearly independent if and only if $Ls(S) = V$.

Proof: Let $v \in V$. As S is linearly independent, using Corollary, the set $S \cup \{v\}$ is linearly independent if and only if $v \in V \backslash Ls(S)$.
Thus, the required result follows.

Let $V = Ls(S)$ for some set S with $|S| = |T|$. Then, using the Theorem 2, we see that if $T \subseteq V$ is linearly independent then $|T| \leq k$. Hence, a maximal linearly independent subset of V can have at most k vectors. Thus, we arrive at the following important result.

Theorem 2.4.3: Let V be a vector space over F and let S and T be two finite maximal linearly independent subsets of V. Then, $|S| = |T|$.
Proof: By Theorem 2, S and T are maximal linearly independent if and only if
$Ls(S) = V = Ls(T)$.
Now, use the previous paragraph to get the required result.
Let V be a finite dimensional vector space. Then, by Theorem 3.4.6, the number of vectors in any two maximal linearly independent set is the same.

We would now use this number to now define the dimension of a vector space.

Definition 2.4.5: Let V be a finite dimensional vector space over F. Then, the ***number*** of vectors in any maximal linearly independent set is called the **dimension of *V***, denoted$dim(V)$.
By convention,$dim(\{0\}) = 0$

Examples 9:

a. As $\{1\}$ is a maximal linearly independent subset of R, dim(R) = 1.

b. As $\{e_1, \dots, e_n\}$ is a maximal linearly independent subset in$R^n, dim(R^n) = n.$

c. As $\{e_1, \dots, e_n\}$ is a maximal linearly independent subset in C^n over C, $dim(C^n) = n.$

d. Using 9c, $\{e_1, \dots, e_n, ie_1, \dots, ie_n\}$ is a maximal linearly independent subset in C^n over R. Thus, as a real vector space, $dim(C^n) = 2n.$

e. As $\{e_{1j} | 1 \leq i \leq m, 1 \leq j \leq n\}$ is a maximal linearly independent subset of $M_{m,n}(C)$ over C, $R^n, dim\left(M_{m,n}(C)\right) = mn.$

SELF-ASSESSMENT EXERCISE(S)

1. Use different sets of scalars to construct different vectors. You might build a few new linear combinations of u_1, u_2, u_3, u_4
2. Let$S = \{(1,1,1,1)^T, (1,-1,1,2)^T, (1,1,-1,1)^T\} \subseteq R^4$. Does $(1,-1,1,2)^T \in Ls(S)$?
   Determine conditions on x, y, z and u such that $(x, y, z, u)^T \in Ls(S)$
3. Prove that if two vectors are linearly dependent, one of them is a scalar multiple of the other.
4. Are the vectors:
   $a_1 = (1,1,2,4), a_2 = (2,-1,-5,2), a_3 = (1,-1,-4,0), a_4 = (2,1,1,6)$ linearly independent in R^4?
5. Find a basis for the subspace of R^4 spanned by the four vectors of Exercise 2 above.
6. Show that the vectors $a_1 = (1,0,-1), a_2 = (1,2,1), a_3 = (0,-3,2)$form a basis for R^3. Express each of the standard basis vectors as linear combinations of a_1, a_2 and a_3.
7. Find three vectors in R^3which are linearly dependent, and are such thatany two of them are linearly independent.

2.5 Summary

The largest collection of vectors that can be obtained as linear combination of vectors is called 'linear span and that $\{1\}$ is a maximal linearly independent subset of R, dim(R) = 1.

Also, if V be a vector space over F and S a linearly independent set in V, then, S is maximal linearly independent if and only if $Ls(S) = V$.

The **subspace *spanned*** by S is defined to be the intersection W of all subspaces of V which contain S.

The vector $\alpha_1 u_1 + \alpha_2 u_2 + \alpha_3 u_3 + \cdots + \alpha_n u_n$ is the linear combination of n-vectors $u_1, u_2, u_3, \ldots, u_n$ from the column vector C^m and *n*-scalars $\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_n$

The number of vectors in any maximal linearly independent set of a finite dimensional vector space *V* over *F* is called the dimension of *V*.

2.6 References/Further Readings

Robert A. Beezer (2014). A First Course in Linear Algebra. Congruent Press Gig Harbor, Washington, USA 3(40).

Arbind K Lal Sukant Pati (2018). Linear Algebra through Matrices.

Peter J. Cameron (2008). Notes on Linear Algebra.

UNIT 3 LINEAR TRANSFORMATIONS I

Unit Structure

3.1 Introduction

You have already learnt about vector space and several concepts related to it. In this unit we initiate the study of certain mappings between two vector spaces, called linear transformations. The importance of these mappings can be realized from the fact that, in the calculus of several variables, every continuously differentiable function can be replaced, to a first approximation, by a linear one. This fact is a reflection of a general principle that every problem on the change of some quantity under the action of several factors can be regarded, to a first approximation, as a linear problem. It often turns out that this gives an adequate result. Also, in physics it is important to know how vectors behave under a change of the coordinate system. This requires a study of linear transformations.

In this unit we study linear transformations and their properties, as well as two spaces associated with a linear transformation and their properties, as well as two spaces associated with a linear transformation, and their dimensions. Then, we prove the existence of linear transformations with some specific properties, as discuss the notion of an isomorphism between two vector spaces, which allows us to say that all finite-dimensional vector spaces of the same dimension are the “same”, in a certain sense.

Finally, we state and prove the Fundamental Theorem of Homomorphism and some of its corollaries, and apply them to various situations.

3.2 Learning Outcomes

By the end of this unit, you should be able to:

- Verify the linearity of certain mappings between vector spaces;
- Construct linear transformations with certain specified properties;
- Define the Range and the Kernel of Linear Transformation
- Calculate the rank and nullity of a linear operator;
- Prove and apply the Rank Nullity Theorem;
- Define an isomorphism between two vector spaces;
- Show that two vector spaces are isomorphic if and only if they have the same dimension;
- Prove and use the fundamental theorem of homomorphism.

3.3 Linear Transformations

By now you are familiar with vector spaces R^2 and R^3. Now consider the mapping
$f: R^2 \to R^3 | f(x, y) = (x, y, 0)$. f is a well-defined function.
Also notice that

i. $f\big((a, b) + (c, d)\big) = f\big((a + c, b + d)\big) = (a + c, b + d, 0) = (a, b, 0) + (c, d, 0)$ for $(a, b), (c, d) \in R^2$ and

ii. For any $\alpha \in R$ and $(a, b) \in R^2$, $f\big((\alpha a, \alpha b)\big) = (\alpha a, \alpha b, 0) = \alpha f\big((a, b)\big)$.

So, we have a function f between two vector spaces such that (i) and (ii) above hold true.

i. says that the sum of two plane vectors is mapped under f to the sum to sum of their images under f.

ii. says that a line in the plane R^2 is mapped under f to a line in R^2. Properties i) and ii) together say that f is linear, a term that we now define.

Definition 3.3.1: Let U and V be vector spaces over a field F. A **linear transformation** (or *linear operator*) from U to V is a function $T: U \to V$, such that

$\boldsymbol{LT_1}$**:** $T(u_1 + u_2) = T(u_1) + T(u_2)$, for $u_1, u_2 \in U$ and

$\boldsymbol{LT_2}$**:** $T(\alpha u) = \alpha T(u)$ for $\alpha \in F$ and $u \in U$.

Conditions (i) and (ii) above can be combined to give the following equivalent condition.

$\boldsymbol{LT_3}$: $T(\alpha_1 u_1 + \alpha_2 u_2) = \alpha_1 T(u_1) + \alpha_2 T(u_2)$, for $\alpha_1, \alpha_2 \in F u_1, u_2 \in U$.

What we are saying is that [$LT_1$and LT_2] implies LT_3. This can be easily shown as follows:
We will show that $LT_3 \rightarrow LT_1$; $LT_3 \rightarrow LT_2$.
Now, $LT_3$is true for all $\alpha_1, \alpha_2 \in F$.
Therefore, it is certainly true for$\alpha_1 = \alpha_2$, that is, LT_1 holds.
Now, to show thatLT_2 is true,
Consider $T(\alpha u)$ for any$\alpha \in F$and$u \in U$.
We have $T(\alpha u) = T(\alpha u + 0 \cdot u) = \alpha T(u) + 0 \cdot T(u) = \alpha T(u)$, thus proving that LT_2 holds.
You can try and prove the converse now, that is, what the following exercise is all about!
E1) Show that the conditions $LT_1$and LT_2together imply LT_3.

Before going further, let us note two properties of any linear transformation, $T: U \rightarrow V$, which follow from LT_1(or LT_2or LT_3).
$\boldsymbol{LT_4}$**:**$T(0) = 0$.
Let's see why this is true.
Since $(0) = T(0 + 0) = T(0) + T(0)$, by LT_1, we subtract $T(0)$ from both sides to get

$$T(0) = 0.$$

$\boldsymbol{LT_5}$**:**$T(-u) = -T(u)$ for all $u \in U$. Why is this so?
Well, since, $0 = T(0) = T(u - u) = T(u) + T(-u)$;
We have $T(-u) = -T(u)$.

E2) Can you show how $LT_4$and LT_5 will follow from LT_2?
Now let us look at some common linear transformations.
Example 1: Consider the vector space U over a field F, and the function $T: U \rightarrow V$, defined by $T(u) = u$ for all $u \in U$. Show that T is a linear transformation. (This transformation is called the **identity transformation,** and is denoted by Iu, or just$I$, if the underlying vector space is understood).

Solution: For any $\alpha, \beta \in F$ and $u_1, u_2 \in U$, we have

$$T(\alpha u_1 + \beta u_2) = \alpha u_1 + \beta u_2 = \alpha T(u_1) + \beta T(u_2)$$

Hence, LT_3 holds, and T is a linear transformation.

Example 2: Let $T: U \rightarrow V$be defined by $T(u) = 0$for all $u \in U$. Check that T is a linear transformation. (It is called the **Null** or **Zero Transformation,** and is denoted by **0**).
Solution: For any $\alpha, \beta \in F$ and $u_1, u_2 \in U$,
we have $T(\alpha u_1 + \beta u_2) = 0 = \alpha \cdot 0 + \beta \cdot 0$
Therefore, T is linear transformation.
Example 3: Consider the function $Pr_1: R_n \rightarrow R$, defined by $Pr_1[(x_1, \dots, x_n)] = x_i$. Show that this is a linear transformation. (This is called the projection on the first coordinate).

Similarly, we can define $Pr_i: R_n \to R$ by $Pr_i[(x_1, \ldots, x_{i-1}, x_i \ldots, x_n)] = x_i$ to be the **projection** on the i^{th}**coordinate** for $i = 2, \ldots, n$.
For instance, $Pr_2: R_3 \to R$ by $Pr_2[(x, y, z)] = y$
Solution: We will use LT_3 to show that projection is a linear operator.
For $\alpha, \beta \in R$ and $(x_1, \ldots \ldots \ldots x_n), (y_1, \ldots \ldots \ldots y_n) \in Rn$, we have
P$Pr_1[\propto (x_1, \ldots \ldots, x_n) + \beta(y_1, \ldots \ldots, y_n)] = Pr_1(\alpha x_1 + \beta y_1, \alpha x_2 + \beta y_2, \ldots \ldots, \alpha x_n + \beta y_n)$

$$= \alpha x_1 + \beta y_1$$
$$= \alpha Pr_1[x_1, \cdots \cdots, x_n] + \beta Pr_1[(y_1, \cdots \cdots, y_n)]$$

Thus Pr_1 (and similarly Pr_i) is a linear transformation.

Before going to the next example, we make a remark about projections.
Remark: Consider the function$P: R_3 \to R_2: P(x, y, z) = (x, y)$, this is a projection from R^3 on to the *xy*-plane. Similarly, the functions f and g, from$R_3 \to R_2$, defined by $f(x, y, z) = (x, z)$, and $g(x, y, z) = (y, z)$are projections from R^3onto the xz-plane and the yz-plane, respectively.
In general, any function $\theta: R_n \to R_m (n > m)$, which is defined by dropping any $(n - m)$ coordinate, is a projection map.
Now let us see another example of a linear transformation that is very geometric in nature.

Example 4: Let $T: R^2 \to R^2$ be defined by $T(x, y) = (x, -y) \ \forall \ x, y \in R$. Show that T is a linear transformation. (This is the **reflection** in the *x*-axis that we show in Fig. 2).

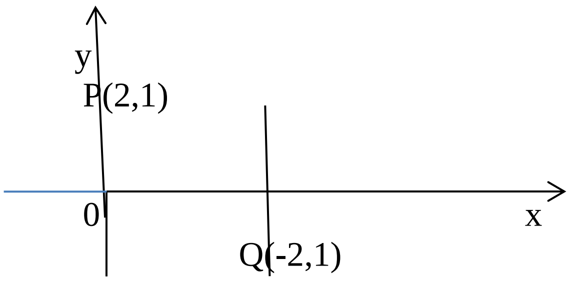

Fig 2: Q is the reflection of P in the X-axis.

Solution: For $\alpha, \beta \in R$ and $(x_1, y_1), (x_2, y_2) \in R^2$, we have

$$T[\alpha(x_1, y_1) + \beta(x_2, y_2)] = T(\alpha x_1 + \beta x_1, \alpha y_1 + \beta y_2)$$
$$= (\alpha x_1 + \beta x_2, -\alpha y_1 - \beta y_2)$$
$$= \alpha(x_1, -y_1) + \beta(x_2, -y_2)$$
$$= \alpha T(x_1, y_1) + \beta T(x_2, y_2)$$

Therefore, T is a linear transformation.
So, far we have given examples of linear transformations. Now, we give an example of a very important function which is not linear. This example's importance lies in its geometric applications.

Example 5: Let u_0 be a fixed non-zero vector in U. Define $T: U \to U$ by $T(u) = u + u_0, \forall u \in U$. Show that T is not a linear transformation. (T is called the translation by u_0. See Fig 3 for a geometrical view).

Solution: T is not a linear transformation since $LT4$ does not hold.
This is because $T(0) = u_0 \neq 0$

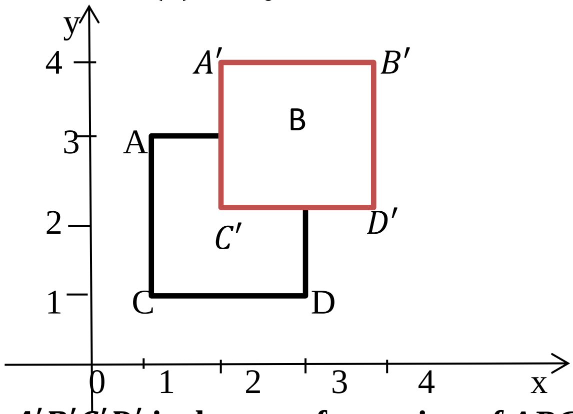

Fig. 3:** $A'B'C'D'$* **is the transformation** ***of ABCD by **(1,1).**

Now, try the following Exercises.

E3) Let $T: R_2 \to R_2$ be the reflection in the *y*-axis. Find an expression for T as in Example 4.
Is T a linear operator?

E4) For a fixed vector (a_1, a_2, a_3) in R_3, define the mapping $T: R_3 \to R$ by $T(x_1, x_2, x_3) = a_1x_1 + a_2x_2 + a_3x_3$. Show that T is a linear transformation.
Note that $T(x_1, x_2, x_3)$ is the dot product of (x_1, x_2, x_3) and (a_1, a_2, a_3).

E5) Show that the map $T: R_3 \to R_3$ defined by
$T(a_1, a_2, a_3) = (x_1 + x_2 - x_3, 2x_1 - x_2, x_2 + 2x_3)$ is a linear operator.
Let us consider the real vector space P_n of all polynomials of degree less than or equal to n.

E6) Let $f \in P_n$ be given by $F(x) = \alpha_0 + \alpha_1 x + a_2 x^2 \ldots \ldots + \alpha_n x^n$, $\alpha_i \in R\ \forall i$.
We define $(Df)(x) = \alpha_1 + 2\alpha_2 x + \cdots \ldots + n\alpha_n x^{n-1}$. Show that $D: P_n$ is a linear transformation. (Observe that Df is nothing but the derivative of *f* and *D* is called the **differentiation operator**).
There is also the concept of a quotient space.
We now define a very useful linear transformation, using this concept.

Example 6: Let *W* be a subspace of a vector space *U* over a field *F*. *W* gives rise to the quotient space *U/W*. Consider the map $T: U \to U/W$ defined by $T(u) = u + W$. Show that T is a linear transformation.

Solution: For any $\alpha, \beta \in F$ and $u_1, u_2 \in U$, we have

$$T(\alpha u_1 + \beta u_2) = \alpha u_1 + \beta u_2 + W$$
$$= (\alpha u_1 + W) + (\beta u_2 + W)$$
$$= \alpha(u_1 + W) + \beta(u_2 + W)$$

$$= \alpha T(u_1) + \beta T(u_2)$$

Thus, T is a linear transformation.
Now solve the following exercise.

Example 7: Let $u_1 = (1, -1), u_2 = (2, -1), u_3 = (4, -3), v_1 = (1,0), v_2 = (0, 1)$ and $v_3 = (1, 1)$ be six vectors in R_2.
Can you define a linear transformation $T: R_2 \to R_2$ such that $(u_1) = v_i, i = 1,2,3$?
(Hint: Note that $2u_1 + u_2 = u_3$ and $v_1 + v_2 = v_3$).
You have already seen that a linear transformation $T: U \to V$ must satisfy $T(\alpha_1 u_1 + \alpha_2 u_2) = \alpha_1 T(u_1) + \alpha_2 T(u_2)$, for $\alpha_1, \alpha_2 \in F$ and $u_1, u_2 \in U$.
More generally, we can show that,
$LT6$: $T(\alpha_1 u_1 + \cdots + \alpha_n u_n) = \alpha_1 T(u_1) + \cdots + \alpha_n T(u_n)$; where $\alpha \in F$ and $u_i \in U$.
This shall be shown by induction, that is, we assume the above relation for $n = m$, and prove it for $m + 1$.
Now, $T(\alpha_1 u_1 + \cdots \alpha_m u_m + \alpha_{m+1} u_{m+1}) = T(u + \alpha_{m+1} u_{m+1})$
where, $u = \alpha_1 u_1 + \cdots + \alpha_m u_m$
$= T(u) + \alpha_{m+1} T(u_{m+1})$, Since the result holds for $n = 2$

$$= T(\alpha_1 u_1 + \cdots \alpha_m u_m) + \alpha_{m+1} T(u_{m+1})$$
$$= \alpha_1 T(u_1) + \cdots + T(\alpha_m) T(u_m) + \alpha_{m+1} T(u_{m+1})$$

Since we have assumed the result for $n = m$.
Thus, the result is true for $n = m + 1$. Hence, by induction, it holds true for all n.

Let us now come to a very important property of any linear transformation $T: U \to V$.
In the earlier unit, we mentioned that every vector space has a basis. Thus, *U* has a basis.
We will now show that *T* is completely determined by its values on a basis of *U*. More precisely, we have:

Theorem 3.1: Let S and T be two linear transformations from *U* to *V*, where, $dim_1\, U = n$. Let $(e_1, \ldots \ldots, e_n)$ be a basis of *U*. Suppose $S(e_i)$for $i = 1, \ldots, n$. Then, $S(u) = T(u)$ for all $u \in U$.

Proof: Let $u \in U$. Since $(e_1, \ldots \ldots, e_n)$is a basis of *U*, *u* can be uniquely written as:
$u = \alpha_1 e_1 + \ldots \ldots + \alpha_n e_n$, where the α_i are scalars.
Then, $S(u) = S(\alpha_1 e_1 + \ldots \ldots + \alpha_n e_n)$

$$= \alpha_1 S(e_1) + \ldots \ldots + \alpha_n S(e_n) \text{ by LT6}$$
$$= \alpha_1 T(e_1) + \ldots \ldots + \alpha_n T(e_n)$$
$$= \alpha_1 (\alpha_1 e_1 + \ldots \ldots + \alpha_n e_n) \text{ by LT6}$$
$$= T(u).$$

What we have just proved is that once we know the values of T on a basis of U, then we can find $T(u)$ for any $u \in U$.
Note: Theorem 3.1 is true even when U is not finite -dimensional.
The proof, in this case, is on the same lines as above.
Let us see how the idea of Theorem 3.1 helps us to prove the following useful result.

Theorem 3.2: Let V be a real vector space and $T: R \to V$ be a linear transformation. Then there exists $v \in V$ such that $T(\alpha) = \alpha v$, $\forall \alpha \in R$.

Proof: A basis for R is (1).
Let $T(1) = V \in V$, then, for any $\alpha \in R$, $T(\alpha) = \alpha T(1) = \alpha v$;
$T(\alpha)$ is a vector space of dimension one, whose basis is [T(1)].

Now try the following exercise, for which you will need Theorem 3.1.

E8) We define a linear operator $T: R^2 \to R^2: T(1,0) = (0,1)$ and $T(0,5) = (1,0)$.
What is i) $T(3,5)$and ii)$T(5,3)$?

Now we shall prove a very useful theorem about linear transformations, which is linked to Theorem 3.1

Theorem 3.3: Let $(e_1 \; \ldots., e_n)$ be a basis of *U* and let $v_1 \; \ldots., v_n$be any *n* vectors in *V*. Then there exists one and only one linear transformation $T: U \to V$ such that
$T(e_1) = v_1; \;\; i \; = \; 1, \ldots, n.$

Proof: Let $u \in U$. Then *u* can be uniquely written as $u = \alpha_1 e_1 \; + \; \ldots\ldots + T\alpha_n e_n$.
Define $T(u) \; = \; \alpha_1 v_1 \; + \; \ldots. + \alpha_n v_n$
T defines a mapping from *U* to *V* such that $T(e_1) = \; v_1$for all $i = 1, \ldots, n$
Let us now show that *T* is linear,
Let a, b be scalars and $u, u' \in U$. Then there exist scalars $\alpha_1, \ldots, \alpha_n, \beta_1, \ldots, \beta_n$ such that $u = \alpha_1 e_1 + \cdots + \alpha_n e_n$ and $u' = \beta_1 e_1 + \cdots + \beta_n e_n$
Then, $au \; + bu' = (a\alpha_1 \; + b\beta_1)e_1 \; + \cdots + (a\alpha_n \; + b\beta_n)e_n$
Hence, $T(au + bu') = (a\alpha_1 \; + b\beta_1)v1 \; + \cdots + (a\alpha_n \; + b\beta_n)v_n \; = a(\alpha_1 v_1 + \cdots + \alpha_n v_n) \; + \; b(\beta_1 v_1 + \cdots + \beta_n v_n) = aT(u) + bT(u')$
Therefore, *T* is a linear transformation with the property that $T(e_i) = v_i$ for all i Theorem 3.1 now implies that *T* is the only linear transformation with the above properties.
Let us now see how Theorem 3.3 can be used.

Example 8: $e_1 = (1,0,0)$, $e_2 = (0,1,0)$ and $e_1 = (0,0,1)$ form the standard basis of R^3. Let (1,2), (2,3) and (3,4) be three vectors in R^2. Obtain the linear transformation $T: R^3 \rightarrow R^2$such that $T(e_1)$ = (1,2), T(e_2) = (2,3) and T(e_3) = (3,4).

Solution: By Theorem 3.3, we know that $T: R^3 \rightarrow R^2$ such that T(e_1) = (1,2), T(e_2) = (2,3), and T(e_3) = (3,4). We want to know what $T(x)$ is, for any $x = (x_1, x_2, x_3) \in R^3$.
Now, $X = x_1e_1 + x_2e_2 + x_3e_3$
Hence, $T(X) = x_1T(e_1) + x_2T(e_2) + x_3T(e_3)$
$= x_1(1,2) + x_2(2,3) + x_3(3,4)$
$= (x_1{+}2x_2 + 3x_3, 2x_1 + \ 3x_2 + 4x_3)$
Therefore, $T(x_1, x_2, x_3) = (x_1{+}2x_2 + 3x_3, 2x_1 + \ 3x_2 + 4x_3)$ is the definition of the linear transformation T.

E9) Consider the complex field ℂ. It is a vector space over ℝ**,**
a) What is its dimension over ℝ? Give a basis of ℂover ℝ.
b) Let $\alpha, \beta \in R$. Give the linear transformation which maps the basis elements of ℂ obtained in (a), onto α and β, respectively.
Let us now look at some vector spaces that are related to a linear operator.

3.3.1 Spaces Associated with a Linear Transformation

In Unit 1, you found that given any function, there is a set associated with it, namely, it's range. We will now consider two sets which are associated with any linear transformation, T. These are the range and the kernel of T.

3.3.1.1 The Range Space and the Kernel

Let U and V be vector spaces over a field F. Let $T: U \rightarrow V$be a linear transformation. We shall define the range of T as well as the Kernel of T. At first, you will see them as sets.
We will prove that these sets are also vector spaces over F.

Definition 3.3.2: The **range** of T, denoted by $\boldsymbol{R(T)}$, is the set $\{T(x): x \in U\}$ such that the **kernel** (or null space) of T denoted by **Ker T,** is the set $\{x \in U: T(x) = 0\}$.
Note that $R(T) \subseteq V$and $Ker\ T \subseteq U$.
To clarify these concepts, consider the following examples:

Example 9: Let $I: U \rightarrow V$be the identity transformation (see Example1). Find $R(I)$ and $Ker\ I$.

Solution: $R(I) = \{I(v): v \in V\} = \{v: v \in V\} = V$.
Also, $Ker\ I = \{v \in V: I(v) = 0\} = \{v \in V: v = 0\} = \{0\}$

Example 10: Let $T: R^3 \to R$ be defined by $(x_1, x_2, x_3) = 3x_1 + x_2 + 2x_3$. Find $R(T)$ and $Ker\ T$.

Solution: $R(T) = \{x \in R: x_1, x_2, x_3 \in R \text{with } 3x_1 + x_2 + 2x_3 = x\}$
For example, $0 \in R(T)$, Since $0 = 3.0 + 0 + 2.0 = T(0,0,0)$
Also, $I \in R(T)$, since $I = 3 \cdot \frac{1}{3} + 0 + 2 \cdot 0 = T\left(\frac{1}{3}, 0,0\right)$, or $I = 3 \cdot 0 + 1 + 2 \cdot 0 = T(0,1,0)$ or $I = 3.0 + 1 + 2.0 = T(0, 1, 0)$; or $I = T(0, 0, ½)$ $or\ I = T\ (\frac{1}{6}, \frac{1}{2}, 0)$.
Now can you see that R(T) is the whole real line R?
This is because, for any $\alpha \in R, \alpha = \alpha \cdot 1 = \alpha T\left(\frac{\alpha}{3}, 0,0\right) = T\left(\frac{\alpha}{3}, 0,0\right) \in R(T)$

$$Ker\ T = \{(x_1, x_2, x_3) \in R^3: 3x_1 + x_2 + 2x_3 = 0\}$$

For example, $(0,0,0) \in \ker T$, but$(1, 0, 0) \notin Ker\ T$.
Therefore, $Ker\ T \neq R^3$. In fact, Ker T is the plane $3x_1 + x_2 + 2x_3 = 0$ in R^3.

Example 11: Let$T: R^3 \to R^3$ be defined by

$$T(x_1, x_2, x_3) = x_1 - x_2 + 2x_3, 2x_1 + x_2, -x_1 - 2x_2 + 2x_3$$

Find $R(T)$ and $Ker\ T$.

Solution: To find $R(T)$, we must find conditions on $y_1, y_2, y_3 \in R$so that $(y_1, y_2, y_3) \in R(T)$, i.e., we must find some $(x_1, x_2, x_3) \in R^3$ so that $(y_1, y_2, y_3) = T(x_1, x_2, x_3) = (x_1 - x + 2x_3, 2x_1 + x_2, x_1 - 2x_2 + 2x_3)$
This means

$$x_1 - x_2 + 2x_3 = y_1 \quad \ldots\ldots\ldots\ldots\ldots (1)$$
$$2x_1 + x_2 = y_2 \quad \ldots\ldots\ldots\ldots\ldots (2)$$
$$-x_1 - 2x_2 + 2x_3 = y_3 \quad \ldots\ldots\ldots\ldots\ldots (3)$$

Subtracting 2 times (1) from (2) and adding (1) and (3) to obtain

$$3x_2 - 4x_3 = y_2 - 2y_1 \quad \ldots\ldots\ldots\ldots\ldots\ldots (4)$$
$$-3x_2 + 4x_3 = y_1 + y_3 \quad \ldots\ldots\ldots\ldots\ldots\ldots (5)$$

Adding Equations (4) and (5) we get

$y_2 - 2y_1 + y_1 + y_3 = 0$, that is, $y_2 + y_3 = y_1$

Thus, $(y_1, y_2, y_3) \in R(T) \Rightarrow y_2 + y_3 = y_1$.
On the other hand, if $y_2 + y_3 = y_1$, we can choose

$$x_3 = 0\ ; \qquad x_2 = \frac{y_2 - 2y_1}{3}\ ; \qquad x_1 = y_1 + \frac{y_2 - 2y_1}{3}$$

Then, we see that $T(x_1, x_2, x_3) = (y_1, y_2, y_3)$
Thus, $y_2 + y_3 = y_1 \Rightarrow (y_1, y_2, y_3) \in R(T)$
Hence, $R(T) = \{(y_1, y_2, y_3) \in R^3: y_2 + y_3 = y_1\}$
Now $(x_1, x_2, x_3) \in Ker\ T$ if and only if the following equations are true:

$$x_1 - x_2 + 2x_3 = y_1$$

$$2x_1 + x_2 = y_2$$

gives

$$-x_1 - 2x_2 + 2x_3 = y_3 + 2x_3 = 0$$
$$2x_1 + x_2 = 0$$
$$-x_1 - 2x_2 + 2x_3 = 0$$

Of course, $x_1 = 0,\ x_2 = 0, x_3 = 0$ is a solution.

Are there other solutions?

To answer this, we proceed as in the first part of this example.
We see that $3x_2 + 4x_3 = 0 \Rightarrow\Rightarrow x_3 = \frac{3}{4}x_2$
Also, $2x_1 + x_2 = 0 \Rightarrow\Rightarrow x_1 = -\frac{1}{2}x_2$
Thus, we can give arbitrary values to x_2 and calculate $x_1$and x_3 in terms of x_2.
Therefore, $KerT = \{(-\frac{\alpha}{2}, \alpha, (\frac{3}{4})\alpha) : \alpha \in R\}$.
In this example, we see that finding $R(T)$ and $Ker\ T$ amounts to solving a system of equations. In subsequent unit, you will learn a systematic way of solving a system of linear equations by the use of matrices and determinants.
The following exercises will help you in getting used to $R(T)$ and $Ker\ T$.
E10) Let T be the zero-transformation given in Example 2. Find $KerT$ and $R(T)$. Does $I \in R(T)$?
E11) Find $R(T)$ and $KerT$ for each of the following operators:

a) $T: R^3 \to R^2: T(x, y, z) = (x, y)$
b) $T: R^3 \to R: T(x, y, z) = z$
c) $T: R^3 \to R^3: T(x_1, x_2, x_3) = (x_1 + x_2 + x_3, x_1 + x_2 + x_3, x_1 + x_2 + x_3))$.

(Note that the operators in (a) and (b) are projections onto the xy-plane and the z-axis, respectively).
Now that you are familiar with the sets $R(T)$ and $KerT$, we will prove that they are vector spaces.

Theorem 3.4: Let *U* and *V* be vector spaces over a field *F*. Let $T: U \to V$ be a linear transformation. Then Ker *T* is a subspace of *U* and *R*(T) is a subspace of *V*.

Proof: Let $x_1, x_2 \in KerT \subseteq U$ and $\alpha_1, \alpha_2 \in F$.
Now, by definition, $T(x_1) = T(x_2) = 0$
Therefore, $\alpha_1 T(x_1) + \alpha_2 T(x_2) = 0$
But $\alpha_1 T(x_1) + \alpha_2 T(x_2) = T(\alpha_1 x_1 + \alpha_2 x_2)$
Hence, $T(\alpha_1 x_1 + \alpha_2 x_2) = 0$
This means that $\alpha_1 x_1 + \alpha_2 x_2 \in KerT$.
Thus, by Theorem 3.2.3 of Unit 1, Ker*T* is a subspace of *U*.

Let $y_1, y_2 \in R(T) \subseteq V$, and $\alpha_1, \alpha_2 \in F$, then, by definition of R(T), there exist $x_1, x_2 \in U$ such that $T(x_1) = y_1$and $T(x_2) = y_2$

So, $\alpha_1 y_1 + \alpha_2 y_2 = \alpha_1 T(x_1) + \alpha_2 T(x_2)$

$= T(\alpha_1 x_1 + \alpha_2 x_2)$

Therefore, $\alpha_1 y_1 + \alpha_2 y_2 \in R(T)$, which proves that $R(T)$ is a subspace of V.

Now that we have proved that $R(T)$ and $Ker\ T$ are vector spaces, you know, from Unit 1, that they must have a dimension. We shall study these dimensions now.

3.3.1.2 Rank and Nullity

Consider any linear transformation, $T: U \to V$, assuming that $\dim U$ is finite. Then KerT, being a subspace of U, has finite dimension and $\dim(KerT) \leq \dim U$.

Also note that $R(T) = T(U)$, the image of *U* under *T*, a fact you will need to use in solving the following exercise.

E12) Let $\{e_1, \cdots, e_n\}$ be a basis of U. Show that $R(T)$ is generated by $\{T(e_1), \ldots, T(e_1)\}$.

From E12), it is clear that, if $\dim U = N$, then $\dim R(T) \leq n$.

Thus, $dim\ R(T)$ is finite, and the following definition is meaningful.

Definition 3.3.3: The **rank** of *T* is defined to be the **dimension** of *R*(*T*), the range space of *T*.

The **nullity** of *T* is defined to be the **dimension** of Ker *T*, the kernel (or the null space) of *T*.

Thus, $\boldsymbol{rank}\ (\boldsymbol{T}) = dim\ R(T)$ and $nullity\ (T) = dim\ Ker\ T$.

We have already seen that $rank\ (T) \leq dim\ U$and $nullity\ (T) \leq \dim U$.

Example 12: Let $T: U \to V$be the zero-transformation given in example 2. What are the rank and nullity of T?

Solution: In Exercise 11, you saw that R(T) = (0) and Ker *T* = *U*,

Therefore, rank (*T*) = 0 and nullity (*T*) = dim U.

Note that rank (*T*) + nullity (*T*) = dim U, in this case.

E13) If *T* is the identity operator on *V*, find rank (*T*) and nullity (*T*).

E14) Let D be the differentiation operator in E6). Give a basis for the range space of D and for Ker D. What are rank (D) and nullity (D)?

In the above example and exercises you will find that for $T: U \to V$, then rank (T) + nullity (T) = dim U.

In fact, this is the most important result about rank and nullity of a linear operator. We shall now state and prove this result.

Theorem 3.5: Let *U* and *V* be vector spaces over a field F and dim $U = n$ andlet$T: U \to V$be a linear operator. Then rank (*T*) + nullity (*T*) = n.

Proof: Let $nullity\,(T) = m$, that is, $dim\,Ker\,T = m$, Let $\{e_1, \cdots, e_m\}$be a basis of Ker T. We know that Ker T is a subspace of U, thus, by a theorem in Unit 1, we can extend this basis to obtain a basis$(e_1, \cdots, e_m, e_{m+1}, \cdots e_n)$of U.
We shall show that {T$(e_{m+1}), \dots \dots,$ T(e_n) }is a basis of R(*T*).
Then, our result will follow because dim R(*T*) will be $n - m = n$ – nullity (*T*).
Let us first prove that {T(e_{m+1}), ……, T(e_n)} spans, or generates, R(*T*).
Let $y \in R(T)$, then, by definition of $R(T)$, there exists $x \in U$ such that $T(x) = y$.
Let $= c_1e_1 + \cdots + c_me_m + c_{m+1}e_{m+1} + \cdots + c_ne_n$; $c_i \in F$ for all i.
Then, $y = T(x) = c_1T(e_1) + \cdots + c_mT(e_m) + c_{m+1}T(e_{m+1}) + \cdots + c_nT(e_n)$
Because $T(e_1) = \cdots = T(e_m) = 0$, since $T(e_1) \in KerT$ T for all$i = 1, \dots., m$.
Therefore, any$y \in R(T)$ is a linear combination of {T(e_{m+1}), ……, T(e_n)}.
Hence, R(T) is spanned by {T(e_{m+1}), ……, T(e_n)}.
It remains to show that the set {T(e_{m+1}), ……, T(e_n)} is linearly independent.
For this, suppose there exist $a_{m+1}, \cdots, a_n$ with $a_{m+1}T(e_{m+1}) + \cdots + a_nT(e_n) = 0$.
Then, $T(a_{m+1}e_{m+1} + \cdots + a_ne_n) = 0$
Hence, $(a_{m+1}e_{m+1} + \cdots + a_ne_n) \in KerT$, which is generated by $\{e_1, \cdots, e_m\}$.
Therefore, there exist $e_1, \cdots, e_m \in F$ such that

$$a_{m+1}e_{m+1} + \cdots + a_ne_n = a_1e_1 + \cdots + a_me_m$$
$$\Rightarrow (-a_1)e_1 + \cdots + (-a_m)e_m + a_{m+1}e_{m+1} + \cdots + a_ne_n = 0$$

Since $\{e_1, \cdots, e_m\}$is a basis of *U*, it follows that this set is linearly independent.
Hence, $-a_1 = 0, \cdots\cdots, -a_m = 0, a_{m+1} = 0, \cdots\cdots, a_n = 0$.
In particular, $a_{m+1} = \cdots\cdots = a_n = 0$, which we wanted to prove.
Therefore, dim $R(T) = n - m = n$ – nullity (*T*), that is, rank (*T*) + nullity (*T*) = n.
Let us see how this theorem can be useful.

Example 13: Let $T: R^3 \to R$ be the map given by $L(x, y, z) = x + y + z$. What is nullity (L)?

Solution: In this case it is easier to obtain $R(L)$, rather than Ker L.
Since $L(1,0,0) = 1 \neq 0, R(L) \neq \{0\}$, and hence dimR(L) ≠ {0}.
Also, $R(L)$ is a subspace of R.

Thus, $dim\ R(L) \leq\ dim\ R = 1$.
Therefore, the only possibility for $dim\ R(L)$ is dim $R(L) = 1$.
By Theorem 5, $dim\ Ker\ L + dim\ R(L) = 3$.
Hence, $dim\ Ker\ L = 3 - 1 = 2$. That is, nullity(L) = 2.
E15) Give the rank and nullity of each of the linear transformations in E11.
E16) Let U and V be real vector spaces and $T: U \rightarrow V$ be a linear transformation, where $dim\ U = 1$. Show that $R(T)$ is either a point or a line.
Before ending this section, we will prove a result that links the rank (or nullity) of the composite of two linear operators with the rank (or nullity) of each of them.

Theorem 3.6: Let V be a vector space over a field F. Let S and T be linear operators from V to V. Then

a) $rank\ (ST)\ min\ \left(rank\ (S), rank\ (T)\right)$
b) $nullity\ (ST) \geq max\ \left(nullity\ (S), nullity\ (T)\right)$

Proof: We shall prove (a)
Note that $(ST)(v) = S(T(v))$ for any $v \in V$
Now, for any $y \in R(ST), \exists\ v \in V$ such that,

$$Y = (ST)(v) = S(T(v)) \qquad \ldots\ldots\ldots\ldots\ldots\ldots (1)$$

Now, (1) $\Longrightarrow y \in R(S)$
Therefore, $R(ST) \subseteq R(S)$
This implies that $rank(ST) \leq rank(S)$.
Again, (1) $\Longrightarrow y \in S(R(T))$, since $T(v) \in R(T)$.
$\therefore R(ST) \subseteq S(R(T))$, so that $dim\ R(ST) \leq dim\ S(R(T)) \leq dim\ R(T)$(since $dim\ L(U) \leq U$ for any linear operator (0).
Therefore, $rank\ (ST) \leq rank\ (T)$.
Thus, $rank\ (ST) \leq min\ (rank\ (S), rank\ (R))$.
The proof of this theorem will be complete, once you solve the following exercise.
E17) Prove (b) of Theorem 6 using the Rank Nullity Theorem
Next to be discussed are some linear operators that have special properties.

3.3.2 Some Types of Linear Transformations

Let us recall, from basic mathematics, that there can be different types of functions, some of which are one-one, onto or invertible. We can also define such types of linear transformations as follows:

Definition 3.3.4: Let $T: U \longrightarrow V$ be a linear transformation.
a) T is called **one-one** (or injective) if, for $u_1, u_2 \in U$ with $u_1 \notin u_2$, we have $T(u_1) \notin T(u_2)$. If T is injective, we also say T is 1 – 1.
Note that T is **1 – 1** if$T(u_1) = T(u_2) \Longrightarrow u_1 = u_2$.
b) T is called onto (or **surjective)** if, for each $v \in V,\ \exists\ u \in U$such that $T(u) = v$, that is $R(T) = V$.
Can you think of examples of such functions? The identity operator is both one-one and onto. Why is this so? Well, $I: V \to V$ is an operator such that, if $v_1, v_2 \in V$with $v_1 \neq v_2$ then $I(v_1) \neq I(v_2)$. Also, $R(I) = V$, so that I is onto.
E18) Show that the zero operator $0: R \to R$ is not one – one

Theorem 3.7: $T: U \to V$is one-one if and only if $KerT = (0)$.

Proof: First assume T is one – one.
Let $u \in Ker\ T$, then $T(u) = 0 = T(0)$.
This means that $u = 0$. thus, $Ker\ T = (0)$.
Conversely, let $Ker\,T = (0)$. Suppose $u_1, u_2 \in U$ with $T(u_1) = T(u_2) \Longrightarrow T(u_1 - u_2) = 0$,
$\Rightarrow u_1 - u_2 \in KerT \Rightarrow u_1 - u_2 = 0 \Rightarrow u_1 = u_2$. Therefore, T is 1 – 1.

Suppose now that T is a one – one and onto linear transformation from a vector space U to a vector space V. Then, from Unit 1 (Theorem 4), we know that T^{-1} exists. But is T^{-1}linear? The answer to this question is 'yes', as is shown in the following theorem.

Theorem 3.8: Let U and V be vector spaces over a field F. Let $T: U \to V$ be a none-one and onto linear transformation. Then, $T^{-1}: U \to V$ is a linear transformation
In fact, $T^{-1}$is also 1 – 1 and onto.

Proof: Let $y_1, y_2 \in V$ and $\alpha_1, \alpha_2 \in F$. Suppose$T^{-1}(y_1) = x_1$and $T^{-1}(y_2) = x_2$, then, by definition, $y_1 = T(x_1)$ and $y_2 = T(x_2)$.
Now, $\alpha_1 y_1 + \alpha_2 y_2 = \alpha_1 T(x_1) + \alpha_2 T(x_2) = T(\alpha_1 x_1 + \alpha_2 x_2)$
Hence, $T^{-1}(\alpha_1 y_1 + \alpha_2 y_2) = \alpha_1 x_1 + \alpha_2 x_2 = \alpha_1 T^{-1}(y_1) + \alpha_2 T^{-1}(y_2)$.
This shows that $T^{-1}$is a linear transformation.
We will now show that T^{-1} is 1 -1, for this, suppose $y_1, y_2 \in V$such that $T^{-1}(y_1) = T^{-1}(y_2)$ Let $x_1 = T^{-1}(y_1)$ and $x_2 = T^{-1}(y_2)$.
Then $T(x_1) = y_1$and $T(x_2) = y_2$.
We know that $x_1 = x_2$. Therefore, $T(x_1) = T(x_2)$, that is, $y_1 = y_2$.
Thus, we have shown that $T^{-1}(y_1) = T^{-1}(y_2) \Rightarrow y_1 = y_2$, proving that $T^{-1}$is 1 – 1.
$T^{-1}$is also surjective because, for any $u \in U,\ \exists\ T(u) = v \in V$such that $T^{-1}(v) = u$.

Theorem 8 says that a one-one and onto linear transformation is **invertible**, and the inverse is also a one-one and onto linear transformation.

This theorem immediately leads us to the following definition.

3.4 Theorems of Vector Spaces

3.4.1 Isomorphism Theorems of Vector Spaces

Definition 3.4.1: Let U and V be vector spaces over a field F, and let $T: U \to V$ be a one-one and onto linear transformation. The T is called an **Isomorphism** between U and V. In this case we say that U and V are **isomorphic vector spaces.** This is denoted by $U \approx V$.

An obvious example of an isomorphism is the identity operator. Can you think of any other? The following exercise may help.
E19) Let $T: R^3 \to R^3 | \, T(x, y, z) = (x + y, y, z)$. Is$T$an isomorphism? Why? Define T^{-1}, if it exists.
E20) Let $T: R^3 \to R^2 | \, T(x, y, z) = (x + y, y + z)$. Is T an isomorphism?
In all these exercises and examples, have you noticed that if T is an isomorphism between U and V then $T^{-1}$is an isomorphism between V and U?
Using these properties of an isomorphism we can get some useful results, like the following:

Theorem 3.9: Let $T: U \to V$be an isomorphism. Suppose $\{e_1, \cdots, e_n\}$is a basis of U. then $\{T(e_1), \cdots, T(e_n)\}$ is a basis of V.

Proof: First we show that the set $\{T(e_1), \cdots, T(e_n)\}$spans $\{T(e_1), \cdots, T(e_n)\}$. Since T is onto, $R(T) = V$. Thus, from E12) you know that $\{T(e_1), \cdots, T(e_n)\}$spans V.
Let us now show that $\{T(e_1), \cdots, T(e_n)\}$is linearly independent.
Suppose there exist scalars $c_1, \ldots, c_n$, such that $c_1T(e_1) + \cdots + c_nT(e_n) = 0$ ....(1)
We must show that $c_1 = \cdots = c_n = 0$
Now, (1) implies that $T(c_1e_1 + \cdots + c_ne_n) = 0$
Since T is one-one and $T(0) = 0$, we conclude that $c_1e_1 + \cdots + c_ne_n = 0$.
But $\{e_1, \cdots, e_n\}$ is linearly independent
Therefore, $c_1 = \cdots = c_n = 0$
Thus, we have shown that$\{T(e_1), \cdots, T(e_n)\}$ is a basis of V.

Remark: The argument showing the linear independence of $\{T(e_1), \cdots, T(e_n)\}$in the above theorem can be used to prove that any one-

one linear transformation $T: U \to V$maps any linearly independent subset of Uonto a linearly independent subset of V(see E22).
We now give an important result equating 'isomorphism' with '1-1' and with 'onto' in the finite-dimensional case.

Theorem 3.10: Let $T: U \to V$be a linear transformation where U, V are of the same finite dimension. Then the following statements are equivalent.

a) T is 1 – 1
b) T is onto.
c) T is an isomorphism.

Proof: To prove the result we shall prove $a) \Longrightarrow b) \Longrightarrow c) \Longrightarrow a)$.
Let $dimU = dimV = n$.
Now a) implies that $Ker\ T\ = 0$ (from Theorem 3.7),
Hence, nullity$(T) = (0)$
Therefore, by Theorem 3.5, $rank(T) = n$, that is, $dimR(T) = n = dimV$.
But $R(T)$ is a subspace of V, thus, by the remark following Theorem 3.2, we get R $(T) = V$, that is, T is onto, i.e., b) is true, so $a) \Rightarrow b)$.
Similarly, if b) holds then $rank\ (T) = n$, and hence, nullity $(T) = 0$.
Consequently, $Ker\ T = \{0\}$, and T is one-one.
Hence, T is one-one and onto, that is$T$ is an isomorphism.
Therefore, b) implies c).
That a) follows from 9c) is immediate from the definition of an isomorphism.
Hence, our result is proved.
Caution: Theorem 10 is true for **finite-dimensional spaces** U and V, of the same **dimension.** It is not true, otherwise.
Consider the following counter-example.

Example 14: (To show that the spaces have to be finite-dimensional):
Let V be the real vector space of all polynomials. Let $D: V \to V$ be defined by

$$D(a_0 + a_1x + \cdots + a_rx^{r-1}) = a_1 + 2a_2x + \cdots + ra_rx^{r-1}.$$

Then show that D is onto but not 1-1.

Solution: Note that V has infinite dimension, a basis being $\{1, x, x, ...\}$.
D is onto because any element of V is of the form $a_0 + a_1x + \cdots + a_nx_n = D$
D is not 1–1 because, for example, $1 \neq 0$ but $D(1) = D(0) = 0$.
The following exercise shows that the statement of Theorem 10 is false if dim U ≠ dim V.
E21) Define a linear operator $T: R^3 \to R^2$ such that T is onto but T is not 1 - 1.
Note that $dimR^3 \neq dimR^2$

Let us use Theorems 3.9 and 3.10 to prove our next result.

Theorem 3.11: Let $T: V \to V$ be a linear transformation and let $\{e_1, \dots, e_n\}$ be a basis of V. Then T is one-one and onto if and only if $\{T(e_1), \cdots, T(e_n)\}$is linearly independent.

]Proof: Suppose T is one-one and onto. Then T is an isomorphism
Hence, by Theorem 3.9, $\{T(e_1), \cdots, T(e_n)\}$is a basis.
Therefore, $\{T(e_1), \cdots, T(e_n)\}$ is linearly independent.
Conversely, suppose $\{T(e_1), \cdots, T(e_n)\}$is linearly independent. Since $\{e_1, \dots, e_n\}$ is a basis of V, $dimV = n$. Therefore, any linearly independent subset of *n* vectors is a basis of V. Hence, $\{T(e_1), \cdots, T(e_n)\}$ is a basis of$V$.
Then, any element v of V is of the form $v = \sum_{i=1}^{n} c_1 T(e_1) = T \sum_{i=1}^{n} c_1 e_1$ where $c_1, \dots, c_n$ are scalars.
Thus, T is onto, and we can use Theorem 3.10 to say that T is an isomorphism.
Here are some exercises now.
E22) Let $T: U \to V$ be a one-one linear transformation and let $\{u_1, \dots, u_k\}$ be a linearly independent subset of U.

a) show that the set $\{T(e_1), \cdots, T(e_n)\}$ is linearly independent.

b) Is it true that every linear transformation maps every linearly independent set of vectors into a linearly independent set?

c) Show that every linear transformation maps a linearly dependent set of vectors onto a linearly dependent set of vectors.

E23) Let $T: R^3 \to R^3$be defined by $T(x_1, x_2, x_3) = (x_1 + x_3, x_2 + x_3, x_1 + x_2)$. Is T invertible? If yes, find a rule for T^{-1} like the one which defines T.

We have seen, in Theorem 3.9, that if $T: U \to V$ is an isomorphism, then T maps a basis of U onto a basis of V. Therefore, $dimU = dimV$. In other words, if U and V are isomorphic then $dimU = dim\ V$.
The natural question arises whether the converse is also true. That is, if $dimU = dimV$, both being finite, can we say that U and V are isomorphic? The following theorem shows that this is indeed the case.

Theorem 3.12: Let U and V be finite-dimensional vector spaces over F. the U and V are isomorphic if and only if $dim\ U = dim\ V$.

Proof: We have already seen that if U and $V$are isomorphic then $dimU = dimV$. Conversely, suppose $dim\ U = dim\ V = n$.
We shall show that U and $V$are isomorphic. Let $\{e_1, \dots, e_n\}$ be a basis of U and $\{f_1, \dots, f_n\}$ be a basis of V.
By Theorem 3.3, there exists a linear transformation $T: U \to V$ such that $T(f_1) = f_i; i = 1, \cdots, n$

We shall show that T is 1-1.
Let $u = c_1e_1 + \cdots + c_ne_n$ be such that $T(u\ ) = 0$
Then, $0 = T(u\ ) = c_1T(e_1) + \cdots + c_nT(e_n) = c_1f_1 + \cdots + c_nf_n$
Since $\{f_1, \dots, f_n\}$ is a basis of V, we conclude that $c_1 = \cdots = c_n = 0$.
Hence, $u = 0$.
Thus, Ker T = (0) and, by Theorem 3.7, T is one – one.
Therefore, by Theorem 3.10, T is an isomorphism, and $U = V$.
An immediate consequence of this theorem follows:

Corollary: Let V be a real (or complex) vector space of dimension n. Then V is isomorphic to R (or C) respectively.

Proof: Since $dimR_n = n = dim_R V$, we get $V \approx R^n$.
Similarly, if $dim_c V = n$, then $V \approx c_n$.

Remark: Let V be a vector space over F and let$B = \{e_1, \dots, e_n\}$be a basis of V. Each $v \in V$ can be uniquely expressed as $v = \sum_{i=1}^{n} \alpha_i\, e_i$. Recall that, $\alpha_1, \dots, \alpha_n$are called the coordinates of v with respect to B.
Define $\theta: V \to F^n | \theta(v) = (\alpha_1, \dots, \alpha_n)$. Then θ is an isomorphism from V to F^n. This is because θ is 1 -1, since the coordinates of v with respect to B are uniquely determined.
Thus, $V \approx F^n$.
We end this section with an exercise.
E24) Let $T: U \to V$be a one-one linear mapping. Show that T is onto if and only if $dimU = dimV$. (Of course, you must assume that U and V are finite dimensional spaces).
Now let us look at isomorphism between quotient spaces.

3.4.2 Homomorphism Theorems of Vector Spaces

Linear transformations are also called **vector space homomorphisms.** There is a basic theorem which uses the properties of homomorphisms to show the isomorphism of certain quotient spaces. It is simple to prove, but is very important because it is always being used to prove more advanced theorems on vector spaces. (In the Abstract Algebra course, we will prove this theorem in the setting of groups and rings).

Theorem 3.13: Let V and W be vector spaces over a field F and $T: V \to W$ be a linear transformation. Then $V/\boldsymbol{KerT} \approx \boldsymbol{R(T)}$ **.**

Proof: You know that Ker T is a subspace of V, so that V/Ker T is a well-defined vector space over F. Also, $R(T) = \{T(v): V \in V\}$.
To proof the theorem, let us define $V/Ker\ T \to R(T)$ by $\theta(v + Ker\ T) = T(v)$
Firstly, we must show that θ is a well-defined function, that is, if

$v + KerT = v\ + KerT$, then $(v + Ker\ T) = \theta(v + Ker\ T)$,
that is, $T(v) = T(v)$.
Now, $v + Ker\ T = v' + KerT \Longrightarrow (v - v') \in KerT$

$T(v - v') = 0 \Longrightarrow T(v) = T(v')$ and hence, θ is well defined.

Next, we check that θ is a linear transformation.
For this, let $a, b \in F$ and $v, v' \in V$, then,

$$\theta\{a(v + Ker\ T) + b(v' + KerT)\} = \theta(av + bv' + KerT)$$
$$= T(av + bv')$$
$$= aT(v) + bT(v'),\text{ since } T \text{ is linear.}$$
$$= a(v + KerT) + b(v + KerT).$$

Thus, θ is a linear transformation.
We shall end the proof by showing that θ is an isomorphism.
θ is 1-1, because

$$\theta(v + Ker\ T) = \theta \Longrightarrow T(v) = 0 \Longrightarrow v \in KerT \Longrightarrow v + Ker\ T$$
$$= 0\ (in\ V/Ker\ T).$$

Thus, $Ker\ \theta = \{0\}$.
θ is onto (because any element of $R(T)$ is $T(v) = (v) = (v + Ker\ T)$)
So, we have proved that θ is an isomorphism.
This proves that $V/Ker\ T = R(T)$.
Let us consider an immediate useful application of Theorem 3.13.

Example 14: Let V be a finite-dimensional space and let S and T be linear transformations from V to V. Show that $Rank(ST) = rank(T) - dim(R(T) \cap Ker\ S)$.

Solution: We have $V \xrightarrow{T} V \xrightarrow{S} V$, ST is the composition of the operators S and T which you have studied in elementary mathematics courses. Now, we apply Theorem 3.13 to the homomorphism $\theta: T(v) \to ST(v): \theta\big(T(v)\big) = (ST)(v)$
Now, $Ker\theta = \{x \in T(V) | S(x) = 0\} = Ker\ S \cap T(V) = Ker\ S \cap R(T)$.
Also $R(\theta) = ST(V)$, since any element of $ST(V)$ is $(ST)(v) = \theta(T(v))$.
Thus, $\frac{T(V)}{Ker\ S \cap T(V)} \approx ST(V)$
Therefore, $dim \frac{T(V)}{Ker\ S \cap T(V)} \approx dim\ ST(V)$
That is, $dim\ T(V) - dim(Ker\ S \cap T(V)) = dim\ ST(V)$, which is what we had to show.
E25) Using Example 14 and the Rank Nullity Theorem, show that

$$nullity(ST) = nullity(T) + dim(R(T) \cap Ker\ S)$$

Now let us see another application of Theorem 3.13.
Example 15: Show that $R^3/R \approx R^2$.

Solution: Note that we can consider **R** as a subspace of R^3 for the following reason:

Any element $a \in R$ is equated with the element $(\alpha, 0,0)$ of R^3.
Now, we define a function $f: R^3 \to R^2 | f(\alpha, \beta, \gamma) = (\beta, \gamma)$, then f is a linear transformation and $Ker\ f = \{(\alpha, 0,0) | \alpha \in R\} \approx R$.
Also, f is onto, since any element (α, β) of R^2 is $f(0, \alpha, \beta)$.
Thus, by Theorem 13, $R^3 \ / \ R \approx R^2$.
Note: In general, for any $n \geq m$, $R^n \ / \ R^m \approx R^{n-m}$.
Similarly, $C^{n-m} \approx C^n \ / \ C^m$ for $n \geq m$.
The next result is a corollary to the Fundamental Theorem of Homomorphism. But, before studying it, read unit 1 for definition of the sum of spaces.

Corollary 1: Let A and B be subspaces of a vector space V. then $A + B/B \approx A/A \cap B$.

Proof: we define a linear function $T: A \to \frac{A+B}{B}$by $T(a) = a + B$
$T$is well defined because$a + B$ is an element of $\frac{a+B}{B}$(since $a = a + 0 \in A + B$).
T is a linear transformation because, for $\alpha_1, \alpha_2 \in F$ and $a_1, a_2 \in A$, we have

$$T(\alpha_1 a_1 + \alpha_2 a_2) = \alpha_1 a_1 + \alpha_2 a_2 + B = \alpha_1(a_1 + B) + \alpha_2(a_2 + B)$$
$$= \alpha_1 T(a_1) + \alpha_2 T(a_2)$$

Now we will show that T is surjective. Any element of$\frac{A+B}{B}$ is of the form $a + b + B$, where $a \in A$ and $b \in B$.
Now, $a + b + B = a + B + b + B = a + B + B$, since $b \in B$
$= a + B$, since B is the zero element of$\frac{A+B}{B}$
$= T(a)$ which proves that T is surjective.

$$\therefore RT = \frac{A+B}{B}$$

We will now prove that $Ker\ T = A \cap B$:
$a \in Ker\ T$, then $a \in A$ and$T(a) = 0$. This means that $a + B = B$, the zero element of $\frac{A+B}{B}$.
Hence, $b \in B$(by E23), therefore, $a \in A \cap B$.
Thus, $Ker\ T \subseteq A \cap B$. On the other hand, $a \in A \cap B \Longrightarrow a \in A$ and $a \in B \Longrightarrow a \in A$ and$a + B = B \Longrightarrow a \in A$and $T(a) = T(0) = 0$
$\Longrightarrow a \in KerT$.
This proves that $A \cap B = Ker\ T$.
Now using Theorem 3.13, we get

$$A/Ker\ T \approx R(T)$$

That is, $A/(A \cap B) \approx (A + B)/B$.
E26) Using the corollary above, show that $A \oplus B/B \approx A$, ($\oplus$ denotes the direct sum of defined in earlier).
There is yet another interesting corollary to the Fundamental Theorem of Homomorphism.

Corollary 2: Let W be a subspace of a vector space V. Then, for any subspace U of V containing W, $\frac{V/W}{U/W} \approx V / U$.

Proof: To start with let us define a function $T: V/W \to V/U: T(V + U) = V + U$.
Now try E27.
E27 a) Check that T is well defined
b) Prove that T is a linear transformation
c) What are the spaces Ker T and R(T)?

SELF-ASSESSMENT EXERCISE(S)

E1) For any $\alpha_1, \alpha_2 \in F$ and $u_1, u_2 \in U$, we know that $\alpha_1 u_1 \in U$ and $\alpha_2 u_2 \in U$.
Therefore, by LT1.

$$T(\alpha_1 u_1 + \alpha_2 u_2) = T(\alpha_1 u_1) + T(\alpha_2 u_2)$$
$$= \alpha_1 T(u_1) + \alpha_2 T(u_1)$$

(by LT2)
Thus, LT3 is true.

E2) By LT2, $T(0, u) = 0 \cdot T(u)$ for any $u \in U$, thus, $T(0) = 0$.
Similarly, for any $u \in U$, $T(-u) = T((-1)u) = (-1)T(u) = -T(u)$.

E3) $T(x, y) = (-x, y), (x, y) \in R$. (See the geometric view in Fig.4) T is a linear operator.
This can be proved the same way as we did in Example4.

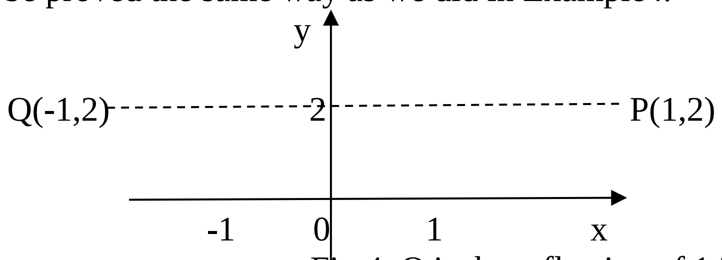

Fig.4: Q is the reflection of 1 in the y-axis

E4) $T((x_1, x_2, x_3) + (y_1, y_2, y_3)) = T(x_1 + y_1, x_2 + y_2, x_3 + y_3)$
$$= a_1(x_1 + y_1) + a_2(x_2 + y_2) + a_3(x_3 + y_3)$$
$$= (a_1 x_1 + a_2 x_2 + a_3 x_3) + (a_1 y_1 + a_2 y_2 + a_3 y_3)$$
$$= T(x_1, x_2, x_3) + T(y_1, y_2, y_3)$$

Also, for any $\alpha \in R$,

$$T(\alpha(x_1, x_2, x_3)) = a_1 \alpha x_1 + a_2 \alpha x_2 + a_3 \alpha x_3$$
$$= \alpha(a_1 x_1 + a_2 x_2 + a_3 x_3) = \alpha T(x_1, x_2, x_3)$$

Thus, LT1 and LT2 hold for T.
E5) Required to show that the map $T: R_3 \to R_3$ defined by $T(a_1, a_2, a_3) = (x_1 + x_2 - x_3, 2x_1 - x_2, x_2 + 2x_3)$ is a linear operator.

We will check if LT1 and LT2 hold firstly

$$T((x_1, x_2, x_3) + (y_1, y_2, y_3)) = T(x_1 + y_1, x_2 + y_2, x_3 + y_3)$$
$$= (x_1 + y_1 + x_2 + y_2 - x_3 - y_3, 2x_1 + 2y_1 - x_2 - y_2, x_2 + y_2 + 2x_3 + 2y_3)$$
$$= (x_1 + x_2 - x_3, 2x_1 - x_2, x_2 + 2x_3) + (y_1 + y_2 - y_3, 2y_1 - y_2 + 2y_3)$$
$$= T(x_1, x_2, x_3) + T(y_1, y_2, y_3)$$

Which shows that LT1 holds.

Also, for any $\alpha \in R$,

$$T(\alpha(x_1, x_2, x_3)) = T(\alpha x_1, \alpha x_2, \alpha x_3)$$
$$= (\alpha x_1 + \alpha x_2 - \alpha x_3, 2\alpha x_1 - \alpha x_2, \alpha x_2 + 2\alpha x_3)$$
$$= \alpha(x_1 + x_2 - x_3, 2x_1 - x_2, x_2 + 2x_3) = \alpha T(x_1, x_2, x_3)$$

Shows that LT2 holds

E6) Let us consider the real vector space $P_n$of all polynomials of degree less or equal to n. We want to show that$D(\alpha f + \beta g) = \alpha D(f) + \beta D(g)$, for any $\alpha, \beta \in R$ and $f, g \in P_n$

Let $f \in P_n$be given by $f(x) = a_0 + a_1 x + a_2 x^2 \dots \dots + a_n x^n$, and $g(x) = b_0 + b_1 x + b_2 x^2 \dots \dots + b_n x^n a_i, b_i \in R \; \forall i$.

Then $(\alpha f + \beta g)(x) = (\alpha a_0 + \beta b_0) + (\alpha a_1 + \beta b_1)x + \cdots + (\alpha a_n + \beta b_n)$

$$\therefore [D(\alpha f + \beta g)](x)$$
$$= (\alpha a_0 + \beta b_1) + 2(\alpha a_2 + \beta b_2)x + \cdots + n(\alpha a_n + \beta b_n)x^{n-1}$$
$$= \alpha(a_1 + 2a_2 x + \cdots \dots + n a_n x^{n-1}) + \beta(b_1 + 2b_2 x + \cdots \dots + n b_n x^{n-1})$$

Thus, $D(\alpha f + \beta g) = \alpha Df + \beta Dg$, showing that D is a linear map.

E7) No, because, if T exists, then

$$T(2u_1 + u_2) = 2T(u_1) + T(u_2)$$

But $2u_1 + u_2 = u_3$

On the other hand, $2T(u_1) + T(u_2) = 2v_1 + v_2 = (2,0) + (0,1)$
$= (2,0) \neq v_3$

Thus, LT3 is violated; hence, no such T exists.

E8) Note that $\{(1,0), (0,5)\}$ is a basis for R^2

Now, $(3,5) = 3(1,0) + (0,5)$

$\therefore T(3,5) = 3T(1,0) + T(0,5) = 3(0,1) + (1,0) = (1,3)$ …………… (i)

Similarly, $(5,3) = 5(1,0) + \frac{3}{5}(0,5)$

$\therefore T(5,3) = 5T(1,0) + \frac{3}{5}T(0,5) = 5(0,1) + \frac{3}{5}(1,0) = \left(\frac{3}{5}, 5\right)$ ………… (ii)

From (i) and (ii), we see that $T(5,3) \neq T(3,5)$

E9) a) $dim_R C = 2$, a basis being $\{1, i\}, i = \sqrt{-1}$

b) Let $T: C \to R$ be such that $T(1) = \alpha$ and $T(i) = \beta$

Then, for any element$x + iy \in C (x, y \in R)$

We have $T(x + iy) = xT(1) + yT(i) = x\alpha + y\beta$

Thus, T is defined by $T(x+iy) = x\alpha + y\beta \ \forall \ x+iy \in C$

E10)$T: U \to V$ such that $T(U) = 0 \ \forall \ u \in U$

$$\therefore KerT = \{u \in U | T(U) = 0\} = U$$
$$R(T) = \{T(u) | u \in U\} = \{0\}$$
$$\therefore 1 \notin R(T)$$

E11) a) $R(T) = \{T(x,y,z) | (x,y,z) \in R^3\} = \{(x,y) | (x,y,z) \in R^3\} = R^2$

$$KerT = \{(x,y,z) | T(x,y,z) = 0\} = \{(x,y,z) | (x,y) = (0{,}0)\}$$
$$= \{(0{,}0,z) | z \in R\}$$

Therefore, Ker T is the z-axis.

b) $R(T) = \{z | (x,y,z) \in R^3\} = R$

$KerT = \{(x,y,0) | x,y \in R\} = xy$ −plane in R^3.

Because, for any $x \in R$, $(x,x,x) = T(x,0{,}0)$

Therefore, RT is generated by $\{(1{,}1{,}1)\}$

$KerT = \{(x_1,x_2,x_3) | x_1 + x_2 + x_3 = 0\}$, which is the plane $x_1 + x_2 + x_3 = 0$ in R^3.

E12) Any element of $R(T)$ is of the form $T(u), u \in U$. Since $\{e_1, e_2, e_3\}$ generates U, $\exists$ scalars $\alpha_1, \alpha_2, \cdots, \alpha_n$such that$u = \alpha_1 e_1 + \alpha_2 e_2 + \cdots + \alpha_n e_n$

Then $T(u) = \alpha_1 T(e_1) + \alpha_2 T(e_2) + \cdots + \alpha_n T(e_n)$, that is, $T(u)$ is in the linear span of $\{T(e_1), T(e_2), \cdots, T(e_n)\}$.

$\therefore \{T(e_1), T(e_2), \cdots, T(e_n)\}$ generates $R(T)$.

E13)$T: V \to V: T(v_0 = v)$.

Since $R(T) = V$ and $KerT = (0)$, we see that $rank(T) = dimV$, $nullity(T) = 0$

E14) $R(D) = \{a_1 + 2a_2 x + \cdots + na_n x^{n-1} | a_1, \cdots, a_n \in R\}$

$R(D) \subseteq P^{n-1}$.

But any element $b_0 + b_1 x + \cdots + b_{n-1} x^{n-1}$ in P^{n-1} is

$$D\left(b_0 x_1 + \frac{b_1}{2} x_2 + \cdots + \frac{b_{n-1}}{n} x_n\right) \in R(D)$$

Therefore, $R(D) = P_{n-1}$

Hence, a basis for $R(D)$ is $\{1, x, \cdots, x^{n-1}\}$ and $rank(D) = n$

$$KerD = \{a_0 + a_1 x + \cdots + a_n x^n | a_1 + 2a_2 x + \cdots + na_n x^{n-1} = 0, a_i \in R \ \forall \ i\}$$
$$= \{a_0 + a_1 x + \cdots + a_n x^n | a_1 = 0, a_2 = 0, \cdots, a_n = 0, a_i \in R \ \forall \ i\}$$
$$= \{a_0 | a_0 = R\} = R$$

Therefore, a basis for Ker D is {1}.

$\Longrightarrow nullity(D) = 1$.

E15) a) We have shown that $R(T) = R^2$.

$\therefore Rank(T) = 2$ Therefore, $nullity(T) - dim\ R = 3 – 2 = 1$.

b) $Rank(T) = 1, nullity(T) = 2$

c) $R(T)$ is generated by $\{(1{,}1{,}1)\}$,

$\therefore Rank(T) = 1$ and $Nullity(T) = 2$.

E16) Now, $rank(T) + nullity(T) = dimU = 1$

Also, $rank(T) \geq 0$,
Therefore, the only values $rank(T)$ can take are 0 and 1.
If $rank(T) = 0$, then, $dimR(T) = 0$.
Thus, $R(T) = \{0\}$, that is, $R(T)$ is a point.
If $rank(T) = 1$, then $dimR(T) = 1$, That is, $R(T)$ is a vector space over R generated by a single element, v, say. Then $R(T)$ is the line $R_v = \{\alpha v | \alpha \in R\}$
E17) By Theorem 5, $nullity(ST) = dimV - rank(ST)$.
By (a) of Theorem 6, we know that
$-\ rank(ST) - rank(S)$ and $-rank(ST) - rank(T)$.
$\therefore\ nullity(ST) \geq dimV - rank(S)$ and $nullity(ST) \geq dimV - rank(T)$.
Thus, $nullity(ST) \geq nullity(S)$ and $nullity(ST) \geq nullity(T)$.
That is, $nullity(ST) \geq max\ \{nullity(S), nullity(T)\}$.
E18) Since $1 \notin 2$, but $0(1) = 0(2) = 0$, we find that 0 is not 1– 1.
E19) Firstly note that T is a linear transformation.
Secondly, T is 1-1 because $T(x, y, z) = T(x', y', z')(x, y, z) = (x', y', z')$
Thirdly, T is onto because any $(x, y, z) \in R^3$ can be written as $T(x - y, y, z)$
therefore, T is an isomorphism and$T^{-1}: R^3 \rightarrow R^3$ exists and is defined by $T^{-1}(x, y, z) = (x - y, y, z)$.
E20) T is not an isomorphism because T is not 1-1, since $(1, -1, 1) \in Ker\ T$.
E21) The linear operator in E11) (a) suffices.
E22) a) Let $\alpha_1, \cdots, \alpha_k \in F$ such that $\alpha_1 T(u_1) + \cdots + \alpha_k T(u_k) = 0$

$$\Longrightarrow T(\alpha_1 u_1 + \cdots + \alpha_k u_k) = 0 = T(0)$$

$$\Rightarrow \alpha_1 u_1 + \cdots + \alpha_k u_k = 0, \text{ since } T \text{ is 1-1}$$

$$\Longrightarrow \alpha_1 = 0, \cdots, \alpha_k = 0 \quad , \quad \text{since } \{u_1, \cdots, u_k\} \text{ is linearly independent}$$

$\therefore \{T(u_1), \cdots, T(u_k)\}$ is linearly independent.
b) No. For example, the zero operator maps every linearly independent set to $\{0\}$, which is not linearly independent.
c) Let $T: U \rightarrow V$ be a linear operator, and $\{u_1, \cdots, u_n\}$ be a linearly dependent set of vectors in *U*. We have to show that $\{T(u_1), \cdots, T(u_k)\}$ is linearly dependent.
Since $\{u_1, \cdots, u_n\}$is linearly dependent, there exists scalars $\alpha_1, \cdots, \alpha_n$, not all zero, such that $\alpha_1 u_1 + \cdots + \alpha_n u_n = 0$.
Then $\alpha_1 T(u_1) + \cdots + \alpha_k T(u_k) = T(0) = 0$, so that $\{T(u_1), \cdots, T(u_k)\}$ is linearly dependent.
E23) T is a linear transformation now, if $(x, y, z) \in Ker\ T$, then $T(x, y, z) = (0,0,0)$.

$$\therefore x + y = 0 = y + z = x + z \Longrightarrow x = 0 = y = z$$

$$\Rightarrow Ker\ T = \{(0,0,0)\}$$

$\Longrightarrow$ T is 1 – 1, therefore, by Theorem 10, T is invertible.
To define $T^{-1}: R^3 \rightarrow R^3$ suppose $T^{-1}(x, y, z) = (a, b, c)$.

Then $T(a,b,c) = (x,y,z)$

$$\Rightarrow (a+b, b+c, a+c) = (x,y,z)$$
$$\Rightarrow a+b = x, b+c = y, a+c = z$$
$$\Rightarrow a = \frac{x+z-y}{2};\ b = \frac{x+y-z}{2};\ c = \frac{y+z-x}{2}$$

$\Rightarrow T^{-1}\left(\frac{x+z-y}{2}, \frac{x+y-z}{2}, \frac{y+z-x}{2}\right)$ for any $(x,y,z) \in R^3$.

E24) $T: U \to V$ is 1 -1. Suppose T is onto. Then T is an isomorphism and $dim\, U = dim\, V$, by Theorem 3.12. Conversely suppose $dim\, U = dim\, V$. Then T is onto by theorem 10.

E25) The Rank Nullity Theorem and Example 14 give

$$dim\, V - nullity(ST) = dim\, V - nullity(T) - dim(R(T) \cap Ker\, S)$$
$$\Rightarrow nullity(ST) = nullity(T) + dim\,(R(T) \cap Ker\, S)$$

E26) In the case of the direct sum $A \oplus B$, we have $A \cap B = \{0\}$

Therefore, $\frac{A\oplus B}{B} \approx A$

E27) a) $v + W = v' + W \Rightarrow v - v' \in W \subseteq U \Rightarrow v - v' \in U \Rightarrow v + U = v' + U$

$$\Rightarrow T(v+W) = T(v'+W)$$

$\therefore T$ is well defined.

b) For any $v + W, v + W$ in V/W and scalar a, b, we have

$$T(a(v+W) + b(v'+W)) = T(av + bv' + W) = av + bv' + U$$
$$= a(v+U) + b(v'+U)$$
$$= aT(v+W) + bT(v'+W)$$

∴T is a linear operator.

c) $Ker\, T = \{v + W | v + U = U\}$, since U is the "zero" for V/U.

$$= \{v + W | v \in U\} = U/W$$

$R(T) = \{v + U | v \in V\} = V/U$.

3.6 Summary

In this unit, we have covered the following points that:

(1) A linear transformation from a vector space U over F to a vector space V over F is a function $T: U \to V$ such that,

LT1: $T(u_1 + u_2) = T(u_1) + T(u_2)\ \forall\, u_1, u_2 \in U$, and

LT2: $(\alpha u) = \alpha T(u)$, for $\alpha \in F$ and $u \in U$.

These conditions are equivalent to the single condition

LT3: $T(\alpha u + \beta u) = \alpha T(u_1) + \beta T(u_2)$ for $\alpha, \beta \in F$ and $u_1, u_2 \in U$.

(2) Given a linear transformation, $T: U \to V$.

i. The Kernel of T is the vector space $\{u \in U | T(u) = 0\}$, denoted by $Ker\, T$

ii. The range of T is the vector space $\{T(u) | u \in U\}$, denoted by $r(T)$

iii. The rank of $T = dim_1 R(T)$

iv. The nullity of $T = dim_1 Ker\ T$.

(3) Let U and V be finite-dimensional vector spaces over F and $T: U \to V$ be a linear transformation. Then, $rank(T) + nullity\ (T) = dim\ U$.

(4) Let $T: U \to V$ be a linear transformation then T is one-one if

$$T(u_1) = T(u_2) \Longrightarrow u_1 = u_2\ \forall\ u_1, u_2 \in U$$

i. T is onto if, for any $v \in V\ \exists\ u \in U$ such that $T(u) = v$.

ii. T is an isomorphism (or is invertible) if it is one-one and onto, and then U and V are called isomorphic spaces. This is denoted by $U \approx V$.

(5) $T: U \to V$ is

i. one-one if and only if $Ker\ T = (0)$

ii. onto if and only if $R(T) = V$

(6) Let $U$and V be **finite-**dimensional vector spaces with the **same dimension.**

Then $T: U \to V$ is 1-1 iff T is onto iff T is an isomorphism

(7) Two finite dimensional vector spaces U and V are isomorphic if and only if $dim\ U = dim\ V$.

(8) Let V and W be vector spaces over a field F, and $T: V \to W$ be a linear transformation.

Then $V/Ker\ T \approx R(T)$.

Let *U* and *V* be vector spaces over a field *F*. A **linear transformation** (or *linear operator*) from *U* to *V* is a function $T: U \to V$ which satisfies some conditions, LT1 to LT6.

We have defined the range of T as well as the Kernel of T and have been seen as sets.

We have also proved that these sets are vector spaces over F.

In general, any function $\theta: R_n \to R_m (n > m)$, which is defined by dropping any $(n - m)$ coordinate, is a projection map.

3.7 References/Further Readings

Robert A. Beezer (2014). A First Course in Linear Algebra. Congruent Press Gig Harbor, Washington, USA 3(40).

Arbind K Lal Sukant Pati (2018). Linear Algebra through Matrices.

Peter J. Cameron (2008). Notes on Linear Algebra.

UNIT 4 LINEAR TRANSFORMATION II

Unit Structure

4.1 Introduction

In the last unit you were introduced to linear transformations and their properties. We shall now show that the set of all linear transformations form a vector space U to a vector space V, forms a vector space itself, and its dimension is dim U (dim V). In particular, we define and discuss the dual space of a vector space.

In Unit 1, we defined the composition of two functions. Over here we shall discuss the composition of two linear transformations and show that it is again a linear operator. Note that we use the terms 'linear transformation' interchangeably.

Finally, we shall study polynomials with coefficients from a field F, in a linear operator $T: V \to V$. It shall be seen that every such T satisfies a polynomial equation $g(x) = 0$, that is, if we substitute T for x in $g(x)$, we get the zero transformation. The minimal polynomial of an operator shall then be defined and some of its properties discussed. These ideas will crop up again in Module 4 Unit 2.

It is advisable that you revise Units 1 to 3 before going further.

4.2 Learning Outcomes

By the end of this unit, you will be able to:

- Prove and use the fact that $L(U,V)$ is a vector space of dimension $(dimU)(dimV)$;

- Use dual bases, whenever convenient;
- Obtain the composition of two linear operators, whenever possible;
- Obtain the minimal polynomial of a linear transformation $T: V \to V$ in some simple cases;
- Obtain the inverse of an isomorphism $T: V \to V$if its minimal polynomial is known.

4.3 Main Content

4.3.1 The Vector Space $L(U, V)$

By now you must be quite familiar with linear operators, as well as vector spaces. In this section we consider the set of all linear operators from one vector space to another, and show that it forms a vector space.
Let U, V be vector spaces over a field F. Consider the set of all linear transformations from U to V. We denote this set by $L(U, V)$.

We will now define addition and scalar multiplication in $L(U, V)$ so that $L(U, V)$ becomes a vector space.

Suppose $S, T \in L(U, V)$ (that is, S and T are linear operators from U to V). Also, define
$(S+T): U \to V$ by $(S+T)(u) = S(u) + T(u) \; \forall \; u \in U$.
Now, for $\alpha_1, \alpha_2 \in F$ and $u_1, u_2 \in U$, we have

$$\begin{aligned}(S+T)(a_1u_1 + a_2u_2) &= S(\alpha_1u_1 + \alpha_2u_2) + T(\alpha_1u_1 + \alpha_2u_2) \\ &= \alpha_1S(u_1) + \alpha_2S(u_2) + \alpha_1T(u_1) + \alpha_2T(u_2) \\ &= \alpha_1(S(u_1) + T(u_1)) + \alpha_2(S(u_2) + T(u_2)) \\ &= \alpha_1(S + T)(u_1) + \alpha_2(S+T)(u_2)\end{aligned}$$

Hence, $(S+T) \in L(U, V)$.
Next, suppose $S \in L(U, V)$ and $\alpha \in F$. We define $\alpha S: U \to V$ as follows: $(\alpha S)(u) = \alpha S(u) \; \forall \; u \in U$. Is αS a linear operator?
To answer this take $\beta_1, \beta_2 \in F$ and $u_1, u_2 \in U$.
Then, $(\alpha S)(\beta_1u_1 + \beta_2u_2) = \alpha S(\beta_1u_1 + \beta_2u_2)$

$$\begin{aligned}&= \alpha(\beta_1S(u_1) + \beta_2S(u_2)) \\ &= \beta_1(\alpha S)(u_1) + \beta_2(\alpha S)(u_2)\end{aligned}$$

Hence, $\alpha S \in L(U, V)$.
So, we have successfully defined addition and scalar multiplication on $L(U, V)$.
E1) Show that the set $L(U, V)$is a vector space over F with respect to the operations of addition and multiplication by scalars defined above.
(Hint: The zero vector in this space is the zero transformation).

Notation: For any vector space *V* we denote $L(U,V)$by $A(V)$.
Let *U* and *V* be vector spaces over *F* of dimensions *m* and *n*, respectively. We have already observed that $L(U,V)$is a vector space over *F*. therefore, it must have a dimension.
We now show that the dimension of $L(U,V)$ is mn.

Theorem 4.1: Let *U*, *V* be vector spaces over a field *F* of dimensions *m* and *n*, respectively, then $L(U,V)$is a vector space of dimension mn.

Proof: Let $\{e_1, \dots . e_m\}$ be a basis of *U* and $\{f_1, \dots . f_n\}$ be a basis of *V*.
By Theorem 3 of Unit 3, there exists a unique linear transformation $E_{11} \in L(U,V)$, such that

$$E_{11}(e_1) = f_1, E_{11}(e_2) = 0, \cdots \cdots, E_{11}(e_m) = 0$$

Similarly, $E_{12} \in L(U,V)$ such that

$$E_{12}(e_1) = 0, E_{12}(e_2) = f_1, \cdots \cdots, E_{12}(e_m) = 0$$

In general, there exist $E_{ij} \in L(U,V)$ for $i = 1, \dots ., n$; $j = 1, \dots ., m$, such that

$$E_{ij}(e_j) = f_i \text{ and } E_{ij}(e_k) = 0 \text{ for } j \neq k.$$

To get used to these E_{ij}, try the following exercise before continuing the proof.
E2) Clearly define E_{2m}, $E_{32}$and E_{mn}.
Now, let us go on with the proof of Theorem 1.
If $u = c_1 e_1 + \dots . + c_m e_m$, where $c_i F \ \forall \ i$, then $E_{ij}(u) = c_j f_i$.
We complete the proof by showing that $\{E_{ij}: i = 1, \dots, m\}$ is a basis of $L(U,V)$.
Let us first show that set is linearly independent over *F*. for this, suppose

$$\sum_{i=1}^{n} \sum_{j=1}^{m} c_{ij} E_{ij} = 0 \qquad \dots\dots\dots\dots\dots\text{(i)}$$

where $c_{ij} \in F$. we must show that $c_{ij} = 0$ for all i, j.
(i) implies that

$$\sum_{i=1}^{n} \sum_{j=1}^{m} c_{ij} E_{ij}(e_k) = 0 \ \forall \ k = 1, \cdots, m \qquad \dots\dots\dots\dots\dots\text{(ii)}$$

Thus, by definition of E_{ij}'s, we obtain

$$\sum_{j=1}^{n} c_{ik} f_i = 0 \qquad \dots\dots\dots\dots\dots\dots\dots\text{(iii)}$$

But, $\{f_1, \dots . f_n\}$is a basis for *V* thus, $c_{ik} = 0$ for all $i = 1, \dots ., n$.
But this is true for all $k = 1, \dots ., m$.
Hence, we conclude that $c_{ij} = 0 \ \forall \ i, j$.
Therefore, the set of E_{ij}'s is linearly independent.

Next, we show that the set $\{E_{ij} | i = 1, \dots, n, j = 1, \dots, m\}$ spans $L(U,V)$.
Suppose $T \in L(U,V)$.
Now, for each *j* such that $1 \leq j \leq m$, $T(e_j) \in V$.
Since $\{f_1, \dots . f_n\}$ is a basis of V, there exist scalars $c_{ij}, \cdots, c_{in}$ such that

$$T(e_j) = \sum_{i=1}^{n} c_{ij} f_i \qquad \dots\dots\dots\dots\dots\dots \text{(i)}$$

we shall prove that

$$T = \sum_{i=1}^{n} \sum_{j=1}^{m} c_{ij} E_{ij} \qquad \ldots\ldots\ldots\ldots\ldots\ldots \text{(ii)}$$

By Theorem 3.1 of Unit 3 it is enough to show that, for each k with $1 \leq k \leq m$,

$$T(e_k) = \sum_{i=1}^{n} \sum_{j=1}^{m} c_{ij} E_{ij}(e_k) \qquad \ldots\ldots\ldots\ldots\ldots\ldots \text{(iii)}$$

Now, $\sum_{i=1}^{n} \sum_{j=1}^{m} c_{ij} E_{ij}(e_k) = \sum_{i=1}^{n} c_{ik} f_i = T(e_k)$ by (ii), this implies (iii).
Thus, we have proved that the set of mn elements $\{E_{ij} | i = 1, \ldots, n, j = 1, \ldots, m\}$ is a basis for $L(U, V)$.
Let us see some ways of using this theorem.

Example 1: Show that $L(R^2, R) \approx R^2$ is a plane.

Solution: $L(R^2, R)$ is a real vector space of dimension 2 x 1 = 2.
Thus, by 10 of Unit 3 Theorem; $L(R^2, R) \approx R^2$, the real plane.

Example 2: Let U, V be vector spaces of dimensions m and n, respectively. Suppose W is a subspace of V of dimension $p (\leq n)$.
Let $X = \{T \in L(U, V) : T(u) \in W \ \forall \ u \in U\}$. Is X a subspace of $L(U, V)$? If yes, find its dimension.

Solution: $X = \{T \in L(U, V) | T(U) \subseteq W\} = L(U, W)$. Thus, X is also a vector space.
Since it is a subset of $L(U, V)$, it is a subspace of $L(U, V)$. By Theorem 1, $dimX = mp$.
E3) What can be a basis for $L(R^2, R)$and for $L(R, R^2)$?
Notice that both these spaces have the same dimension over R.
After having looked at $L(U, V)$, we now discuss this vector space for the particular case when $V = F$.

4.3.2 The Dual Space

The vector space $L(U, V)$, discussed in Unit 2, has a particular name when $V = F$.

Definition 4.3.1: Let U be a vector space over F. Then the space $L(U, F)$is called the dual space of U^*, and is denoted by U.
In this section we shall study some basic properties of U*. The elements of U have a specific name, which we now give.

Definition 4.3.2: A linear transformation $T: U \rightarrow F$ is called a **linear functional**. Thus, a linear functional on U is a function $T: U \rightarrow F$ such that
$T(\alpha_1 u_1 + \alpha_2 u_2) = \alpha_1 T(u_1) + \alpha_2 T(u_2)$, for $\alpha_1, \alpha_2 \in F$ and $\alpha_1, \alpha_2 \in U$.

For example, the map, $f: R^3 \to R | f(x_1, x_2, x_3) = a_1x_1 + a_2x_2 + a_3x_3$, where $a_1, a_2, a_3 \in R$ are fixed, is a linear functional on R^3. You have already seen this in Unit 3 (E4).
We now come to a very important aspect of the dual space.
We know that the space V*, of linear functional on V, is a vector space.
Also, if dim V = m, then dim V* = m, by Theorem 1 (Remember, dim *F* = 1).
Hence, we see that dim V = dim V*.
From Theorem 12 of unit 3, it follows that the vector spaces V and, V* are isomorphic.
We now construct a special basis for V*.
Let $\{e_1, \ldots . e_n\}$ be a basis for *V*, by Theorem3 of Unit 3, for each $i = 1, \ldots, m$, there exists a unique linear functional f_i on *V* such that
$f_j(e_j) = \delta_{ij} = \begin{cases} 1, i = j \\ 0, i \neq j \end{cases}$ where, δ_{ij} is the Kronecker delta function.
We shall prove that the linear functional $f_1, \ldots . f_m$, constructed above, form a basis of V*.
Since dim V = dim V* = m, it is enough to show that the set $\{f_1, \ldots . f_m\}$ is linearly independent. For this we suppose $e_1, \ldots . e_m \in F$ such that $c_1f_1 + \cdots + c_mf_m = 0$.
We must show that $c_i = 0$, for all i.
Now, $\sum_{j=1}^{n} c_jf_j = 0$
$\Rightarrow \sum_{j=1}^{n} \left(c_jf_j(e_i)\right) = 0$ for each i

$$\Rightarrow \sum_{j=1}^{n} c_j \left(f_j(e_i)\right) = 0 \; \forall \; i$$

$$\Rightarrow \sum_{j=1}^{n} c_j\delta_{ij} = 0 \; \forall \; i \quad \Rightarrow c_i = 0 \; \forall \; i$$

Thus, the set $\{f_1, \ldots . f_m\}$is a set of *m* linearly independent elements of a vector space V* of dimension m, Thus, from Unit 3 (Theorem 3.5, Corrolary 1), it forms a basis of V*.

Definition 4.3.3: The basis $\{f_1, \ldots . f_m\}$of V* is called the dual basis of the basis $\{e_1, \ldots . e_m\}$ of *V*.
We now come to the result that shows the convenience of using a dual basis.

Theorem 4.2: Let *V* be a vector space over *F* of dimension *n*, $\{e_1, \ldots . e_n\}$ be a basis of *V* and $\{f_1, \ldots . f_n\}$ be the dual basis of $\{e_1, \ldots . e_n\}$.Then, for each $f \in V*$, $f = \sum_{i=1}^{n} f(e_j)f_i$ and, for each $v \in V$,$v = \sum_{i=1}^{n} f_i(v)e_i$
Proof: Since $\{f_1, \ldots . f_n\}$ is a basis of V*, for $f \in V*$ there exist scalars$c_1, \ldots . c_n$ such that

$$f = \sum_{i=1}^{n} c_i f_i$$

Therefore, $f(e_j) = \sum_{i=1}^{n} c_i f_i(e_j)$

$$= \sum_{j=1}^{n} c_j \delta_{ij}$$

$$= c_j$$

This implies that $c_i = f(e_i) \ \forall \ i = 1, \dots, n,$ therefore, $f = \sum f_i$.
Similarly, for $v \in \sum V$, there exist scalars $a_1, \dots . a_n$ such that

$$v = \sum_{i=1}^{n} a_i e_i$$

Hence, $f_j(v) = \sum_{i=1}^{n} a_i f_i(e_j)$

$$= \sum_{i=1}^{n} a_j \delta_{ij}$$

$$= a_j$$

and we obtain

$$v = \sum_{i=1}^{n} f_i (v) e_i$$

Let us see an example of how this theorem works.
Example 3: Consider the basis $e_1 = (1,0,-1)$, $e_2 = (1,1,0)$ of C^3 over C. Find the dual basis of $\{e_1, e_2, e_3\}$.
Solution: Any element of C^3 is $v = (z_1, z_2, z_3)$, $z_i \in C$. Since $\{e_1, e_2, e_3\}$is a basis, we have $\alpha_1, \alpha_2, \alpha_3 \in C$. Since

$$V = \{z_1, z_2, z_3\} = \alpha_1 e_1 + \alpha_2 e_2 + \alpha_3 e_3$$
$$= (\alpha_1 + \alpha_2 + \alpha_3, \alpha_2 + \alpha_3, -\alpha_1 + \alpha_2)$$

Thus, $\alpha_1 + \alpha_2 + \alpha_3 = z_1$

$$\alpha_2 + \alpha_3 = z_2$$
$$-\alpha_1 + \alpha_2 = z_3$$

These equations can be solved to get

$$\alpha_1 = z_1 - z_2$$
$$\alpha_2 = z_1 - z_2 + z_3$$
$$\alpha_3 = 2z_2 - z_1 - z_3$$

Now, by Theorem 2,$v = f_1(v)e_1 + f_2(v)e_2 + f_3(v)e_3$, where $\{f_1, f_2, f_3\}$ is the dual basis.
Also, $v = \alpha_1 e_1 + \alpha_2 e_2 + \alpha_3 e_3$.
Hence, $f_1(v) = \alpha_1, f_2(v) = \alpha_2, f_3(v) = \alpha_3 \ \forall \ v \in C^3$.
Thus, the dual basis of $\{e_1, e_2, e_3\}$ is $\{f_1, f_2, f_3\}$, where f_1, f_2, f_3 will be defined as follows:

$$f_1(z_1, z_2, z_3) = \alpha_1 = z_1 - z_2$$
$$f_1(z_1, z_2, z_3) = \alpha_2 = z_1 - z_2 + z_3$$
$$f_1(z_1, z_2, z_3) = \alpha_3 = 2z_2 - z_1 - z_3$$

E5) What is the dual basis for the basis $\{1, x, x^2\}$ of the space $P_2 = \{a_0 + a_i x + a_2 x^2 | a_i \in R\}$?

4.3.2.1 The Dual of the Dual Space

Let *V* be an *n*-dimensional vector space. It has already been shown that V and V* are isomorphic because dim V = dim V*.
The **dual of V*** is called the **second dual of V** denoted by **V****.
Let us show that $V \approx V **$.
Now any element of *V*** is a linear transformation from *V** to *F*.
Also, for any $v \in V$ and $f \in V *$, $f(v) \in F$.
So, we define a mapping $\varphi: V \to V ** : v \to \varphi v$, where $(\varphi v)(f) = f(v) \forall f \in V *$ and $v \in V$. (Over here we will use $\varphi(v)$ and φv interchangeably).
Note that, for any $v \in V, \varphi v$ is a well-defined mapping from $V * \to F$.
We have to check that it is a linear mapping.
Now, for $c_1, c_2 \in F$ and $f_1, f_2 \in V *$.

$$\begin{aligned}(\varphi v)(c_1 f_1 + c_2 f_2) &= (c_1 f_1 + c_2 f_2)(v)\\ &= c_1 f_1(v) + c_2 f_2(v)\\ &= c_1(\varphi v)(f_1) + c_2(\varphi v)(f_2)\end{aligned}$$

Therefore, $\varphi v \in L(V *, F) = V **, \forall\, v$
Furthermore, the map $\theta: V \to V **$ is linear.
This can be seen as follows:
For $c_1, c_2 \in F$ and $v_1, v_2 \in V$.

$$\begin{aligned}\theta(c_1 v_1 + c_2 v_2)(f) &= f(c_1 v_1 + c_2 v_2)\\ &= c_1 f(v_1) + c_2 f(v_2)\\ &= c_1(\theta v_1)(f) + c_2 f(\theta v_2)(f)\\ &= (c_1 \theta v_1 + c_2 \theta v_2)(f)\end{aligned}$$

This is true $\forall\ f \in V *$.
Thus, $\theta(c_1 v_1 + c_2 v_2) = c_1 \theta(v_1) + c_2 \theta(v_2)$.
Now that we have shown that θ is linear, we want to show that it is actually an isomorphism. We will show that θ is 1-1.
By Theorem 3.7 of Unit 3, it suffices to show that $\theta(v) = 0$ implies $v = 0$.
Let $\{f_1, \cdots, f_n\}$be the dual basis of a basis $\{e_1, \cdots, e_n\}$ of V.
By Theorem 2, we have $v = \sum_{i=1}^{n} f_i\,(v) e_i$
Now, $\theta(v) = 0 \Longrightarrow (\theta v)(f_i) = 0\ \forall\ i = 0,1, \cdots, n$
Hence, it follows that θ is 1-1, thus, θ is an isomorphism (Unit 3, Theorem 3.10).
What we have just proved is the following theorem.
Theorem 4.3: The map $\theta: V \to V **$, defined by $(\varphi v)(f) = f(v)\ \forall\ v \in V$ and$f \in V *$, is an isomorphism.
We now give an important corollary to this theorem.

Corollary: Let ψ be a linear functional on V^* (i.e., $\psi \in V^{**}$). Then there exists a unique $v \in V$ such that $\psi(f) = f(v)$ for all $f \in V^*$.

Proof: By Theorem 3, since $\theta$is an isomorphism, it is onto and 1-1, thus, there exists a unique $v \in V$ such that $\theta(v) = \psi$. This by definition, implies that $\theta(v)(f) = (v)(f) = f(v)$ for all $f \in V^*$.

Now, use the second dual try to prove the following exercise.
E6) Show that each basis of $V*$ is the dual of some basis of V.
In the following section we look at the composition of linear operators, and the vector space$A(v)$, where V is a vector space over F.

4.3.3 Composition of Linear Transformations

Do you remember the definition f the composition of functions, which you studied in Unit 1? Let us now consider the particular case of the composition of two linear transformations. Suppose $T: U \to V \to W$, defined by $S_0T(u) = S(T(u))\ \forall\ u \in U$.

This is diagrammatically represented in Fig. 1.

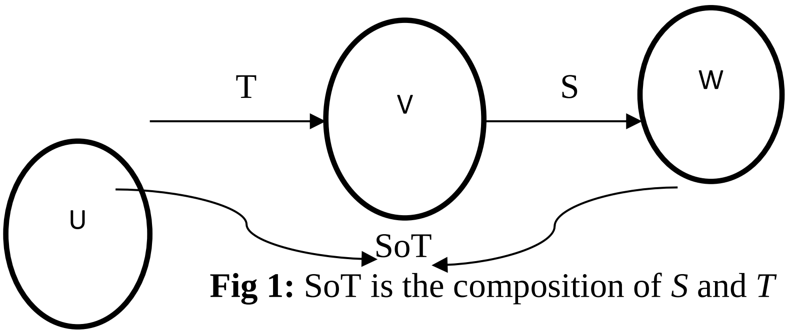

Fig 1: SoT is the composition of S and T

The first question which comes to our mind is whether $S \circ T$ is linear. The affirmative answer is given by the following result.

Theorem 4.4: Let U, V, W be vector spaces over F. Suppose $S \in L(V, W)$ and $T \in L(U, V)$. Then $S \circ T \in L(U, W)$.

Proof: All we need to prove is the linearity of the map $S \circ T$. Let $\alpha_1, \alpha_2 \in F$ and $u_1, u_2 \in U$. Then

$$S \circ T(\alpha_1 u_1 + \alpha_2 u_2) = S\big(T(\alpha_1 u_1 + \alpha_2 u_2)\big)$$
$$= S\big(\alpha_1 \mathrm{T}(\mathrm{u}_1) + \alpha_2 T(u_2)\big), \text{ since } T \text{ is linear}$$

$= \alpha_1 S(T(u_1)) + \alpha_2 S(T(u_2))$, since S is linear

$= \alpha_1 S \circ T(u_1) + \alpha_2 S \circ T(u_2)$

This shows that $S \circ T \in L(U, W)$.

Try the following exercises now

E7) Let I be the identity operator on V. Show that $S \circ I = I \circ S = S$ for all $S \in A(V)$.

E8) Prove that $S \circ 0 = 0 \circ S = 0$ for all $S \in A(V)$, where 0 is the null operator.

Let's now make an observation.

Remark: Let $S: V \to V$ be an invertible linear transformation (ref. unit 3), that is an isomorphism. Then, by Unit 3, Theorem 3.8, $S^{-1} \in L(V, V) = A(V)$.

Since $S^{-1} \circ S(v) = v$ and $S \circ S^{-1}(v) = v$ for all $v \in V$.

$S \circ S^{-1} = S^{-1} \circ S = I$, where I, denotes the identity transformation on V.

This remark leads us to the following interesting result.

Theorem 4.5: Let V be a vector space over a field F. A linear transformation $S \in A(V)$ is an isomorphism if an only if there exists $T \in A(V)$ such that $S \circ T = I = T \circ S$.

Proof: Let us first assume that S is an isomorphism. Then, the remark above tells us that there exists$S^{-1} \in A(V)$ such that $S \circ S^{-1} = I = S^{-1} \circ S$.

Thus, we have $T(= S^{-1})$ such that $S \circ T = T \circ S = I$.

Conversely, suppose T exists in $A(V)$, such that $S \circ T = I = T \circ S$. Show that S is 1-1 and onto.

We first show that S is 1 -1, that is, Ker S = {0}.

Now, $x \in Ker\ S \Longrightarrow S(x) = 0 \Longrightarrow T \circ S(x) = T\{0\} = 0 \Longrightarrow I(x) = 0 \Longrightarrow x = 0$.

Thus, Ker S = {0}.

Next, we show that S is onto, that is, for any $v \in V, \exists\ u \in V$ such that $S(u) = v$. Now, for any $v \in V$, $v = I(v) = ST(v) = S(T(v)) = S(u)$, where $u = T(v) \in V$. thus, S is onto.

Hence, S is 1-1 and onto, that is, S is an isomorphism.

Use Theorem 4.5 to solve the following exercise.

E9) Let $S(x_1, x_2) = (x_1, x_2)$ and $T(x_1, x_2) = (-x_1, x_2)$.Find $S \circ T$ and $T \circ S$. Is S(orT) invertible?

Now, let us look at some examples involving the composite of linear operators.

Example 4: Let $T: R^2 \rightarrow R^2$ and $S: R^3 \rightarrow R^2$ be defined by $T(x_1, x_2) = (x_1, x_2, x_1 + x_2)$ and $S(x_1, x_2, x_3) = (x_1, x_2)$. Find $S \circ T$and $T \circ S$.

Solution: First, note that $T \in L(R^2, R^3)$ and $S \in L(R^3, R^2)$. Therefore,$S \circ T$and$T \circ S$ are both well defined linear operators.
Now, $S \circ T(x_1, x_2, x_3) = S(T(x_1, x_2)) = S(x_1, x_2, x_1 + x_2) = (x_1, x_2)$.
Hence $S \circ T$= the identity transformation of $R^2 = I_{R^2}$.
Now, $T \circ S(x_1, x_2, x_3) = T(S(x_1, x_2, x_3)) = T(x_1, x_2) = (x_1, x_2, x_1 + x_2)$.
In this case $S \circ T \in A(R^2)$, while $T \circ S \in A(R^3)$.
Clearly, $S \circ T \neq T \circ S$.
Also, note that $S \circ T = I$, but $T \circ S \neq I$.

Remark: Even if $S \circ T$ and $T \circ S$ both being to $A(V)$, $S \circ T$ may not be equal to $T \circ S$. Such an example is given below.

Example 5: Let $S, T \in A(R^2)$, be defined by $T(x_1, x_2) = (x_1 - x_2, x_1 - x_2)$ and
$S(x_1, x_2) = (0, x_2)$. Show that $S \circ T \neq T \circ S$.

Solution: You can check that $S \circ T(x_1, x_2) = (0, x_1 - x_2)$ and $T \circ S(x_1, x_2) = (x_1 - x_2)$.
Thus, there exists $(x_1, x_2) \in R^2$ such that $S \circ T(x_1, x_2) \neq T \circ S(x_1, x_2)$ (For instance, $S \circ T(1,1) \neq T \circ S(1,1)$).
That is, $S \circ T \neq T \circ S$.

Note: Before checking whether $S \circ T$ is a well-defined linear operator, you must be sure that both S and T are well defined linear operators.
Now try to solve the following exercise.

E10) Let $T(x_1, x_2) = (0, x_1, x_2)$ and $S(x_1, x_2, x_3) = (x_1, x_2, x_2 + x_3)$. Find $S \circ T$ and $T \circ S$. When is $S \circ T = T \circ S$?

E11) Let $T(x_1, x_2) = (2x_1, x_1 + x_2)$for $(x_1, x_2) \in R^2$ and
$S(x_1, x_2, x_3) = (x_1 + 2x_2, 3x_1 - x_2, x_3)$ for $(x_1, x_2, x_3) \in R^3$.
Are $S \circ T$ and $T \circ S$ defined ? If yes, find them.

E12) Let U, V, W, Z be vector spaces over F. Suppose $T \in L(U, V), S \in L(V, W)$ and
$R \in L(W, Z)$. Show that $(R \circ S) \circ T = R \circ (S \circ T)$.

E13) Let $S, T \in A(V)$ and S be invertible.
Show that $rank(ST) = rank(TS) = rank(T)(ST)$.

So far, we have discussed the composition of linear transformation and seen that if $S, T \in A(V)$, then $S \circ TA(V)$, where V is a vector space of dimension n.

Thus, we have introduced another binary operation (Unit 1) in $A(V)$, namely, the ***composition of operators***, denoted by **o**.
Remember, we already have the binary operations given in the previous unit.
In the following theorem, we state some simple properties that involve all these operations.

Theorem 4.6: Let $R, S, T \in A(V)$ and let $\alpha \in F$. Then
a) $R \circ (S + T) = R \circ S + R \circ T$ and $(S + T) \circ R = S \circ R + T \circ R$.
b) $\alpha(S \circ T) = \alpha S \circ T = S \circ \alpha T$.

Proof:
a) For any $v \in V$,

$$R \circ (S + T) = R((S + T)(v))$$
$$= R(S(v) + T(v)) = (R \circ S)(v) + (R \circ T)(v)$$
$$= (R \circ S + R \circ T)(v)$$

Hence, $R \circ (S + T) = R \circ S + R \circ T$.
Similarly, we can prove that $(S + T) \circ R = S \circ R + T \circ R$
b) For any $v \in V$,

$$\alpha(S \circ T)(v) = \alpha\left(S\big(T(v)\big)\right)$$
$$= (\alpha S)\big(T(v)\big) = (\alpha S \circ T)(v)$$

Therefore, $\alpha(S \circ T) = \alpha S \circ T$.
Similarly, we can show that $\alpha(S \circ T) = S \circ \alpha T$.

Notification: Subsequently, we shall be writing ST in place of $S \circ T$.
Thus, $ST(u) = S(T(u)) = (S \circ T)u$.
Also, if $T \in A(V)$, then $T^0 = I, T^1 = T, T^2 = T \circ T$ and, in general,

$$T^n = T^{n-1} \circ T = T \circ T^{n-1}$$

The properties of $A(V)$, stated in theorems 4.1 and 4.6 are very important and will be used implicitly again and again. To get used to $A(V)$ and the operations in it, try the following exercises.
E14) Consider $S, T: R^2 \to R^2$ defined by $S(x_1, x_2) = (x_1, -x_2)$ and $T(x_1, x_2) = (x_1 + x_2, x_2 - x_3)$. What are $S + T, ST, TS, S \circ (S - T)$ and $(S - T) \circ S$?

E15) Let $S \in A(V)$, $dimV = n$ and $rank(S) = r$ and
Let $M = \{T \in A(V) | ST = 0\}$, $N = \{T \in A(V) | TS = 0\}$
a) Show that M and N are subspaces of $A(V)$.
b) Show that $M = L(V, Ker\ S)$.
c) What is $dim\ M$?

By now you must have got used to handling the elements of $A(V)$. The next section deals with polynomials that are related to these elements.

4.3.4 Minimal Polynomial Theorem

Recall that a polynomial in one variable x over F is of the form

$$p(x) = a_0 + a_1x + \cdots + a_nx^n, \text{ where } a_0, a_1, \cdots, a_n \in F.$$

If $a_n \neq 0$, then $p(x)$ is said to be of **degree n**.
If $a_n = 1$, then $p(x)$ is called a **monic polynomial of degree n**.
For example, $x^2 + 5x + 6$ is a monic polynomial of degree 2.
The set of all polynomials in x with coefficients in F is denoted by $F[x]$.
Definition 4.3.4: For a polynomial p, as above, and an operator $T \in A(V)$, we define

$$p(T) = a_0I + a_1T + \cdots + a_nT^n.$$

Since each of $I, T, \ldots, T^n \in A(V)$, we find $P(T) \in A(V)$. We say $P(T) \in F[T]$. If q is another polynomial in x over F, then $P(T)q(T) = q(T) = P(T)$, that is, $P(T)$ and $q(T)$ commute with each other. This can be seen as follows:
Let $q(T) = b_0I + b_1T + \cdots + b_mT^m$.
Then $P(T)q(T) = (a_0I + a_1T + \cdots + a_nT^n)(b_0I + b_1T + \cdots + b_mT^m)$

$$= a_0b_0I + (a_0b_1 + a_1b_0)T + \cdots + a_nb_mT^{n+m}$$
$$= (b_0I + b_1T + \cdots + b_mT^m)(a_0I + a_1T + \cdots + a_nT^n)$$
$$= q(T)P(T)$$

E16) Let $p, q \in F[x]$ such that $p(T) = 0, q(T) = 0$. Show that $(p + q)(T) = 0$.
Note that $(p + q)(x)$ means $p(x) + q(x)$.

E17) Check that $(2I + 3S + S^3)$ commutes with $(S + 2S^4)$, for $S \in A(R^n)$
We now go on to prove that given any $T \in A(V)$ we can find a polynomial $g \in F[x]$ such that
$g(T) = 0$, that is, $g(T)(v) = 0 \; \forall \; v \in V$.

Theorem 4.7: Let V be a vector space over F of dimension n and $T \in A(V)$. Then there exists a non-zero polynomial g over F such that $g(T) = 0$ and the degree of g is at most n^2.

Proof: We have already seen that $A(V)$ is a vector space of dimension n^2.
Hence, the set $\{I, T, T^2, \cdots, T^{n^2}\}$ of $n^2 + 1$ vectors of $A(V)$, must be linearly dependent (ref. Unit 2, Theorem 2.7). Therefore, there must exist $a_0, a_1, \cdots, a_{n^2} \in F$ (not all zero) such that

$$a_0I + a_1T + \cdots + a_{n^2}T^{n^2} = 0.$$

Let g be the polynomial of degree at most n^2, such that $(T) = 0$.
The following exercise will help you in getting used to polynomials in x and T.

E18) Give an example of polynomials $g(x)$ and $h(x)$ in $r[x]$, for which $g(I) = 0$ and
$h(0) = 0$, where I and 0 are the identity and zero transformations in $A(R)$.

E19) Let $T \in A(V)$then we have a map θ from $F[x]$ to $A(V)$ given by $\theta(p) = p(T)$ show that, for $a, b \in F$ and $p, q \in F[x]$,

a) $\theta(ap + bq) = a(p) + b(q)$

b) $\theta(pq) = \theta(p)\theta(q)$.

In Theorem 4.7, we have proved that there exists some $g \in F[x]$ with $g(T) = 0$.
But, if $g(T) = 0$, then $(g)(T) = 0$, for any $\alpha \in F$.
Also, if $deg\ g \leq n^2$. Thus, there are infinitely many polynomials that satisfy the conditions in theorem 4.7. But if we insist on some more conditions on the polynomial g, then we end up with one and only one polynomial which will satisfy these conditions and the conditions in Theorem 4.7.

Theorem 4.8: Let $T \in A(V)$, then there exists a unique monic polynomial p of smallest degree such that $p(T) = 0$.

Proof: Consider the set $S = \{g \in F[x] | g(T) = 0\}$.
This set is non-empty since, by Theorem 4.7, there exists a non-zero polynomial g, of degree at most n^2, such that $g(T) = 0$.
Now consider the set $D = \{degf | f \in S\}$.
Then D is a subset of $N \cup \{0\}$, and therefore, it must have a minimum element, m.
Let $h \in S$ such that $deg\ h = m$, then $h(T) = 0$ and $deg\ h \leq deg\ g\ \forall\ g \in S$.
If $h = a_0 + a_1x + \cdots + a_mx^m, a_m \neq 0$, then $p = x^{m-1}h$ is a monic polynomial such that $p(T) = 0$.
Also, $deg\ p = deg\ h \leq deg\ g\ \forall\ g \in S$.
Thus, we have shown that there exists a monic polynomial p, of least degree, such that
$p(T) = 0$.

We now show that p is unique, that is, if q is any monic polynomial of smallest degree such that $q(T) = 0$, then $p = q$. But this is easy.
Firstly, since $deg\ p = deg\ g\ \forall\ g \in S, deg\ p \leq deg\ q$.
Similarly, $deg\ q \leq deg\ p$. $\therefore deg\ p = deg\ q$.
Now, suppose $p(x) = a_0 + a_1x + \cdots + a_{n-1}x^{n-1} + a_nx^n$ and

$$q(x) = b_0 + b_1x + \cdots + b_{n-1}x^{n-1} + b_nx^n$$

Since $p(T) = 0$ and $q(T) = 0$, we get $(p- q)(T) = 0$.
But, $p- q = (a- b) +\ ...\ +\ldots$

Hence, $(p - q)$ is a polynomial of degree strictly less than the degree of p, such that $(p-q)(T) = 0$.
That is, $p-qS$ with $deg\ (p-q) < deg\ p$.
This is a contradiction to the way we chose p, unless $pq = 0$, that is, $p = q$.
P is the unique polynomial satisfying the conditions of Theorem 4.8.
This theorem immediately leads us to the following definition.

Definition 4.3.5: For $T \in A(V)$, the unique monic polynomial p of smallest degree such that $p(T) = 0$ is called the **minimal polynomial** of T.
Note that the minimal polynomial p, of T, is uniquely determined by the following three properties.
1) p is a monic polynomial over F
2) $p(T) = 0$
3) if $g \in F(x)$ with $g(T) = 0$, then $deg\ p \leq deg\ g$.

Consider the following example and exercises.

Example 6: For any vector space V, find the minimal polynomials for I, the identity transformation, and 0, the zero transformation.

Solution: Let $p(x) = x-1$ and $q(x) = x$. Then p and q are monic such that $p(I) = 0$ and $q(0) = 0$. Clearly no non-zero polynomials of smaller degree have the above properties. Thus $x-1$ and x are the required polynomials.
E20) Define $T: R^3 \rightarrow R^3 | T(x_1, x_2, x_3) = (0, x_1, x_2)$. Show that the minimal polynomial of T is x^3.
E21) Define $T: R^n \rightarrow R^n | T(x_1, \cdots, x_{n-1})$. What is the minimal polynomialof T?
Does E20 help you?
E22) Let $T: R^3 \rightarrow R^3$ be defined by $T(x_1, x_2, x_3) = (3x_1, x_1 - x_2, 2x_1 + x_2 + x_3)$.
Show that $(T^2-I)(T-3I) = 0$. What is the minimal polynomial of T ?
We will now state and prove a criterion by which we can obtain the minimal polynomial of linear operator T, once we know any polynomial $f \in F[x]$ with $f(T) = 0$. It says that the minimal polynomial must be a factor of any such f.

Theorem 4.9: Let $T \in A(V)$ and let $p(x)$ be the minimal polynomial of T. Let $f(x)$ be any polynomial such that$f(T) = 0$.Then there exists a polynomial g(x)such that $f(x) = p(x)g(x)$.
Proof: The division algorithm states that given $f(x)$ and $p(x)$, there exist polynomials $g(x)$ and $h(x)$ such that $f(x) = p(x)g(x) + h(x)$, where $h(x) = 0$ or $deg\ h\ (x) < deg\ p(x)$. Now, $0 = f(T) = p(T)g(T) + h(T) = h(T)$, since $p(T) = 0$.

Therefore, if $h(x) \neq 0$, then $h(T) = 0$, and $deg\ h(x) < deg\ p(x)$.
This contradicts the fact that $p(x)$ is the minimal polynomial of T.
Hence, $h(x) = 0$ and we get $f(x) = p(x)g(x)$.
Using this theorem, can you obtain the minimal polynomial of T in E22 more easily? Now we only need to check if $T - I, T + I$ or $T - 3I$ are 0.

Remark: If $dim\ V = n$ and $T \in A(V)$, we have seen that the degree of the minimal polynomial $\boldsymbol{p}$ of $T \leq n^2$.
We will study a systematic method of finding the minimal polynomial of T, and some applications of this polynomial. But now we will only illustrate one application of the concept of the minimal polynomial by proving the following theorem.

Theorem 4.10: Let $T \in A(V)$, then T is invertible if and only if the constant term in the minimal polynomial of T is not zero.

Proof: Let $p(x) = a_0 + a_1x + \cdots + a_{m-1}x^{m-1} + a_mx^m$ be the minimal polynomial of T. Then, $a_0I + a_1T + \cdots + a_{m-1}T^{m-1} + a_mT^m = 0$

$$T(a_1I + a_2T + \cdots + a_{m-1}T^{m-2} + T^{m-1}) = -a_0I$$

……………………. (1)
Firstly, we will show that if T^{-1} exists, then $a_0 \neq 0$.
On the contrary, suppose $a_0 = 0$.
Then (1) implies that $T(a_1I + a_2T + \cdots + a_{m-1}T^{m-2} + T^{m-1}) = 0$.
Multiplying both sides by T^{-1} on the left, we get

$$a_1I + a_2T + \cdots + a_{m-1}T^{m-2} + T^{m-1} = 0$$

This equation gives us a monic polynomial $q(x) = a_1 + \cdots + x^{m-1}$ such that $q(T) = 0$ and $deg\ q < deg\ p$.
This contradicts the fact that p is the minimal polynomial of T.
Therefore, if T^{-1} exists the constant term in the minimal polynomial of T cannot be zero.

Conversely; suppose the constant term in the minimal polynomial of T is not zero, that is, $a \neq 0$. Then dividing equation (1) on both sides by $(-a_0)$, we get

$$T((-a_1/a_0)I + \cdots + (-I/a_0)\ T^{m-1}) = I$$

Let $S = (-a_1/a_0)I + \cdots + (-I/a_0)\ T^{m-1}$,
Then we have $ST = I$ and $TS = I$.
This shows, by Theorem 4.5, that T^{-1} exists and $T^{-1} = S$.
E23) Let P_n be the space of all polynomials of degree $\leq n$. Consider the linear operator
$D: P_2 \to P_2$ given by $D(a_0 + a_1x + a_2x^2) = a_1 + 2a_2x$.
(Note that D is just the differentiation operator.)
Show that $D^4 = 0$. What is the minimal polynomial of D? Is D invertible?

E24) Consider the reflection transformation given in Unit 3, Example 4, find its minimal polynomial. Is T invertible? If so, find its inverse.
E25) Let the minimal polynomial of $S \in A(V)$ be $x^n, n \geq I$. Show that there exists $v_0 \in V$ such that the set $\{v_0, S(v_0), \ldots., S^{n-1}(v_0)\}$ is linearly independent.
We shall now end the unit by summarizing what we have covered in it.

SELF-ASSESSMENT EXERCISE(S) SOLUTIONS

Solutions/Answers
E1) We have to check that VS1 – VS10 are satisfied by $L(U, V)$.
We have already shown that VS1 and VS6 are true.
VS2: For any $L, M, N \in L(U, V)$, we have $\forall\ u \in U$,
$$[(L + M) + N](u) = (L + M)(u) + N(u)$$
$$= [L(u) + M(u)] + N(u)$$
$= L(u) + [M(u) + N(u)]$, since addition is associative in V.
$$= [L + (M + N)](u)$$
$\therefore (L + M) + N = L + (M + N)$.
VS3:$0: U \to V: 0(u) = 0\ \forall\, u \in U$ is the zero element of $L(U, V)$.
VS4: For any $S \in L(U, V), (-1)S = -S$, is the additive inverse of S.

VS5: Since addition is commutative in$V, S + T = T + S\ \ \forall\ S, T\ in\ L(U, V)$.

VS7:$\forall\ \alpha \in F$ and $S, T \in L(U, V)$,
$\alpha(S + T) = (\alpha S + \alpha T)(u)\ \ \forall\ \ u \in U$,
$\therefore \alpha(S + T) = \alpha S + \alpha T$.

VS8: $\forall\ \alpha, \beta \in F$ and $S, T \in L(U, V)$, then
$(\alpha + \beta)S = \alpha S + \beta S$.

VS9: $\forall\ \alpha, \beta \in F$ and $S \in L(U, V),\ (\alpha\beta)S = \alpha(\beta S)$.

VS10: $\forall\ \ S \in L(U, V)$, $1 \cdot S = S$.

E2) $E_{2m}(e_m) = f_2$ and $E_{2m}(e_i) = 0$ for $i \neq m$.
$E_{32}(e_i) = f_3$ and $E_{23}(e_i) = 0$ for $i = 2$.
$$E_{mn}(e_i) = \begin{cases} f_m,\ \ if\ i = n \\ 0, \quad otherwise \end{cases}$$
E3) Both spaces have dimension 2 over R.
A basis for $L(R^2, R)$ is $\{E_{11}, E_{12}\}$, where,
i.$E_{11} = (1,0) = 1$
ii.$E_{11} = (0,1) = 0$
iii.$E_{12} = (1,0) = 0$
iv.$E_{12} = (0,1) = 1$

A basis for $L(R, R^2)$ is $\{E_{11}, E_{21}\}$, where,

i. $E_{11}(1) = (1,0)$
ii. $E_{11}(0) = (0,1)$
iii. $E_{21}(0) = (1,0)$
iv. $E_{12}(1) = (0,1)$

E4) Let $f: R^3 \to R$ be any linear functional. Let $f(1,0,0) = a_1$, $f(0,1,0) = a_2$, $f(0,0,1) = a_3$. Then, for any $x \leftrightharpoons (x_1, x_2, x_3)$, we have $x = x_1(1,0,0) + x_2(0,1,0) + x_3(0,0,1)$,

$$\therefore f(x) = x_1 f(1,0,0) + x_2 f(0,1,0) + x_3 f(0,0,1)$$
$$= a_1 x_1 + a_2 x_2 + a_3 x_3$$

E5) Let the dual basis be $\{f_1, f_2, f_3\}$.
Then, for any $v \in P_2$, $v = f_1(v) \cdot 1 + f_2(v) \cdot x + f_3(v) \cdot x_2$
:. If $v = a_0 + a_1 x + a_2 x^2$, then $f_1(v) = a_0, f_2(v) = a_1, f_3(v) = a_2$.
That is, $f_1(a_0 + a_1 x + a_2 x^2) = a_0$, $f_2(a_0 + a_1 x + a_2 x^2) = a_1$, $f_3(a_0 + a_1 x + a_2 x^2) = a_2$, for any $(a_0 + a_1 x + a_2 x^2) \in P_2$.
E6) Let $\{f_1, \cdots, f_n\}$ be a basis of V*. Let its dual basis be $\{\varphi_1, \cdots, \varphi_n\}$, $\varphi_i \in V^{**}$. Let $e_i \in V$ such that $\theta(e_i) = \varphi_i$ (ref. Theorem 3) for $i = 1, \dots, n$.
Then $\{e_1, \dots, e_n\}$ is a basis of V, since θ^{-1} is an isomorphism and maps a basis to $\{e_1, \dots, e_n\}$.
Now, $f_i(e_j) = \theta(e_j)(f_i) = \varphi_j(f_i) = \delta_{ji}$, by definition of a dual basis.
$\therefore \{f_1, \cdots, f_n\}$is the dual of $\{e_1, \dots, e_n\}$.
E7) For any $S \in A(V)$ and for any $v \in V$,
$S \circ I(v) = S(v)$ and $I \circ S(v) = I(S(v)) = S(v)$.
$\therefore S \circ I = S = I \circ S$.
E8) For any $S \in A(V)$ and for any $v \in V$,
$S \circ 0(v) = S(0) = 0$, and $0 \circ S(v) = 0(S(v)) = 0$.
$\therefore S \circ 0 = 0 \circ S = 0$.

E9) $S \in A(R), T \in A(R^2)$.

$$S \circ T(x_1, x_2) = S(-x_2, x_1) = (x_1, x_2)$$

$T \circ S(x_1, x_2) = T(x_1, -x_2) = (x_1, x_2) \ \forall (x_1, x_2) \in R^2$.
$\therefore S \circ T = T \circ S = I$, and hence, both S and T are invertible.
E10) $T \in L(R^2, R^3), S \in L(R^3, R^2)$.
$\therefore S \circ T \in A(R), T \circ S \in A(R^3)$.
$\therefore S \circ T$ and $T \circ S$ can never be equal.
Now $S \circ T(x_1, x_2) = S(0, x_1, x_2) = (x_1, x_1 + x_2) \ \forall (x_1, x_2) \in R^2$
Also, $T \circ S(x_1, x_2, x_3) = T(x_1 + x_2, x_2 + x_3)$
$= (0, x_1 + x_2, x_2 + x_3) \ (x_1, x_2, x_3) \in R^3$.

E11) Since $T \in A(R^2)$ and $S \in A(R^3)$, $S \circ T$ and $T \circ S$ are not defined.

E12) Both $(R \circ S) \circ T$ and $R \circ (S \circ T)$ are in $L(U, Z)$.
For any $u \in U, [(R \circ S) \circ T](u) = (R \circ S)[T(u)]$
$= R[S(T(u))] = R[S \circ T)(u)] = [R \circ (S \circ T)](u)$.
$\therefore (R \circ S) \circ T = R \circ (S \circ T)$.

E13) By Unit 3, Theorem 3.6, $rank\ (S \circ T) \leq rank\ (T)$.
Also, $rank\ (T) = rank(I \circ T) = rank((S^{-1} \circ S) \circ T)$
$= rank\ (S^{-1} \circ (S \circ T)) \leq rank(S \circ T)$ (by Theorem 3.6).
Thus, $rank\ (S \circ T) \leq rank\ (T) \leq rank(S \circ T)$.
$\therefore rank\ (S \circ T) = rank\ (T)$.
Similarly, you can show that rank $(T \circ S) = rank\ (T)$.

E14) $(S + T)(x, y) = (x, -y) + (x + y, y - x) = (2x + y - x)$
$S \circ T(x, y) = S(x + y, y - x) = (x + y, x - y)$
$T \circ S(x, y) = T(x, -y) = (x - y, -(x + y\ ))$
$[S \circ (S - T)](x, y) = S(-y, x - 2y) = (-y, 2y - x)$
$[(S - T) \circ S](x, y) = (S - T)(x, -y) = (x, y) - (x - y, -(x + y))$
$= (y, 2y + x)\ \forall\ (x, y) \in R^2$.

E15) a) We first show that if $A, B \in M$ and $\alpha, \beta \in F$, then $(\alpha A + \beta B) \in M$.
Now, $S(\alpha A + \beta B) = S \circ \alpha A + S \circ \beta B$, by Theorem 4.6.
$= \alpha(S \circ A) + \beta(S \circ B)$, again, by Theorem 6
$= \alpha_0 + \beta_0$, since $A, B \in M$
$= 0$
$\therefore \alpha A + \beta B \in M$ and M is a subspace of $A(V)$.
Similarly, you can show that N is a subspace of $A(V)$.

b) For any $T \in M$, $ST(v) = 0\ \forall\ v \in V$.
$\therefore T(v) \in Ker\ S\ \ \forall\ \ v \in V$.
$\therefore R(T)$, the range of T, is a subspace of $Ker\ S$.
$\therefore\ TL(V, Ker\ S)$
$\therefore M \subseteq L(V, Ker\ S)$
Conversely, for any $T \in L(V, Ker\ S)$, $T \in A(V)$ such that $S(T(v)) = 0\ \forall\ v \in V$.
$ST\ =\ 0$
$\therefore T \in M, \therefore L(V, Ker\ S) \subseteq M$
$\therefore$ We have proved that $M = L(V, Ker\ S)$.
$\therefore dim\ M = (dim\ V)(nullity\ S)$, by Theorem 4.1
$= n(n - r)$ by the Rank Nullity Theorem.

E16)$(p + q)(T) = p(T) + q(T) = 0 + 0 = 0$.

E17) $(2I + 3S + S^3)(S + 2S^4) = (2I + 3S + S^3)S + (2I + 3S + S^3)(2S^4)$

$$= 2S + 3S^2 + S^4 + 4SS^4 + 6S^5 + 2S^7$$

$$= 2S + 3S^2 + 5S^4 + 6S^5 + 2S^7$$

Also, $(S + 2S^4)(2I + 3S + S3) = 2S + 3S^2 + 5S^4 + 6S^5 + 2S^7$

$$\therefore (S + 2S^4)(2I + 3S + S^3) = (2I + 3S + S^3)(S + 2S^4).$$

E18) Consider $g(x) = x - 1 \in R[x]$

Then, $g(I) = I - 1I = 0$.

Also, if $h(x) = x$, then $h(0) = 0$.

Notice that the degrees of g and h are both $I \leq dim\ R^3$.

E19) Let $p = a_0 + a_1x + \cdots + a_nx^n$, $q = b_0 + b_1x + \cdots + b_mx^m$.

a) Then $ap + bq = a(a_0 + a_1x + \cdots + a_nx^n) + b(b_0 + b_1x + \cdots + b_mx^m)$

$$= aa_0 + aa_1x + \cdots + aa_nx^n + bb_0 + bb_1x + \cdots + bb_mx^m$$

$\therefore \phi(ap + bq)$

$$= aa_0I + aa_1T + \cdots + aa_nT^n + bb_0I + bb_1T + \cdots + bb_mT^m$$

$$= ap(T) + bq(T) = a\phi(p) + b\phi(q)$$

b) $pq = (a_0 + a_1x + \cdots + a_nx^n)(b_0 + b_1x + \cdots + b_mx^m)$

$$= a_0b_0 + (a_1b_0 + a_0b_1)x + \cdots + a_nb_mx^{n+m}$$

$\therefore \phi(pq) = a_0b_0I + (a_1b_0 + a_0b_1)T + \cdots + a_nb_mT^{n+m}$

$$= (a_0I + a_1T + \cdots + a_nT^n)(b_0I + b_1T + \cdots + b_mT^m)$$

$$= \phi(p)\phi(q).$$

E20) Let $T \in A(R^3)$ and$p(x) = x^3$,then p is a monic polynomial.

Also, $p(T)(x_1, x_2, x_3) = T^3(x_1, x_2, x_3)$

$$= T^2(0, x_1, x_2) = T(0,0, x_1)$$

$$= (0,0,0) \quad \forall \quad (x_1, x_2, x_3) \in R^3$$

$$\therefore p(T) = 0$$

We must also show that no monic polynomial q of smaller degree exists such that $q(T) = 0$.

Suppose $q = a + bx + x^2$ and $q(T) = 0$

Then $(aI + bT + T^2)(x_1, x_2, x_3) = (0, 0, 0)$

$$\Leftrightarrow a(x_1, x_2, x_3) + b(0, x_1, x_2) + (0,0, x_1) = (0, 0, 0)$$

$$\Leftrightarrow ax_1 = 0, ax_2 + bx_1 = 0, ax_3 + bx_2 + x_1 = 0\ \forall\ (x_1, x_2, x_3) \in R^3$$

$$\Leftrightarrow a = 0, b = 0, x_1 = 0$$

But x_1 can be non-zero

Therefore, q does not exist

Hence,p is a minimal polynomial of T.

E21) Consider $p(x) = x^n$, then $p(T) = 0$ and no non-zero polynomial q of lesser degree exists such that $q(T) = 0$. This can be checked in the solution of E20.

E22)$(T^2 - I)(T - 3I)(x_1, x_2, x_3) = (T^2 - I)((3x_1, x_1 - x_2, 2x_1 + x_2 + x_3) - (3x_1, 3x_2, 3x_3))$

$= (T^2 - I)(0, x_1 - 4x_2, 2x_1 + x_2 - 2x_3)$

$= T(0, x_1 - 4x_2, 3x_1 - 3x_2 - 2x_3) - (0, x_1 - 4x_2, 2x_1 + x_2 - 2x_3)$

$= (0,0,0) \quad \forall \quad (x_1, x_2, x_3) \in R^3$

Therefore, $(T^2 - I)(T - 3I) = 0$

Suppose $\exists \ q = a + bx + x^2$ such that $q(T) = 0$

Then $q(T)a(x_1, x_2, x_3) = (0,0,0) \ \forall \ (x_1, x_2, x_3) \in R^3$

This means that

$$a + 3b + 9 = 0,$$
$$(b + 2)x_1 + (a + b + 1)x_2 = 0,$$
$$(2b + 9)x_1 + bx_2 + (a + b + 1)x_3 = 0.$$

Eliminating *a* and *b*, we find that these equations can be solved provided

$$5x_1 - 2x_2 - 4x_3 = 0$$

But they should be true for any $(x_1, x_2, x_3) \in R^3$

Therefore, the equations can't be solved, and *q* does not exist.

Hence, the minimal polynomial of *T* is $(x^2 - I)(x - 3)$.

E23) $D^4(a_0 + a_1x + a_2x^2) = D^3(a_1 + 2a_2x)$

$= D^2(2a_2) = D(0)$

$= 0 \ \forall \ (x_1, x_2, x_3) \in p^2$

$\therefore D^4 = 0.$

The minimal polynomial of *D* can be *D*, D^2, D^3 or D^4.

Check that $D^3 = 0$, but $D^2 \neq 0$. :.

The minimal polynomial of *D* is $p(x) = x^3$.

Since *p* has no non-zero constant term, then *D* is not an isomorphism.

E24) $T: R^2 \to R^2 | T(x, y) = (x, -y)$. Check that $T^2 - I = 0$.

Therefore, the minimal polynomial *p* must divide $x^2 - I$.

$\Longrightarrow P(x)$ can be $x - 1, x + 1$ or $x^2 - 1$.

Since $T - I \neq 0$ and $T + I \neq 0$, we see that $p(x) = x^2 - 1$.

By Theorem 4.10, *T* is invertible.

Now $T^2 - I = 0$

Therefore,$T(-T) = 1$, hence,$T - 1 = -T$.

E25) Since the minimal polynomial of *S* is x^n, $S^n = 0$ and $S^{n-1} \neq 0$, then $\exists \ v_0 \in V$ such that $S^{n-1}(v_0) \neq 0$.

Let $a_1, a_2, \cdots, a_n \in F$, such that $a_1v_0 + a_2S(v_0) + \cdots + a_nS^{n-1}(0) = 0$ ……… (1)

Then, applying S^{n-1} to both sides of this equation, we have

$$a_1S^{n-1}(v_0) + \cdots + a_nS^{2n-1}(v_0) = 0,$$

…………………….. (2)

$\Rightarrow a_1S^{n-1}(v_0) = 0$, since $S^n = 0, \ S^{n+1} = \cdots = S^{2n-1}$

$\Longrightarrow a_1 = 0$

Now (1) reduces to $a_2 S(v_0) + \cdots + a_n S^{n-1}(v_0) = 0$.
Applying $S^{\mathrm{n}-2}$ to both sides we get $a_2 = 0$.
In this way we get $a_1 = 0 \; \forall \; i = 1, \ldots, n$.
Therefore, the set $\{v_0, S(v_0), \cdots, S^{n-1}(v_0)\}$ is linearly independent.

We conclude that the composition of two linear transformations is a linear operator.

Note that we use the terms 'linear transformation' interchangeably.
In a linear operator $T: V \to V$, T satisfies a polynomial equation $g(x) = 0$, that is, if we substitute T for x in $g(x)$, we get the zero transformation.

4.5 SUMMARY

In this unit we covered the following points.

i. $L(U, V)$ the vector space of all linear transformations from U to V is of dimension $(dim\ U)(dim\ V)$.
ii. The dual space of a vector space V is $L(U, F) = V*$, and is isomorphic to V.
iii. If $\{e_1, \ldots\ldots, e_n\}$ is a basis of V and $\{f_1, \ldots.., f_n\}$ is its dual basis,
iv. then $f = \sum_{i=1}^{n} f(e_i) f_i \forall f \in V*$ and $v = \sum_{i=1}^{n} f_i(v)\, e_i \;\; \forall \;\; v \in V$.
v. Every vector space is isomorphic to its second dual.
vi. Suppose $S \in L(V, W)$ and $T \in L(U, V)$. Then their composition $S \circ T \in L(U, W)$.
vii. $S \in A(V) = L(V, V)$ is an isomorphism if and only if there exists $T \in A(V)$ such that $S \circ T = I = T \circ S$.
viii. For $T \in A(V)$ there exists a non-zero polynomial $g \in F[x]$, of degree at most n^2, such that $g(T) = 0$, where $dim\ V = n$.
ix. The minimal polynomial of T and f is a polynomial p, of smallest degree such that $p(T) = 0$.
x. If p is the minimal polynomial of T and f is a polynomial such that $f(T) = 0$, then there exists a polynomial $g(x)$ such that $f(x)g(x)$.
xi. Let $T \in A(V)$.Then T^{-1}exists if and only if the constant term in the minimal polynomial of T is not zero.

4.6 References/Further Readings

Robert A. Beezer (2014). A First Course in Linear Algebra. Congruent Press Gig Harbor, Washington, USA 3(40).

Arbind K Lal Sukant Pati (2018). Linear Algebra through Matrices.

Peter J. Cameron (2008). Notes on Linear Algebra.

MODULE 2

You have studied Vector Spaces in Module One. A simple means of representing them, namely, by matrices (plural form of 'matrix') shall be studied in this module. We shall show that, given a linear transformation, a matrix associated to it can be obtained, and vice versa. Also, certain properties of a linear transformation can be studied more easily if the associated matrix is studied instead. For example, you shall see that it is often easier to obtain the characteristic roots of a matrix than of a linear transformation.

The units under this module include:

Unit 1 Matrices I
Unit 2 Matrices II
Unit 3 Matrices III

UNIT 1 MATRICES I

Unit Structure

1.1 Introduction

Matrices were introduced by the English Mathematician, Arthur Cayley, in 1858. He came upon this notion in connection with linear substitutions.

Matrix theory now occupies an important position in pure as well as applied mathematics. In physics, one comes across such terms as matrix mechanics, scattering matrix, spin matrix, annihilation and creation matrices. In economics we have the input-output matrix and the payoff matrix; in statistics we have the transition matrix; and in engineering, the stress matrix, strain matrix, and many other matrices.

Matrices are intimately connected with linear transformations. In this unit we will bring out this link. We will first define matrices and derive algebraic operations on matrices from the corresponding operations on linear transformations. We will also discuss some special types of matrices. One type, a triangular matrix, will be used often in the subsequent units.

To realize the deep connection between matrices and linear transformations, you should go back to the exact sport in Units 1 and 2 to which frequent references are made.

This unit may take you a little longer to study, than previous ones, but don't let that worry you. The material in it is actually very simple.

1.2 Learning Outcomes

By the end of this unit, you will be able to:

- Define and give examples of various types of matrices;
- Obtain a matrix associated to a given linear transformation
- Define a linear transformation, if you know its associated matrix;
- Evaluate the sum, difference, product and scalar multiples of matrices;
- Obtain the transpose and conjugate of a matrix;

1.3 Matrices

1.3.1 Matrix Description

Consider the following system of three simultaneous equations in four unknowns:

$$a - 3b + 2c - d = 0$$
$$\frac{1}{2}a + b - 2d = 0$$
$$3b - 4c = 0$$

The coefficients of the unknowns, x, y, z and t can be arranged in rows and columns to form a rectangular array as follows:

$$\begin{bmatrix} 1 & -3 & 2 & -1 \\ \frac{1}{2} & 1 & 0 & -1 \\ 0 & 3 & -4 & 0 \end{bmatrix} \begin{matrix} \text{(Coefficients of the first equation)} \\ \text{(Coefficients of the second equation)} \\ \text{(Coefficients of the third equation)} \end{matrix}$$

Such a rectangular array (or arrangement) of numbers is called a matrix. A matrix is usually enclosed within square brackets [ ] or round brackets ( ).

$$\begin{bmatrix} 1 & -3 & 2 & -1 \\ \frac{1}{2} & 1 & 0 & -1 \\ 0 & 3 & -4 & 0 \end{bmatrix}$$

The numbers appearing in the various positions of a matrix are called the **entries** (or **elements)** of the matrix. Note that the same number may appear at two or more different positions of a matrix. For example, 1 appears in 3 different positions in the matrix given above.

In the matrix above, the three horizontal rows of entries have 4 elements each. These are called the **rows** of this matrix. The four vertical rows of entries in the matrix, having 3 elements each, are called its **columns.** Thus, this matrix has three rows and four columns described as a matrix of size 3 x 4 ("3 by 4" or "3 cross 4"), or simply a "3 x 4" matrix. The rows are counted from top to bottom and the columns are counted from left to right. Thus, the first row is (1, -3, 2, -1), the second row is (½, 1, 0, 1), and the third row is (0, 3, -4, 0).

Similarly, the first column is $\begin{bmatrix} 1 \\ \frac{1}{2} \\ 0 \end{bmatrix}$, The second column is $\begin{bmatrix} 1 \\ \frac{1}{2} \\ 0 \end{bmatrix}$, and the third column is $\begin{bmatrix} 1 \\ \frac{1}{2} \\ 0 \end{bmatrix}$

Note that each row is a (1×4) matrix and each column is a (3×1) matrix,

We shall now define a matrix of any size.

Let us see what we mean by a matrix of size $(m \times n)$, where m and n are any two natural numbers.

Definition 1.3.1: Let F be a field. A rectangular array of mn elements of F arranged in m rows and n columns is called a **matrix of size** $(m \times n)$**;** or an $(m \times n)$ **matrix,** over ***F*****.**

$$\begin{bmatrix} \boldsymbol{a_{11}} & \boldsymbol{a_{12}} & \cdots & \cdots & \boldsymbol{a_{1n}} \\ \boldsymbol{a_{21}} & \boldsymbol{a_{22}} & \cdots & \cdots & \boldsymbol{a_{2n}} \\ \vdots & \vdots & \cdots & \cdots & \vdots \\ \vdots & \vdots & \cdots & \cdots & \vdots \\ \boldsymbol{a_{m1}} & \boldsymbol{a_{m2}} & \cdots & \cdots & \boldsymbol{a_{mn}} \end{bmatrix}$$

You must remember that the mn entries need not be distinct.
The element at the intersection of the i^{th} row and the j^{th}column is called the $(i,j)^{th}$elements.
For example, in the $(m \times n)$ matrix above, the $(2,n)^{th}$elements is a_{2n} which is the intersection of the 2^{nd}row and the n^{th} column .
A brief notation for this matrix is $[a_{ij}]_{mn}$, or simply $[a_{ij}]$.
Matrices are also denoted by capital letters *A, B, C,* etc.
The set of all$m \times n$matrices over F is denoted by$M_{m\times n}(F)$, thus, $[1,\sqrt{2}] \in M_{1\times 2}(F)$.
If $m = n$, the matrix is called a **square matrix**.
In an $(m \times n)$ matrix, each row is a $(1 \times n)$ matrix and is also called a **row vector**.
Similarly, each column is an $(m \times 1)$ matrix and is also called a **column vector**.
Let us look at a situation in which a matrix can arise.

Example 1:
In the B.Sc.(Hons.) Mathematics Programme of the National Open University of Nigeria (N.O.U.N.), there are 25 male and 11 female students in Year I; 18 male and 10 female students in Year II; 15 male and 8 female students in Year III and 12 male and 6 female students in Year IV. How does this information give rise to a matrix?

Solution:
One of the ways in which we can arrange this information in the form of a matrix is as follows:

$$\begin{array}{c} \begin{matrix} I & II & III & IV & (B.Sc.) \end{matrix} \\ \begin{bmatrix} 25 & 18 & 15 & 12 \\ 11 & 10 & 8 & 6 \end{bmatrix} \begin{matrix} Male \\ Female \end{matrix} \end{array}$$

This is a (2×4) matrix.

Another way could be the (4×2) matrix
$$\begin{matrix} & M \quad F & \\ & \begin{bmatrix} 25 & 11 \\ 18 & 10 \\ 15 & 8 \\ 12 & 6 \end{bmatrix} & \begin{matrix} B.Sc.I \\ B.Sc.II \\ B.Sc.III \\ B.Sc.IV \end{matrix} \end{matrix}$$

Either of these matrix representations immediately shows us how many male/female students are in any of the classes.
To get used to matrices and their elements, you can try the following exercises.

Suppose $A = \begin{bmatrix} 1 & 3 & 5 \\ 8 & 4 & 1 \\ -1 & 0 & 2 \end{bmatrix}, B = \begin{pmatrix} 1 & 3 & 7 & -2 \\ 8 & 4 & 1 & 0 \\ -1 & 0 & 2 & 5 \end{pmatrix},$

$$C = \begin{pmatrix} 3 & 1 & 5 & -2 \\ 4 & 8 & 1 & 0 \\ 0 & -1 & 2 & 5 \end{pmatrix},\ D = \begin{pmatrix} 2 & 1 & 7 \\ 0 & 4 & 3 \\ -1 & 8 & 1 \end{pmatrix}$$

a) Obtain the $(2,3)^{th}$ elements of matrices A,B,C and D
b) State the major difference(s) between matrices A and B
c) Enumerate the differences and similarities observed in B and C
d) What are the elements in the fourth row of B?

1.3.2 Algebra of Matrices

Definition 1.3.2: The $m \times n$ matrix, $O = O_{m\times n}$, defined by $[O]_{ij} = 0$, for all $1 \le i \le m$, $1 \le j \le n$ is known as **zero matrix.**

Matrix Equality

Definition 1.3.3: The $m \times n$ matrices A and B are said to be **equal** (written as $A = B$) provided $[A]_{ij} = [B]_{ij}$ for all $1 \le i \le m$, $1 \le j \le n$.

Two matrices are said to be **equal** if

i. They have the same size, that is, they have the same number of rows as well as the same number of columns and,
ii. Their elements, at all the corresponding positions, are the same.
iii. The following example will clarify what we mean by equal matrices.

Example 2:

If $\begin{pmatrix} 1 & 0 \\ 2 & 3 \end{pmatrix} = \begin{pmatrix} x & y \\ z & 3 \end{pmatrix}$, then $x = 1$, $y = 0$, $z = 2$

Remark: Firstly, both matrices are of the same size, namely, (2×2) and for these matrices to be equal the corresponding elements of both must be equal for all i, j.

Matrix Addition (or Subtraction)

Given the $m \times n$ matrices A and B, define the **sum (or difference)** of A and B as an $m \times n$ matrix, written as $A \pm B$ and defined by

$[A \pm B]_{ij} = [A]_{ij} \pm [B]_{ij};\ \ 1 \le i \le m, 1 \le j \le n$

So, matrix addition takes two matrices of the same size and combines them (in a natural way) to create a new matrix of the same size.

Definition 1.3.4: Let A and B be the following two m x n matrices

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & \cdots & a_{mn} \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} b_{11} & b_{12} & \cdots & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & \cdots & b_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ b_{m1} & b_{m2} & \cdots & \cdots & b_{mn} \end{pmatrix}$$

Then the sum of A and B is defined to be the matrix

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

In other words, $A + B$ is the m x n matrix whose $(i, j)^{th}$ element is the sum of the $(i, j)^{th}$ element of A and the $(i, j)^{th}$element of B.

Let us see an example of how two matrices are added.

Example 3:

Let $A = \begin{pmatrix} 1 & 0 & 5 \\ 2 & -3 & 4 \end{pmatrix}$ and $B = \begin{pmatrix} 7 & 3 & 1 \\ -1 & 6 & 2 \end{pmatrix}$

Then $A + B = \begin{pmatrix} 1 & 0 & 5 \\ 2 & -3 & 4 \end{pmatrix} + \begin{pmatrix} 7 & 3 & 1 \\ -1 & 6 & 2 \end{pmatrix}$

$$= \begin{pmatrix} 1 + 7 & 0 + 3 & 5 + 1 \\ 2 + (-1) & -3 + 6 & 4 + 2 \end{pmatrix} = \begin{pmatrix} 8 & 3 & 6 \\ 1 & 3 & 6 \end{pmatrix}$$

Also, $A - B = \begin{pmatrix} 1 - 7 & 0 - 3 & 5 - 1 \\ 2 - (-1) & -3 - 6 & 4 - 2 \end{pmatrix} = \begin{pmatrix} -6 & -3 & 4 \\ 3 & -9 & 2 \end{pmatrix}$

1.3.3 Matrix of a Linear Transformation

We shall now obtain a matrix that corresponds to a given linear transformation and see how easy it is to go from matrices to linear transformations, and back.

Let U and V be vector spaces over a field F, of dimensions n and m, respectively.

Let $B_1 = \{e_1, \ldots, e_n\}$ be an ordered basis of U, and $B_2 = \{f_1, \ldots, f_m\}$ be an ordered basis of V, (By **an ordered basis we mean that the order in which the elements of the basis are written is fixed**. Thus, an ordered basis $\{e_1, e_2\}$ is not equal to an ordered basis $\{e_2, e_1\}$).

Given a linear transformation $T: U \to V$, we will associate a matrix to it. For this, we consider $T(e_1), \ldots, T(e_n)$, which are all elements of V and hence, they are linear combinations of $f_1, \ldots, f_m$. Thus, there exist mn scalars α_{ij}, such that

$$T(e_1) = \alpha_{11}f_1 + \alpha_{21}f_2 + \ldots + \alpha_{m1}f_m$$
$$\vdots$$
$$T(e_j) = \alpha_{1j}f_1 + \alpha_{2j}f_2 + \ldots + \alpha_{mj}f_m$$
$$\vdots$$
$$T(e_n) = \alpha_{1n}f_1 + \alpha_{2n}f_2 + \ldots + \alpha_{mn}f_m$$

From these n equations we form an $m \times n$ matrix whose first column consists of the coefficients of the first equation; second column consists of the coefficients of the second equation, and so on. This matrix

$$\begin{pmatrix} \alpha_{11} & \alpha_{12} & \cdots & \cdots & \alpha_{1n} \\ \alpha_{11} & \alpha_{22} & \cdots & \cdots & \alpha_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \alpha_{m1} & \alpha_{m2} & \cdots & \cdots & \alpha_{mn} \end{pmatrix}$$

is called the matrix of T with respect to the bases B_1 and B_2. Notice that the coordinate vector of $T(e_j)$ is the j^{th} column of A.

We use the notation $[T]_{B_1,B_2}$ for this matrix.

Thus, to obtain $[T]_{B_1,B_2}$ we consider $T(e_j)\ \forall\ e_j \in B_1$, and write them as linear combinations of the elements of B_2.

If $T \in L(V,V)$, B is a basis of V and we take $B_1 = B_2 = B$, then $[T]_{B,B}$ is called the matrix of T with respect to the basis B, and can also be written as $[T]_B$.

Remark: Why do we insist on order bases? What happens if we interchange the order of the elements in B to $\{e_n, \ldots, e_{n-1}\}$? The matrix $[T]_{B_1,B_2}$ also changes, the last column becoming the first column now. Similarly, if we change the positions of the f_i's in B_2, the rows of $[T]_{B_1,B_2}$ will get interchanged.

Thus, to obtain a unique matrix corresponding to T, we must insist on B_1 and B_2 being ordered bases. Henceforth, while discussing the matrix of a linear mapping, we will always assume that our bases are ordered bases. We will now give an example, followed by some exercises.

Example 4: Consider the linear operator $T: R^3 : T(x,y,z) = (x,y)$. Choose bases B_1 and B_2 of R^3 and R^2, respectively. Then obtain $[T]_{B_1,B_2}$.

Solution: Let $B_1 = \{e_1, e_2, e_3\}$, where $e_1 = (1,0,0), e_2 = (0,1,0), e_3 = (0,0,1)$ and $B_2 = \{f_1, f_2\}$, where $f_1 = \{1,0\}, f_2 = \{0,1\}$,

Note that B_1 and B_2 are the standard bases of R^3 and R^2, respectively.

$$T(e_1) = (1,0) = f_1 = 1 \cdot f_1 + 0 \cdot f_2$$
$$T(e_2) = (0,1) = f_2 = 0 \cdot f_1 + 1 \cdot f_2$$
$$T(e_3) = (0,0) = 0 \cdot f_1 + 0 \cdot f_2.$$

Thus, $[T]_{B_1,B_2} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$

E1) Choose two other bases B'_1 and B'_2 of R^3 and R^2, respectively. (In Module 1 Unit 4, you came across a lot of bases of both these vector spaces). For T in the example above, give the matrix $[T]_{B'_1,B'_2}$.

What **E1** shows us is that the matrix of a transformation depends on the bases that we use for obtaining it. The next two exercises also being out the same fact.

E2) Write the matrix of the linear transformation $T: R^3 \to R^2 : T(x, y, z) = (x + 2y + 2z, 2x + 3y + 4z)$ with respect to the standard bases of R^3 and R^2.

E3) What is the matrix of T, in E5, with respect to the bases $B'_1 = \{(1,0,0),(0,1,0),(1,-2,1)\}$and $B'_2 = \{(1,2),(2,3)\}$?

The next exercise is about an operator that you have come across often

E4) Let V be the vector space of polynomials over R of degree < 3, in the variable t.

Let $D: V \to V$ be the differential operator given in Unit 2 (E3, when n = 3). Show that the matrix of D with respect to the basis $\{1, t, t^2, t^3\}$ is $\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix}$.

So far, given a linear transformation, we have obtained a matrix from it. This works the other way also. That is, given a matrix we can define a linear transformation corresponding to it.

Example 5: Describe $T: R^3 \to R^3$ such that $[T]_B = \begin{pmatrix} 1 & 2 & 4 \\ 2 & 3 & 1 \\ 3 & 1 & 2 \end{pmatrix}$

where B is the standard basis of R^3

Solution: Let $B = \{e_1, e_2, e_3\}$, Now, we are given that

$$T(e_1) = 1 \cdot e_1 + 2 \cdot e_2 + 3 \cdot e_3$$
$$T(e_2) = 2 \cdot e_1 + 3 \cdot e_2 + 1 \cdot e_3$$
$$T(e_3) = 4 \cdot e_1 + 1 \cdot e_2 + 2 \cdot e_3.$$

You know that any element of R^3 is $(x, y, z) = xe_1 + ye_2 + ze_3$

Therefore, $T(x, y, z) = T(xe_1 + ye_2 + ze_3)$

$xT(e_1) + yT(e_2) + zT(e_3)$, since T is linear

$$= x(e_1 + 2e_2 + 3e_3) + y(2e_1 + 3e_2 + e_3) + z(4e_1 + e_2 + 2e_3)$$
$$= (x + 2y + 4z)e_1 + (2x + 3y + z)e_2 + (3x + y + 2z)e_3$$
$$= (x + 2y + 4z, 2x + 3y + z, 3x + y + 2z)$$

$\therefore T: R^3 \to R^3$ is defined by

$$T(x, y, z) = (x + 2y + 4z, 2x + 3y + z, 3x + y + 2z)$$

Try the following exercises now.

E5) Describe $T: R^3 \to R^2$such that $[T]_{B_1,B_2} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}$ where B_1 and B_2 are the standard bases of $R^3$and R^2, respectively.
E6) Find the linear operator $T: \mathbb{C} \to \mathbb{C}$ whose matrix, with respect to the basis $\{1, i\}$ is $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$.
(Note that $\mathbb{C}$, the field of complex numbers, is a vector space over R, of dimension 3)
Now we are in a position to define the sum of matrices and multiplication of a matrix by a scalar.

1.3.4 Matrix Scalar Multiplication

In Unit 3 of Module 1, you have studied about the sum and scalar multiples of linear transformations. In the following theorem we will see what happens to the matrices associated with the linear transformations that are sums or scalar multiples of given linear transformations.

Theorem 1.1: Let U and V be vector spaces over F, of dimensions n and m, respectively. Let B_1 and B_2 be arbitrary bases of U and V, respectively.
(Let us abbreviate $[T]_{B_1,B_2}$ to $[T]$ during this theorem.) Let $S, T \in L(U, V)$ and $\alpha \in F$. Suppose $[S] = [a_{ij}], [T] = [b_{ij}]$, then $[S + T] = [a_{ij} + b_{ij}]$ and $[\alpha S] = [\alpha a_{ij}]$

Proof: Suppose $B_1 = \{e_1, e_2, \cdots, e_n\}$ and $B_2 = \{f_1, f_2. \cdots, f_m\}$. Then all the matrices to be considered here will be of size $m \times n$.
Now, by our hypothesis,
$S(e_j) = \sum_{i=1}^{m} a_{ij} f_i \;\; \forall\, j = 1, \cdots, n$ and $T(e_j) = \sum_{i=1}^{m} b_{ij} f_i \;\; \forall\, j = 1, \cdots, n$.
Therefore, $(S + T)(e_j) = S(e_j) + T(e_j)$ (by definition of S + T)

$$= \sum_{i=1}^{m} a_{ij} f_i + \sum_{i=1}^{m} b_{ij} f_i$$

$$= \sum_{i=1}^{m} (a_{ij} + b_{ij}) f_i$$

Thus, by definition of the matrix with respect to B_1 and B_2, we get $[S + T] = a_{ij} + b_{ij}$.
Given the $m \times n$ matrix A and a scalar α in set of complex number, the scalar multiple of A is an $m \times n$ matrix, written as αA and defined by
$[\alpha A]_{ij} = \alpha [A]_{ij}$; $1 \leq i \leq m, 1 \leq j \leq n$
Now, $(\alpha S)(e_j) = \alpha \left(S(e_j) \right)$ (by definition of αS)

$$= \alpha \sum_{i=1}^{m} a_{ij} f_i = \sum_{i=1}^{m} (\alpha a_{ij}) f_i$$

Thus, $[\alpha S] = [\alpha a_{ij}]$

From example above, if $\alpha = 5$ and $A = \begin{pmatrix} 1 & 0 & 5 \\ 2 & -3 & 4 \end{pmatrix}$, then

$$\alpha A = 5 \begin{pmatrix} 1 & 0 & 5 \\ 2 & -3 & 4 \end{pmatrix} = \begin{pmatrix} 5(1) & 5(0) & 5(5) \\ 5(2) & 5(-3) & 5(4) \end{pmatrix} = \begin{pmatrix} 5 & 0 & 25 \\ 10 & -15 & 20 \end{pmatrix}$$

1.3.5 Vector Space Properties of matrices

Now that matrix addition and scalar multiplication have been defined, a number of properties of each operation can be stated and proved, as well as a few properties that relate how they interact. This will ultimately lead us to prove that the set of all $m \times n$ matrices over F is a vector space over F.

Theorem 1.2: Vector Space Properties of Matrices

Suppose that M_{mn} is the set of all $m \times n$ matrices with addition and scalar multiplication as defined.

i. **Additive Closure of Matrices**
ii. If $A, B \in M_{mn}$, then$A + B \in M_{mn}$.
iii. **Scalar Closure of Matrices**
iv. If $\alpha \in \mathbb{C}$ and$A \in M_{mn}$, then$\alpha A \in M_{mn}$
v. **Commutativity of Matrices**
vi. If $A, B \in M_{mn}$, then $A + B = B + A$.
vii. **Additive Associativity of Matrices**
viii. If $A, B, C \in M_{mn}$, then $A + (B + C) = (A + B) + C$.
ix. **Zero Matrix**
x. There is a matrix, *0*, called the zero matrix, such that $A + 0 = A$ for all$A \in M_{mn}$
xi. **Additive Inverses of Matrices**
xii. If $A \in M_{mn}$, then there exists a matrix $(-A) \in M_{mn}$, so that $A + (-A) = 0$.
xiii. Scalar Multiplication Associativity of Matrices
xiv. If $\alpha, \beta \in ComplexNo.$ and $A \in M_{mn}$, then $\alpha(\beta A) = (\alpha\beta)A$
xv. **Distributivity across Matrix Addition of Matrices**
xvi. If $\alpha, \beta \in ComplexNo.$ and $A \in M_{mn}$, then $\alpha(A + B) = \alpha A + \alpha B$
xvii. **Distributivity across Scalar Addition of Matrices**
xviii. If $\alpha, \beta \in ComplexNo.$ and $A \in M_{mn}$, then $(\alpha + \beta)A = \alpha A + \beta A$.
xix. One Matrices
xx. If $A \in M_{mn}$, then $1A = A$

Proof:
While some of these properties seem very obvious, they all require proof. However, the proofs are not very interesting, and a bit tedious.
We will prove one version of distributivity very carefully, and you can test your proof-building skills on some of the others.
For any i, j; $1 \leq i \leq m, 1 \leq j \leq n$,
$[(\alpha + \beta)A]_{ij} = (\alpha + \beta)[A]_{ij}$ (Matrix Scalar Multiplication)
$= \alpha[A]_{ij} + \beta[A]_{ij}$ (Distributivity in Complex Numbers)
$= [\alpha A]_{ij} + [\beta A]_{ij}$ (Matrix Scalar Multiplication)
$= [\alpha A + \beta A]_{ij}$ (Matrix Addition)
These properties imply that $M_{mn}(F)$ is a vector space over F.
Now that we have shown that $M_{mn}(F)$ is a vector space over F, we know it must have a dimension.

1.3.6 Dimension of $M_{mn}(F)$ over F

The following theorem shall be proved to explain the dimension of $M_{mn}(F)$ over F. But, before you go further, check whether you remember the definition of a vector space isomorphism (Unit 3).

Theorem 1.3: Let U and V be vector spaces over F of dimensions n and m, respectively. Let B_1 and B_2 be a pair of bases of U and V, respectively. The mapping $\phi: L(U, V) \rightarrow M_{mn}(F)$, given by $\phi(T) = [T]_{B_1, B_2}$ is a vector space isomorphism.

Proof: The fact that ϕ is a linear transformation follows from Theorem 1.1.
We proceed to show that the map is also 1-1 and onto.
For the rest of the proof, we shall denote $[S]_{B_1, B_2}$by $[S]$ only, and take $B_1 = \{e_1, e_2, \cdots, e_n\}$ and $B_2 = \{f_1, f_2. \cdots, f_m\}$.
ϕ is 1-1: Suppose $S, T \in L(U, V)$ be such that $\phi(S) = \phi(T)$.
Then $[S] = [T]$.
Therefore, $S(e_j) + T(e_j) \; \forall \; e_j \in B_1$.
Thus, by Unit 3 (Theorem, 3.1), we have $S = T$.
ϕ is on 0: if $A \in M_{mn}(F)$ we want to construct $T \in L(U, V)$ such that $\phi(T) = A$.
Suppose $A = [a_{ij}]$. Let $v_1, \ldots, v_n \in V$ such that

$$v_1 = \sum_{i=1}^{m} a_{ij} f_i \; ; \; j = 1, \cdots, n$$

Then, by Theorem 3.3 of Unit 3, there exists a linear transformation $T \in L(U, V)$ such that

$$T(e_j) = v_1 = \sum_{i=1}^{m} a_{ij} f_i$$

Thus, by definition, $\phi(T) = A$
Therefore, ϕ is a vector space isomorphism.
A corollary to this theorem gives us the dimension of $M_{mn}(F)$.

Corollary: Dimension of $M_{mn}(F) = mn$.
Proof: Theorem 1.2 tells us the $M_{mn}(F)$ is isomorphic $to\ L(U,V)$.
Therefore, $dim_F M_{mn}(F) = dim_F(L(U,V)$ (by Theorem 12 of Unit 5) = mn, from Unit 6 (Theorem 1).
Why do you think we chose such a roundabout way for obtaining dim $M_{mn}(F)$?
We could as well have tried to obtain mn linearly independent m x n matrices and show that they generate $M_{mn}(F)$. But that would be quite tedious (see E16). Also, we have done so much work on $L(U,V)$ so why not use that! And, doesn't the way we have used seem neat?
Now, let's look at for some exercises related to Theorem 1.2.
E7) At most, how many matrices can there be in any linearly independent subject of $M_{2\times 3}(F)$?
E8) Are the matrices [1, 0] and [1, -1] linearly independent over R?
E9) Let E_{ij} be an $m \times n$ matrix whose $(i,j)th$ element is 1 and the other elements are 0. Show that $\{E_{ij}: \leq i \leq m;\ 1 \approx\leq j \leq h\}$ is a basis of $M_{mn}(F)$ over F.
Conclude that $dim_F M_{mn}(F) = mn$.
Now we move on to the next section, where we see some ways of getting new matrices from given ones.

1.3.7 New Matrices from Old

Transpose of a Matrix
Definition 1.3.5: Given a $m \times n$ matrix A, its **transpose** is the $n \times m$ matrix A^T given by

$$[A^T]_{ij} = [A]_{ji} \text{ ; for } 1 \leq i \leq m, 1 \leq j \leq n$$

Example 5:

Given a 3×4 matrix $G = \begin{bmatrix} 3 & 1 & 2 & -4 \\ 0 & -2 & 5 & 7 \\ -1 & 6 & 8 & 2 \end{bmatrix}$, the transpose of matrix G would be obtained by interchanging (rewriting or swapping) the columns by the rows, thus,

$$G^T = \begin{bmatrix} 3 & 0 & -1 \\ 1 & -2 & 6 \\ 2 & 5 & 8 \\ -4 & 7 & 2 \end{bmatrix}$$ which is a 4×3 matrix.

Definition 1.3.6: In the case where a matrix is equal to its transpose, that is, given a matrix A, then $A = A^T$, we have a **symmetric matrix**.
A square matrix B such that $B^T = -B$, is called **a skew–symmetric matrix**

Example 6: $A = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$, then $A^T = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$

$B = \begin{pmatrix} 0 & 2 \\ 2 & 0 \end{pmatrix}$, then $B^T = \begin{pmatrix} 0 & -2 \\ 2 & 0 \end{pmatrix} = -\begin{pmatrix} 0 & 2 \\ -2 & 0 \end{pmatrix} = -B$

Informally, a matrix is symmetric if we can "flip" it about the main diagonal (upper-left corner, running down to the lower-right corner) and have it look unchanged.

Example 7:
Consider any (2×2)matrix A. Calculate $A + A^T$and $A - A^T$. Which of them is symmetric and which is skew-symmetric?

Solution: $A = \begin{pmatrix} 3 & 4 \\ 2 & 1 \end{pmatrix}$, then $A^T = \begin{pmatrix} 3 & 2 \\ 4 & 1 \end{pmatrix}$

$A + A^T = \begin{pmatrix} 3+3 & 4+2 \\ 2+4 & 1+1 \end{pmatrix} = \begin{pmatrix} 6 & 6 \\ 6 & 2 \end{pmatrix}$ This is a (2×2) symmetric matrix

$A - A^T = \begin{pmatrix} 3-3 & 4-2 \\ 2-4 & 1-1 \end{pmatrix} = \begin{pmatrix} 0 & 2 \\ -2 & 0 \end{pmatrix}$ This is a (2×2) skew-symmetric matrix.

For a (3×3)matrix A

$A = \begin{pmatrix} 1 & 2 & 4 \\ 2 & 3 & 1 \\ 3 & 4 & 5 \end{pmatrix}$, then $A^T = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 4 & 1 & 5 \end{pmatrix}$

$$A + A^T = \begin{pmatrix} 1+1 & 2+2 & 4+3 \\ 2+2 & 3+3 & 1+4 \\ 3+4 & 4+1 & 5+5 \end{pmatrix} = \begin{pmatrix} 2 & 4 & 7 \\ 4 & 6 & 5 \\ 7 & 5 & 10 \end{pmatrix} (3 \times 3)$$

Symmetric matrix

$$A - A^T = \begin{pmatrix} 1-1 & 2-2 & 4-3 \\ 2-2 & 3-3 & 1-4 \\ 3-4 & 4-1 & 5-5 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & -3 \\ -1 & 3 & 0 \end{pmatrix} (3 \times 3)$$

Skew-symmetric matrix

NOTE: For a matrix to be **symmetric** it has to be a **square matrix,** that is, the number of rows must be equal to the number of columns.

Example 8:

If matrix $H = \begin{bmatrix} 6 & 3 & -7 & 8 \\ 3 & 1 & 4 & -2 \\ -7 & 4 & 0 & 1 \\ 8 & -2 & 1 & 5 \end{bmatrix} = H^T$, then H is symmetric

1.3.8 Theorems Involving Some Types of Matrices

Theorem 1.4: Suppose that *A* is a symmetric matrix. Then *A* is square.
Proof:
Suppose A is a $m \times n$matrix.
Because *A* is symmetric, we know by the definition symmetric matrix that $A = A^T$.
So, in particular, matrix equality requires that A and A^T must have the same size.
But the size of A^T is $n \times m$ and because A has *m* rows and A^T has *n* rows, we conclude that $m = n$, hence A must be square by the definition of a square matrix.

Theorem 1.5: Suppose that *A* and *B* are $\boldsymbol{m \times n}$ matrices, then$\boldsymbol{(A + B)^T = A^T + B^T}$.
Proof:
The statement to be proved is an equality of matrices, for $1 \le i \le m$, $1 \le j \le n$

$$[(A + B)^T]_{ij} = [A + B]_{ji} \quad \text{(Transpose of a matrix)}$$
$$= [A]_{ji} + [B]_{ji} \quad \text{(Matrix addition)}$$
$$= [A^T]_{ij} + [B^T]_{ij} \quad \text{(Transpose of a matrix)}$$
$$= [A^T + B^T]_{ij} \quad \text{(Matrix addition)}$$

Since the matrices $(A + B)^T$and $A^T + B^T$agree at each entry, then the definition of matrix equality tells us the two matrices are equal.

Theorem 1.6: Let $\alpha \in C$ and A is an $m \times n$ matrix, then $(\alpha A)^T = \alpha A^T$
Proof:
The statement to be proved is an equality of matrices, for $1 \le i \le m$, $1 \le i \le n$

$$[(\alpha A)^T]_{ji} = [\alpha A]_{ij} \quad \text{(Transpose of a matrix)}$$
$$= \alpha [A]_{ji} \quad \text{(Matrix Scalar Multiplication)}$$
$$= \alpha [A^T]_{ji} \quad \text{(Transpose of a matrix)}$$
$$= [\alpha A^T]_{ji} \quad \text{(Matrix Scalar Multiplication)}$$

Since the matrices $[\alpha A]^T$and αA^Tagree at each entry, definition of matrix equation is used which implies that the two matrices are equal.

Theorem 1.7: Suppose that A is a $m \times n$matrix. Then $(A^T)^T = A$
Proof:
For $1 \le i \le m$, $1 \le j \le n$

$$[(A^T)^T]_{ij} = [A^T]_{ji} \quad \text{(Transpose of a matrix)}$$
$$= [A]_{ij} \quad \text{(Transpose of a matrix)}$$

Since the matrices $(A^T)^T$and A agree at each entry, by the definition of matrix equality, the two matrices are equal.

Definition 1.3.7: Let A be $m \times n$ matrix. Then the **conjugate** of A, written as $\bar{A}$ is an $m \times n$ matrix defined by

$$\left[\overline{A}\right]_{ij} = \overline{[A]_{ij}}$$

If A is a matrix over the complex field C, then the matrix obtained by replacing each entry of A by its complex conjugate is called the conjugate of A, and is denoted by $\overline{A}$.

Three properties of conjugates, which are similar to those of the transpose, are

i. $\overline{A + B} = \overline{A} + \overline{B}$

ii. $\overline{\left(\overline{A}\right)} = A$

iii. $\overline{(A^T)} = \left(\overline{A}\right)^T$

Example 9:

If $A = \begin{bmatrix} 3 - i & 2 & 1 + 4i \\ -3 - 5i & 3 - 2i & i \end{bmatrix}$, then $\bar{A} = \begin{bmatrix} 3 + i & 2 & 1 - 4i \\ -3 + 5i & 3 + 2i & -i \end{bmatrix}$

Theorem 1.8: Suppose that A and B are $m \times n$matrices. Then$\overline{A + B} = \overline{A} + \overline{B}$

Proof:

For $1 \leq i \leq m, 1 \leq j \leq n$

$$\begin{aligned} \left[\overline{A + B}\right]_{ij} &= \overline{[A + B]_{ij}} && \text{(Complex conjugate of a matrix)} \\ &= \overline{[A]_{ij} + [B]_{ij}} && \text{(Matrix addition)} \\ &= \overline{[A]_{ij}} + \overline{[B]_{ij}} && \text{(Addition of complex conjugate)} \\ &= \left[\overline{A}\right]_{ij} + \left[\overline{B}\right]_{ij} && \text{(Complex conjugate of a matrix)} \\ &= \left[\overline{A} + \overline{B}\right]_{ij} && \text{(Matrix addition)} \end{aligned}$$

Example 10: Let $B = \begin{bmatrix} 1 - i & 2 - 3i & 4 + i \\ -3 & 2i & 2 - i \end{bmatrix}$, then $\overline{B} = \begin{bmatrix} 1 + i & 2 + 3i & 4 - i \\ -3 & -2i & 2 + i \end{bmatrix}$

Theorem 1.9: Conjugate of the Conjugate of a Matrix

Suppose that A is a $m \times n$ matrix, then$\overline{\left(\overline{A}\right)} = A$.

Example 11: From example 10 above,

If $\bar{A} = \begin{bmatrix} 3 + i & 2 & 1 - 4i \\ -3 + 5i & 3 + 2i & -i \end{bmatrix}$, then

$$\bar{\bar{A}} = \begin{bmatrix} 3 - i & 2 & 1 + 4i \\ -3 - 5i & 3 - 2i & i \end{bmatrix} = A$$

Theorem 1.10: Matrix Conjugation and Transposes

Suppose that *B* is an $m \times n$ matrix, then $\overline{(A^T)} = \left(\overline{A}\right)^T$.

Example 12:

From example, $A^T = \begin{bmatrix} 3-i & -3-5i \\ 2 & 3-2i \\ 1+4i & i \end{bmatrix}$, then

$$\overline{A^T} = \begin{bmatrix} 3+i & -3+5i \\ 2 & 3+2i \\ 1-4i & -i \end{bmatrix}$$

Also, from example, $\left(\overline{A}\right)^T = \begin{bmatrix} 3+i & -3+5i \\ 2 & 3+2i \\ 1-4i & -i \end{bmatrix}$

Hence, $\overline{(A^T)} = \left(\overline{A}\right)^T$

Definition 1.3.8: A square matrix *A* for which $\overline{A}^T = A$ is called a **Hermitian matrix.**

A square matrix A is called a **Skew-Hermitian matrix** if $\overline{A}^T = -A$.

For example, the matrix

Let $D = \begin{bmatrix} 1 & 1+i \\ 1-i & 2 \end{bmatrix}$, then $\overline{D} = \begin{bmatrix} 1 & 1-i \\ 1+i & 2 \end{bmatrix}$ and

$\overline{D}^T = \begin{bmatrix} 1 & 1+i \\ 1-i & 2 \end{bmatrix} = D$

Hence *D* is a Hermitian matrix

$F = \begin{bmatrix} i & 1+i \\ -1+i & 0 \end{bmatrix}$ then $\overline{F} = \begin{bmatrix} -i & 1-i \\ -1-i & 0 \end{bmatrix}$ and

$\overline{F}^T = \begin{bmatrix} -i & -1-i \\ 1-i & 0 \end{bmatrix} = -\begin{bmatrix} i & 1+i \\ -1+i & 0 \end{bmatrix} = -F$

Hence *F* is a Skew-Hermitian matrix.

Note: For a real matrix *A*, *A* is Hermitian if *A* is symmetric.

Similarly, *A* is skew-Hermitian if *A* is skew-symmetric.

Diagonal Matrix

Definition 1.3.9: Let *U* and *V* be vector spaces over *F* of dimension *n* such that $B_1 = \{e_1, \ldots, e_n\}$ and $B_2 = \{f_1, \ldots, f_n\}$ are based of *U* and *V*, respectively. Also let $d_1, \ldots, d_n \in F$.

Consider the transformation $T: U \to V / T(a_1e_1 + \ldots + a_ne_n) = a_1d_1f_1 + \ldots + a_nd_nf_n$ then $T(e_1) = d_1f_1$, $T(e_2) = d_2f_2, \ldots, T(e_n) = d_nf_n$, such a matrix is called a **diagonal matrix**.

Let us see what this means:

$$[T](B_1 \cdot B_2) = \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & d_n \end{pmatrix}$$

Let $A[a_{11}]$ be a square matrix, the entries $a_{11}, a_{12}, \ldots, a_{mn}$ are called the **diagonal entries** of A. This is because they lie along the diagonal, from left to right, of the matrix.
All the other entries of *A* are called the **off-diagonal entries** of *A*.
A square matrix whose off-diagonal entries are zero (i.e., $a_{ij} = 0$, $\forall\ i \neq j$) is called a **diagonal matrix**.
The diagonal matrix above is denoted by $\boldsymbol{diag(d_1, d_2, \ldots, d_n)}$.

Note: The $d_i's$ may or may not be zero.
If all the $d_i's$ are zero, the $(n \times n)$zero matrix is obtained, which corresponds to the zero operator.
If $d_i = 1,\ i = 1, \ldots, n$, the $(n \times n)$ identity matrix, I_n (or *I*, when the size is understood).
If $\alpha \in F$, the linear operator $\alpha I: R^n \to R^n$ such that $\alpha I(v) = \alpha v$ for all $v \in R^n$, is called a **Scalar operator.** Its matrix with respect to any basis is $\alpha I = diag(\alpha, \alpha, \ldots, \alpha)$. Such a matrix is called a **Scalar matrix** which is a diagonal matrix whose diagonal entries are all equal.

Triangular Matrix
Definition 1.3.10: Let $B = \{e_1, e_2, \ldots, e_n\}$ be a basis of a vector space *V*. Let $S \in L(V, V)$ be an operator such that

$$\begin{array}{lllllllll} S(e_1) & = & a_{11}e_1 & & & & & & \\ S(e_2) & = & a_{12}e_1 & + & a_{22}e_2 & & & & \\ \vdots & \vdots & \vdots & & & & & & \\ \vdots & \vdots & \vdots & & & & & & \\ S(e_n) & = & a_{1n}e_1 & + & a_{2n}e_2 & + & \cdots & + & a_{nn}e_n \end{array}$$

Then, the matrix of *S* with respect to *B* is

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ 0 & a_{22} & \ldots & \ldots & a_{2n} \\ \vdots & \vdots & \ldots & \ldots & \vdots \\ \vdots & \vdots & \ldots & \ldots & \vdots \\ 0 & 0 & 0 & \ldots & a_{nn} \end{pmatrix}$$

A square matrix *A* such that $a_{ij} = 0\ \forall i > j$ is called an ***Upper Triangular Matrix.***

Example 13:
$\begin{pmatrix} 1 & 3 \\ 0 & 2 \end{pmatrix}$; $\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$; $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ are all upper triangular, while $\begin{pmatrix} 0 & 3 \\ 0 & 0 \end{pmatrix}$ is **strictly upper triangular**.

Note that every strictly upper triangular matrix is an upper triangular matrix.

Definition 1.3.11: Let $T: V \to V$ be an operator such that $T(e_j)$ is a linear combination of $e_j, e_{j+1}, \ldots, e_n \forall\ j$.

The matrix of T with respect to B is $\begin{pmatrix} b_{11} & 0 & 0 & \cdots & 0 \\ b_{21} & b_{22} & \ldots & \ldots & 0 \\ \vdots & \vdots & \ldots & \ldots & \vdots \\ \vdots & \vdots & \ldots & \ldots & \vdots \\ b_{n1} & b_{n2} & b_{n3} & \ldots & b_{nn} \end{pmatrix}$; $b_{ij} = 0 \ \forall i < j$, such a matrix is called a **Lower Triangular Matrix.**

If $b_{ij} = 0 \ \forall i \leq j$, then B is said to be a **Strictly Lower Triangular Matrix**.

The matrix $\begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 5 & 0 & 0 & 0 & 0 \\ 3 & -1 & 0 & 0 & 0 \\ -2 & 4 & 2 & 0 & 0 \\ 1 & 0 & 3 & -1 & 0 \end{pmatrix}$ is a strictly lower triangular matrix.

Of course, it is also lower triangular!

Remark: Given (3×3) upper triangular matrix, say $G = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{pmatrix}$, its transpose is

$G^T = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 4 & 0 \\ 3 & 5 & 6 \end{pmatrix}$ is a lower triangular matrix

In fact, for any $(n \times n)$upper triangular matrix A, its transpose is a lower triangular matrix, and vice versa.

Class work:

1. If an upper triangular matrix A is symmetric, then show that it must be a diagonal matrix.
2. Show that the diagonal entries of a skew-symmetric matrix are all zero, but the converse is not true.

1.3.9 Matrix Multiplication

We have already discussed scalar multiplication. Now we see how to multiply two matrices. Again, the motivation for this operation comes from linear transformations.

1.3.9.1 Matrix of the Composition of Linear Transformations

Let U, V and W be vector spaces over F, of dimension p, n and m, respectively. Let B_1, B_2, B_3 be bases of these respective spaces. Let $T \in L(U, V)$ and $S \in L(V, W)$. Then $(ST) \in L(U, W)$

Suppose $[T]B_1 \cdot B_2 = B = \left[b_{jk}\right]_{n\times p}$ and $[S]B_2 \cdot B_3 = A = \left[a_{ij}\right]_{m\times n}$
What is then the matrix $[ST]B_1 \cdot B_3$?
To answer this, let $B_1 = \{e_1, e_2, \ldots, e_n\}$; $B_2 = \{f_1, f_2, \ldots, f_n\}$; $B_3 = \{g_1, g_2, \ldots, g_n\}$
Since $T(e_k) = \sum_{j=1}^{n} b_{jk} f_j$; for all $k = 1,2,\ldots,p$ and
$S(f_j) = \sum_{i=1}^{m} a_{jk} g_j$; for all $j = 1,2,\ldots,n$
$\therefore S \circ T(e_k) = S(T(e_k)) = S\sum_{j=1}^{n} b_{jk} f_j = b_{1k}S(f_1) + b_{2k}S(f_2) + \cdots + b_{nk}S(f_n)$

$$= b_{1k}\sum_{i=1}^{m} a_{11}g_i + b_{2k}\sum_{i=1}^{m} a_{12}g_i + \cdots + b_{nk}\sum_{i=1}^{m} a_{1n}g_i$$

$= \sum_{i=1}^{m}(a_{i1}b_{1k} + a_{i2}b_{2k} + \ldots + a_{in}b_{nk})g_i$ (Collection of the coefficients of g_i)
Thus, $[ST]B_1 \cdot B_3 = [c_{ik}]_{m\times p}$; where $c_{ik} = \sum_{i=1}^{m} a_{ij}b_{jk}$
Matrix $[c_{ik}]$ is defined as the product AB
Let $A = [a_{ik}]_{m\times n}$, $B = [b_{ik}]_{m\times n}$ be two matrices over F, of sizes $m \times n$ and $n \times p$, respectively.
We define AB to be the $m \times p$ matrix C whose $(i,k)^{th}$ entry is

$$c_{ik} = \sum_{\substack{i=1\\ j=1}}^{n} a_{ij}b_{jk} = a_{i1}b_{1k} + a_{i2}b_{2k} + \ldots + a_{in}b_{nk}$$

In order to obtain the $(i,k)^{th}$ element of AB, take the i^{th} row of matrix A and the k^{th} column of B which are both *n-tuples*.
Multiply their corresponding elements and add up all these products.
For example, if the 2nd row of $A = [1 \quad 2 \quad 3]$ and the 3rd column of $B = \begin{bmatrix}1\\5\\6\end{bmatrix}$, then the $(2{,}3)^{th}$ entry of $AB = 1 \times 4 + 2 \times 5 + 3 \times 6 = 32$
Note that two matrices A and B can only be multiplied if the number of columns of A equals the number of rows of B.
The following illustration may help in explaining what we do to obtain the product of two matrices:

$$AB = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1j} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2j} & \ldots & a_{2n} \\ \vdots & \vdots & \ldots & \vdots & \ldots & \vdots \\ \vdots & \vdots & \ldots & \vdots & \ldots & \vdots \\ \vdots & \vdots & \ldots & \vdots & \ldots & \vdots \\ a_{m1} & a_{m2} & \ldots & \ldots & \ldots & a_{mn} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} & \ldots & b_{1k} & \ldots & b_{1p} \\ b_{21} & b_{22} & \ldots & b_{2k} & \ldots & b_{2p} \\ \vdots & \vdots & \ldots & \vdots & \ldots & \vdots \\ \vdots & \vdots & \ldots & \vdots & \ldots & \vdots \\ \vdots & \vdots & \ldots & \vdots & \ldots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nk} & \cdots & b_{np} \end{pmatrix}$$

$$= \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1k} & \cdots & c_{1p} \\ c_{21} & c_{22} & \cdots & c_{2k} & \cdots & c_{2p} \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ \vdots & \vdots & \cdots & c_{ik} & \cdots & c_{ip} \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ c_{m1} & c_{m2} & \cdots & c_{mk} & \cdots & c_{mp} \end{pmatrix}$$

where $c_{ik} = \sum a_{ij} b_{jk}$

Note: This is a very new kind of operation so take your time in trying to understand it.
To get you used to matrix multiplication we consider the product of a row and a column matrix.

Let $A = [a_1 \quad a_2 \quad \cdots \quad a_n]$ be $(1 \times n)$ matrix and $B = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$ be $(n \times 1)$ matrix.
The product $AB = [a_1 b_1 + a_2 b_2 + \cdots \ldots + a_n b_n]$ which is a (1×1)matrix

Example 14: Let $A = [-1 \quad 2 \quad 0 \quad 3]$ and $B = \begin{bmatrix} 3 \\ 1 \\ 4 \\ -2 \end{bmatrix}$.

$$AB = [(-1)(3) + (2)(1) + (0)(4) + (3)(-2)] = (-3 + 2 + 0 - 6) = (-7)$$

Example 15: Let $C = \begin{bmatrix} 1 & -5 & 2 \\ -2 & 6 & 1 \\ 3 & 0 & -4 \end{bmatrix}$ and $D = \begin{bmatrix} 2 & -3 \\ -4 & 0 \\ 3 & 1 \end{bmatrix}$.

$$CD = \begin{bmatrix} 1 \cdot 2 + (-5)(-4) + 2.3 & 1(-3) + (-5)0 + 2.1 \\ (-2) \cdot 2 + 6(-4) + 1.3 & (-2)(-3) + 6.0 + 1.1 \\ 3.2 + 0(-4) + (-4)3 & 3(-3) + 0.0 + 1(-4)1 \end{bmatrix}$$
$$= \begin{bmatrix} 28 & -1 \\ -25 & 7 \\ -6 & -13 \end{bmatrix}$$

Notice that *DC* is not defined because the number of columns of $D = 2 \neq$ number of rows of $C = 3$.
Note that; if *CD* is defined, DC may not necessarily be defined.
In fact, even if *CD* and *DC* are both defined it is possible that $CD \neq DC$.
Consider the following example:

$C = \begin{bmatrix} 3 & 2 & 4 \\ 1 & -1 & 0 \end{bmatrix}$ and $D = \begin{bmatrix} 2 & 3 \\ 1 & 0 \\ -1 & 1 \end{bmatrix}$

$$CD = \begin{bmatrix} 3\cdot 2 + 2\cdot 1 + 4(-1) & 3\cdot 3 + 2\cdot 0 + 4\cdot 1 \\ 1\cdot 2 + (-1)1 + 0(-1) & 1\cdot 3 + (-1)0 + 0\cdot 1 \end{bmatrix}$$
$$= \begin{bmatrix} 4 & 13 \\ 1 & 3 \end{bmatrix}$$
$$DC = \begin{bmatrix} 2\cdot 3 + 3\cdot 1 & 2\cdot 2 + 3(-1) & 2\cdot 4 + 3\cdot 0 \\ 1\cdot 3 + 0\cdot 1 & 1\cdot 2 + 0(-1) & 1\cdot 4 + 0\cdot 0 \\ (-1)3 + 1\cdot 1 & (-1)2 + 1(-1) & (-1)4 + 1\cdot 0 \end{bmatrix}$$
$$= \begin{bmatrix} 9 & 1 & 8 \\ 3 & 2 & 4 \\ -2 & -3 & -4 \end{bmatrix}$$

From example above, CD is a 2x2 matrix while DC is a 3x3 matrix.
CD and DC are defined but of different dimensions, thus $CD \neq DC$.
Another point of difference between multiplication of numbers and matrix multiplication is that $A \neq 0$, $B \neq 0$ but AB can be zero.

Example 16: If $A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$, $B = \begin{pmatrix} 1 & 0 \\ -1 & 0 \end{pmatrix}$

$$AB = \begin{pmatrix} 1\cdot 1 + 1(-1) & 1\cdot 0 + 0\cdot 0 \\ 1\cdot 1 + 1(-1) & 1\cdot 0 + 1\cdot 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

So, you see that the product of two non-zero matrices can be zero.

Let's check if *BA* is equal to *AB*:

$$BA = \begin{pmatrix} 1\cdot 1 + 0\cdot 1 & 1\cdot 1 + 0\cdot 1 \\ (-1)1 + 0\cdot 1 & (-1)0 + 1\cdot 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$$

This shows that *AB* is necessarily not equal to *BA*;

Hence matrix multiplication is not commutative.
The following exercises will give you some practice in matrix multiplication.

Exercise:

1) If $C = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix}$, $D = \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 0 & 0 \end{pmatrix}$. Find (i) $C + D$ (ii) CD (iii) DC (iv) Is $CD = DC$?

2) Let$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, $B = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$. Calculate i) $(A + B)^2$ ii) $A^2 + 2AB + B^2$ iii) Is i) = ii)
Hint:$A^2 = A \cdot A$

1.3.9.2 Properties of a Matrix Product

We will now state 5 properties concerning matrix multiplication. (Their proofs could get a little technical, and we prefer not to give them here).

i. **Associative Law:** If A, B, C are $m \times n$, $n \times p$and $p \times q$matrices, respectively, over F, then $(AB)C = A(BC)$, i.e., matrix multiplication is associative.
ii. **Distributive Law:** If A is an $m \times n$matrix and B, C are $n \times p$matrices, then
iii. $A(B + C) = AB + AC$.
iv. Similarly, if, A and B are $m \times n$matrices, and C is an $n \times p$matrix, then
v. $(A + B)C = AC + BC$.
vi. **Multiplicative identity:** Given the identity matrix I_n, which acts as the multiplicative identity for matrix multiplication. We have $AI_n = A$, $AI_m = A$, for every $m \times n$ matrix A.
vii. **If** $a \in F$, and A, B are $m \times n$ and $n \times p$ matrices over F, respectively then
viii. $\alpha(AB) = (\alpha A)\, B = A\,(\alpha B)$
ix. **If** A, B are $m \times n$, $n \times p$matrices over F, respectively, then $(AB)^T = B^T A^T$.

(This says that the operation of taking the transpose of a matrix is anti-commutative).

These properties can help you in solving the following examples.

Example 17:

a. For $A = \begin{pmatrix} 3 & -4 \\ -1 & 0 \\ 2 & -2 \end{pmatrix}$ and $B = \begin{pmatrix} -1 & 3 & -6 \\ 2 & 4 & 0 \end{pmatrix}$. Show that $2(AB) = (2A)B$

b. Show that $(A + B)^T = A^2 + AB + BA + B^2$ for any two $n \times n$matrices A and B.

So**lution:**

a. $$2AB = 2\begin{pmatrix} 3 & -4 \\ -1 & 0 \\ 2 & -2 \end{pmatrix}\begin{pmatrix} -1 & 3 & -6 \\ 2 & 4 & 0 \end{pmatrix}$$

$$= 2\begin{pmatrix} 3(-1) + (-4)2 & 3(3) + (-4)4 & 3(-6) + (-4)0 \\ (-1)(-1) + 0 \cdot 2 & (-1)3 + 0 \cdot 4 & (-1)(-6) + 0 \cdot 0 \\ 2(-1) + (-2) \cdot 2 & 2 \cdot 3 + (-2) \cdot 4 & 2(-6) + (-2) \cdot 0 \end{pmatrix}$$

$$= 2\begin{bmatrix} -11 & -7 & -18 \\ 1 & -3 & 6 \\ -6 & -2 & -12 \end{bmatrix}$$

$$= \begin{bmatrix} -22 & -14 & -36 \\ 2 & -6 & 12 \\ -12 & -4 & -24 \end{bmatrix}$$

Also, $(2A)B = \begin{pmatrix} 6 & -8 \\ -2 & 0 \\ 4 & -4 \end{pmatrix}\begin{pmatrix} -1 & 3 & -6 \\ 2 & 4 & 0 \end{pmatrix}$

$$= \begin{pmatrix} 6(-1)+(-8)2 & 6(3)+(-8)4 & 6(-6)+(-8)0 \\ (-2)(-1)+0\cdot 2 & (-2)3+0\cdot 4 & (-2)(-6)+0\cdot 0 \\ 4(-1)+(-4)\cdot 2 & 4\cdot 3+(-4)\cdot 4 & 4(-6)+(-4)\cdot 0 \end{pmatrix}$$

$$= \begin{bmatrix} -6-16 & 18-32 & -36-0 \\ 2+0 & -6+0 & 12+0 \\ -4-8 & 12-16 & -24-0 \end{bmatrix}$$

$$= \begin{bmatrix} -22 & -14 & -36 \\ 2 & -6 & 12 \\ -12 & -4 & -24 \end{bmatrix}$$

Hence $2(AB) = (2A)B$

Now; attempt (b) part of the example

Now we shall go on to introduce you to the concept of an invertible matrix.

Solutions /Answers to Exercises

E1) Suppose $B'_1 = \{(1,0,1),(0,2,-1)\}$ and $B'_2 = \{0,1),(1,0)\}$

Then $T(1,0,1) = (1) = 0\cdot(0,1) + 1\cdot(1,0)$

$T(0,2,-1) = (0,2) = 2\cdot(0,1) + 0\cdot(1,0)$

$T(1,0,0) = (1,0) = 0\cdot(0,1) + 1\cdot(1,0)$

$[T]_{B'_1B'_2} = \begin{pmatrix} 0 & 2 & 0 \\ 1 & 0 & 1 \end{pmatrix}$

E2) $B_1 = \{e_1, e_2, e_3\}$ and $B_2 = \{f_1, f_2\}$ are the standard bases (given in Example 3).

$T(e_1) = T(1,0,0) = (1,2) = f_1 + 2f_2$

$T(e_2) = T(0,1,0) = (2,3) = 2f_1 + 3f_2$

$T(e_3) = T(0,0,1) = (2,4) = 2f_1 + 4f_2$

$[T]_{B_1B_2} = \begin{pmatrix} 1 & 2 & 2 \\ 2 & 3 & 4 \end{pmatrix}$

E3) $T(1,0,0) = (1,2) = 1\cdot(1,2) = 0\cdot(2,3)$

$T(0,1,0) = (2,3) = 0\cdot(1,2) + 1\cdot(2,3)$

$T(1,-2,1) = (-1,0) = 3\cdot(1,2) - 2\cdot(2,3)$

$[T]_{B'_1B'_2} = \begin{pmatrix} 1 & 0 & 3 \\ 0 & 1 & -2 \end{pmatrix}$

E4) Let $B = \{1, t, t^2, t^3\}$.

Then, $D(1) = 0 = 0\cdot 1 + 0\cdot t + 0\cdot t^2 + 0\cdot t^3$

$D(t) = 1 = 1\cdot 1 + 0\cdot t + 0\cdot t^2 + 0\cdot t^3$

$D(t^2) = 2t = 0\cdot 1 + 2\cdot t + 0\cdot t^2 + 0\cdot t^3$

$D(t^3) = 3t^2 = 0\cdot 1 + 2\cdot t + 3\cdot t^2 + 0\cdot t^3$

Therefore, $[D]_B$ is the given matrix.

E5) We know that

$$T(e_1) = f_1$$
$$T(e_2) = f_1 + f_2$$
$$T(e_3) = f_2$$

Therefore, for any $(x, y, z) \in R^3$.

$$T(x, y, z) = T(xe_1 + ye_2 + ze_3) = xT(e_1) + yT(e_2) + zT(e_3)$$
$$= xf_1 + y(f_1 + f_2) + zf_2 = (x + y)f_1 + (y + z)f_2$$

That is, $R: R^3 \to R^2: T(x, y, z) = (x + y, y + z)$

E6) We are given that

$$T(1) = 0 \cdot 1 + 1 \cdot i = i$$
$$T(i) = (-1) \cdot 1 + 0 \cdot 1 = -1$$

Therefore, for any $a + ib \in C$, we have

$$T(a + ib) = aT(1) + bT(i) = ai - b$$

E7) Since $dim\ M_{2\times3}(R)$ is 6, any linearly independent subset can have 6 elements, at most.

E8) Let $\alpha, \beta \in R$ such that $\alpha[1,0] + \beta[1, -1] = [0,0]$ Then $\alpha + \beta, -\beta = [0, 0]$.

Thus, $\alpha = 0, \beta = 0$.

Therefore, the matrices are linearly independent.

SELF-ASSESSMENT EXERCISE(S)

a. Show that $(A + B)^T = A^2 + AB + BA + B^2$ for any two $n \times n$matrices A and B.

b. Let $A = \begin{pmatrix} 2 & -1 & 0 \\ 1 & 0 & -3 \\ 0 & 0 & 0 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & -4 & 0 \\ 2 & 0 & 3 \\ 4 & 0 & -2 \end{pmatrix}$. Find $(AB)^T$ and $B^T A^T$, are they equal?

c. Let *A*, *B* be two symmetric *n* x *n* matrices over *F*. Show that *AB* is symmetric if and only if *AB* = *BA*.

d. Let *A*, *B* be two diagonal *n* x *n* matrices over *F*. Show that *AB* is also a diagonal matrix.

e. Is the matrix $\begin{bmatrix} 1 & 0 \\ 2 & -1 \end{bmatrix}$ invertible? If so, find its inverse

Conclusion

Matrix theory has been seen to occupy a very important position in pure and applied mathematics as well as in economics where the use of input-output matrix for solving some national economy problems.

Matrices have been seen to be intimately connected with linear transformations.

Algebraic operations on matrices were derived from the corresponding operations on linear transformations. We will also discuss. One type, a triangular matrix, which is one of the special types of matrices, was considered.

1.4 Summary

We briefly sum up what has been done in this unit.

1) The term "matrices" have been defined and described and the method of associating matrices with linear transformations was explained.
2) Sums of matrices and multiplication of matrices by scalars have been showed with numerical examples.
3) The proof of $M_{m\times n}(F)$ as a vector space of dimensionmn over F was explained
4) the transpose of a matrix, the conjugate of a complex matrix, the conjugate transpose of a complex matrix, a diagonal matrix, identity matrix, scalar matrix and lower and upper triangular matrices have been defined.
5) Multiplication of matrices was defined and its connection with the composition of linear transformations has been shown. Some properties of the matrix product were also listed and applied.

1.5 References/Further Readings

Robert A. Beezer (2014). A First Course in Linear Algebra Congruent Press, Gig Harbor, Washington, USA 3(40).

Arbind K Lal and Sukant Pati (2018). Linear Algebra through Matrices.

Peter J. Cameron (2008). Notes on Linear Algebra.

K.A. Stroud and Dexter J. Booth (2007). Engineering Mathematics 6th Edition. Palgrave Macmillan.

Abolarin, O.E. and Ishola, Y.C. (2005). Business Mathematics II. Jube Evans Books and Publications

UNIT 2 MATRICES II

Unit Structure

2.1 Introduction

The last unit has introduced us to the basic descriptions and definitions of a matrix and various types of matrices, as well as discussing the method of associating matrices with linear transformations along with some theorems.

In this unit, invertible matrices shall be defined, discussed and determined. The meaning of the matrix of a change of basis shall be highlighted and shown to be invertible. Theorems relating invertible matrices with change of bases shall be stated and proved.

2.2 Learning Outcomes

- Determine if a given matrix is invertible;
- Obtain the inverse of a matrix;
- Discuss the effect that the change of basis has on the matrix of a linear transformation.
- State the invertible theorem
- State and prove the conditions for a matrix to be invertible.

2.3 Invertible Matrices

2.3.1 Inverse of a Matrix

Just as we defined the operations on matrices by considering them on linear operators first, we give a definition of invertibility for matrices based on considerations of invertibility of linear operators.

It may help you to recall what we mean by an invertible linear transformation.

A linear transformation $T: U \to V$ is invertible if
(a) T is 1 – 1 and onto, or, equivalently,
(b) There exists a linear transformation $S: V \to U$ such that $S \circ T = I_u, T \circ S = I_v$.
In particular, $T \in (V, V)$ is said to be invertible if there exist $S \in L(V, V)$ such that $ST = TS = I.$
We have the following theorem involving the matrix of an invertible linear operator.

Theorem 2.1: Let V be an *n*-dimensional vector space over a field F, and B be a basis of V.
Let $T \in (V, V)$,T is invertible if there exists $A \in M_n(F)$ such that $[T]_B A = I_n$.

Proof: Suppose T is invertible, then there exist $S \in L(V, V)$ such that $ST = TS = I$.
Then, by a Theorem, $[TS]_B = [ST]_B = I$, that is, $[T]_B[S]_B = [S]_B[T]_B = I$
Take $A = [S]_B$, then $[T]_B A = I = A[T]_B$
Conversely, suppose there exist a matrix A such that $[T]_B A = A[T]_B = I$.
Let $S \in L(V, V)$ be such that $[S]_B = A$. (S exists because of Theorem 2), then

$$[T]_B[S]_B = [S]_B[T]_B = I = [I]_B$$

Thus, $[TS]_B[ST]_B = [I]_B$
So, by Theorem, $ST = ST = I$. that is, T is invertible.
Theorem 1 motivates us to give the following definition.

Definition 1: A matrix $A \in M_n(F)$ is said to be **invertible** if there exists $B \in M_n(F)$ such that $AB = BA = I_n$.

N.B: Only a square matrix can be invertible.
I_n is an example of an invertible matrix, since $I_n \cdot I_n = I_n$
On the other hand, the $n \times n$ zero matrix **0** is not invertible, since $0A = 0 \neq I_n$, for any A.
Note that Theorem 1says that T is invertible iff $[T]_B$is invertible.
We give another example of an invertible matrix now:

Example 1: Is $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ invertible?

Solution: Suppose $A$is invertible, then there exists$B = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ such that $AB = I = BA$

Now$AB = I \Rightarrow \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a+c & b+d \\ c & d \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

$\Rightarrow a + c = 1;\ c = 0, d = 1, b + d = 0$

$\Rightarrow a = 1, b = -1$ and $d = 1$

$\therefore B = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$.

Let us check $BA = I$

$$BA = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1+0 & 1-1 \\ 0+0 & 0+1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I$$

Therefore, A is invertible.

We now show that if an inverse of a matrix exists, it must be unique.

Theorem 2.2: Suppose $A \in M_n(F)$ is invertible. There exists a unique matrix $B \in M_n(F)$ such that $AB = BA = I$.

Proof: Suppose $B, C \in M_n(F)$ are two matrices such that $AB = BA = I$, and $AC = CA = I$, then $B = BI = B(AC) = (BA)C = IC = C$.
Because of Theorem1 we can make the following definition.

Definition 2: Let Abe an invertible matrix. The unique matrix B such that $AB = BA$ is I is called the inverse of A and is denoted byA^{-1}.
Let us take an example.

Example 2: Calculate the product AB, where $A = \begin{bmatrix} 1 & a \\ 0 & 1 \end{bmatrix}$, $B = \begin{bmatrix} 1 & b \\ 0 & 1 \end{bmatrix}$.
Calculate A^{-1}

Solution:

$$AB = \begin{bmatrix} 1 & a \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & b \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & a+b \\ 0 & 1 \end{bmatrix}$$

Now, how can we use this to obtain A^{-1} ?
Well, if $AB = I$, then $a + b = 0$.
So, if we take $B = \begin{bmatrix} 1 & -a \\ 0 & 1 \end{bmatrix}$, we would have $AB = BA = I$.

Thus $A^{-1} = \begin{bmatrix} 1 & -a \\ 0 & 1 \end{bmatrix}$

Next is a few observations about the matrix inverse, in the form of a theorem.

Theorem 2.3:

a) If A is invertible, then

(i) A^{-1} is invertible and $(A^{-1})^{-1} = A$,

(ii) A^T is invertible and $(A^T)^{-1} = (A^{-1})^T$.

b) If $A, B \in M_n(F)$ are invertible, then AB is invertible and $(AB)^{-1} = B^{-1}A^{-1}$

Proof:

(a) By definition

$AA^{-1} = A^{-1}A = I$ ………………………. (1)

(i) Equation (1) shows that A^{-1} is invertible and $(A^{-1})^T = A$

(ii) If we take transposes in Equation (1) and use the property that $(AB)^T = B^T A^T$,

we obtain

$$(A^{-1})^T A^T = A^T (A^{-1})^T = I^T = I$$

So A^T is invertible and $(A^T)^{-1} = (A^{-1})^T$

(b) To prove this, we will use the associativity of matrix multiplication.

Now $(AB)(B^{-1}A^{-1}) = [A(BB^{-1})]A^{-1} = AA^{-1} = I$

$$(B^{-1}A^{-1})(AB) = B^{-1}[(A^{-1}A)B] = BB^{-1} = I$$

So,AB is invertible and$(AB)^{-1} = B^{-1}A^{-1}$

We now relate matrix invertibility with the linear independence of its rows or columns.

When we say that the *m* rows of $A = [a_{ij}] \in M_{m\times n}(F)$ are linearly independent, what do we mean?

Let $R_1, \dots, R_m$ be the m row vectors $[a_{11,} a_{12}, \dots, a_{1n}]$, $[a_{21,} a_{22}, \dots, a_{21n}], \dots, [a_{m1,} a_{m2}, \dots, a_{mn}]$ respectively. We say that they are linearly independent if, whenever $\exists\ a_{1,} a_2, \dots, a_m \in F$ such that $a_1R_1 + a_2R_2 + \cdots + a_mR_m = 0$, then, $a_{1,} = 0, \dots\dots, a_m = 0$.

Similarly, the *n* columns $C_1, C_2, \dots, C_n$ of A are linearly independent if $b_1C_1 + \cdots + b_nC_n = 0,\ \Rightarrow b_1 = 0, b_2 = 0, \dots, b_n = 0$ where $b_1, \dots, b_n \in F$.

Thus, the theorem follows:

Theorem 2.4: Let $A \in M_n(F)$, then the following conditions are equivalent

(a) A is invertible

(b) The columns of A are linearly independent.

(c) The rows of A are linearly independent.

Proof: We first prove $(a) \Leftrightarrow (b)$, using Theorem,
Let V be an n-dimensional vector space over F and $B = \{e_1, e_2, \ldots, e_n\}$ be a basis of V.
Let $T \in L(V, V)$ be such that $(T)_B = A$, then A is invertible iff T is invertible iff $T(e_1), T(e_2), \ldots, T(e_n)$ are linearly independent (see theorem under linear independence).

Now we define the map $\theta: V \to M_{n+1}(F): (a_1 e_1 + \ldots + a_n e_n) = \begin{pmatrix} a_1 \\ \vdots \\ \vdots \\ a_n \end{pmatrix}$.

Let $C_{1,} C_2, \ldots, C_n$ be the columns of A. Then $\theta\big(T(e_1)\big)C_i$ for all $i = 1, \ldots, n$.
Since θ is an isomorphism, $T(e_1), T(e_2), \ldots, T(e_n)$ are linearly independent iff $C_{1,} C_2, \ldots, C_n$ are linearly independent.
Thus, A is invertible iff $C_{1,} C_2, \ldots, C_n$ are linearly independent.
Thus, we have proved $(a) \Leftrightarrow (b)$.
Now, the equivalence of (a) and (c) follows because A is invertible iff A^T is invertible iff the columns of A^T are linearly independent (as we have just shown) iff the rows of A are linearly independent (since the columns of A^T are the rows of A).
So, we have shown that $(a) \Leftrightarrow (c)$.
Thus, the theorem is proved.

Example 3: Let $A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \in M_3(R)$. Determine whether or not A is invertible.

Solution: Let $R_{1,} R_2, R_3$ be the rows of A. We will show that they are linearly independent.
Suppose $xR_1 + yR_2 + zR_3 = 0$, where $x, y, z \in R$. Then,
$x(1,0,1) + y(0,1,1) + z(1,1,1) = (0,0,0)$.
This gives us the following equations.

$$x + y = 0$$
$$y + z = 0$$
$$x + y + z = 0$$

On solving these we have $x = 0,\ y = 0, z = 0$.
Thus, by Theorem 2.2, A is invertible.

Exercise: Check if $B = \begin{pmatrix} 2 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 3 & 0 \end{pmatrix} \in M_3(Q)$.

We will now see how we associate a matrix to a change of basis.
This association will be made use of very often in the next unit.

2.3.2 Matrix of Change of Basis

Definition 3: Let V be an n-dimensional vector space over the field F. Let $B = \{e_1, e_2, \ldots, e_n\}$ and $B' = \{e'_1, e'_2, \ldots, e'_n\}$ be two bases of V. The **transition matrix** M from B to B' is the $n \times n$ matrix whose jth column is the coordinate representation $[e''_j]_B$ of the jth vector of B' relative to B or simply called the matrix of the change of basis from B to B' denoted by $M_{B,B'}$.

Proposition: Let B and B' be bases for the n-dimensional vector space V over the field K. Then, for any vector $e \in V$, the coordinate representations of e with respect to B *and* B' are related by $[e]_B = M[e]_{B'}$

Proof:
Let a_{ij} be the i,j entry of the matrix M. By definition, we have $e'_j = \sum_{i=1}^{n} a_{ij} e_i$
Take an arbitrary vector $e \in V$ and let
$[e]_B = [c_1, \ldots, c_n]^T$, $\qquad [e]_{B'} = [d_1, \ldots, d_n]^T$
This means, by definition, that

$$e = \sum_{i=1}^{n} c_i e_i = \sum d_j e'_j$$

Substituting the formula for e'_j into the second equation, we have

$$e = \sum_{j=1}^{n} d_j \left(\sum_{i=1}^{n} a_{ij} e_i \right)$$

Reversing the order of summation, we get

$$e = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} a_{ij} d_j \right) e_i$$

Now we have two expressions for e as a linear combination of the vectors e_i.
By the uniqueness of the coordinate representation, they are the same, that is,

$$c_i = \sum_{j=1}^{n} a_{ij} d_j$$

In matrix form, this says $\begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} = M \begin{bmatrix} d_1 \\ \vdots \\ d_n \end{bmatrix}$ or $[e]_B = M[e]_{B'}$ as required.
Note that M is the matrix of the transformation $T \in L(V, V)$ such that $T = e_j^{\prime\prime}$ with respect to the basis B. Since $\{e'_1, e'_2, \ldots, e'_n\}$ is a basis of V, from Unit 3we see that T is 1 – 1 and onto. Thus, T is invertible. So M is invertible.

Thus, the matrix of the change of basis from B to B' is invertible

Note: a) $M_{B,B'} = I_n$, this is because, in this case $e_j' = e_j \forall j = 1, \ldots, n$.

b) $M_{B,B'} = I(e_j') = [I]B', B$, this is because, in this case $e_j' = \sum_{i=1}^{n} a_{ij} e_i \; \forall j = 1, \ldots, n$.

Now suppose A is any invertible matrix. By Theorem 2.2, $\exists \;\; T \in L(V, V)$ such that

$[T]_R = A$.

Since A is invertible, T is invertible. Thus, T is 1 – 1 and onto.

Let $f_i = T(e_i) \forall j = 1, \ldots, n$, then $B' = \{f_1, f_2, \ldots, f_n\}$ is also a basis of V, and the matrix of change of basis from B to B' is A.

Theorem 2.5: Let $B = \{e_1, e_2, \ldots, e_n\}$ be a fixed basis of V. The mapping $B' \to M_{B,B'}$ is a 1-1 and onto correspondence between the set of all bases of V and the set of invertible $n \times n$ matrices over F.

Let us see an example of how to obtain $M_{B,B'}$.

Example 4: In R^2, $B = \{e_1, e_2\}$ is the standard basis. Let B' be the basis obtained by rotating B through an angle θ in the anti-clockwise direction (see Fig. 1).

Then $B' = \{e_1', e_2'\}$, where $e_1' = (\cos\theta, \sin\theta), e_2' = (-\sin\theta, \cos\theta)$. Find $M_{B,B'}$.

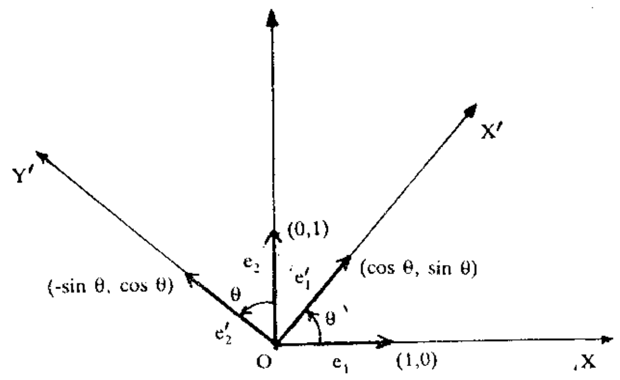

Fig. 1: Change of basis.

Solution:

$e_1' = \cos\theta\,(1.0) + \sin\theta\,(0.1)$ and $e_2' = -\sin\theta\,(1.0) + \cos\theta\,(0.1)$

Thus, $M_{B,B'} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$.

The next corollary summarizes how transition matrices behave.

Here, I denotes the **identity matrix**, that is, the matrix having 1's on the main diagonal and 0's everywhere else.

Given a matrix M, we denote by M^{-1} the *inverse* of M, the matrix Q satisfying $MQ = QM = I$.

Not every matrix has an inverse, we say that P is **invertible** or **non-singular** if it has an inverse.

Corollary: Let B, B', B'' be bases of the vector space *V*, then

a) $M_{B,B} = I$

b) $M_{B',B} = \left(M_{B,B'}\right)^{-1}$

c) $M_{B',B} = M_{B,B'} M_{B'B''}$

The proof follows from the preceding Proposition.
For example, for (b) we have
$[e]_B = M_{B,B'}[e]_{B'}$, $\qquad [e]_{B'} = M_{B',B}[e]_{B'}$
So, $[e]_B = M_{B,B'} M_{B',B}[e]_B$
By the uniqueness of the coordinate representation, we have $M_{B,B'} M_{B',B} = I$

Remark: To express the coordinate representation with respect to the new basis in terms of the old one, we need the inverse of the transition matrix: $[e]_{B'} = M_{B,B'}^{-1}[e]_B$

Example5: Consider the vector space R^2, with the two bases
$B = \left(\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}\right),\ B' = \left(\begin{bmatrix}1\\1\end{bmatrix}, \begin{bmatrix}2\\3\end{bmatrix}\right)$
The transition matrix is $M_{B,B'} = \begin{bmatrix}1 & 2\\1 & 3\end{bmatrix}$ whose inverse is calculated to be
$M_{B',B} = \begin{bmatrix}3 & -2\\-1 & 1\end{bmatrix}$
So, the theorem tells us that, for any $x, y \in R$, we have
$\begin{bmatrix}x\\y\end{bmatrix} = x\begin{bmatrix}1\\0\end{bmatrix} + y\begin{bmatrix}0\\1\end{bmatrix} = (3x - 2y)\begin{bmatrix}1\\1\end{bmatrix} + (-x + y)\begin{bmatrix}2\\3\end{bmatrix}$ as is easily checked.

Definition 4: The $m \times n$matrices *A* and *B* are said to be ***equivalent*** if *B* = *PAQ*, where *P* and *Q* are invertible matrices of sizes $m \times n$and $n \times n$respectively.

Theorem 2.6: Given any$m \times n$matrix *A*, there exist invertible matrices *P* and *Q* of sizes$m \times m$ and$n \times n$respectively, such that *PAQ* is in the canonical form for equivalence.

Remarks: The relation "equivalence" defined above is an equivalence relation on the set of all $m \times n$matrices; that is, it is reflexive, symmetric and transitive.

When mathematicians talk about a "canonical form" for an equivalence relation, they mean a set of objects which are representatives of the equivalence classes: that is, every object is equivalent to a unique object in the canonical form.

We have shown this for the relation of equivalence defined earlier, except for the uniqueness of the canonical form. This is our job for the next section.

2.3.2 The Invertible Matrix Theorem

This section consists of a single important theorem containing many equivalent conditions for a matrix to be invertible. Most of the theorem have been proved earlier in the unit, so, you are encouraged to refer to them to see the beauty of the theorem.

Theorem 2.7:
Let A be an $n \times n$ matrix, and let $T: R^n \rightarrow R^n$ be the matrix transformation $T(x) = Ax$.
The following statements are equivalent:

a) A is invertible.
b) A has n pivots.
c) $Nullity(A) = \{0\}$.
d) The columns of A are linearly independent.
e) The columns of A span R^n.
f) $Ax = b$ has a unique solution for each $b \in R^n$.
g) T is invertible.
h) T is one-to-one.
i) T is onto.

There are two kinds of *square* matrices:

1. invertible matrices, and
2. non-invertible matrices.

For invertible matrices, all of the statements of the invertible matrix theorem are true.

For non-invertible matrices, all of the statements of the invertible matrix theorem are false.

The following conditions are also equivalent to the invertibility of a square matrix A.

They are all simple restatements of conditions in the invertible matrix theorem.

a. The reduced row echelon form of A is the identity matrix I_n.
b. $Ax = 0$ has no solutions other than the trivial one.
c. $nullity(A) = 0$.
d. The columns of A form a basis for R^n.
e. $Ax = b$ is consistent for all $b \in R^n$.

f. $Col(A) = R^n$.
g. $Dim\ Col(A) = n$
h. $rank(A) = n$.

Let us see some common situations in which the invertible matrix theorem is useful.

Consider the matrix $\boldsymbol{A} = \begin{pmatrix} 1 & 2 & -1 \\ 2 & 4 & 7 \\ -2 & -4 & 1 \end{pmatrix}$. Is the matrix invertible?

The second column is a multiple of the first. The columns are linearly dependent, so A does not satisfy condition of the theorem. Therefore, A is not invertible.

Example 6: Let A be an $n \times n$ matrix and let $T(x) = Ax$. Suppose that the range of T is R^n. Show that the columns of A are linearly independent.

Solution: The range of T is the column space of A, so A satisfies condition 5 of the Theorem. Therefore, A also satisfies condition 4, which says that the columns of A are linearly independent.

Example 7: Let B be an 3×3 matrix such that $\boldsymbol{A}\begin{pmatrix} 1 \\ 7 \\ 0 \end{pmatrix} = \boldsymbol{A}\begin{pmatrix} 2 \\ 0 \\ -1 \end{pmatrix}$

If we set $b = A\begin{pmatrix} 1 \\ 7 \\ 0 \end{pmatrix} = A\begin{pmatrix} 2 \\ 0 \\ -1 \end{pmatrix}$, then $Ax = b$ has multiple solutions, so it does not satisfy condition 6 of the Theorem.

Therefore, it does not satisfy condition 5, so the columns of A do not span R^3.

Therefore, the column space has dimension strictly less than 3, the rank is at most 2.

Example 8: Suppose that A is an $n \times n$ matrix such that $Ax = b$ is inconsistent some vector b. Show that $Ax = b$ has infinitely many solutions for some (other) vector b.

Solution: By hypothesis, A does not satisfy condition 6 of the Theorem Therefore, it does not satisfy condition 3, so Nullity(A) is an infinite set.

If we take $b = 0$, then the equation $Ax = b$ has infinitely many solutions.

SELF-ASSESSMENT EXERCISES

1. Let B be the standard basis of R^3 and B' be another basis such that $M_{B,B'} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 3 \end{pmatrix}$. What are the elements of B?
2. When are two $m \times n$ matrices A and B said to be equivalent?
3. Let $B = \begin{pmatrix} 2 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 3 & 0 \end{pmatrix} \in M_3(Q)$. Determine whether or not B is invertible.

Conclusion

Only a square matrix can be invertible. A matrix is P is **invertible** or **non-singular** if it has an inverse.

Given two bases B *and* B', for the n-dimensional vector space V over the field K, then, for any vector $e \in V$, the coordinate representations of e with respect to B *and* B' are related by $[e]_B = M[e]_{B'}$.

The reader should be comfortable translating any of the statements in the invertible matrix theorem into a statement about the pivots of a matrix.

2.4 Summary

The concept of an invertible matrix has been explained in this unit.

We defined the matrix of a change of basis, and discussed the effect of change of bases on the matrix of a linear transformation.

We have also learnt that if an inverse of a matrix exists, it must be unique.

2.5 References/Further Readings

Robert A. Beezer (2014). A First Course in Linear Algebra Congruent Press, Gig Harbor, Washington, USA 3(40).
Arbind K Lal Sukant Pati (2018). Linear Algebra through Matrices.

Peter J. Cameron (2008). Notes on Linear Algebra.

K.A. Stroud and Dexter J. Booth (2007). Engineering Mathematics 6th Edition. Palgrave Macmillan.

Abolarin, O.E. and Ishola, Y.C. (2005). Business Mathematics II. Jube Evans Books and Publications

UNIT 3 MATRICES – III

Unit Structure

3.1 Introduction

In Unit 1, you were introduced to matrices and shown how a system of linear equations can give us a matrix. An important reason for which linear algebra arose is the theory of simultaneous linear equations. A system of simultaneous linear equations can be translated into a matrix equation, and solved by using matrices.

The study of the rank of a matrix is a natural forerunner to the theory of simultaneous linear equations. It is in terms of rank that we can find out whether a simultaneous system of equations has a solution or not. In this unit, you shall be studying the rank of a matrix and its relationship with the inverse of the matrix. Then we discuss row operations on a matrix and use them for obtaining the rank and inverse of a matrix. Finally, we apply this knowledge to determine the nature of solutions of a system of linear equations. The method of solving a system of linear equations that we give here is by “successive elimination of variable”. It is also called the Gaussian elimination process.

In the next unit we shall discuss concepts that are intimately related to matrices.

3.2 Learning Outcomes

By the end of this unit, you will be able to:

- Define and obtain the rank of a matrix;
- Reduce a matrix to the echelon form;
- Obtain the inverse of a matrix by row-reduction;
- Solve a system of simultaneous linear equations by the method of successive elimination of variables.

3.3 Matrices

3.3.1 Elementary Operations on a Matrix

3.3.1.1 Elementary Row and column operations

Let A be a $m \times n$ matrix over a field K, we define certain operations on A called row denoted by $R_1, \dots, R_n$ and column operations denoted by $C_1, \dots, C_n$.

There are three types of elementary row operations:
Type 1: Interchange the i^{th} and j^{th} rows, for $j \neq i$
Type 2: Multiply the i^{th} row by a non-zero scalar, c.
Type 3: Add a multiple of the j^{th} row (R_j) to the i^{th} row (R_i) for $j \neq i$
Example 1: Let $A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \end{bmatrix}$
Then interchanging the two rows, we have $R_{12}(A) = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 2 & 3 \end{bmatrix}$
Multiplying R_2 by a non-zero scalar, e.g., $c = 3$;

$$R_{12}(A) = \begin{bmatrix} 1 & 2 & 3 \\ 0 \times 3 & 1 \times 3 & 2 \times 3 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 3 & 6 \end{bmatrix}$$

Add multiple of R_2 to R_1; $\mathrm{R}_1 + 2R_{12}(A) = \begin{bmatrix} 1 + 2(0) & 2 + 2(1) & 3 + 2(2) \\ 0 & 1 & 2 \end{bmatrix}$

$$= \begin{bmatrix} 1 & 4 & 7 \\ 0 & 1 & 2 \end{bmatrix}$$

There are also three types elementary column operations:
Type 1 Interchange the i^{th} and j^{th} columns, for $j \neq i$.
Type 2 Multiply the i^{th} column by a non-zero scalar.
Type 3 Add a multiple of the j^{th} column (C_j) to the i^{th} column (C_i), where $j \neq i$.
Example 2: Let $A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \end{bmatrix}$

Then interchanging two columns, we have $C_{12}(A) = \begin{bmatrix} 2 & 1 & 2 \\ 1 & 0 & 3 \end{bmatrix}$

Multiplying $C_3$by a non-zero scalar, e.g., $c = 4$; $C_3(A) = \begin{bmatrix} 1 & 2 & 3 \times 4 \\ 0 & 1 & 2 \times 4 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 12 \\ 0 & 1 & 8 \end{bmatrix}$

Add multiple of C_2 to C_1; $\mathrm{C}_1 + 5C_2(A) = \begin{bmatrix} 1+5(2) & 2 & 3 \\ 0+5(1) & 1 & 2 \end{bmatrix} = \begin{bmatrix} 11 & 2 & 3 \\ 5 & 1 & 2 \end{bmatrix}$

By applying these operations, we can reduce any matrix to a particularly simple form:

We will now prove a theorem which we will use later for obtaining the rank of a matrix easily.

Theorem 3.1: Elementary operations on a matrix do not alter its rank.

Proof: The way we will prove the statement is to show that the row space remains unchanged under row operations and the column space remains unchanged under column operations. This means that the row rank and the column rank remain unchanged. This immediately shows, by Theorem 3.1, that the rank of the matrix remains unchanged.

Now, let us show that the row space remains unaltered.

Let $R_1, \ldots, R_n$ be the rows of a matrix *A*, then the row space of *A* is generated by $\{R_1 \ldots R_i \ldots R_j \ldots R_m\}$

On applying R_{ij} to *A*, the rows of *A* remain the same. Only their order gets changed. Therefore, the row space of by $R_{ij}(A)$ is the same as the rows space of *A*.

If we apply $R_1(a)$, for by $a \in K, a \neq 0$, then any linear combination of by $R_1, \ldots, R_m$ is by $a_1R_1 + \ldots + a_mR_m = a_1 + \ldots + a_iR_i + \ldots + a_mR_m$, which is a linear combination of by $R_1 \ldots aR_i \ldots R_m$.

Thus, $|\{R_1 \ldots R_i \ldots R_m\}| = [\{R_1 \ldots aR_i \ldots R_m\}]$.

That is, the row space of *A* is the same as the row space of $R_i(a)(A)$.

If we apply $R_{ij}(a)$for $a \in K$, then any linear combination

$$b_1R_1 + \ldots + b_iR_j + \ldots + b_mR_m$$
$$= b_1R_1 + \ldots + b_i(R_i + aR_j) + \ldots + (b_j - b_ia)R_j + \ldots + b_mR_m$$

Thus, $[\{R_1 \ldots, \ldots R_m\}] = [\{R_1, \ldots, R_i + aR_j, \ldots, R_j, \ldots, R_m\}]$.

Hence, the row space of A remains unaltered under any elementary row operations.

We can similarly show that the column space remains unaltered under elementary column operations.

Elementary operations lead to the following definition.

Definition 1: A matrix obtained by subjecting I_n to an elementary row or column operation is called an **elementary matrix**.

For example, $C_{12}(I_3) = C_{12}$ = an elementary matrix.

$I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $C_{12}(I_3) = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $C_{13}(I_3) = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$,

$C_{23}(I_3) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$

$R_{12}(I_3) = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $R_{13}(I_3) = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$, $R_{23}(I_3) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$

From the above, we have

$C_{12}(I_3) = R_{12}(I_3)$; $C_{13}(I_3) = R_{13}(I_3)$; $C_{23}(I_3) = R_{23}(I_3)$.

Since there are six types of elementary operations, we have six types of elementary matrices, but not all of them are different.

Exercise: Check that i) $C_{21}(3)(I_4) = R_{12}(3)(I_4)$ ii) $C_2(2)(I_4) = R_2(2)(I_4)$

iii) $C_{23}(I_4) = R_{23}(I_4)$.

In general, $C_{ij}(I_n) = R_{ij}(I_n)$; $C_i(a)(I_n) = R_i(a)(I_n)$ for $a \neq 0$ and

$C_{ij}(a)(I_n) = R_{ij}(a)(I_n)$ for $i \neq j$ and $a \in K$.

Thus, there are only three types of elementary matrices.

We denote

i. $R_{ij}(I) = C_{ij}(I)$ by E_{ij}
ii. $R_i(a)(I) = C_i(a)(I)$ (for $a \neq 0$) by $E_i(a)$ and
iii. $R_{ij}(a)(I) = C_{ij}(a)(I)$ by $E_{ij}(a)$for $i \neq j$ and $a \in K$.

E_{ij},$E_i(a)$ and $E_{ij}(a)$ are called the elementary matrices corresponding to the pairs (R_{ij}, C_{ij}), $(R_i(a), C_i(a))$ and $(R_{ij}(a), C_{ij}(a))$ respectively.

Let's now see what happens to the matrix $A = \begin{pmatrix} 0 & 1 & 2 \\ 3 & 0 & 0 \\ 2 & 1 & 0 \end{pmatrix}$

If we multiply it on the left by $E_{12} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 0 & 1 & 2 \\ 3 & 0 & 0 \\ 2 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 1 & 2 \\ 2 & 1 & 0 \end{pmatrix} = R_{12}(A)$$

Similarly, $AE_{12} = C_{12}(A)$.

Again, consider $E_3(2)A = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 2 \end{pmatrix}\begin{pmatrix} 0 & 1 & 2 \\ 3 & 0 & 0 \\ 2 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 2 \\ 3 & 0 & 0 \\ 4 & 2 & 0 \end{pmatrix} = R_3(2)(A)$

Similarly, $AE_3(2) = C_3(2)(A)$

Finally, $E_{13}(5)A = \begin{pmatrix} 0 & 1 & 2 \\ 3 & 0 & 0 \\ 2 & 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 & 5 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 2 \\ 3 & 0 & 15 \\ 2 & 1 & 10 \end{pmatrix} = R_{31}(5)(A)$

But $AE_{13}(5) = \begin{pmatrix} 0 & 1 & 2 \\ 3 & 0 & 0 \\ 2 & 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 & 5 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 2 \\ 3 & 0 & 15 \\ 2 & 1 & 10 \end{pmatrix} = C_{31}(5)(A)$

What you have just seen are example of a general phenomenon. We will now state this general result formally. (Its proof is slightly technical, and so, we skip it.)

Theorem 3.2: For any matrix A

a. $R_{ij}(A) = E_{ij}(A)$.

b. $R_i(a)(A) = E_i(a)(A)$; for $a \neq 0$

c. $R_{ij}(a)(A) = E_{ij}(a)(A)$

d. $C_{ij}(A) = AE_{ij}$

e. $C_i(a)(A) = AE_i(a)$; for $a \neq 0$

f. $C_{ij}(a)(A) = AE_{ij}(a)$

An immediate corollary to this theorem shows that all the elementary matrices are invertible (see unit 2).

Corollary: An elementary matrix is invertible. In fact,

a. $E_{ij}E_{ij} = I$

b. $E_i(a^{-1})E_i(a) = I$ for $a \neq 0$

c. $E_{ij}(-a)E_{ij}(a) = I$

Proof: we prove (a) only and leave the rest to you.

Now, by definition of $R_{ij}$from Theorem,

$E_{ij}E_{ij} = R_{ij} = R_{ij}\big(R_{ij}(I)\big) = I,$

The corollary tells us that the elementary matrices are invertible and the inverse of an elementary matrix is also an element matrix of the same type.

Now we will introduce you to a very nice type of matrix, which any matrix can be transformed to by applying elementary operations.

3.3.1.2 Row-Reduced Echelon Matrices

Consider the matrix $\begin{pmatrix} 1 & 0 & 9 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$

In this matrix, the three non-zero rows come before the zero row, and the first non-zero entry in each non-zero row is 1. Also, below this 1, are only zero. This type of matrix has a special name, which we now give.

Definition 2: An $m \times n$ matrix A is called a **row-reduced echelon** matrix if

a) The non-zero rows come before the rows,

b) In each non-zero row, the first non-zero entry is 1, and

c) The first non-zero entry in every non-zero row (after the first row) is to the right of the first non-zero entry in the preceding row.

The matrix $\begin{bmatrix} 0 & 1 & 3 & 4 & 9 & 7 & 8 & 0 & -1 & 0 & 1 \\ 0 & \bar{0} & 0 & 0 & 1 & 5 & 6 & 10 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 7 & 0 & 12 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 10 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$ is a(6×11)row-echelon matrix.

Every matrix can be transformed to the row echelon form by a series of elementary row operations. We say that the matrix is reduced to the echelon form.

Example 3: Consider the following matrix:

Let $A = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 2 & -1 & -1 & 1 \\ 0 & 1 & 2 & 0 & 3 & 1 \\ 0 & 0 & 0 & 1 & 4 & 1 \\ 0 & 2 & 4 & 1 & 10 & 2 \end{bmatrix}$

Reduce A to the row echelon form.

Solution: The first column of A is zero. The second is non-zero. The $(1,2)^{th}$element is 0.

But we want 1 at this position.

Apply R_{12} to A and get $A_1 = \begin{bmatrix} 0 & 1 & 2 & -1 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 2 & 0 & 3 & 1 \\ 0 & 0 & 0 & 1 & 4 & 1 \\ 0 & 2 & 4 & 1 & 10 & 2 \end{bmatrix}$

The $(1,2)^{th}$entry has become 1. Now we subtract multiples of the first row from other rows so that the $(2,2)^{th}$, $(3,2)^{th}$, $(4,2)^{th}$and $(5,2)^{th}$entries become zero. So, we apply $R_{ij}(-1)$ and $R_{51}(-2)$to obtain

$$A_2 = \begin{bmatrix} 0 & 1 & 2 & -1 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 2 & 0 & 4 & 0 \\ 0 & 0 & 0 & 1 & 4 & 1 \\ 0 & 0 & 0 & 3 & 12 & 2 \end{bmatrix}$$

Now, beneath the entries of the first row we have zeros in the first 3 columns, and in the fourth column we find non-zero entries. We want 1 at the $(2,4)^{th}$position, so we interchange the 2nd and 3rd rows to obtain

$$A_3 = \begin{bmatrix} 0 & 1 & 2 & -1 & -1 & 1 \\ 0 & 0 & 0 & 1 & 4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 4 & 1 \\ 0 & 0 & 0 & 3 & 12 & 0 \end{bmatrix}$$

We now subtract suitable multiples if the 2nd row from the 3rd, 4th and 5th rows so that the $(3,4)^{th}$,$(4,4)^{th}$, $(5,4)^{th}$entries all become zero.

$$\begin{matrix} R_{42}(-1) \\ \therefore R_{42}(-3) \\ A_3 \overset{R}{\approx} B \end{matrix} \begin{bmatrix} 0 & 1 & 2 & -1 & -1 & 1 \\ 0 & 0 & 0 & 1 & 4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$A_3 \overset{R}{\approx} B$ means that on apply the operation R to *A* we obtain matrix *B*.

Now we have zero below the entries of the 2nd row, except for the 6th column. The$(3,6)^{th}$element is 1. We subtract suitable multiples of the 3rd row from the 4th and 5th rows so that the $(4,6)^{th}$elements become zero.

$$\therefore A_4 \overset{R_{43}(-1)}{\approx} \begin{bmatrix} 0 & 1 & 2 & -1 & -1 & 1 \\ 0 & 0 & 0 & 1 & 4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Now we have achieved a row echelon matrix.
Notice that we applied 7 elementary operations to *A* to obtain this matrix. In general, we have the following theorem.

Theorem 3.3: Every matrix can be reduced to a row-reduced echelon matrix by a finite sequence of elementary row operations.
The **proof** of this result is just a repetition of the process that you went through in Example 4 for practice; we give you the following exercise.

Example 4: Reduce the matrix $\begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 3 & 1 & 0 \end{bmatrix}$ to echelon form.

Solution: $\begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 3 & 1 & 0 \end{bmatrix} \Rightarrow \overset{\approx}{R_{31}}(-3) \begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 0 & -5 & 0 \end{bmatrix} \Rightarrow \overset{\approx}{R_{32}}(5) \begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

Theorem 3.3 leads us to the following definition.

Definition 3: If a matrix *A* is reduced to a row-reduce echelon matrix *E* by a finite sequence of elementary row operations, then *E* is called a **row-reduced echelon form** (or, the row echelon form) of *A*.

3.3.2 Rank of a matrix

Theorem 3.4: Let A be an $m \times n$ matrix over the field *K*. Then it is possible to change A into B by elementary row and column operations, where *B* is a matrix of the same size satisfying $B_{ii} = 1$ for $0 \leq i \leq r$ for $r \leq min\{m, n\}$ and all other entries of *B* are zero.

If A can be reduced to two matrices, B and B', both of the above form, where the numbers of non-zero elements are r and r'respectively, by different sequences of elementary operations, then $r = r'$, and so $B = B'$.

Definition 4: The number *r* in the above theorem is called the ***rank*** of *A*; while a matrix of the form described for *B* is said to be in the ***canonical form for equivalence***.

We can write the canonical form matrix in "block form" as$B = \begin{bmatrix} I_r & O \\ O & O \end{bmatrix}$, where I_r is an $r \times r$ identity matrix and 0 denotes a zero matrix of the appropriate size (that is, $r \times (n-r), (m-r) \times r$ and $(m-r) \times (n-r)$ respectively for the three 0's). Note that some or all of these 0's may be missing: for example, if *r* = *m*, we just have $[I_m \quad O]$.

Proof:
We outline the proof that the reduction is possible.
To prove that we always get the same value of r, we need a different argument.

The proof is by induction on the size of the matrix *A*: in other words, we assume as inductive hypothesis that any smaller matrix can be reduced as in the theorem. Let the matrix *A* be given. We proceed in steps as follows:

- If *A* = 0(the all-zero matrix), then the conclusion of the theorem holds, with *r* = 0; no reduction is required. So, assume that $A \neq 0$.
- If$A_{11} \neq 0$, then skip this step. If $A_{11} = 0$, then there is a non-zero element A_{ij} somewhere in *A*; by swapping the first and i^{th}rows, and the first and j^{th} columns, if necessary (Type 3 operations), we can bring this entry into the (1,1) position.
- Now we can assume that $A_{11} \neq 0$. Multiplying the first row by A_{11}^{-1}, (row operation Type 2), we obtain a matrix with $A_{11} = 1$.
- Now by row and column operations of Type 1, we can assume that all the other elements in the first row and column are zero. For if $A_{1j} \neq 0$, then subtracting A_{1j} times the first column from the j^{th} gives a matrix with$A_{1j} = 0$. Repeat this until all non-zero elements have been removed.

Now let *B* be the matrix obtained by deleting the first row and column of *A*.

Then B is smaller than A and so, by the inductive hypothesis, we can reduce B to canonical form by elementary row and column operations. The same sequence of operations applied to A now finish the job.

Example 5: Here is a small example. Let $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$

We have $A_{11} = 1$, so we can skip the first three steps. Subtracting twice the first column from the second, and three times the first column from the third, gives the matrix $\begin{bmatrix} 1 & 0 & 0 \\ 4 & -3 & -6 \end{bmatrix}$.

Now subtracting four times the first row from the second gives $\begin{bmatrix} 1 & 0 & 0 \\ 0 & -3 & -6 \end{bmatrix}$

From now on, we have to operate on the smaller matrix $[-3 \ -6]$, but we continue to apply the operations to the large matrix.

Multiply the second row by $-1/3$ to get $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 2 \end{bmatrix}$

Now subtract twice the second column from the third to obtain $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$

We have finished the reduction, and we conclude that the rank of the original matrix A is equal to 2.

We finish this section by describing the elementary row and column operations in a different way.

For each elementary row operation on an n-rowed matrix A, we define the corresponding *elementary matrix* by applying the same operation to the $n \times n$ identity matrix I.

Similarly, we represent elementary column operations by elementary matrices obtained by applying the same operations to the mxm identity matrix.

We don't have to distinguish between rows and columns for our elementary matrices. For example, the matrix $\begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ corresponds to the elementary column operation of adding twice the first column to the second, or to the elementary row operation of adding twice the second row to the first. For the other types, the matrices for row operations and column operations are identical.

Lemma: The effect of an elementary row operation on a matrix is the same as that of multiplying on the left by the corresponding elementary matrix. Similarly, the effect of an elementary column operation is the

same as that of multiplying on the right by the corresponding elementary matrix.
The proof of this lemma is somewhat tedious calculation.

Example 6: We continue our previous example. In order, here is the list of elementary matrices corresponding to the operations we applied to A. (Here 2×2 matrices are row operations while 3×3 matrices are column operations).

$$\begin{bmatrix}1 & -2 & 0\\0 & 1 & 0\\0 & 0 & 1\end{bmatrix}, \begin{bmatrix}1 & 0 & -3\\0 & 1 & 0\\0 & 0 & 1\end{bmatrix}, \begin{bmatrix}1 & 0\\-4 & 1\end{bmatrix}, \begin{bmatrix}1 & 0\\0 & -1/3\end{bmatrix}, \begin{bmatrix}1 & 0 & 0\\0 & 1 & -2\\0 & 0 & 1\end{bmatrix},$$

So, the whole process can be written as a matrix equation:

$$\begin{bmatrix}1 & 0\\0 & -1/3\end{bmatrix}\begin{bmatrix}1 & 0\\-4 & 1\end{bmatrix}A\begin{bmatrix}1 & -2 & 0\\0 & 1 & 0\\0 & 0 & 1\end{bmatrix}\begin{bmatrix}1 & 0 & -3\\0 & 1 & 0\\0 & 0 & 1\end{bmatrix}\begin{bmatrix}1 & 0 & 0\\0 & 1 & -2\\0 & 0 & 1\end{bmatrix} = B$$

or more simply $\begin{bmatrix}1 & 0\\4/3 & -1/3\end{bmatrix}A\begin{bmatrix}1 & -2 & 1\\0 & 1 & -2\\0 & 0 & 1\end{bmatrix} = B.$

where, as before, $A = \begin{bmatrix}1 & 2 & 3\\4 & 5 & 6\end{bmatrix}$, $B = \begin{bmatrix}1 & 0 & 0\\0 & 1 & 0\end{bmatrix}$

An important observation about the elementary operations is that each of them can have its effect undone by another elementary operation of the same kind, and hence every elementary matrix is invertible, with its inverse being another elementary matrix of the same kind. For example, the effect of adding twice the first row to the second is undone by adding −2 times the first row to the second, so that $\begin{bmatrix}1 & 2\\0 & 1\end{bmatrix}^{-1} = \begin{bmatrix}1 & -2\\0 & 1\end{bmatrix}$

Since the product of invertible matrices is invertible, we can state the above theorem in a more concise form. First, one more definition:

3.3.2.1 Row and Column Ranks of a Matrix

Definition 5: Let A be an $m \times n$ matrix over a field F. We say that the ***column rank*** of A is the maximum number of linearly independent columns of A; while the ***row rank*** of A is the maximum number of linearly independent rows of A. (We regard columns or rows as vectors in F^m and F^n respectively).

Now we need a sequence of four lemmas.

Lemma:

(a) Elementary column operations do not change the column rank of a matrix.

(b) Elementary row operations do not change the column rank of a matrix.

(c) Elementary column operations do not change the row rank of a matrix.

(d) Elementary row operations do not change the row rank of a matrix.

Proof
(a) This is clear for Type 3 operations, which just rearrange the vectors. For Types 1 and 2, we have to show that such an operation cannot take a linearly independent set to a linearly dependent set; the *vice versa* statement holds because the inverse of an elementary operation is another operation of the same kind.

So, suppose that $v_1,\ldots,v_n$are linearly independent.
Consider a Type 1 operation involving adding c times the j^{th} column to the i^{th}; the new columns are $v'_1,\ldots,v'_n$where $v'_k = v_k$for $k \neq i$while $v'_1 = v_i + cv_j$
Suppose that the new vectors are linearly dependent, then there are scalars $a_1,\ldots,a_n$, not all zero, such that $0 = a_1v'_1+\ldots+a_nv'_n$

$$= a_1v_1 + \cdots + a_i(v_i + cv_j) + \cdots + a_jv_j + \cdots + a_nv_n$$
$$= a_1v_1 + \cdots + a_iv_i + \cdots + (a_j + ca_i)v_j + \cdots + a_nv_n$$

Since $v_1,\ldots,v_n$are linearly independent, we conclude that

$$a_1 = 0, \ldots, a_i = 0, \ldots, a_jca_i = 0, \ldots, a_n = 0$$

from which we see that all the $a_k$are zero, contrary to assumption. So, the new columns are linearly independent.
The argument for Type 2 operations is similar but easier.
(b) It is easily checked that, if an elementary row operation is applied, then the new vectors satisfy exactly the same linear relations as the old ones (that is, the same linear combinations are zero). So, the linearly independent sets of vectors don't change at all.
(c) Same as (b), but applied to rows.
(d) Same as (a), but applied to rows.

Theorem 3.4: For any matrix A, the row rank, the column rank, and the rank are all equal. In particular, the rank is independent of the row and column operations used to compute it.

Proof: Suppose that we reduce *A* to canonical form *B* by elementary operations, where *B* has rank *r*. These elementary operations don't change the row or column rank, by our lemma; so the row ranks of *A* and *B* are equal, and their column ranks are equal. But it is trivial to see that, if $B = \begin{bmatrix} I_r & O \\ O & O \end{bmatrix}$, then the row and column ranks of *B* are both equal to r.
So, the theorem is proved.
We can get an extra piece of information from our deliberations.
Let *A* be an invertible $n \times n$ matrix. Then the canonical form of *A* is just *I*:
its rank is equal to *n*. This means that there are matrices P and Q, each a product of elementary matrices, such that $PAQ = I_n$.

From this we deduce that $A = P^{-1}I_nQ^{-1} = P^{-1}Q^{-1}$;

Corollary: ***Every invertible square matrix is a product of elementary matrices.***
In fact, we learn a little bit more. We observed when we defined elementary matrices, that they can represent either elementary column operations or elementary row operations. So, when we have written *A* as a product of elementary matrices, we can choose to regard them as representing column operations, and we see that *A* can be obtained from the identity by applying elementary column operations. If we now apply the inverse operations in the other order, they will turn *A* into the identity (which is its canonical form). In other words, the following is true:

Corollary: ***If A is an invertible $n \times n$ matrix, then A can be transformed into the identity matrix by elementary column operations alone (or by elementary row operations alone).***

3.3.2.2 Row and Column Spaces of a Matrix

Consider any $m \times n$ matrix *A*, over a field *F*. We can associate two vector spaces with it, in a very natural way. Let us see what they are.
Let$A = [a_{ij}]$, *A* has *m* rows, say, $R_1, R_2, \ldots, R_m$, where

$$R_1 = (a_{11}, a_{12}, \ldots, a_{1n}), R_2 = (a_{21}, a_{22}, \ldots, a_{2n}), \ldots, R_m = (a_{m1}, a_{m2}, \ldots, a_{mn})$$

Thus, $R_i \in \begin{bmatrix} R_1 \\ R_2 \\ \vdots \\ R_m \end{bmatrix}$ and $i \in A$

The subspace of F^n generated by the row vectors $R_1, \ldots, R_m$of *A*, is called the **row space** of A, and is denoted by ***RS (A)***.

Example 7: If $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$, does $(0,0,1) \in RS(A)$?

Solution: The row space of A is the subspace of R_3 generated by (1, 0, 0) and (0, 1, 0). Therefore$RS(A) = \{(a, b, 0) | a, b \in R\}$
$\therefore (0,0,1) \notin RS(A)$.

Definition 6: The **dimension of the row space** of *A* is called the **row rank** of *A*, and is denoted by $\boldsymbol{p_r(A)}$.
Thus, $p_r(A)$ = maximum number of linear independent rows of A.
In Example 1, $p_r(A) = 2$= number of rows of A. But consider the next example.

Example 8: If $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 0 \end{bmatrix}$, find $p_r(A)$.

Solution: The row space of A is the subspace of R_2 generated by (1, 0), (0, 1) and (2, 0). But (2, 0) already lies in the vector space generated by (1, 0) and 0, 1), since $(2,0) = 2(1,0)$. Therefore, the row of A is generated by the linear independent vectors (1, 0) and (0, 1).
Thus, $p_r(A) = 2$.
So, in Example 2, $p_r(A)$ <number of rows of A
In general, for $m\ x\ n$ matrix A, RS (A) is generated by m vectors.
Therefore, $p_r(A) \leq m$..
Also, $RS(A)$is a subspace of F^n and$dim_F\, F^n = n$. Therefore, $p_r(A) \leq n$.
Thus, for any $m \times n$ matrix A, $0 \leq p_r(A) \leq min(m, n)$.

Example 9: Show that $A = 0 \Leftrightarrow p_r(A) = 0$.
Just as we have defined the row space of A, we can define the column space of A.
Each column of A is an m-tuple, and hence belongs toF^m.
The column of A is denoted by $C_1, \dots, C_n$.
The **subspace of F^m** generated by $\{C_1, \dots, C_n\}$ is called the **column space of A** and is denoted by **CS (A).**
Definition 7: The **dimension** of CS (A) is called **the column rank** of A, and is denoted of $p_c(A)$.
Again, since CS (A) is generated by n vectors and is a subspace of F^m, we obtain
$0 \leq p_c(A) \leq min(m, n)$.
Example 10: Obtain the column rank and row rank of $A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix}$
Solution:
The column space of A is the subspace of R^2 generated by (1, 0), (0, 2), (1, 1).
Now, $dim_R\, C\, S(A) \leq dim_R\, R^2 = 2$.
Also (1, 0) and (0, 2), are linearly independent,
$\therefore$ {(1, 0), (0, 2)} is a basis of CS(A), and $p_c(A) = 2$.
The row space of A is the subspace of R^3 generated by (1, 0, 1) and (0, 2, 1).
These vectors are linearly independent and hence, form a basis of $RS(A)$
$\therefore p_r(A) = 2$.
In Example 2 you may have noticed that the row rank and column rank of $A$are equal. In fact, in Theorem 1, we prove that $p_r(A) = p_c(A)$for any matrix A. But first, we prove a lemma.
Lemma 1: Let A, B be two matrices over F such that AB is defined. Then
a) $CS(AB) \subseteq CS(A)$
b) $RS(AB) \subseteq RS(B)$
Thus,$p_c(AB) \leq p_c(A)$, $p_r(AB) \leq p_r(B)$.

Proof:

(a) Suppose $A = [a_{ij}]$ is an $n \times p$ matrix. Then, from the previous section, you know that the j^{th} column of $C = AB$ would be

$$\begin{bmatrix} c_{1j} \\ c_{2j} \\ \vdots \\ c_{mj} \end{bmatrix} = \begin{bmatrix} \sum_{k=1}^{n} a_{1k}b_{kj} \\ \sum_{k=1}^{n} a_{2k}b_{kj} \\ \vdots \\ \sum_{k=1}^{n} a_{mk}b_{kj} \end{bmatrix} = \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} b_{1j} + \cdots + \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix} b_{nj}$$

$$= C_1 b_{1j} + \cdots + C_n b_{nj}$$

Where $C_1, \ldots, C_n$ are the columns of A.

Thus, the columns of AB are linear combinations of the columns of A.

$$AB \in CS(A).$$

So, $CS(AB) \subseteq CS(A)$

Hence, $p_c(AB) \leq p_c(A)$.

b) By a similar argument as above, we have $RS(AB) \subseteq RS(B)$and so $p_r(AB) \leq p_r(B)$.

Example 11: Prove (b) of Lemma 1.

The i^{th} row of C = AB is

$$[c_{i1} \quad c_{i2} \quad \ldots \quad c_{ip}] = \left[\sum_{k=1}^{n} a_{ik}b_{k1} \quad \sum_{k=1}^{n} a_{ik}b_{k2} \quad \ldots \quad \sum_{k=1}^{n} a_{ik}b_{kp}\right]$$

$= a_{i1}[b_{11} \quad b_{12} \quad \ldots \quad b_{1p}] + a_{i2}[b_{21} \quad b_{22} \quad \ldots \quad b_{2p}] + \cdots +$
$a_{in}[b_{n1} \quad b_{n2} \quad \ldots \quad b_{np}]$,

a linear combination of the rows of B

$$\therefore RS(AB) \subseteq RS(B)$$

$\therefore p_r(AB) \leq p_r(B)$.

We will now use Lemma 1 for proving the following theorem.

Theorem 3.5: $p_c(A) = p_r(A)$for any matrix **A** over **F.**

Proof: Let$A \in M_{m\times n}(F)$. Suppose $p_r(A) = r$ and$p_c(A) = t$.

Now, $RS(A) = \{R_1, R_2, \ldots, R_m\}$where $R_1, R_2, \ldots, R_m$are the rows of A.

Let $\{e_1, e_2, \ldots, e_r\}$ be a basis of RS(A). Then R_1 is a linear combination of $e_1, e_2, \ldots, e_r$, for each $i = 1, \ldots, m$.

Let $R_i = \sum_{j=1}^{r} b_{ij}e_j$, $i = 1,2, \ldots, m$, where $b_{ij} \in F$ for $1 \leq i \leq m$, $1 \leq j \leq r$

We can write these equations in matrix form as

$$\begin{bmatrix} R_1 \\ \vdots \\ \vdots \\ R_m \end{bmatrix} = \begin{bmatrix} b_{11} \cdots\cdots b_{1r} \\ \cdots\cdots\cdots \\ \cdots\cdots\cdots \\ b_{m1} \cdots\cdots b_{mr} \end{bmatrix} \begin{bmatrix} e_1 \\ \vdots \\ \vdots \\ e_r \end{bmatrix}$$

So,$A = BE$, where $B = [b_{ij}]$ is an $m \times r matrix\ and\ E\ is\ the r \times n$matrix with rows $e_1, e_2, \ldots, e_r (e_1 \in F^n$, for each $i = 1, \ldots, r)$

So, $t = p_c(A) = (BE) \leq p_c(B)$ (by Lemma 1)

$\leq min(m, r)$

$\leq r$

Thus, $t \leq r$

Just as we got $A = BE$ above, we have $A = [f_1, \ldots, f_t]D$, where $\{f_1, \ldots, f_t\}$is a basis of the column space of *A*, and D is a $(t\ x\ n)$ matrix.

Thus, $r = p_r(A) \leq p_r(D) \leq t$ (by Lemma 1)

So, we have $r \leq t$ and $t \leq r$. This gives $r = t$.

Theorem 3.1 allows us to make the following definition.

Definition 8: The integer $p_c(A) = p_r(A)$is called the rank of *A*, and is denoted by $p(A)$.

You will see that theorem 3.1 is very helpful if we want to prove any fact about $p(A)$. If it is easier to deal with the rows of *A* we can prove the fact for $p(A)$.

Similarly, if it is easier to deal with the columns of A, we can prove the fact for $p_c(A)$. While proving Theorem 3 we have used this facility that theorem 3.1 gives us.

Use theorem 1 to solve the following exercises.

1) If A, B are two matrices such that AB is defined then show that $p(AB) \leq min(\ \mathrm{p}(A), p(B))$.

2) Suppose $C \neq 0 \in M_{m\times1}(F)$, and $R \neq 0 \in M_{1\times n}(F)$, then show that the rank of the $m \times n$ matrix CR is 1. (Hint: use Example 4).

Does the term 'rank' seem familiar to you? Do you remember studying about the rank of a linear transformation in Unit 2? We now see if the rank of a linear transformation is related to the rank of its matrix. The following theorem brings forth the precise relationship.

Let us now look at some ways of transforming a matrix by playing around with its rows. The idea is to get more and more entries of the matrix to be zero. This will help us in solving systems of linear equations.

Theorem 3.6: Let *E* be a row-reduced echelon form of *A*. Then the rank of *A* = number of non-zero rows of *E*.

Proof: We obtain *E* from *A* by applying elementary operations.

By a Theorem; p(A) = p(E).

Also, p(*E*) = the number of non-zero rows of *E*, by Theorem 4.

Thus, we have proved the theorem. ##

Let us look at some examples to actually see how the echelon form of a matrix simplifies matters.

Example 12: Find p(A), where$A = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 5 & 6 \end{bmatrix}$ by reducing it to its row-reduced echelon form.

Solution: $\begin{bmatrix} 1 & 2 & 3 \\ 1 & 5 & 6 \end{bmatrix} \Rightarrow \underset{\approx}{R_2(-1)} \begin{bmatrix} 1 & 2 & 3 \\ 0 & 3 & 3 \end{bmatrix} \Rightarrow \underset{\approx}{R_2(1/3)} \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 1 \end{bmatrix}$
which is the desired row-reduced echelon form. This has 2 non-zero rows. Hence, p(A) =2.
Exercise13: Obtain the row-reduced echelon form of the matrix
$A = \begin{bmatrix} 1 & 2 & 0 & 5 \\ 2 & 1 & 7 & 6 \\ 4 & 5 & 7 & 10 \end{bmatrix}$. Hence determine the rank of the matrix.
By now must have got used to obtaining row echelon forms. Let us discuss some ways of applying this reduction.

3.3.3 Applications of Row-Reduction of Matrices

In this section we shall see how to utilize row-reduction for obtaining the inverse of a matrix, and for solving a system of leaner equations.

3.3.3.1 Inverse of a Matrix

In previous theorem, you discovered that applying a row transformation to a matrix *A* is the same as multiplying it on the left by a suitable elementary matrix. Thus, applying a series of row transformations to *A* is the same as pre-multiplying *A* by a series of elementary matrices.

This means that after the *nth* row transformation we obtain the matrix $E_n E_{n-1} \ldots E_2 E_1 A$, where $E_1, E_2, \ldots, E_n$ are elementary matrices.

Now, how do we use this knowledge for obtaining the inverse of an invertible matrix? Suppose we have an *n* x *n* invertible matrix *A*. We know that *A* = *IA*, where $I = I_n$.

Now, we apply a series of elementary row operations $E_1, E_2, \ldots, E_n$to *A* so that *A* gets transformation to I_n.
Thus,

$$I = E_s E_{s-1} \ldots E_2 E_1 A = E E_{s-1} \ldots E_2 E_1 (IA)$$
$$= (E E_{s-1} \ldots E_2 E_1) A = BA$$

Where $B = E \ldots E_1 I$. Then *B* is the inverse of *A*.
Note that we are reducing *A* to *I* and not only to the echelon form.
We illustrate this below.

Example 13: Determine if the matrix $A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 3 & 1 & 2 \end{bmatrix}$ is invertible.

If it is invertible, find its inverse.

Solution:

Now, $A = IA = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 3 & 1 & 2 \end{pmatrix}$

To transform *A* we will be pre-multiplying it by elementary matrices. We will also be pre-multiplying IA by these matrices. Therefore, as A is transformed to *I*, the same transformations are done to 1 on the right-hand side of the matrix equation given above.

Now $\begin{pmatrix} 1 & 2 & 3 \\ 0 & -1 & -5 \\ 0 & -5 & -7 \end{pmatrix} A \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ -3 & 0 & 1 \end{pmatrix}$

Applying $R_{21}(-2)$and $R_{31}(-3)$to A

$$= \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 5 \\ 0 & 5 & 7 \end{pmatrix} A \begin{pmatrix} 1 & 0 & 0 \\ 2 & -1 & 0 \\ 3 & 0 & -1 \end{pmatrix}$$

Applying $R_2(-1)$and $R_3(-1)$

$$= \begin{pmatrix} 1 & 0 & -7 \\ 0 & 1 & 5 \\ 0 & 0 & -18 \end{pmatrix} A \begin{pmatrix} -3 & 2 & 0 \\ 2 & -1 & 0 \\ -7 & 5 & -1 \end{pmatrix}$$

Applying $R_{12}(-2)$and $R_{32}(-5)$

$$= \begin{pmatrix} 1 & 0 & -7 \\ 0 & 1 & 5 \\ 0 & 0 & -18 \end{pmatrix} A \begin{pmatrix} -3 & 2 & 0 \\ 2 & -1 & 0 \\ -7 & 5 & -1 \end{pmatrix}$$

Applying $R_3(-{}^1/_{18})$

$$= \begin{pmatrix} 1 & 0 & -7 \\ 0 & 1 & 5 \\ 0 & 0 & -18 \end{pmatrix} \begin{pmatrix} -3 & 2 & 0 \\ 2 & -1 & 0 \\ {}^7/_{18} & {}^{-5}/_{18} & {}^1/_{19} \end{pmatrix}$$

$= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} {}^{-5}/_{18} & {}^1/_{18} & {}^7/_{18} \\ {}^1/_{18} & {}^7/_{18} & {}^{-5}/_{18} \\ {}^7/_{18} & {}^{-5}/_{18} & {}^1/_{18} \end{pmatrix}$ Applying $R_{21}(-2)$and $R_{31}(-3)$to A

Hence, A is invertible and its inverse is $B = \frac{1}{18}\begin{bmatrix} -5 & 1 & 7 \\ 1 & 7 & -5 \\ 7 & -5 & 1 \end{bmatrix}$

To make sure that we haven't made a careless mistake at any stage, check the answer by multiplying B with A. your answer should be I

Exercise: Show that $\begin{bmatrix} 0 & 1 & 3 \\ 2 & 3 & 5 \\ 3 & 5 & 7 \end{bmatrix}$ is invertible. Find its inverse.

Let us now look at another application of row-reduction.

3.3.3.2 Solution of System of Linear Equations

Any system of m linear equations, in n unknowns $x_1, \ldots, x_n$ is

$$\begin{array}{ccccccc} a_{11}x_1 & + & \ldots & + & a_{1n} & = & b_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{m1}x_1 & + & \ldots & + & a_{mn}x_n & = & b_m \end{array}$$

where all the x_{ij} and b_i are scalars

This can be written in matrix form as $AX = B$, where $A = [a_{ij}]$, $X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$, $B = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$

Definition 6: If *B* = *0*, the system is called **homogenous, otherwise, it is non-homogeneous**.

In this situation, we are in a position to say how many linearly independent solutions the system of equations has.

Theorem 3.7: The number of linearly impendent solutions of the matrix equation $AX = 0$ is $n - r$, where *A* is an $m \times n$ matrix and $r = p(A)$.

Proof: In Unit 1,you studied that given the matrix *A*, we can obtain a linear transformation $T: F^n \to F^m$ such that $[T]_{B,B} = A$, where B and B' are bases of F^n and F^m, respectively.

Now, $X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ is a solution of $AX = 0$ if and only if it lies in $\ker T$ (since $T(X) = AX$).

Thus, the number of linearly independent solutions is

$dim\ ker\ T\ nullity(T) = n - rank(T)$ (Unit 2, Module 1)

Also; $rank(T) = p(A)$. (Theorem 3.2)

Thus, the number of linearly independent solutions is $n - p(A)$.

This theorem is very useful in finding out whether a homogeneous system has any non-trivial solution or not.

Example 14: Consider the system of 3 equations in 3 unknowns:

$$3x - 2y + z = 0$$
$$x + y = 0$$
$$x - 3z = 0$$

How many solutions does it have which are linearly independent over R?

Solution: Here is our coefficient matrix, $A = \begin{bmatrix} 3 & -2 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & -3 \end{bmatrix}$

Thus, $n = 3$. We have to find r. For this, we apply the row-reduction method:

We obtain A ~ $\begin{bmatrix} 1 & 0 & -3 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{bmatrix}$, which is in echelon form and has rank 3.

Thus, p(A) = 3.

Thus, the number of linearly independent solutions is 3 – 3 = 0. This means that this system of equations has no non-zero solution.

In the Example, the number of unknowns was equal to the number of equations, that is, $n = m$. What happens if $\boldsymbol{n} > m$?

A system of m homogeneous equations in n unknowns has a non-zero solution if $n > m$, why? Well, if $n > m$, then the rank of the coefficient matrix is less than or equal to m, and hence, less than. So, $n - r > 0$ Therefore, at least one non-zero solution exists.

Exercise: Give a set of linearly independent solutions for the system of equations:

$$x + 2y + 3z = 0$$
$$2x + 4y + z = 0$$

Now consider the general equation *AX* = *B*, where *A* is an $m \times n$ matrix. We form the augmented matrix |*AB*|. This is an $m \times (n + 1)$ matrix whose last column is the matrix B. Here, we also include the case *B* = *0*. Interchanging equations, multiplying an equation by a non-zero scalar, and adding to any equation scalar times some other equation does not alter the set of solutions of the system of equations. In other words, if we apply elementary row operations on |AB| then the solution set does not change.

The following result tells us under what conditions the system *AX* = *B* has a solution.

Theorem 3.8: The system of linear equations given by the matrix equation *AX* = *B* has a solution if $p(A) = p(|AB|)$.

Proof: *AX* = *B* represents the system:

$$\begin{matrix} a_{11}x_1 & + & \dots & + & a_{1n} & = & b_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{m1}x_1 & + & \dots & + & a_{mn}x_n & = & b_m \end{matrix}$$

This is the same as

$$\begin{array}{ccccccccc} a_{11}x_1 & + & \dots & + & a_{1n} - b_1 & = & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{m1}x_1 & + & \dots & + & a_{mn}x_n - b_m & = & 0 \end{array}$$

which is represented by $|AB|\begin{bmatrix} X \\ -1 \end{bmatrix} = 0$.

Therefore, any solution of AX = B is also a solution of $|AB|\begin{bmatrix} X \\ -1 \end{bmatrix} = 0$ and vice versa.

By Theorem 3.8; this system has a solution if and only if $n + 1 > p(|AB|)$.

Now if the equation $|AB|\begin{bmatrix} X \\ -1 \end{bmatrix} = 0$ has a solution, say $\begin{bmatrix} c_1 \\ \vdots \\ c_{n-1} \end{bmatrix}$ then

$c_1C_1 + c_2C_2 + \cdots + c_nC_n = B$, $C_1, \cdots, C_n$ are the columns of *A*. That is, B is a linear combination of the $C_i's$ therefore, $RS([AB]) = RS(A)$.

Conversely, if $p([AB])$then the number of linearly independent columns of A and |AB| are the same. Therefore, B must be a combination of the columns $C_1, \cdots, C_n$ of *A*.

Let $B = a_1C_1 + \cdots + a_nC_n$; $a_i \in F, \forall\ i\ a_i \in F, \forall\ i$

Then a solution of $AX = B$ is$X = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$

Thus, AX =B has a solution if and only if $p(A) = p(|AB|)$.

Remark: If A is invertible then the system of $AX = B$ has the unique solution $X = A^{-1}B$.

3.3.3.3 Successive (or Gaussian) Elimination Method

Now, once we know that the system given by of$AX = B$is consistent, how do we find a solution? We utilize the method of successive (or Gaussian) elimination. This method is attributed to the famous German mathematician, Carl Friedrich Gauss (1777-1855).

Gauss was called the "prince of mathematicians" by his contemporaries. He did a great amount of work in pure mathematics as well as probability theory of errors, geodesy, mechanics, electro-magnetism and optics.

To apply the method of Gaussian elimination, we first reduce |AB| to its row echelon form.

Then, we write out the equations and solve them, which is simple.
Let us illustrate the method:

Example 15: Solve the following system by using the Gaussian elimination process.

$$x + 2y + 3z = 1$$
$$2x + 4y + z = 2$$

Solution: The given system is the same as
We first reduce the coefficient matrix to echelon form:

$$\begin{pmatrix} 1 & 2 & 3 & 1 \\ 2 & 4 & 1 & 2 \end{pmatrix} \approx R_1(-2)\begin{pmatrix} 1 & 2 & 3 & 1 \\ 0 & 0 & -5 & 0 \end{pmatrix}$$
$$\approx R_2\left(-1/5\right)\begin{pmatrix} 1 & 2 & 3 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

This gives us an equivalent to x = 2y and z = 0.
$x + 2y + 3z = 1$ and $z = 0$
These are again equivalent to $x = 1 - 2y$ and $z = 0$.
We get the solution in terms of a parameter.
Put$y = \alpha$, then $x = 1 - 2\alpha$, $y = \alpha$ and $z = 0$ is a solution, for any scalar α, thus, the solution set is $\{(1 - 2\alpha, \alpha, 0) | \alpha \in R\}$.
Now let us look at an example where *B =0*, that is, the system is homogeneous

Example 16: Obtain a solution a solution set of the simultaneous equations.

$$x_1 + 2x_2 + 5x_4 = 0$$
$$2x_1 + x_2 + 7x_3 + 6x_4 = 0$$
$$4x_1 + 5x_2 + 7x_3 + 16x_4 = 0$$

Solution: The matrix of coefficients is $A = \begin{bmatrix} 1 & 2 & 0 & 5 \\ 2 & 1 & 7 & 6 \\ 4 & 5 & 7 & 16 \end{bmatrix}$

The given system is equivalent to of $AX = 0$.

A row-reduced echelon form of this matrix is $\begin{bmatrix} 1 & 2 & 0 & 5 \\ 0 & 1 & -7/3 & 4/3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

Then the given system is equivalent to

$$x_1 + 2x_2 + 5x_4 = 0 \Rightarrow x_1 = -2x_2 - 5x_4$$
$$x_2 - \tfrac{7}{3}x_3 + \tfrac{4}{3}x_4 = 0 \Rightarrow x_2 = \tfrac{7}{3}x_3 - \tfrac{4}{3}x_4$$
$$\Rightarrow x_1 = -\tfrac{14}{3}x_3 - \tfrac{7}{3}x_4$$

which is the solution in terms of *z* and *t*.
Thus, the solution set of the given system of equations, in terms of two parameters α and β is $\left\{\left(-\frac{14}{3}\alpha - \frac{7}{3}\beta, \frac{7}{3}\alpha - \frac{4}{3}\beta, \alpha, \beta\right) | \alpha, \beta \in R\right\}$.
This is a two-dimensional vector subspace of $R^4$with basis

$$\left\{\left(-\frac{14}{3}, \frac{7}{3}, 1,0\right), \left(\frac{-14}{3}, \frac{-4}{3}, 0,1\right)\right\}$$

For practice, attempt the following exercise.

Example 17: Use the Gaussian method to obtain solution sets of the following system of
equations:

$$4x_1 - 3x_2 + x_3 = 7$$
$$x_1 - 2x_2 - 2x_3 = 3$$
$$3x_1 - x_2 + 2x_3 = -1$$

Solution: The augmented matrix is $\begin{bmatrix} 4 & -3 & 1 & 7 \\ 1 & -2 & -2 & 3 \\ 3 & -1 & 2 & -1 \end{bmatrix}$

Its row-reduced echelon form is $\begin{bmatrix} 1 & -2 & 21 & 3 \\ 0 & 1 & 9/5 & -1 \\ 0 & 0 & 1 & 5 \end{bmatrix}$

The given system is equivalent to

$$x_1 - 2x_2 - 2x_3 = 3$$
$$x_2 + \frac{9}{5}x_3 = -1$$
$$x_3 = 5$$

We can solve this system to get the unique solution $x_1 = -7, x_2 = -10, x_3 = 5$

Theorem 3.9: If a set of linear equations undergoes any of the following operations, then the resulting set of equations has exactly the same solution set as the original set of equations:

i. Multiplication of any equation by a non-zero constant
ii. Interchange of two equations
iii. Addition to an equation the result of multiplying another equation by a constant

Proof: Part iii) shall be proved, i) and ii) are left for the students to attempt.

Let$(s_1, s_2 \cdots, s_n)$ be a solution to

$$a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$
$$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n = b_i$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$
$$a_{p1}x_1 + a_{p2}x_2 + \cdots + a_{pn}x_n = b_p$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$
$$a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m$$

Suppose we add to equation (p) the result of multiplying equation (i) by r, then the new system has a different equation (p):

$$(ra_{i1} + a_{p1})x_1 + (ra_{i2} + a_{p2})x_2 + \cdots + (ra_{in} + a_{pn})x_n = rb_i + b_p$$

Since the rest of the equations are interchanged, $(s_1, s_2 \cdots, s_n)$ is still a solution for those equations.

Substituting $s_1 = x_1 \cdots, s_n = x_n$ in the new equation (p) and multiplying, we have

$r(a_{i1}s_1 + \cdots + a_{in}s_n) + (a_{i1}s_1 + \cdots + a_{in}s_n) = rb_i + b_p$and

$(s_1, s_2 \cdots, s_n)$ is a solution of the new equations (p).

Conversely, suppose that $(t_1, t_2 \cdots, t_n)$ is any solution of the new system, that is, suppose that

$$a_{11}t_1 + \cdots + a_{1n}t_n = b_1$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$
$$(ra_{i1} + a_{p1})t_1 + \cdots + (ra_{in} + a_{pn})t_n = rb_i + b_p$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$
$$a_{m1}x_1 + \cdots + a_{mn}x_n = b_m$$

Multiply equation (i) by $-r$ and add the result of equation (p) to obtain the original system with $(t_1, t_2 \cdots, t_n)$ is a solution.

To streamline our work, in the next example we use abbreviations to indicate the algebraic operation done, For example,

-

Abbreviation	Meaning
2E1	Multiply Eq. (1) by 2
$E3 \rightleftharpoons E4$	Interchange Eq. (3) and (4)
-2E1+E5	Add the result of multiplying Eq. (1) by -2 to Eq. (5), and replace Eq. (5) by the result

Example 18: Solve the system of equations:

$$x_1 + 2x_2 + 3x_3 = 4$$
$$4x_1 + 5x_2 + 6x_3 = 7$$
$$7x_1 + 8x_2 + 9x_3 = 10$$

Solution:

$$\begin{matrix} x_1 + 2x_2 + 3x_3 = 4 \\ 4x_1 + 5x_2 + 6x_3 = 7 \\ 7x_1 + 8x_2 + 9x_3 = 10 \end{matrix} \quad \xrightarrow[-7E1+E3]{-4E1+E2} \quad \begin{matrix} x_1 + 2x_2 + 3x_3 = 4 \\ -3x_2 - 6x_3 = -9 \\ -6x_1 - 12x_3 = -18 \end{matrix}$$

$$\xrightarrow{-2E2+E3} \begin{matrix} x_1 + 2x_2 + 3x_3 = 4 \\ -3x_2 - 6x_3 = -9 \\ 0 = 0 \end{matrix} \xrightarrow{\frac{1}{3}E2} \begin{matrix} x_1 + 2x_2 + 3x_3 = 4 \\ x_2 + 2x_3 = 3 \\ 0 = 0 \end{matrix}$$

Now we have only two equations in three unknowns.

In the second equation, we can let $x_3 = k$, where k is any complex number.

Then $x_2 = 3 - 2k$.

Substituting $x_3 = k$ and $x_3 = 3 - 2k$ into the first equation, we have

$$x_1 = 4 - 2x_2 - 3x_3 = 4 - 2(3 - 2k) - 3(k) = -2 + k$$

Thus, the general solution is

$$(-2 + k, 3 - 2k, k) \text{ or } \begin{matrix} x_1 = -2 + k \\ x_2 = 3 - 2k \\ x_3 = k \end{matrix}$$

Thus, we see that the system has an infinite number of solutions.

Note that specific solutions can be generated by choosing specific values for k
We are now near the end of this unit.

SELF-ASSESSMENT EXERCISE(S)

1. Determine if the matrix $A = \begin{bmatrix} 1 & 4 & 2 \\ 2 & 1 & 4 \\ 4 & 2 & 1 \end{bmatrix}$ is invertible. If it is invertible, find its inverse.
2. Consider the system of 3 equations in 3 unknowns:

$$\begin{aligned} 2x + y - z &= -1 \\ x - y + z &= 5 \\ 3x - y - 2z &= 0 \end{aligned}$$

How many solutions does it have which are linearly independent over R?
3. Solve the following systems of equations:

a) $$\begin{aligned} x_1 + 2x_2 + 3x_3 &= 4 \\ 4x_1 + 5x_2 + 6x_3 &= 7 \\ 7x_1 + 8x_2 + 9x_3 &= 12 \end{aligned}$$

b) $$\begin{aligned} x_1 + 2x_2 + 3x_3 &= 0 \\ 4x_1 + 5x_2 + 6x_3 &= 0 \\ 7x_1 + 8x_2 + 9x_3 &= 0 \\ 10x_1 + 11x_2 + 12x_3 &= 0 \end{aligned}$$

c) $$\begin{aligned} x_1 - 2x_2 + x_3 &= 0 \\ 3x_1 - x_2 + x_3 &= 0 \\ -x_1 + 4x_2 - x_3 &= 0 \end{aligned}$$

Each system is in echelon form.
For each, say whether the system has a unique solution, no solution, or infinitely many solutions.

a)
$$\begin{aligned} -3x + 2y &= 0 \\ -2y &= 0 \end{aligned}$$

b) $$\begin{aligned} x + y &= 4 \\ y - z &= 0 \end{aligned}$$

c) $$\begin{aligned} x_1 + x_2 &= 4 \\ x_2 - x_3 &= 0 \\ 0 &= 0 \end{aligned}$$

d) $$\begin{aligned} x_1 + x_2 - 3x_3 &= -1 \\ x_2 - x_3 &= 2 \\ x_3 &= 0 \\ 0 &= 0 \end{aligned}$$

Conclusion

In conclusion, we could see that calculating the rank of a matrix can be used to detect whether a simultaneous system of equations has a solution or not. In this unit, we have established the relationship between rank of a matrix with the inverse of the matrix.

3.4 Summary

In this unit we covered the following points.

- The row rank, column rank and rank of a matrix have been defined, and shown to be equal.
- The rank of a linear transformation is equal to the rank of its matrix.
- Elementary row and column operations have been defined
- You have been shown how to reduce a matrix to the row-reduced echelon form.
- the echelon form has been used to obtain the inverse of a matrix.
- The number of linearly independent solutions of a homogenous system of equations given by the matrix equation $AX = 0$ is $n - r$, where r = rank of A and n = number of columns of A.
- The system of linear equations given by the matrix equation $AX = B$ is consistent if and only if $p(A) = p([AB])$.
- A system of linear equations could be solved by the process of successive elimination of variables, which is, the Gaussian method.

3.5 References/Further Readings

Robert A. Beezer (2014). A First Course in Linear Algebra Congruent Press, Gig Harbor, Washington, USA 3(40).

Arbind K Lal Sukant Pati (2018). Linear Algebra through Matrices.

Peter J. Cameron (2008). Notes on Linear Algebra.

K.A. Stroud and Dexter J. Booth (2007). Engineering Mathematics 6th Edition. Palgrave Macmillan.

Abolarin, O.E. and Ishola, Y.C. (2005). Business Mathematics II. Jube Evans Books and Publications

Jim Hefferon (2020) LINEAR ALGEBRA. hefferon.net/linearalgebra. Fourth edition.

MODULE 3

Unit 1 Determinants I
Unit 2 Determinants II

UNIT 1 DETERMINANTS I

Unit Structure

1.1 Introduction
1.2 Learning Outcomes
1.3 Main Content
    1.3.1 Defining Determinant
    1.3.2 Determinant as Area and Volume:
    1.3.3 Properties of Determinants
    1.3.4 Minor, Cofactors and Adjoint of a Matrix
    1.3.5 Finding the Inverse of a Matrix
    1.3.6 Solution of System of Linear Equations using Determinant
1.4 Summary
1.5 References/Further Readings

1.1 Introduction

Having considered matrices in the last module, it is important to note the fundamental difference between a matrix and the determinant of a matrix. While a matrix is an array of numbers (elements), each of which has its own distinct position in the array; the determinant of a matrix is a number produced by combining the elements of the matrix in a prescribed manner. Note that determinants of square matrices only can be defined. In Unit 3 of module 2, we discussed the successive elimination method for solving a system of linear equations. This unit introduces you to another method, which depends on the concept of a determinant function. Determinants were used by the German mathematician Leibniz (1646 – 716) and the Swiss Mathematician, Vander Monde (1735-1796) who gave the first systematic presentation of the theory of determinants.

Several ways of developing the theory of determinants have been considered in section 3along with the study of the properties of determinants and certain other basic facts about them. We go on to give applications in solving a system of linear equations (Cramer's Rule) and obtaining the inverse of a matrix.

We also advise you to revise the units of module 2 before starting this unit.

1.2 Learning Outcomes

By the end of this unit, you will be able to:

- Define the determinant of a matrix.
- Evaluate the determinant of a square matrix using properties of determinants.
- Obtain the minor, cofactors and adjoint of a square matrix.
- Compute the inverse of an invertible matrix, using its adjoint.
- Apply Cramer’s rule to solve system of linear equations.

1.3 Main Content

1.3.1 Defining Determinant

There are many ways of introducing and defining the determinant function from$M_n(F)$ to F. In this section we give one of them, the classical approach. This was given by the French Mathematician Laplace (1749 -1827), and still very much in use.

The determinant is a function defined on square matrices; its value is a scalar. It has some very important properties: perhaps most important is the fact that a matrix is invertible if and only if its determinant is not equal to zero.

We denote the determinant function by det, so that det(A) is the determinant of A. For a matrix written out as an array, the determinant is denoted by replacing the square brackets by vertical bars,$|\ |$, for example, $det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$.

It is to be noted that there are two other methods of obtaining determinants – via permutations and via multi-linear forms. We shall not be doing these methods in this course. For the purpose of actual calculation of determinants the method that already given is normally used. The other methods are used to prove various properties of determinants.

So far, we have looked at determinant algebraically only; there is a geometrical interpretation of determinants also, which shall be treated now.

1.3.2 Determinants as Area and Volume:

Let $u = (a_2, a_2)$ and $v = (b_2, b_2)$ be two vectors in R^2. Then, the magnitude of the area of the parallelogram spanned by *u* and *v* (see fig. 1) is the absolute value of

$$det(u, v) = \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}.$$

Thus, if $u_1, u_2, \cdots, u_n$ are n vectors in R^n, then the absolute is the magnitude of the volume of the *n*-dimensional box spanned by $u_1, u_2, \cdots, u_n$.

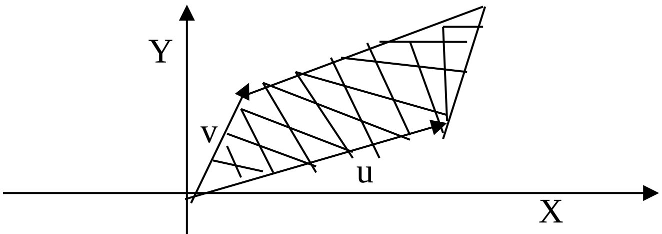

Fig.1 The shaded area is det(u,v)

Example 1: What is the magnitude of the volume of the box in R^3, spanned by i, j and k?

$$\begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} = 1$$

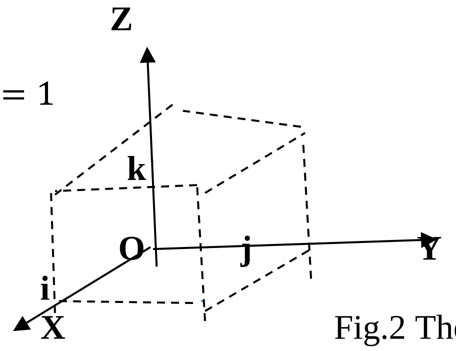

Fig.2 The magnitude of volume

1.3.3 Properties of Determinants

In this section we shall state some properties of determinants, mostly proof. We will take examples and check that these properties hold for them.

Now, for any $A \in M_n(F)$, then we have the following 8 properties, P1 – P8.

P1: If every element of a row (column) of a square matrix A is zero then$|A| = 0$.
P2: If A is an upper (lower) triangular matrix or it is a diagonal matrix, its determinant is the
product of its diagonal elements.
P3: If $A^T$is the transpose of matrix A, then $|A^T| = |A|$
P4: If two rows (columns) of A are identical, then $|A| = 0$
P5: If matrix B is obtained from A by multiplying its i^{th} row (column) by a non-scalar K,
then $|B| = K|A|$
P6: If matrix B is obtained from A by adding to its row the product of K (a scalar) and its j^{th}
row, then $|B| = |A|$.
The theorem also holds when row is replaced by column throughout.
P7: A square matrix A is non-singular if and only if $|A| \neq 0$.
P8: The determinant of the product of two matrices is equal to the product of the determinants
of the two matrices, that is, $|AB| = |A||B|$
We would apply the properties P1 – P8 to some matrices to prove the properties.

Example 2: Solve the following determinants:

i) $|A| = \begin{vmatrix} 1 & 4 & 2 \\ 3 & 0 & 5 \\ 1 & 4 & 2 \end{vmatrix}$ ii) $|B| = \begin{vmatrix} 1 & 3 & 1 & 1 \\ 5 & 1 & 5 & 2 \\ 0 & 7 & 0 & 3 \\ -2 & 0 & -2 & 4 \end{vmatrix}$ iii) $|C| = \begin{vmatrix} 1 & 4 & 2 \\ 0 & 3 & 5 \\ 0 & 0 & 2 \end{vmatrix}$

iv) $|D| = \begin{vmatrix} -3 & 0 & 0 \\ 3 & 4 & 0 \\ 1 & 4 & 2 \end{vmatrix}$ v) $|E| = \begin{vmatrix} -1 & 2 & 0 \\ 5 & 4 & 0 \\ 1 & 4 & 0 \end{vmatrix}$ vi) $|F| = \begin{vmatrix} 4 & 3 & -7 \\ 0 & 0 & 0 \\ 1 & 8 & 12 \end{vmatrix}$

Solution:

i) since the 1st and 3rd rows coincide, then $|A| = 0$
ii) since the 1st and 3rd columns coincide, then$|B| = 0$
iii) C is an upper triangular matrix, hence $|C| = 1x2x3 = 6$
iv) D is a lower triangular matrix, hence $|D| = -3 \times 4 \times 2 = -24$
v) All the elements in the 3rd column are zero, then $|E| = 0$
vi) All the elements in the 2nd row are zero, then $|F| = 0$

Example 3: Let $A = \begin{vmatrix} a & b & b & b \\ b & a & b & b \\ b & b & a & b \\ b & b & b & a \end{vmatrix}$, find $|A|$

Solution: $|A| = \begin{vmatrix} a & b & b & b \\ b & a & b & b \\ b & b & a & b \\ b & b & b & a \end{vmatrix}$

Add the 2nd, 3rd and 4th rows to the 1st row

$$|A| = \begin{vmatrix} a+3b & a+3b & a+3b & a+3b \\ b & a & b & b \\ b & b & a & b \\ b & b & b & a \end{vmatrix}$$

Subtract the 1st column from every other column

$$= \begin{vmatrix} a+3b & 0 & 0 & 0 \\ b & a-b & 0 & 0 \\ b & 0 & a-b & 0 \\ b & 0 & 0 & a-b \end{vmatrix}$$

Expanding along the 1st row gives

$$|A| = (a+3b)\begin{vmatrix} a-b & 0 & 0 \\ 0 & a-b & 0 \\ 0 & 0 & a-b \end{vmatrix} \quad \text{(a diagonal matrix)}$$

Hence $|A| = (a+3b)(a-b)^3$

Example 4: Show that $\begin{vmatrix} 1 & 1 & 1 & 1 \\ x_1 & x_2 & x_3 & x_4 \\ x_1^2 & x_2^2 & x_3^2 & x_4^2 \\ x_1^3 & x_2^3 & x_3^3 & x_4^3 \end{vmatrix} = \prod_{i<j}(x_i - x_j); 1 \le i \le j \le 4$.

This is known as the Vandermonde's determinant of order 4.

Solution:

Step1: subtract the 1st column from every other column

$$= \begin{vmatrix} 1 & 0 & 0 & 0 \\ x_1 & x_2 - x_1 & x_3 - x_1 & x_4 - x_1 \\ x_1^2 & x_2^2 - x_1^2 & x_3^2 - x_1^2 & x_4^2 - x_1^2 \\ x_1^3 & x_2^3 - x_1^3 & x_3^3 - x_1^3 & x_4^3 - x_1^3 \end{vmatrix}$$

Step 2: expand along the first row and factorizing the entries

$$= 1\begin{vmatrix} x_2 - x_1 & x_3 - x_1 & x_4 - x_1 \\ x_2^2 - x_1^2 & x_3^2 - x_1^2 & x_4^2 - x_1^2 \\ x_2^3 - x_1^3 & x_3^3 - x_1^3 & x_4^3 - x_1^3 \end{vmatrix}$$

$$= \begin{vmatrix} x_2 - x_1 & x_3 - x_1 & x_4 - x_1 \\ (x_2 - x_1)(x_2 + x_1) & (x_3 - x_1)(x_3 + x_1) & (x_4 - x_1)(x_4 + x_1) \\ (x_2 - x_1)(x_2^2 + x_1^2 + x_2 x_1) & (x_3 - x_1)(x_3^2 + x_1^2 + x_3 x_1) & (x_4 - x_1)(x_4^2 + x_1^2 + x_4 x_1) \end{vmatrix}$$

Step 3: Take out $(x_2 - x_1), (x_3 - x_1)$ and $(x_4 - x_1)$ from columns 1, 2 and 3 respectively

$$= (x_2 - x_1)(x_3 - x_1)(x_4 - x_1)\begin{vmatrix} 1 & 1 & 1 \\ x_2 + x_1 & x_3 + x_1 & x_4 + x_1 \\ x_2^2 + x_1^2 + x_2x_1 & x_3^2 + x_1^2 + x_3x_1 & x_4^2 + x_1^2 + x_4x_1 \end{vmatrix}$$

Step 4: subtracting the first column from the second and third columns

$$= (x_2 - x_1)(x_3 - x_1)(x_4 - x_1)\begin{vmatrix} 1 & 1 & 1 \\ x_2 + x_1 & x_3 - x_2 & x_4 - x_2 \\ x_2^2 + x_1^2 + x_2x_1 & x_3^2 - x_2^2 + x_3x_1 - x_2x_1 & x_4^2 - x_2^2 + x_4x_1 - x_2x_1 \end{vmatrix}$$

Step 5: expand by the first row and factorizing the entries

$$= (x_2 - x_1)(x_3 - x_1)(x_4 - x_1)[(x_3 - x_2)(x_4^2 - x_2^2 + x_4x_1 - x_2x_1) - (x_4 - x_2)(x_3^2 - x_2^2 + x_3x_1 - x_2x_1)]$$

$$= (x_2 - x_1)(x_3 - x_1)(x_4 - x_1)\{(x_3 - x_2)[(x_4 - x_2)(x_4 + x_2) + x_1(x_4 - x_2)] - (x_3 - x_2)[(x_4 - x_2)(x_4 + x_2) + x_1(x_4 - x_2)](x_2x_1)\}$$

$$= (x_2 - x_1)(x_3 - x_1)(x_4 - x_1)\{(x_3 - x_2)[(x_4 - x_2)(x_4 + x_2 + x_1)] - (x_4 - x_2)[(x_3 - x_2)(x_4 + x_2 + x_1)]\}$$

$$= (x_2 - x_1)(x_3 - x_1)(x_4 - x_1)\{(x_3 - x_2)(x_4 - x_2)[(x_4 + x_2 + x_1) - (x_3 + x_2 + x_1)]\}$$

$$= (x_2 - x_1)(x_3 - x_1)(x_4 - x_1)\{(x_3 - x_2)(x_4 - x_2)(x_4 - x_3)\}$$

$$= (x_2 - x_1)(x_3 - x_1)(x_3 - x_2)(x_4 - x_1)(x_4 - x_2)(x_4 - x_3)$$

$$= \prod_{i<j}(x_i - x_j); 1 \leq i \leq j \leq 4$$

Example 5:

Let us define the function $\theta(t)$ by $\theta(t) = \begin{vmatrix} f(t) & g(t) \\ f'(t) & g'(t) \end{vmatrix}$. Show that $\theta'(t) = \begin{vmatrix} f(t) & g(t) \\ f''(t) & g''(t) \end{vmatrix}$

Solution:

$$\theta(t) = \begin{vmatrix} f(t) & g(t) \\ f'(t) & g'(t) \end{vmatrix} = f(t)g'(t) - g(t)f'(t)$$

$\theta'(t) = f'(t)g'(t) + f(t)g''(t) - \{g'(t)f'(t) + g(t)f''(t)\}$ [Since $\frac{d}{dt}(fg) = \frac{df}{dt}g + \frac{dg}{dt}f$]

$$= f(t)g''(t) - g(t)f''(t) = \begin{vmatrix} f(t) & g(t) \\ f''(t) & g''(t) \end{vmatrix}$$

Theorem 1.1: Let $A = [a_{ij}]_n x_n$, then

a) $a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in} = det(A) = a_{1i}C_{1i} + a_{2i}C_{2i} + \cdots + a_{ni}C_{ni}$

b) $a_{i1}C_{j1} + a_{i2}C_{j2} + \cdots + a_{in}C_{jn} = 0 = a_{1i}C_{ji} + a_{2i}C_{2j} + \cdots + a_{ni}C_{nj}$ if $i \neq j$

We will not be proving this theorem here.

We only mention that (a) follows immediately from the definition of det (A), since

$$det(A) = (-1)^{i+1}a_{i1}|A_{i1}| + \cdots + (-1)^{i+n}a_{in}|A_{in}|$$

The following example will help you to get used to calculating determinants.

We will define the determinant function $det: M_n(F) \rightarrow F$ by induction on n. That is, we will define it for $n = 1,2,3, \ldots$ and then define it for any n, assuming the definition for $n - 1$.

When n = 1, for any $A \in M_1(F)$ we have $a = [a]$, for some $a \in F$. In this case we define

$det(A) = det([a]) = a.$

For example, $det(5) = 5$.

When n = 2, for any $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \in M_2(F)$, we define $det(A)$ using the definition for the case $n = 2$ as followings:

$$det(A) = a_{11}\, det([a_{22}]) - a_{12}\, det([a_{21}])$$

Example 6: $det\left(\begin{bmatrix} 0 & 1 \\ -2 & 3 \end{bmatrix}\right) = 0 \times 3 - 1 \times (-2) = 2$

Using the definition for the case $n = 2$ as followings:

When n = 3, for any $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \in M_3(F)$, $det(A)$ is defined as

$$det(A) = |A|$$
$$= a_{11}\, det\left(\begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix}\right) - a_{12}\, det\left(\begin{bmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{bmatrix}\right) + a_{13}\, det\left(\begin{bmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}\right)$$
$$= a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31})$$

Example 7: $det\left(\begin{bmatrix} 1 & 2 & -3 \\ -2 & 3 & 5 \\ -1 & 4 & 0 \end{bmatrix}\right) = 1\begin{vmatrix} 3 & 5 \\ 4 & 0 \end{vmatrix} - 2\begin{vmatrix} -2 & 5 \\ -1 & 0 \end{vmatrix} + (-3)\begin{vmatrix} -2 & 3 \\ -1 & 4 \end{vmatrix}$

$$= 1(0 - 20) - 2\big(0 - (-5)\big) - 3\big(-8 - (-3)\big)$$
$$= 1(-20) - 2(5) - 3(-5) = -20 - 10 + 15 = -15$$

$|A|$ could also be calculated from the second row as follows:

$$det\left(\begin{bmatrix} 1 & 2 & -3 \\ -2 & 3 & 5 \\ -1 & 4 & 0 \end{bmatrix}\right) = -(-2)\begin{vmatrix} 2 & -3 \\ 4 & 0 \end{vmatrix} + 3\begin{vmatrix} 1 & -3 \\ -1 & 0 \end{vmatrix} - 5\begin{vmatrix} 1 & 2 \\ -1 & 4 \end{vmatrix}$$
$$|A| = 2(0 - -12) + 3(0 - 3) - 5(4 - -2) = 24 - 9 - 30 = -15$$

Exercise: Calculate the $|A|$ using the third row. What do you notice?
Now, let us see how this definition is extended to define det(A) for any $n \times n$ matrix A, $n \neq 1$.

$$\text{Let } A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{(n-1)1} & \cdots & \cdots & a_{(n-1)(n-1)} \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

$det(A) = (-1)^{1+1}a_{11}det(A_{11}) + (-1)^{1+2}a_{12}det(A_{12}) + \cdots + (-1)^{1+n}a_{1n}det(A_{1n})$,
where A_{ij} is the $(n-1) \times (n-1)$ matrix obtained from A by deleting the i^{th} row and the j^{th} column, and i is a fixed integer with $1 \leq i \leq n$.
Thus, we see that $det(A) = \sum_{j=1}^{n}(-1)^{i+j}a_{ji}det(A_{ji})$ define the determinant of an $n \times n$ matrix A in terms of the determinants of the $(n-1) \times (n-1)$ matrices a_{ji} ; $i, j = 1,2, \cdots, n$

Note: while calculating |A|, it would be most preferable to expand along a row that has the maximum number of zeros, this cut downs the number of terms to be calculated.

The following example will help you to get used to calculating determinants.

Example 8: If $A = \begin{pmatrix} -3 & -2 & 0 & 2 \\ 2 & 1 & 0 & -1 \\ 1 & 0 & 1 & 2 \\ 2 & 1 & -3 & 1 \end{pmatrix}$, calculate $|A|$.

Solution: The first three rows have one zero each. Let us expand along third row. Observe that $a_{32} = 0$. So, we don't need to calculate A_{32}. Now,
$A_{31} = \begin{pmatrix} -2 & 0 & 2 \\ 1 & 0 & 1 \\ 1 & -3 & 1 \end{pmatrix}$, $A_{33} = \begin{pmatrix} -3 & -2 & 2 \\ 1 & 1 & -1 \\ 1 & 1 & 1 \end{pmatrix}$, $A_{34} = \begin{pmatrix} -3 & -2 & 0 \\ 2 & 1 & 0 \\ 2 & 1 & -3 \end{pmatrix}$

We will obtain $|A_{31}|$, $|A_{33}|$and $|A_{34}|$ by expanding along the second, third and second rows, respectively.

$$|A_{31}| = \begin{vmatrix} -2 & 0 & 2 \\ 1 & 0 & 1 \\ 1 & -3 & 1 \end{vmatrix}$$

$$= (-1)^{2+1} \cdot 1 \begin{vmatrix} 0 & 2 \\ -3 & 1 \end{vmatrix} + (-1)^{2+2} \cdot 0 \begin{vmatrix} -2 & 2 \\ 1 & -1 \end{vmatrix} + (-1)^{2+3} \cdot (-1) \begin{vmatrix} -2 & 0 \\ 1 & -3 \end{vmatrix}$$

(Expansion along the second row)

$$= (-1) \cdot 6 + 0 + (-1) \cdot 6$$
$$= -6 + 6 = 0$$

$$|A_{33}| = \begin{vmatrix} -3 & -2 & 2 \\ 2 & 1 & -1 \\ 2 & 1 & 1 \end{vmatrix}$$

$$= (-1)^{3+1} \cdot 2 \begin{vmatrix} -2 & 2 \\ 1 & -1 \end{vmatrix} + (-1)^{3+2} \cdot 1 \begin{vmatrix} -3 & -2 \\ 2 & 1 \end{vmatrix} + (-1)^{3+3} \cdot (1) \begin{vmatrix} -3 & -2 \\ 2 & 1 \end{vmatrix}$$

(Expansion along the third row)

$$= (1)(2)(0) + (-1)(1)(-1) + (1)(1)(1)$$
$$= 0 + 1 + 1 = 2$$

$$|A_{34}| = \begin{vmatrix} -3 & -2 & 0 \\ 2 & 1 & 0 \\ 2 & 1 & -3 \end{vmatrix}$$

$$= (-1)^{2+1} \cdot 2 \begin{vmatrix} -2 & 0 \\ 1 & -3 \end{vmatrix} + (-1)^{2+2} \cdot 1 \begin{vmatrix} -3 & 0 \\ 2 & -3 \end{vmatrix} + (-1)^{2+3} \cdot (0) \begin{vmatrix} -3 & -2 \\ 2 & 1 \end{vmatrix}$$

(Expansion along the second row)

$$= (-1)(2)(6) + (1)(1)(9) + (-1)(0)(1)$$
$$= -12 + 9 + 0 = -3$$

Thus, the required determinant is given by

$$= a_{31}|A_{31}| + a_{32}|A_{32}| + a_{33}|A_{33}| + a_{34}|A_{34}|$$
$$= (1)(0) - (0) + (1)(2) + (-2)(-3) = 8$$

1.3.4 Minor, Cofactors and Adjoint of a Matrix

Definition: Let A be an $n \times n$ matrix. For $1 \leq i, j \leq n$, we define the (i, j) **minor** of A to be the $(n-1) \times (n-1)$ matrix obtained by deleting the i^{th} row and j^{th} column of A. In other words, if the elements of i^{th} row and j^{th} column of A are removed from A, the **determinant** of the remaining $(n-1)$ square matrix is known as the first minor of det(A) denoted by m_{ij}, or more frequently as minor of a_{ij}.

Example 9: Consider the matrix $\begin{bmatrix} 1 & 2 & -3 \\ -2 & 3 & 5 \\ -1 & 4 & 0 \end{bmatrix}$

Minor of element (1) is $\begin{vmatrix} 3 & 5 \\ 4 & 0 \end{vmatrix}$

Minor of element (2) is $\begin{vmatrix} -2 & 5 \\ -1 & 0 \end{vmatrix}$

Minor of element (-3) is $\begin{vmatrix} -2 & 3 \\ -1 & 4 \end{vmatrix}$ etc.

Now we define the (i,j) **cofactor** of A to be $(-1)^{i+j}$ times the determinant of the (i,j) minor, where $(-1)^{i+j}$ is called the **place sign** of element a_{ij}.

Note: These place signs have a chessboard pattern, starting with sign '+' in the top left corner, that is,
Place sign of a_{11} is $(-1)^{1+1} = +$ **Place sign** of a_{12} is $(-1)^{1+2} = -$
Place sign of a_{13} is $(-1)^{1+3} = +$ **Place sign** of a_{21} is $(-1)^{2+1} = -$
Place sign of a_{22} is $(-1)^{2+2} = +$ **Place sign** of a_{23} is $(-1)^{2+3} = -$
Place sign of a_{31} is $(-1)^{3+1} = +$ **Place sign** of a_{32} is $(-1)^{3+2} = -$
Place sign of a_{33} is $(-1)^{3+3} = +$

$$\begin{matrix} + & - & + \\ - & + & - \\ + & - & + \end{matrix}$$

Cofactor of a_{ij} is the place sign of a_{ij} multiply by the minor of a_{ij}, that is,

Cofactor of a_{11} is $(-1)^{1+1}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} = +(a_{22}a_{33} - a_{23}a_{32})$

Cofactor of a_{12} is $(-1)^{1+2}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} = -(a_{21}a_{33} - a_{23}a_{31})$

Cofactor of a_{13} is $(-1)^{1+3}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} = +(a_{21}a_{32} - a_{22}a_{31})$

Cofactor of a_{21} is $(-1)^{2+1}\begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} = -(a_{12}a_{33} - a_{13}a_{32})$

Cofactor of a_{22} is $(-1)^{2+2}\begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} = +(a_{11}a_{33} - a_{13}a_{31})$

Cofactor of a_{23} is $(-1)^{2+3}\begin{vmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{vmatrix} = -(a_{11}a_{32} - a_{12}a_{31})$

Cofactor of a_{31} is $(-1)^{3+1}\begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} = +(a_{12}a_{23} - a_{13}a_{22})$

Cofactor of a_{32} is $(-1)^{3+2}\begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix} = -(a_{11}a_{23} - a_{13}a_{21})$

Cofactor of a_{33} is $(-1)^{3+3}\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = +(a_{11}a_{22} - a_{12}a_{21})$

The (i,j) **cofactor** of A is denoted by $C_{ij}(A)$.

In the example above, **cofactor** of $1 = C_{11}$would be $+\begin{vmatrix} 3 & 5 \\ 4 & 0 \end{vmatrix} = 3 \cdot 0 - 5 \cdot 4 = -20$,

Cofactor of $(2) = C_{12}$ is $-\begin{vmatrix} -2 & 5 \\ -1 & 0 \end{vmatrix} = -[(-2)0 - 5(-1)] = -5$

Cofactor of $(-3) = C_{13}$ is$+\begin{vmatrix} -2 & 3 \\ -1 & 4 \end{vmatrix} = [(-2)4 - 3(-1)] = -5$

Cofactor of $(-2) = C_{21}$ is $-\begin{vmatrix} 2 & -3 \\ 4 & 0 \end{vmatrix} = -[(2)0 - (-3)4] = -12$

Cofactor of $(3) = C_{22}$ is $+\begin{vmatrix} 1 & -3 \\ -1 & 0 \end{vmatrix} = [1.0 - (-3)(-1)] = -3$

Cofactor of $(5) = C_{23}$ is $-\begin{vmatrix} 1 & 2 \\ -1 & 4 \end{vmatrix} = -[1.4 - 2(-1)] = -6$

Cofactor of $(-1) = C_{31}$ is $+\begin{vmatrix} 2 & -3 \\ 3 & 5 \end{vmatrix} = [(2)5 - (-3)(3)] = 19$

Cofactor of $(4) = C_{32}$ is $-\begin{vmatrix} 1 & -3 \\ -2 & 5 \end{vmatrix} = -[1.5 - (-3)(-2)] = 1$

Cofactor of $(0) = C_{33}$ is $+\begin{vmatrix} 1 & 2 \\ -2 & 3 \end{vmatrix} = -[1.3 - (2)(-2)] = 7$

Therefore, the matrix of cofactors of A denoted by $C(A)$ is $\begin{bmatrix} -20 & -5 & -5 \\ -12 & -3 & -6 \\ 19 & 1 & 7 \end{bmatrix}$.

Finally, the **adjoint** of A is the $n \times n$ matrix $Adj(A)$ whose (i, j) entry is the (j, i) cofactor $C_{ji}(A)$ of A or simply the transpose of $C_{ij}(A)$.

Thus, the **adjoint** of A is the $n \times n$ matrix is the transpose of the matrix of corresponding cofactors of A.

Definition: Let $A = [a_{ij}]$ be any $n \times n$ matrix. Then the adjoint of A is the $n \times n$ matrix, denoted by $Adj(A)$, and defined by

$$Adj(A) = \begin{bmatrix} C_{11} & C_{12} & \cdots & \cdots & C_{1n} \\ C_{21} & C_{22} & \cdots & \cdots & C_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ C_{n1} & C_{n2} & \cdots & \cdots & C_{nn} \end{bmatrix}^T = \begin{bmatrix} C_{11} & C_{21} & \cdots & \cdots & C_{n1} \\ C_{12} & C_{22} & \cdots & \cdots & C_{n2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ C_{1n} & C_{2n} & \cdots & \cdots & C_{nn} \end{bmatrix}$$

For the example above, $Adj(A) = C^T(A) = \begin{bmatrix} -20 & -5 & -5 \\ -12 & -3 & -6 \\ 19 & 1 & 7 \end{bmatrix}$.

Example10: Obtain the adjoint of the matrix$A = \begin{bmatrix} cos\ \theta & 0 & -sin\theta \\ 0 & 1 & 0 \\ sin\ \theta & 0 & cos\ \theta \end{bmatrix}$.

Solution:$|A| = \begin{vmatrix} cos\ \theta & 0 & sin\theta \\ 0 & 1 & 0 \\ -sin\ \theta & 0 & cos\ \theta \end{vmatrix} = cos\ \theta(cos\ \theta - 0) - 0(0 - 0) + sin\theta(sin\theta - 0)$

$$\therefore |A| = cos^2\theta + sin^2\theta = 1$$

$C_{11} = cos\theta;\ C_{12} = 0;\ C_{13} = -sin\theta;\ C_{21} = 0;\ C_{22} = cos^2\theta + sin^2\theta = 1;$

$C_{23} = 0; C_{31} = sin\theta; C_{32} = 0; C_{33} = cos\theta$

Therefore, the matrix of cofactors is $C_{ij}(A) = \begin{bmatrix} cos\ \theta & 0 & -sin\theta \\ 0 & 1 & 0 \\ sin\ \theta & 0 & cos\ \theta \end{bmatrix}$

Thus, adjoint of the matrix is $Adj(A) = \begin{bmatrix} cos\ \theta & 0 & -sin\theta \\ 0 & 1 & 0 \\ sin\ \theta & 0 & cos\ \theta \end{bmatrix}$.

1.3.5 Finding the inverse of a matrix

A method of finding out if a matrix is invertible is the use of adjoint of a matrix.
The following theorem uses the adjoint to give another way of finding out if a matrix A is invertible. It also gives usA^{-1}, if A is invertible.

Theorem 2: Let A be an $n \times n$ matrix over F, then $A \cdot \left(Adj(A)\right) = \left(Adj(A)\right) \cdot A = det(A)I$
Proof: Recall matrix multiplication from Module 2 unit 1.

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} C_{11} & C_{12} & \cdots & \cdots & C_{n1} \\ C_{12} & C_{22} & \cdots & \cdots & C_{n2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ C_{1n} & C_{2n} & \cdots & \cdots & C_{nn} \end{bmatrix}$$

Now, by Theorem 1, we know that $a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in} = det(A)$ and
$a_{i1}C_{j1} + a_{i2}C_{j2} + \cdots + a_{in}C_{jn} = 0$ if $i \neq j$.
Therefore,

$$A \cdot \left(Adj(A)\right) = \begin{bmatrix} det(A) & 0 & \cdots & \cdots & 0 \\ 0 & det(A) & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \vdots & \vdots & 0 \\ 0 & 0 & \cdots & \cdots & det(A) \end{bmatrix}$$

$$= det(A) \begin{bmatrix} 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & \cdots & \cdots & 0 \\ \vdots & \vdots & 1 & \vdots & \vdots \\ 0 & 0 & \vdots & 1 & 0 \\ 0 & 0 & \cdots & \cdots & 1 \end{bmatrix} = det(A)I$$

Similarly, $\left(Adj(A)\right) \cdot A = det(A)I$
An immediate corollary shows us how to calculate the inverse of a matrix, if it exists.

Corollary 1: Let A be an $n \times n$ matrix over F. Then A is invertible if and only if $det(A) \neq 0$, then $A^{-1} = \frac{1}{det(A)} Adj(A)$

Proof: If A is invertible, then A^{-1} exists and $A^{-1}A = I$.
So, by theorem 1, $det(A^{-1})det(A) = det(I) = I$
$\therefore \quad det(A) \neq 0$
Conversely, if $det(A) \neq 0$, then Theorem 2 says that

$$A\left(\frac{1}{det(A)} Adj(A)\right) = I = \left(\frac{1}{det(A)} Adj(A)\right) A$$

Therefore, $A^{-1} = \frac{1}{det(A)} Adj(A)$

For Example 9 above,

$$\begin{vmatrix} 1 & 2 & -3 \\ -2 & 3 & 5 \\ -1 & 4 & 0 \end{vmatrix} = 1(0-20) - 2\big(0-(-5)\big) + (-3)\big(-8 - (-3)\big) = -15$$

$$\therefore A^{-1} = \frac{1}{-15}\begin{bmatrix} -20 & -5 & -5 \\ -12 & -3 & -6 \\ 19 & 1 & 7 \end{bmatrix}.$$

For Example10, $A^{-1} = \frac{1}{1}\begin{bmatrix} cos\,\theta & 0 & -sin\theta \\ 0 & 1 & 0 \\ sin\,\theta & 0 & cos\,\theta \end{bmatrix} = \begin{bmatrix} cos\,\theta & 0 & -sin\theta \\ 0 & 1 & 0 \\ sin\,\theta & 0 & cos\,\theta \end{bmatrix}$

The process of obtaining the inverse of a system of equations using adjoint can be expressed in the following steps:

i. Evaluate the determinant of the matrix, say A
ii. Form a matrix of cofactors of the elements of the matrix A
iii. Write the transpose of C, that is, C^T which is the adjoint of the given matrix
iv. Divide the Adjoint by $|A|$
v. The resulting matrix is the inverse, A^{-1}, of matrix A

1.3.6 Solution of System of Linear Equations using Determinant

In unit 3 of module 2 above, we solved system of linear equations especially when the number of equations is not equal to the number of variables using the method of Gaussian elimination. A method of solving system of linear equations when the number of equations equals the number of variables is the use of determinant. In this section we shall give a rule derived by the mathematician, Cramer, for solving this system of linear equations.

Consider the system of n linear equations in n unknowns, given by

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

This can be written in matrix form:

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

Theorem 3: Let the matrix equation of a system of linear equations be $AX = B$, where

$$A = [a_{11}]_{n\times n};\ X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_n \end{bmatrix} \text{ and } B = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ \vdots \\ b_n \end{bmatrix}$$

Let the columns of A be $C_1, C_2, \cdots, C_n$. If det(A) ≠ 0, the given system has a unique solution, namely,$x_1 = \frac{D_1}{D}, x_2 = \frac{D_2}{D}, \cdots, x_n = \frac{D_n}{D}$,
$D_i = det(C_1, \cdots, C_{i-1}, \boldsymbol{B}, \cdots, C_n)$, that is, determinant of the matrix obtained from A by replacing the i^{th} column by B, and D = det (A).

Proof: Since |A| ≠ 0, the corollary 1 says that A^{-1} exists
Now $AX = B \Longrightarrow A^{-1}AX = A^{-1}B$

$$\Longrightarrow IX = \frac{1}{D}adj(A)B$$

$$X = \frac{1}{D}\begin{bmatrix} C_{11} & C_{21} & \cdots & \cdots & C_{n1} \\ C_{12} & C_{22} & \cdots & \cdots & C_{n2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ C_{1n} & C_{2n} & \cdots & \cdots & C_{nn} \end{bmatrix}\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ \vdots \\ b_n \end{bmatrix}$$

Thus, $$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_n \end{bmatrix} = \frac{1}{D}\begin{bmatrix} C_{11}b_1 + C_{21}b_2 + \cdots + C_{n1}b_1 \\ C_{12}b_1 + C_{22}b_2 + \cdots + C_{n2}b_2 \\ \vdots \\ \vdots \\ C_{1n}b_1 + C_{2n}b_2 + \cdots + C_{nn}b_n \end{bmatrix}$$

Thus, $$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_n \end{bmatrix} = \frac{1}{D}\begin{bmatrix} D_1 \\ D_2 \\ \vdots \\ \vdots \\ D_n \end{bmatrix}$$ this gives us Cramer's rule, namely,

$$x_1 = \frac{D_1}{D}, x_2 = \frac{D_2}{D}, \cdots, x_n = \frac{D_n}{D}$$

The following example and exercise may help you to practice using Cramer's rule

Example 11: Solve the following system using Cramer's rule:

a) $$\begin{aligned} 5x + 2y &= -19 \\ 3x + 4y &= -17 \end{aligned}$$

b) $$\begin{aligned} 2x_1 + 3x_2 - x_3 &= 4 \\ 3x_1 + x_2 + 2x_3 &= 13 \\ x_1 + 2x_2 - 5x_3 &= -11 \end{aligned}$$

Solution:

a) The given system is equivalent to AX = B, where

$$A = \begin{pmatrix} 5 & 2 \\ 3 & 4 \end{pmatrix}, X = \begin{bmatrix} x \\ y \end{bmatrix}, \mathrm{B} = \begin{bmatrix} -19 \\ -17 \end{bmatrix}$$

Now, $|A| = D = 20 - 6 = 14$

$$D_1 = \begin{vmatrix} -19 & 2 \\ -17 & 4 \end{vmatrix} = -76 + 34 = -42$$
$$D_2 = \begin{vmatrix} 5 & -19 \\ 3 & -17 \end{vmatrix} = -85 + 57 = -28$$

Applying the Crammer's rule to have

$$x = \frac{D_1}{D} = \frac{-42}{14} = -3,$$
$$y = \frac{D_2}{D} = \frac{-28}{14} = -2$$

Substitute these values in the given equations to check that we haven't made a mistake in our calculations.

b) The given system is equivalent to AX = B, where

$$A = \begin{pmatrix} 2 & 3 & -1 \\ 3 & 1 & 2 \\ 1 & 2 & -5 \end{pmatrix},\ X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix},\ \text{B} = \begin{bmatrix} 4 \\ 13 \\ -11 \end{bmatrix}$$

Now, $|A| = D = 28$

$$D_1 = \begin{vmatrix} 4 & 3 & -1 \\ 13 & 1 & 2 \\ -11 & 2 & -5 \end{vmatrix} = 56$$
$$D_1 = \begin{vmatrix} 3 & 4 & -1 \\ 1 & 13 & 2 \\ 2 & -11 & -5 \end{vmatrix} = 28$$
$$D_1 = \begin{vmatrix} 2 & 3 & 4 \\ 3 & 1 & 13 \\ 1 & 2 & -11 \end{vmatrix} = 84$$

Applying the Crammer's rule to have

$x_1 = \frac{D_1}{D} = \frac{56}{28} = 2,\ x_2 = \frac{D_2}{D} = \frac{28}{28} = 1$ and $x_3 = \frac{D_3}{D} = \frac{84}{28} = 3$

SELF-ASSESSMENT EXERCISE(S)

1. Solve (i) $\begin{vmatrix} 1 & 3 & 0 \\ 2 & 1 & 2 \\ 1 & 3 & 0 \end{vmatrix}$ (ii) $\begin{vmatrix} 2 & 3 & 5 \\ 1 & 0 & 1 \\ 4 & 6 & 10 \end{vmatrix}$ (iii) $\begin{vmatrix} a & 0 & 0 \\ \alpha & b & 0 \\ \beta & \delta & c \end{vmatrix}$ and (iv) $\begin{vmatrix} a & d & e \\ \alpha & b & f \\ 0 & 0 & c \end{vmatrix}$

2. If $A = \begin{bmatrix} 1 & 4 & 3 \\ 6 & 2 & 5 \\ 1 & 7 & 0 \end{bmatrix}$. Find (a) |A| (b) Adj(A) (c) A^{-1}

3. $B = \begin{bmatrix} 2 & 3 & -1 \\ 0 & 0 & 6 \\ 0 & 0 & 5 \end{bmatrix}$. Find (a) |B| (b) Adj(B) (c) B^{-1}

4. Solve the following systems of equations by the Cramer's rule:

a) $\begin{aligned} a + 2b + c &= 4 \\ 3a - 4b - 2c &= 2 \\ 5a + 3b + 5c &= -1 \end{aligned}$ b) $\begin{aligned} x_1 + 2x_2 - 3x_3 &= 3 \\ 2x_1 - x_2 - x_3 &= 11 \\ 3x_1 + 2x_2 + x_3 &= -5 \end{aligned}$ c) $\begin{aligned} x - 4y - 2z &= 21 \\ 2x + y + 2z &= 3 \\ 3x + 2y - z &= -2 \end{aligned}$

Conclusion

In conclusion, to obtain the inverse of a matrix using adjoint you have to first evaluate the determinant of the matrix, and matrix of cofactors of each of the elements of the matrix and follow the steps highlighted in section3 above. It has also been established that determinants have both algebraic and geometrical interpretation and can be used to find the magnitude of the areas and volumes of solids in terms of vectors in R^2 and R^3.

1.4 Summary

In this unit we have covered the following points

The definition of the determinant of a square matrix as well as the properties (P1-P8) of determinants were stated.

1.7 References/Further Readings

Robert A. Beezer (2014). A First Course in Linear Algebra Congruent Press, Gig Harbor, Washington, USA 3(40).

Arbind K Lal Sukant Pati (2018). Linear Algebra through Matrices.

Peter J. Cameron (2008). Notes on Linear Algebra.

K.A. Stroud and Dexter J. Booth (2007). Engineering Mathematics 6th Edition. Palgrave Macmillan.

Abolarin, O.E. and Ishola, Y.C. (2005). Business Mathematics II. Jube Evans Books and Publications

UNIT 2 DETERMINANTS II

Unit Structure

2.1 Introduction

We continue the concept of determinant in this unit. Note that throughout this unit, F will denote field of characteristic zero, $(M_n(F))$ will denote the set of $n \times n$ matrices over F and $V_n(F)$ will denote the space of all $n \times 1$ matrices over F.

The concept of a determinant must be understood properly because it shall be used continuously again and again. Do spend more time on section 2.3, if necessary.

2.2 Learning Outcomes

- Define a linear transformation on a finite-dimensional non-zero vector space
- Evaluate the determinant of a linear transformation;
- Evaluate the rank of a matrix by using the concept of the determinant rank.

2.3 Main Content

2.3.1 Product Formula

In the last unit, we noted that one of the properties of determinant is the fact that the determinant of the product of two matrices is equal to the product of the determinants of the two matrices.

Having studied matrix multiplication in module 2 unit 1, we shall obtain the determinant of a product of matrices and define the determinant of a linear transformation. A method of obtaining the determinant rank of a matrix shall also be discussed in this unit.

Theorem 1: Let A and B be $n \times n$ matrices over F, then $det(AB) = det(A)det(B)$.

We shall only verify this theorem for some cases using some examples as the proof is slightly complicated.

Example 1: Calculate |A|, |B| and |AB| when $A = \begin{pmatrix} 1 & 0 & 2 \\ 3 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ and $B = \begin{pmatrix} 2 & 10 & 9 \\ 0 & 3 & 8 \\ 0 & 0 & 5 \end{pmatrix}$

Solution: Let us verify theorem 1 for our pair of matrices.
Now, on expanding by the third row (the reasons already given in unit 1), we have |A| = 1.
Also, |B| = 30, which can be immediately seen since B is a triangular matrix

Also, $AB = \begin{pmatrix} 1 & 0 & 2 \\ 3 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 2 & 10 & 9 \\ 0 & 3 & 8 \\ 0 & 0 & 5 \end{pmatrix} = \begin{pmatrix} 2 & 10 & 19 \\ 6 & 33 & 35 \\ 0 & 0 & 5 \end{pmatrix}$

$$|AB| = 5\begin{vmatrix} 2 & 10 \\ 6 & 33 \end{vmatrix} = 30$$

Thus, $|AB| = |A||B| = (1)(30) = 30$

Example 2: Show that $|AB| = |A||B|$ where $A = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 2 & -2 \\ 3 & -3 & 5 \end{pmatrix}$ and $B = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 2 & 0 \\ 5 & -3 & 3 \end{pmatrix}$.

Solution:

$$AB = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 2 & -2 \\ 3 & -3 & 5 \end{pmatrix}\begin{pmatrix} -1 & 0 & 1 \\ -2 & 2 & 0 \\ 5 & -3 & 3 \end{pmatrix} = \begin{pmatrix} -6 & 3 & -2 \\ -14 & 10 & -6 \\ 28 & -21 & 18 \end{pmatrix}$$

$$|AB| = \begin{vmatrix} -6 & 3 & -2 \\ -14 & 10 & -6 \\ 28 & -21 & 18 \end{vmatrix}$$

$$= -6(180 - 126) - 3(-252 + 168) + (-2)(294 - 280)$$

$$= -324 + 252 - 28 = -100$$

$$|B| = \begin{vmatrix} -1 & 0 & 1 \\ -2 & 2 & 0 \\ 5 & -3 & 3 \end{vmatrix} = -6 - 4 = -10$$

$$|A| = \begin{vmatrix} 1 & 0 & -1 \\ 0 & 2 & -2 \\ 3 & -3 & 5 \end{vmatrix} = 10 - 6 + 6 = 10$$

$|\mathrm{A}||\mathrm{B}| = 10(-10) = -100$

Hence $|AB| = |A||B| = (10)(-10) = -100$

$$BA = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 2 & 0 \\ 5 & -3 & 3 \end{pmatrix}\begin{pmatrix} 1 & 0 & -1 \\ 0 & 2 & -2 \\ 3 & -3 & 5 \end{pmatrix} = \begin{pmatrix} 2 & -3 & 6 \\ -2 & 4 & -2 \\ 14 & -15 & 16 \end{pmatrix}$$

This shows that $AB \neq BA$

Theorem 1 can be extended to a product of $m(n \times n)$ matrices $A_1, A_2, \cdots, A_m$.

That is, $det(A_1, A_2, \cdots, A_m) = det(A_1)det(A_2) \cdots det(A_m)$.

Now, you know that in general, $AB \neq BA$, but, $|AB| = |A||B|$

On the other hand, in general$|A + B| \neq |A| + |B|$, using the example above

$$A + B = \begin{pmatrix} 1 & 0 & 2 \\ 3 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} + \begin{pmatrix} 2 & 10 & 9 \\ 0 & 3 & 8 \\ 0 & 0 & 5 \end{pmatrix} = \begin{pmatrix} 3 & 10 & 11 \\ 3 & 4 & 8 \\ 0 & 0 & 6 \end{pmatrix}$$

$$|\mathrm{A} + \mathrm{B}| = \begin{vmatrix} 3 & 10 & 11 \\ 3 & 4 & 8 \\ 0 & 0 & 6 \end{vmatrix} = 6(12 - 30) = -108$$

$$|\mathrm{A}| = \begin{vmatrix} 1 & 0 & 2 \\ 3 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} = 1(1 - 0) = 1$$

$$|\mathrm{B}| = \begin{vmatrix} 2 & 10 & 9 \\ 0 & 3 & 8 \\ 0 & 0 & 5 \end{vmatrix} = 5(6 - 0) = 30$$

Thus, $|A| + |B| = 1 + 30 = 31$

So, $|A + B| = -108 \neq |A| + |B| = 31$

Thus, what we have just done is that determinant is not a linear function.

We now give an immediate corollary to theorem 1.

Corollary 1: If $A \in M_n(F)$ is invertible, then $det(A^{-1}) = \frac{1}{det(A)}$

Proof:

Let $B \in M_n(F)$ such that $AB = I$. Then $det(AB) = det(A)det(B) = det(1) = 1$

Thus, $det(A) \neq 0$ and $det(B) = \frac{1}{det(A)}$

In particular, $det(A^{-1}) = \frac{1}{det(A)}$

Another corollary to Theorem 1 is given below:

Corollary 2: Similar matrices have the same determinant.

Proof: If B is similar to A, then $B = P^{-1}AP$ for some invertible matrix P.

Thus, by Theorem 1, $det(B) = det(P^{-1}AP)$

$$= det(P^{-1})det(A)det(P) = \frac{1}{det(P)} \cdot det(P)$$ (By corollary 1)

We use this corollary to introduce you to the determinant of a linear transformation. At each stage you have seen the very close relationship between linear transformations and matrices. Here too, you will see this closeness.

Definition 1: Let $T: V \to V$ be a linear transformation on a finite-dimensional non-zero vector space V and let $A = [T]_B$ be the matrix of T with respect to a given basis B of V. Then the determinant of T is defined by $det(T) = det(A)$.

This definition is independent of the basis of V that is chosen because; if we choose another basis B' of V we would obtain the matrix $A' = [T]_{B'}$, which is similar to A (see Module 2 Unit 2, Corollary to Theorem 5).

Thus, $det\ (A') = det\ (A)$.

We have the following example and exercises.

Example 3: Find $det(T)$ where we define $T: R^3 \to R^3$ by

$$T(x_1, x_2, x_3) = (3x_1 + x_3, \quad -2x_1 + x_2, \quad -x_1 + 2x_2 + 4x_3)$$

Solution: Let $B = \{(1,0,0), (0,10), (0,0,1)\}$ be the standard ordered basis of R^3.

Now,

$$T(1,0,0) = (3, -2, -1) = 3(1,0,0) - 2(0,1,0) - 1(0,0,1)$$
$$T(0,1,0) = (0,1,2) = 0(1,0,0) + 1(0,1,0) + 2(0,0,1)$$
$$T(0,0,1) = (1,0,4) = 1(1,0,0) + 0(0,1,0) + 4(0,0,1)$$

$$\therefore A = [T]_B = \begin{bmatrix} 3 & 0 & 1 \\ -2 & 1 & 0 \\ -1 & 2 & 4 \end{bmatrix}$$

So, by definition,

$$det\ (T) = det\ (A) = \begin{vmatrix} 3 & 0 & 1 \\ -2 & 1 & 0 \\ -1 & 2 & 4 \end{vmatrix} = 3\begin{vmatrix} 1 & 0 \\ 2 & 4 \end{vmatrix} + 1\begin{vmatrix} -2 & 1 \\ -1 & 2 \end{vmatrix} = 12 - 3$$
$$= 9$$

Now let us see what happens if $B = 0$. Remember, in Unit 8 you saw that $AX = 0$ has $n - r$ linearly independent solutions, where $r = rank\ A$. The following theorem tells us this condition in terms of $det(A)$.

Theorem 2: The homogeneous system $AX = 0$ has a non-trivial solution if and only if

$$det(A) = 0$$

Proof: First assume that $AX = 0$ has a non-trivial solution.

Suppose, if possible, that $det(A) \neq 0$.

Then Cramer's Rule's says that $AX = 0$ has only the trivial solution $X = 0$ (because each $D_i = 0$ in Theorem 3 of the last unit).
This is a contradiction to our assumption.
Therefore, $det(A) = 0$.
Conversely, if $det(A) = 0$, then A is not invertible.
Therefore, the linear mapping:
$A: V_n(F) \to V_n(F): A(X) = AX$ is not invertible.
Therefore, this mapping is not one -to –one, hence, $Ker\ A \neq 0$, that is $AX = 0$ for some non-zero $X \in Vn(F)$.
Thus, $AX = 0$ has a non-trivial solution.
Use theorem 2 to solve the following:

Example 4: Does the system $\begin{matrix} 2x + 3y + z = 0 \\ x - y - z = 0 \\ 4x + 6y + 2z = 0 \end{matrix}$ have a non-zero solution?

Solution: The given system is equivalent to $AX = 0$, where $A = \begin{bmatrix} 2 & 3 & 1 \\ 1 & -1 & -1 \\ 4 & 6 & 2 \end{bmatrix}$
Since the third row of A is twice the first row of A.
Therefore, by P2 and P4 of Section 1.3.3 of unit 1, $|A| = 0$.
Therefore, by Theorem 5, the given system has a non-zero solution.
Let us introduce you to the determinant rank of a matrix, which leads us to another method of obtaining the rank of a matrix

2.3.2 The Determinant Rank

In Unit 2 module 2, you were introduced to the rank of a linear transformation and the rank of a matrix, respectively. Then we related the two ranks. In this section we will discuss the determinant rank and show that it is the rank of the concerned matrix. First, we give a necessary and sufficient condition for *n* vectors in $V_n(F)$ to be linearly dependent.

Theorem 3: Let $x_1, x_2, \cdots, x_n \in V_n(F)$, then $x_1, x_2, \cdots, x_n$ are linearly dependent over the field F if and only if $det(x_1, x_2, \cdots, x_n) = 0$.

Proof: Let $U = (x_1, x_2, \cdots, x_n)$ be the $n \times n$ matrix whose column vectors are $x_1, x_2, \cdots, x_n$. Then $x_1, x_2, \cdots, x_n$ are linearly dependent over F if and only if there exist scalars $\alpha_1, \alpha_2, \cdots, \alpha_n \in F$, not all zero, such that $\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n = 0$.

Now, $U \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} = (x_1 x_2 \cdots x_n) \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix}$

$$= x_1\alpha_1 + x_2\alpha_2 + \cdots + x_n\alpha_n$$
$$= \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n$$

Thus, $x_1, x_2, \cdots, x_n$ are linearly dependent over the field F if and only if $UX = 0$ for some

non-zero $X = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} \in V_n(F)$.

This happens if and only if $det(U) = 0$, by Theorem 1.
Thus, Theorem 2 is proved.
Theorem 3 is equivalent to the statement $(x_1 x_2 \cdots x_n)$ are linearly independent if and only if $det(x_1 x_2 \cdots x_n) \neq 0$.
Now, use theorem 3 to solve the following example:

Example 5: Check if the vectors; $\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$, $\begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}$, $\begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix}$ are linearly independent.

Solution: $\begin{vmatrix} 1 & 0 & 2 \\ 0 & -1 & 3 \\ 1 & 1 & 0 \end{vmatrix} = -3 + 2 = -1 \neq 0$

Hence, the given vectors are linearly independent.

Sub-matrix of *A* is a matrix that can be obtained from *A* by deleting some rows and columns.

Now, consider the matrix $A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 1 & 2 & 3 \end{bmatrix}$

Since two rows of *A* are equal, we know that |A| = 0.
But consider its 2×2 sub-matrix $A_{13} = \begin{bmatrix} 0 & 4 \\ 1 & 2 \end{bmatrix}$ whose determinant is –4 ≠ 0.
In this case we say that the determinant rank of *A* is 2.
In general, we have the following definition:

Definition 2: Let *A* be an $(m \times n)$ matrix. If A ≠ 0, then the **determinant rank** of A is the largest positive integer *r* such that

i. there exists an $(r \times r)$ sub-matrix of *A* whose determinant is non-zero, and
ii. for s > r, the determinant of any $(s \times s)$ sub-matrix of *A* is 0.

Note: The determinant rank *r* of any $m \times n$ matrix is defined, not only of a square matrix.
Also, $r \leq min(m, n)$.

Example 6: Obtain the determinant rank of $A = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$

Solution: Since A is a (3×2) matrix, the largest possible value of its determinant rank can be 2. Also, the sub-matrix $\begin{bmatrix} 1 & 4 \\ 2 & 5 \end{bmatrix}$ of A has determinant (-3) $\neq$ 0.
Therefore, the determinant rank of A is 2.

Example 7: Calculate the determinant rank of A, where $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$

Solution: Here, A is a (2×3) matrix and the largest possible value of its determinant rank can be 2. Also, the sub-matrix$\begin{bmatrix} 1 & 2 \\ 4 & 5 \end{bmatrix}$ of A has determinant (-3) $\neq$ 0.
Therefore, the determinant rank of A is 2.

Now we come to the reason for introducing the determinant rank, this gives us another method for obtaining the rank of a matrix.

Theorem 4: The determinant rank of an $(m \times n)$ matrix A is equal to the rank of A.

Proof: Let the determinant rank of A be r. Then there exists a $(r \times r)$ sub-matrix of A whose determinant is non-zero. By Theorem 3, its column vectors are linearly independent.
It follows by the definition of linear independence, that these column vectors, when extended to the column vectors of A, remain linearly independent.
Thus, A has at least r-linearly independent column vectors,
Therefore, by definition of the rank of a matrix,

$$r \leq rank(A) = p(A) \qquad \text{…..…… (1)}$$

Also, by definition of $p(A)$, we know that the number of linearly independent rows that A has is $p(A)$. These rows form a $(p(A) \times n)$ matrix $p(A)$.
Thus, B will have $p(A)$ linearly independent columns.
Retaining these linearly independent columns of B we obtain a $(p(A) \times p(A))$ sub-matrix C of B.
So, C is a sub-matrix of A whose determinant will be non-zero by theorem 3, since its columns are linearly independent.
Thus, by the definition of the determinant rank of A, we have

$$p(A) \leq r \qquad \text{………. (2)}$$

Combination of (1) and (2) gives $p(A) = r$.
We will use Theorem 4 in the following example.

Example 8: Find the rank of $A = \begin{bmatrix} 2 & 3 & 4 \\ 3 & 1 & 2 \\ -1 & 2 & 2 \end{bmatrix}$

Solution: $|A| = 0$, but $\begin{vmatrix} 2 & 3 \\ 3 & 1 \end{vmatrix} = -7 \neq 0$
Thus, by theorem 4, $p(A) = 2$.

Remark: This example shows that Theorem 4 can simplify the calculation of the rank of a matrix in some cases. We don't have to reduce a matrix to echelon form each time. But at times using this method seems to be as tedious as the row-reduction method, for example,

Example 9: Use Theorem 4 to find the rank of A, where

(a) $A = \begin{pmatrix} 3 & 1 & 2 & 5 \\ 1 & 2 & -1 & 2 \\ 4 & 3 & 1 & 7 \end{pmatrix}$

(b) $A = \begin{pmatrix} 2 & 3 & 5 & 1 \\ 1 & -1 & 2 & 1 \end{pmatrix}$

Solution:

a) The determinant rank of A is less or equals to $3(\leq 3)$.

The determinant rank of the 3×3 sub-matrix $\begin{pmatrix} 3 & 1 & 2 \\ 1 & 2 & -1 \\ 4 & 3 & 1 \end{pmatrix}$ is zero.

Also, the determinant rank of the 3×3 sub-matrix $\begin{pmatrix} 3 & 2 & 5 \\ 1 & -1 & 2 \\ 4 & 1 & 7 \end{pmatrix}$ is zero.

In fact, you can check that all the determinant ranks of the 3×3 sub-matrices are zero.

Now let us look at the 2×2 sub-matrix of A,

Since $\begin{vmatrix} 3 & 1 \\ 1 & 2 \end{vmatrix} = 5 \neq 0$, and $p(A) = 2$

b) The determinant rank of A is less or equals to 2 (≤ 2)

Now, $\begin{vmatrix} 2 & 3 \\ 1 & -1 \end{vmatrix} = -5 \neq 0$, and $p(A) = 2$

Exercise (a) shows how much time can be taken by using this method. On the other hand, Exercise (b) shows how little time it takes to obtain $p(A)$, using the determinant rank. Thus, the method to be used for obtaining $p(A)$ varies from case to case.

SELF-ASSESSMENT EXERCISE(S)

Find the rank and determinant rank of the following matrices:

1) $A = \begin{bmatrix} 3 & 1 & 2 \\ 2 & 3 & 1 \\ -1 & 2 & 2 \end{bmatrix}$

2) $A = \begin{bmatrix} 2 & 2 & -2 \\ 2 & 3 & 1 \\ 1 & 2 & 2 \end{bmatrix}$

3) $A = \begin{bmatrix} 1 & 1 & 2 \\ 2 & 0 & 1 \\ -1 & 2 & 2 \end{bmatrix}$

4) $A = \begin{bmatrix} 2 & 0 & 1 \\ -1 & 4 & -1 \\ -1 & 2 & 0 \end{bmatrix}$

Conclusion

We conclude that the homogeneous system of linear equations $AX = 0$ has a non-zero solution if and only if $det(A) = 0$.

The determinant rank of an $(m \times n)$ matrix A is equal to the rank of A which help to simplify the calculation of the rank of a matrix in some cases and there would be no need to reduce a matrix to echelon form at every time.

2.6 Summary

In this unit we have covered the following points.
With the aid of examples, we have been able to show that $det(AB) = det(A)det(B)$.

We also defined determinant of a linear transformation from U to V, where $dimU = dimV$. Theorems were used to obtain the rank of matrices This unit also defined the determinant rank, and proved that rank of A is equal to determinant rank of A.

2.7 References/Further Readings

Robert A. Beezer (2014). A First Course in Linear Algebra Congruent Press, Gig Harbor, Washington, USA 3(40).

Arbind K Lal Sukant Pati (2018). Linear Algebra through Matrices.

Peter J. Cameron (2008). Notes on Linear Algebra.

K.A. Stroud and Dexter J. Booth (2007). Engineering Mathematics 6^{th} Edition. Palgrave Macmillan.

Abolarin, O.E. and Ishola, Y.C. (2005). Business Mathematics II. Jube Evans Books and Publications

MODULE 4

UNIT 1 EIGENVALUES AND EIGENVECTORS

Unit Structure

1.1 Introduction

Matrices of linear transformations have been studied in Modules 1 and 2. You have had several opportunities, in earlier units to observe that the matrix of a linear transformation depends on the choice of the bases of the concerned vector spaces. In this unit, we shall consider the problem of finding a suitable basis B, of the vector space V, such that the $n \times n$ matrix $[T]_B$ is a diagonal matrix. It is in this context that the study of eigenvalues and eigenvectors plays a central role.

The eigenvalue problem involves the evaluation of all the eigenvalues and eigenvectors of a linear transformation or a matrix. The solution of this problem has basic applications in almost all branches of the sciences, technology and the social science besides its fundamental role in various branches of pure and applied mathematics. The emergence of computers and the availability of modern computing facilities have further strengthened this study, since they can handle very large systems of equations.

1.2 Learning Outcomes

By the end of this unit, you should be able to:

- Obtain the characteristic polynomial of a linear transformation or a matrix;
- Obtain the eigenvalues, eigenvectors and eigenspaces of a linear transformation of a matrix;
- Obtain a basis of a vector space V with respect to which the matrix of a linear transformation $T: V \to V$ is in diagonal form;
- Obtain a non-singular matrix P which diagonalizes a given diagonalizable matrix A.

1.3 Eigenvalues and Eigenvectors

1.3.1 The Algebraic Eigenvalue Problem

The word "*eigenvalue"* is a mixture of German and English; meaning "characteristic value" or "proper value" (here "proper" is used in the sense of "property").

Another term used in older books is "latent root", here "latent" means "hidden", the idea is that the eigenvalue is somehow hidden in a matrix representing a, and has to be extracted by some procedure.

1.3.2 Eigenvalues and Eigenvectors of a Linear Transformation

Definition 1: An eigenvalue of a linear transformation $T: V \to V$ is a scalar such that there exists a non-zero $x \in V$ is called an eigenvector of T with respect to the eigenvalue $\lambda \in K$, if $x \neq 0$ and $T(x) = \lambda x$.
The set $\{x: T(x) = \lambda x\}$ consisting of the zero vector and the eigenvectors with eigenvalue λ, is called the $\lambda -$*eigen-space* of T.

Example 1: Consider the linear mapping $T: R^2 \to R^2$ such that $T(x, y) = (2x, y)$.
Then, $T(1,0) = (2,0) = 2(1,0)$, therefore, that $T(x, y) = (2x, y) = (1,0) \neq (0,0)$.
In the example above, (1,0) is an eigenvector of T with respect to the eigenvalue 2.
Thus, a vector $x \in V$ is an eigenvector of the linear transformation T if

i. x is non-zero, and
ii. $T(x) = \lambda x$ for some scalar $T(x) = \lambda \in K$.

The fundamental algebraic eigenvalue problem deals with the determination of all the eigenvalues of a linear transformation.
Let us look at some examples of how we can find eigenvalues.

Example 2: Obtain an eigenvalue and a corresponding eigenvector for the linear operator $T: R^3 \to R^3$such that$T(x, y, z) = (2x, 2y, 2z)$.
Solution: Clearly, $T(x, y, z) = 2(x, y, z); \quad (x, y, z) \in R^3$
Thus, 2 is an eigenvalue of T.
Any non-zero element of R^3will be an eigenvector of T corresponding to 2.

Example 3: Obtain an eigenvalue and a corresponding eigenvector of

$$T(x, y, z): C^3 \to C^3; \; T(x, y, z) = (ix, iy, z).$$

Solution: Firstly, note that T is a linear operator. Now, if $\lambda \in C$ is an eigenvalue, then there exist $(x, y, z) \neq (0,0,0)$ such that $T(x, y, z) = \lambda(x, y, z) \Rightarrow (ix, iy, z) = (\lambda x, \lambda y, \lambda z)$

$$\Rightarrow ix = \lambda x \; ; \; -iy = \lambda y \; ; \; z = \lambda z$$

These equations are satisfied if $\lambda = i \; ; y = 0 \; ; \; z = 0$.
$\lambda = I$ is an eigenvalue with a corresponding eigenvector being $(1,0,0)$ or $[(x, 0,0)$ for any $x \neq 0]$.
It is also satisfied if $\lambda = -i \; ; x = 0 \; ; \; z = 0$or if$\lambda = 1 \; ; x = 0 \; ; \; y = 0$. There, $-i$ and 1 are also eigenvalues with corresponding eigenvectors (0,*y*,0) and 0,0,*z*) respectively for any $y \neq 0 \; ; \; z \neq 0$.

1.3.3 Vector space corresponding to an eigenvalue of a linear transformation

Suppose $\lambda \in K$ is an eigenvalue of the linear transformation$T: V \to V$. Define the set $W_A = \{x \in V | T(x) = \lambda x\} = \{0\} \cup \{\text{eigenvectors of T corresponding to } \lambda\}$.

Thus, a vector $v \in W$ if and only if $v = 0$ is an eigenvector of T corresponding to λ.
Now, $x \in W_A \Leftrightarrow Tx = \lambda Ix$, I being the identity operator,

$$\Leftrightarrow (T - \lambda I)x = 0 \Leftrightarrow x \in Ker(T - \lambda I)$$

$\therefore W_A = Ker(T - \lambda I)$ and hence, W_A is a subspace of V.
Since λ is an eigenvalue of T, it has an eigenvector, which must be non-zero.
Thus, W_λ is non-zero

Definition 2: For an eigenvalue λ of T, the non-zero subspace W_λ is called the eigen-space of T associated with the eigenvalue.

Example 4: Obtain W_2 for the linear operator given in Example 1.

Solution: $W_2 = \{(x, y, z) \in R^3 | T(x, y, z) = 2(x, y, z)\}$
$\{(x, y, z) \in R^3 | T(2x, 2y, 2z) = 2(x, y, z)\} = R^3$
For *T* in Example 2, obtain the complex vector spaces W_i , W_{-1}, W_1.
$W_i = \{(x, y, z) \in C^3 | T(x, y, z) = i(x, y, z)\}$
$= \{(x, y, z) \in C^3 | (ix, -iy, z) = (x, y, z) = (ix, iy, iz)\}$
$= \{(x, 0,0) | x \in C\}$
Similarly, you can show that $W_{-1} = \{(0, x, 0) | x \in C\}$ and $W_1 = \{(0,0, x,) | x \in C\}$.

As with every other concept related to linear transformations, we can define eigenvalues and eigenvectors for matrices also.

Definition 3: A scalar λ is an eigenvalue of an $(n \text{ x } n)$ matrix *A* over *F* if there exists $X \in V_n(F); X \neq 0$, such that $AX = \lambda X$ are eigenvectors of the matrix *A* corresponding to the eigenvalueλ.

Example 5: Let $B = \begin{pmatrix} 6 & -6 \\ 11 & -12 \end{pmatrix}$.
The vector $v = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$ satisfies $\begin{pmatrix} 6 & -6 \\ 11 & -12 \end{pmatrix} \begin{pmatrix} 2 \\ 3 \end{pmatrix} = 3 \begin{pmatrix} 2 \\ 3 \end{pmatrix}$.
This shows that $\begin{pmatrix} 2 \\ 3 \end{pmatrix}$ is an eigenvector with eigenvalue 3.
The vector$v = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$ also is an eigenvector of *B* with eigenvalue 2.
Similarly, the vector is an eigenvector of *B* with eigenvalue 2.
Generally, ifλ is an eigenvalue of $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$, then we could find a corresponding eigenvector $\begin{pmatrix} x \\ y \end{pmatrix}$ by solving the linear equations.

Example6: If $A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}$. Obtain an eigenvalue and a corresponding eigenvector of A.

Solution:$A \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, this shows that 1 is an eigenvalue and $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ is an eigenvector corresponding to it.
In fact,$A \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = 2 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ and$A \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = 3 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$,

Thus, 2 and 3 are eigenvalues of A, with corresponding eigenvectors$\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ and$\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ respectively.

Example 7: Obtain an eigenvalue and a corresponding eigenvector of

$$B = \begin{pmatrix} 0 & -1 \\ 1 & 2 \end{pmatrix} \in M_2(R).$$

Solution: Suppose$\lambda \in R$is an eigenvalue of *B*,
Then $\exists \begin{pmatrix} x \\ y \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, such that $BX = \lambda X$, that is, $\begin{pmatrix} -y \\ x + 2y \end{pmatrix} = \begin{pmatrix} \lambda x \\ \lambda y \end{pmatrix}$.
So, for what values of λ, *x* and *y*are the equation satisfied?
Note that $x \neq 0$ and $y \neq 0$, because if either is zero then the other will have to be zero.
Solving the equations, we obtain $\lambda = 1$with corresponding eigenvector $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$.
Classwork: Now solve the eigenvalue problem $\begin{pmatrix} 1 & 2 \\ 0 & 3 \end{pmatrix}$.

1.3.4 Eigenspace corresponding to an eigenvalue of a matrix

Just as we defined an eigenspace associated with a linear transformation, we define the eigenspace W_λ, corresponding to an eigenvalue of an $n \times n$ matrix *A*, as follows:

$$W_\lambda = \{X \in V_n(F) | AX = \lambda x\} = \{X \in V_n(F) | A(X - \lambda I)X = 0\}$$

For instance, the eigenspace W_1, in Example 6, is $\begin{pmatrix} x \\ y \\ z \end{pmatrix} \in V_3(R)$

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x \\ 2y \\ 3z \end{pmatrix}$$

The algebraic eigenvalue problem for matrices is to determine all the eigenvalues and eigenvectors of a given matrix. In fact, the eigenvalues and eigenvectors of an n x n matrix *A* are precisely the eigenvalues ad eigenvector of a regarded as a linear transformation from $V_n(F)$ to $V_n(F)$.
We end this section with the following remark:

A scalar λ is an eigenvalue of the matrix *A* if and only if $(A - \lambda I)X = 0$ has a non-zero solution, i.e., if and only if $det(A - \lambda I) = 0$.
Similarly, λ is an eigenvalue of the linear transformation *T* if and only if $det(T - \lambda I) = 0$.
So far, we have been obtaining eigenvalues by observation, or by some calculations that may not give us all the eigenvalues of a given matrix or linear transformation. T
he remark above suggests where to look for all the eigenvalues in the next section we determine eigenvalues and eigenvectors explicitly.

1.3.5 Eigenvalues and Eigenvectors of Matrices

In the previous section we have seen that a scalar λ is an eigenvalue of a matrix A if and only if $det(A - \lambda I) = 0$.
In this section we shall see how this equation helps us to solve the eigenvalue problem.

1.3.6 Characteristic Polynomial

Once we know that λ is an eigenvalue of a matrix A, the eigenvectors can easily be obtained by finding non-zero solutions of the system of equations given by $AX = \lambda X$.

Now, if $A = \begin{bmatrix} a_{11} & a_{12} & \cdots & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \cdots & \vdots \\ \vdots & \vdots & \cdots & \cdots & \vdots \\ a_{n1} & a_{n2} & \cdots & \cdots & a_{nn} \end{bmatrix}$ and $X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_n \end{bmatrix}$

The equation $AX = \lambda X$ becomes

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \cdots & \vdots \\ \vdots & \vdots & \cdots & \cdots & \vdots \\ a_{n1} & a_{n2} & \cdots & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_n \end{bmatrix} = \lambda \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_n \end{bmatrix}$$

Carry out the matrix multiplication to obtain the following system of equations:

$$\begin{gathered} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = \lambda x_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = \lambda x_2 \\ \vdots \\ \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n = \lambda x_n \end{gathered}$$

This is equivalent to the following system

$$\begin{gathered} (a_{11} - \lambda)x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = 0 \\ a_{21}x_1 + (a_{22} - \lambda)x_2 + \cdots + a_{2n}x_n = 0 \\ \vdots \\ \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + (a_{nn} - \lambda)x_n = 0 \end{gathered}$$

This homogeneous system of linear equations has a non-trivial solution if and only if the determinant of the coefficient matrix is equal to zero (by Unit 3, theorem 1).
Thus, λ is an eigenvalue of A if and only if

$$det(A-\lambda I)=\begin{vmatrix} a_{11}-\lambda & a_{12} & \cdots & \cdots & a_{1n} \\ a_{21} & a_{22}-\lambda & \cdots & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \cdots & \vdots \\ \vdots & \vdots & \cdots & \cdots & \vdots \\ a_{n1} & a_{n2} & \cdots & \cdots & a_{nn}-\lambda \end{vmatrix}=0$$

Now, $det(\lambda I-A)=(-1)det(A-\lambda I)$ (that is, multiplying each row by (-1)).
Hence, $det(\lambda I-A)=0$if and only if $det(A-\lambda I)=0$.
This leads us to define the concept of the characteristic polynomial

Definition 4: Let $A=[a_{ij}]$ be any (nxn) matrix. Then the characteristic polynomial of the matrix A is defined by

$$f_A(t)=det(A-\lambda I)=\begin{vmatrix} t-a_{11} & -a_{12} & \cdots & \cdots & -a_{1n} \\ -a_{21} & t-a_{22} & \cdots & \cdots & -a_{2n} \\ \vdots & \vdots & \cdots & \cdots & \vdots \\ \vdots & \vdots & \cdots & \cdots & \vdots \\ -a_{n1} & -a_{n2} & \cdots & \cdots & t-a_{nn} \end{vmatrix}$$
$$=t^n+c_1t^{n-1}+c_2t^{n-2}+\cdots+c_{n-1}t+c_n$$

where the coefficients $c_1, c_2, \cdots, c_n$ depend on the entries a_{ij} of the matrix A.
The equation $f_A(t)=0$ is the characteristic equation of A.
When no confusion arises, we shall simply write $f(t)$ in place of $f_A(t)$.
Consider the following example.

Example 8: Obtain the characteristic polynomial of the matrix $\begin{pmatrix} 1 & 2 \\ 0 & -1 \end{pmatrix}$

Solution: The required polynomial is $\begin{vmatrix} -1 & -2 \\ 0 & t+1 \end{vmatrix}=(t-1)(t+1)=t^2-1$

Example 9: Obtain the characteristic polynomial of the matrix $\begin{pmatrix} 0 & 0 & 2 \\ 1 & 0 & 1 \\ 0 & 1 & -2 \end{pmatrix}$.

Solution: The required polynomial is

$$\begin{vmatrix} t & 0 & -2 \\ -1 & t & -1 \\ 0 & -1 & t+2 \end{vmatrix}=t[t(t+1)-1]-2(1)$$
$$=t(t^2+2t-1)-2=t^3+2t^2-t-2$$

Note that λ is an eigenvalue of A iff $det(\lambda I-A)=f_A(t)=0$, that is, iff λ is a root of the characteristic polynomial $f_A(t)$, defined above. Due to this fact, eigenvalues are also called characteristic root, and eigenvectors are called characteristic vectors.
For example, the eigenvalues of the matrix in Example 8 are the roots of the polynomial t^2-1 which are 1 and -1.

To obtain the eigenvalues of the matrix in example 9 above, factorize the characteristic polynomial is

That is, $t^3 + 2t^2 - t - 2 = t^2(t+2) - (t+2)$

$(t^2 - 1)(t+2) = (t-1)(t+1)(t+2)$

$t = 1,\ t = -1$ and $t = -2$

Now, the characteristic polynomial $f_A(t)$ is a polynomial of degree n. Hence, it can have n roots at the most. Thus, an $n \times n$ matrix has n eigenvalues, at the most.

For example, the 2×2 matrix in Example 8 has two eigenvalues, 1 and – 1, and the matrix in example 9 has 3 eigenvalues, 1, -1 and -2.

Now we will prove a theorem that will help us in unit 2.

Theorem 1: Similar matrices have the same eigenvalues.

Proof: Let an $n \times n$ matrix *B* be similar to an $n \times n$ matrix *A*. Then, by definition, $B = P^{-1}AP$, for some invertible matrix *P*.

Now, the characteristic polynomial of *B*,

$$\begin{aligned} f_B(t) &= det(tI - B) \\ &= det(tI - P^{-1}AP) \\ &= detP^{-1}(tI - A)P \qquad \text{(Since } P^{-1}tIP = tP^{-1}P = tI) \\ &= det(P^{-1})det(tI - A)det(P) \\ &= det(tI - A)det(P^{-1})det(P) \\ &= f_A(t)det(P^{-1}P) \\ &= f_A(t) \qquad \text{(Since } det(P^{-1}P) = det(I) = I) \end{aligned}$$

Thus, the roots of $f_B(t)$ is and $f_A(t)$ is coincide.

Therefore, the eigenvalues of *A* and *B* are the same.

Let us consider some more examples so that the concepts mentioned in this section become absolutely clear to you.

Example 10: Obtain the eigenvectors of the matrix $\begin{pmatrix} 1 & 2 \\ 0 & -1 \end{pmatrix}$.

Solution:

From example 8 above, the characteristic polynomial is $t^2 - 1$ and when equated to zero, we
have $t = 1,\ t = -1$.

The eigenvectors of the matrix with respect to the eigenvalue $\lambda_1 = -1$ are the non-trivial solutions of

$$\begin{pmatrix} 1 & 2 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = -1\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

which gives the equations

$$\left.\begin{aligned} x_1 + 2x_2 &= -x_1 \\ -x_2 &= -x_2 \end{aligned}\right\} \Rightarrow x_1 = -x_2$$

This result merely tells us that whatever the value of x_2 is, the value of x_1 is (-1) times it. Therefore, the eigenvector $x_1 = \begin{pmatrix} k \\ -k \end{pmatrix}$, ($k$ is an integer) is the general form of an infinite number of such eigenvectors.

The simplest eigenvector is therefore is $x_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$, corresponding to $\lambda_1 = -1$.

For $\lambda_2 = 1$, a similar result can be obtained

$$\begin{pmatrix} 1 & 2 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 1\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \Rightarrow \left.\begin{matrix} x_1 + 2x_2 = x_1 \\ -x_2 = x_2 \end{matrix}\right\} \Rightarrow x_2 = 0$$

The eigenvector is therefore is $x_2 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, corresponding to $\lambda_2 = 1$.

Example 11: Find the eigenvalues and eigenvectors of the matrix $\begin{pmatrix} 0 & 0 & 2 \\ 1 & 0 & 1 \\ 0 & 1 & -2 \end{pmatrix}$.

Solution: In solving the example, you found that the eigenvalues of *A* are $\lambda_1 = 1$, $\lambda_2 = -1, \lambda_3 = -2$.

Now we obtain the eigenvectors of *A*.

The eigenvectors of *A* with respect to the eigenvalue $\lambda_1 = 1$ are the non-trivial solutions of

$$\begin{pmatrix} 0 & 0 & 2 \\ 1 & 0 & 1 \\ 0 & 1 & -2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = 1\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

which gives the equations

$$\left.\begin{matrix} 2x_3 = x_1 \\ x_1 + x_3 = x_2 \\ x_2 - 2x_3 = x_3 \end{matrix}\right\} \Rightarrow \begin{matrix} x_1 = 2x_3 \\ x_2 = 3x_3 \\ x_3 = x_3 \end{matrix}$$

This result merely tells us that whatever the value of x_3 is, the value of x_1 is 2 times it and the value of x_2 is 3 times it.

Therefore, the eigenvector $x_1 = \begin{pmatrix} 2k \\ 3k \\ k \end{pmatrix}$, ($k$ is an integer) is the general form of an infinite number of such eigenvectors.

The simplest eigenvector for $\lambda_1 = 1$, is therefore is $X_1 = \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix}$.

For $\lambda_2 = -1$, a similar result can be obtained

The eigenvectors corresponding to $\lambda_2 = -1$ are given by

$$\begin{pmatrix} 0 & 0 & 2 \\ 1 & 0 & 1 \\ 0 & 1 & -2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = -1\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

which gives the equations

$$\left.\begin{aligned} 2x_3 &= -x_1 \\ x_1 + x_3 &= -x_2 \\ x_2 - 2x_3 &= -x_3 \end{aligned}\right\} \Rightarrow \begin{aligned} x_1 &= -2x_3 \\ x_2 &= x_3 \\ x_3 &= x_2 \end{aligned}$$

Therefore, the simplest eigenvector corresponding to $\lambda_2 = -1$ is $X_2 = \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix}$.

For $\lambda_2 = -2$, a similar result can be obtained

The eigenvectors corresponding to $\lambda_2 = -2$ are given by

$$\begin{pmatrix} 0 & 0 & 2 \\ 1 & 0 & 1 \\ 0 & 1 & -2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = -2\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

which gives the equations

$$\left.\begin{aligned} 2x_3 &= -2x_1 \\ x_1 + x_3 &= -2x_2 \\ x_2 - 2x_3 &= -2x_3 \end{aligned}\right\} \Rightarrow \begin{aligned} x_1 &= -x_3 \\ x_2 &= 0 \\ x_3 &= x_3 \end{aligned}$$

Therefore, the simplest eigenvector corresponding to $\lambda_3 = -2$ is $X_3 = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$.

Let $T: V \rightarrow V$ be a linear transformation on a finite-dimensional vector space *V* over the field F. we have seen that $\lambda \in F$ is an eigenvalue of $\Leftrightarrow$ $det(T - \lambda I) = 0 \Leftrightarrow det(\lambda I - T) = 0 \Leftrightarrow det(\lambda I - A) = 0$, where $A = [T]_B$ is the matrix of *T* with respect to a basis *B* of *V*.

Note that $[\lambda I - T]_B = \lambda I - [T]_B$.

This shows that λ is an eigenvalue of T if and only if λ is an eigenvalue of the matrix $A = [T]_B$, where B is a basis of *V*. We define the characteristic polynomial of the linear transformation *T* to be same as the characteristic polynomial of the matrix $A = [T]_B$, where *B* is basis *V*.

This definition does not depend on the basis *B* chosen, since similar matrices have the same characteristic polynomial (Theorem 1), and the matrices of the same linear transformation T with respect to two different ordered bases of *V* are similar.

Just as for matrices, the eigenvalues of *T* are precisely the roots of the characteristic polynomial of *T*.

Example 12: Let $T: V \rightarrow V$ be the linear transformation which maps $e_1 = (1,0)$ to $e_2 = (0,1)$ and e_2 to $-e_1$. Obtain the eigenvalues of T.

Solution: Let $A = [T]_B = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$, where $B = \{e_1, e_2\}$.

The characteristic polynomial of *T* = the characteristic polynomial of *A*

$$\begin{vmatrix} t & 1 \\ -1 & t \end{vmatrix} = t^2 + 1$$, which has no real roots.

Hence, the linear transformation *T* has *n* real eigenvalues.

However, it has two complex eigenvalues i and $-i$

Try the following exercise now.

Example 13: Obtain the eigenvalues and eigenvectors of the differential operator

Solution: $B = \{1, x, x^2\}$ is a basis of P_2

Then $[D]_B = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix}$

Therefore, the characteristic polynomial of D is $\begin{vmatrix} t & -1 & 0 \\ 0 & t & -2 \\ 0 & 0 & t \end{vmatrix} = t^2$

Hence, the only eigenvalue is $\lambda = 0$

The eigenvectors corresponding to $\lambda = 0$ are $\alpha_0 + \alpha_1 x + \alpha_2 x^2$ where, $D(\alpha_0 + \alpha_1 x + \alpha_2 x^2) = 0$, that is, $\alpha_1 + 2\alpha_2 x = 0$

This gives $\alpha_1 = 0$, $\alpha_2 = 0$

Therefore, the set of eigenvectors corresponding to $\lambda = 0$ are $\{\alpha_0 : \alpha_0 \in R, \alpha_0 \neq 0\} = R \backslash \{0\}$

Now that we have discussed a method of obtaining the eigenvalues and eigenvectors of a matrix, let us see how they help in transforming any square matrix into a diagonal matrix.

1.3.7 Diagonalization

We would start this section by stating and proving a theorem that discusses the linear independence.

Theorem 2: Let $T: V \to V$ be a linear transformation on a finite-dimensional vector space V over the field F. Let $\lambda_1, \lambda_2, \cdots, \lambda_m$ be the distinct eigenvalues of T and $v_1, v_2, \cdots, v_m$ be eigenvectors of T corresponding to $\lambda_1, \lambda_2, \cdots, \lambda_m$, respectively, then, $v_1, v_2, \cdots, v_m$ are linearly independent over F.

Proof:

We know that $Tv_i = \lambda_i v_i$, $0 \neq v_i \in V$ for $i = 1,2,\cdots,m$ and $\lambda_i \neq \lambda_j$ for $i \neq j$.

Suppose, if possible, that $\{v_1, v_2, \cdots, v_m\}$ is a linearly dependent set.

Now, the single non-zero vector v_i is linearly independent.

We choose $r(\leq m)$ such that $\{v_1, v_2, \cdots, v_{r-1}\}$ is linearly independent and $\{v_1, v_2, \cdots, v_{r-1}, v_r\}$ is linearly dependent.

Then,

$$v_r = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_{r-1} v_{r-1}$$

……………. (1)

for some $\alpha_1, \alpha_2, \cdots, \alpha_{r-1} \in F$
Multiply (1) by *T*, we have

$$Tv_r = \alpha_1 Tv_1 + \alpha_2 Tv_2 + \cdots + \alpha_{r-1} Tv_{r-1}$$

…………………. (2)
This gives

$$\lambda_r v_r = \alpha_1 \lambda_1 v_1 + \alpha_2 \lambda_2 v_2 + \cdots + \alpha_{r-1} \lambda_{r-1} v_{r-1}$$

……………....... (3)
Now, we multiply (1) by λ_r and subtract it from (3), to get

$$0 = \alpha_1(\lambda_1 - \lambda_r)v_1 + \alpha_2(\lambda_2 - \lambda_r)v_2 + \cdots + \alpha_{r-1}(\lambda_{r-1} - \lambda_r)v_{r-1}$$

………… (4)
Since the set $\{v_1, v_2, \cdots, v_{r-1}\}$ if linearly independent, each coefficient in the above equation must be 0. Thus, we have $\alpha_i(\lambda_i - \lambda_r) = 0$ for $i = 1,2, \cdots, r-1$.
But $\lambda_i \neq \lambda_r$, for $i = 1,2, \cdots, r-1$.
Hence $(\lambda_i - \lambda_r) \neq 0$ for $i = 1,2, \cdots, r-1$, and we must have $\alpha_r = 0$ for $i = 1,2, \cdots, r-1$.

However, this is not possible since (1) would imply that $v_r = 0$, and, being an eigenvector, v_r can never be zero. Thus, we reach a contradiction.

Hence, the assumption we started with must be wrong.
Thus, $\{v_1, v_2, \cdots, v_m\}$ must be linearly independent, and the theorem is proved.

Theorem 2 shall be used to choose a basis for a vector space *V* so that the matrix $[T]_B$ is a diagonal matrix.

Definition 5: A linear transformation $T: V \to V$ on a finite-dimensional vector space V is said to be diagonalizable if there exists a basis $B = \{v_1, v_2, \cdots, v_m\}$ of *V* such that the matrix of *T* with respect to the basis *B* is a diagonal matrix, that is,

$$[T]_B = \begin{bmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ 0 & 0 & \lambda_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

Where $\lambda_1, \lambda_2, \cdots, \lambda_n$ are scalars which need not be distinct.
Suppose $B = \{v_1, v_2, \cdots, v_n\}$ is such a basis showing that T is diagonalizable, then $T(v_i) = \alpha_{ij} v_i$ for $i = 1, \ldots, n$ where α_{ij} of is the i^{th} diagonal entry of the diagonal matrix *A*.
Thus, the basis vectors are eigenvectors.
Conversely, if we have a basis of eigenvectors, then the matrix representing *T* is diagonal.

Proposition: The linear map T on $V$is diagonalizable if and only if there is a basis of V consisting of eigenvectors of T.

Example 14: The matrix$\begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$ is not diagonalizable.
It is easy to see that it's only eigenvalue is 1, and the only eigenvectors are scalar multiples of $(0 \quad 1)^T$. So, we cannot find a basis of eigenvectors.

The next theorem tells us under what conditions a linear transformation is diagonalizable.

Theorem 3: A linear transformation T, on a finite-dimensional vector space V, is diagonalizable if and only if there exists a basis of V consisting of eigenvector of T.

Proof: Suppose that T is diagonalizable. By definition, there exists a basis $B = \{v_1, v_2, \cdots, v_m\}$ of V, such that

$$[T]_B = \begin{bmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ 0 & 0 & \lambda_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_n \end{bmatrix}.$$

By definition of $[T]_B$, we must have $\lambda_1, \lambda_2, \cdots, \lambda_m$;
$Tv_1 = \lambda_1 v_1;\ Tv_2 = \lambda_2 v_2, ; \cdots \cdots ; Tv_n = \lambda_n v_n$
Since basis vectors are always non-zero,$v_1, v_2, \cdots, v_n$ are non-zero.
Thus, we find that $v_1, v_2, \cdots, v_n$ are eigenvectors of T.

Conversely, let $B = \{v_1, v_2, \cdots, v_n\}$ be a basis of V consisting of eigenvectors of T, then, there exist scalars, $\alpha_1, \alpha_2, \cdots, \alpha_n$, not necessarily distinct, such that

$$Tv_1 = \lambda_1 v_1, Tv_2 = \lambda_2 v_2, \cdots \cdots, Tv_n = \lambda_n v_n.$$

But then we have $[T]_B = \begin{bmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ 0 & 0 & \lambda_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_n \end{bmatrix}$ which means that T is diagonalizable.
The next theorem combines theorem 2 and 3

Theorem 4: Let $T: V \to V$ be a linear transformation, where V is an n-dimensional vector space. Assume that T has n distinct eigenvalues, then T is diagonalizable.

Proof: Let$\lambda_1, \lambda_2, \cdots, \lambda_n$ be the n distinct eigenvalues of T.

Then there exist eigenvectors $v_1, v_2, \cdots, v_n$ corresponding to the eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_n$, respectively. By theorem 2, the set $\{v_1, v_2, \cdots, v_n\}$ is linearly independent and has n vectors, where $n = dimV$.

From Unit 3 (corollary to Theorem 5), $B = \{v_1, v_2, \cdots, v_n\}$ is a basis of V consisting of eigenvectors of T.
Thus, by theorem 3, T is diagonalizable.
Just as we have reached the conclusion of Theorem 4 for linear transformations, we define diagonalizability of a matrix, and reach a similar conclusion for matrices.

Definition 6: An $n \times n$ matrix A is said to be diagonalizable if A is similar to a diagonal matrix, that is, $P^{-1}AP$ is diagonal for some non-singular $n \times n$ matrix P.

Note that the matrix A is diagonalizable if and only if the matrix A regarded as a linear transformation $A: V_n(F) \to V_n(F) = AX$, is diagonalizable.

Thus, Theorem 2, 3, and 4 are true for the matrix A regarded as a linear transformation from $V_n(F)$ to $V_n(F)$.

Therefore, given an $n \times n$ matrix A, we know that it is diagonalizable if it has n distinct eigenvalues.
We now give a practical method of diagonalizing a matrix.

Theorem 5: Let A be $n \times n$ matrix having n distinct eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_n$ and
Let $X_1, X_2, \cdots, X_n \in V_n(F)$ be eigenvectors of A corresponding to $\lambda_1, \lambda_2, \cdots, \lambda_n$, respectively. Let $P = (X_1, X_2, \cdots, X_n)$ be the $n \times n$ matrix having $X_1, X_2, \cdots, X_n$ as its column vectors, then $P^{-1}AP = diag(\lambda_1, \lambda_2, \cdots, \lambda_n)$.

Proof: By actual multiplication, you can see that

$$\begin{aligned} AP &= A(X_1, X_2, \cdots, X_n) \\ &= (AX_1, AX_2, \cdots, AX_n) \\ &= (\lambda_1 X_1, \lambda_2 X_2, \cdots, \lambda_n X_n) \\ &= [X_1, X_2, \cdots, X_n] \begin{bmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ 0 & 0 & \lambda_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_n \end{bmatrix} \end{aligned}$$

$$P = diag(\lambda_1, \lambda_2, \cdots, \lambda_n)$$

Now, by Theorem 2, the column vectors of P are linearly independent.

This means that P is invertible (Unit 3).
Therefore, we can pre-multiply both sides of the matrix equation

$$AP = P\, diag(\lambda_1, \lambda_2, \cdots, \lambda_n)$$

Let us see how this theorem works in practice.

Theorem 6: Let $T: V \to V$ be a linear transformation, then the following are equivalent:

a) T is diagonalizable;

b) V is the direct sum of eigenspaces of T;

c) $T = \lambda_1\mu_1 + \lambda_2\mu_2 + \cdots + \mu_n\lambda_n$, where $\lambda_1, \lambda_2, \cdots, \lambda_n$ are the distinct eigenvalues of T, and $\mu_1, \mu_2, \cdots, \mu_n$, are projections satisfying $\mu_1 + \mu_2 + \cdots + \mu_n = I$ and $\mu_i\mu_r = 0$ for $i \neq j$.

Proof:
Let $\lambda_1, \lambda_2, \cdots, \lambda_n$ be the distinct eigenvalues of T, and let $v_{i1}, v_{i2}, \cdots, v_{im}$ be a basis for the λ_i −eigenspace of T. Then T is diagonalizable if and only if the union of these bases is a basis for V.
So (a) and (b) are equivalent.

Now suppose that (b) holds. A proposition and its converse show that there are projections $\mu_1, \mu_2, \cdots, \mu_n$ satisfy the conditions in (c) where the image $Im(\mu_i)$ is the λ_i −eigenspace.

Now in this case it is easily checked that T and $\sum \lambda_i\mu_i$ agree on every vector in V, so they are equal.
Hence (b) implies (c).

Finally, if $T = \sum \lambda_i\mu_i$, where the μ_i satisfy the conditions of (c), then V is the direct sum of the spaces $Im(\mu_i)$ and $Im(\mu_i)$ is the λ_i −eigenspace.
So (c) implies (b), hence the proof.

Example 15: The matrix $A = \begin{bmatrix} -6 & 6 \\ -12 & 11 \end{bmatrix}$ is diagonalizable since the eigenvectors $\begin{bmatrix} 3 \\ 4 \end{bmatrix}$ and $\begin{bmatrix} 2 \\ 3 \end{bmatrix}$ are linearly independent, and so form a basis for R.
Indeed, we see that $\begin{bmatrix} -6 & 6 \\ -12 & 11 \end{bmatrix}\begin{bmatrix} 3 & 4 \\ 2 & 3 \end{bmatrix} = \begin{bmatrix} 3 & 4 \\ 2 & 3 \end{bmatrix}\begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$, so that $P^{-1}AP$ is diagonal, where P is the matrix, whose columns are the eigenvectors of A.
Furthermore, one can find two projection matrices whose column spaces are the eigenspaces, namely $P_1 = \begin{bmatrix} 9 & -6 \\ 12 & -8 \end{bmatrix}$, $P_2 = \begin{bmatrix} -8 & 6 \\ -12 & 9 \end{bmatrix}$.
The reader can check directly that

i) ${P_1}^2 = P_1; {P_2}^2 = P_2$

ii) $P_1P_2 = P_2P_1 = 0$

iii) $P_1 + P_2 = I$
iv) $2P_1 + 3P_2 = A$

Proposition: Let $A = \sum_{i=1}^{r} \lambda_i P_i$ be the expression for the diagonalizable matrix A in terms of projections P_i satisfying the conditions of Theorem 6, that is, $\sum_{i=1}^{r} P_i = I$ and $P_i P_j = 0$ for $i \neq j$, then

(a) for any positive integer m, we have $A^m = \sum_{i=1}^{r} \lambda_i^m P_i$
(b) for any polynomial $f(x)$, we have $f(A) = \sum_{i=1}^{r} f(\lambda_i) P_i$.

Proof:
(a) The proof is by induction on m, the case $m = 1$ being the given expression.
Suppose that the result holds for $m = k - 1$. Then

$$A^k = A^{k-1}A$$
$$= \left(\sum_{i=1}^{r} \lambda_i^{k-1} P_i\right)\left(\sum_{i=1}^{r} \lambda_i P_i\right).$$

When we multiply out this product, all the terms $P_i P_j$ are zero for $i \neq j$ and we obtain
simply $\sum_{i=1}^{r} \lambda_i^{k-1} \lambda_i P_i$ as required. So, the induction goes through.
(b) If $f(x) = \sum a_m x^m$, we obtain the result by multiplying the equation of part (a) by a_m and
summing over m.
(Note that, for $m = 0$, we use the fact that $A^0 = I = \sum_{i=1}^{r} P_i = \sum_{i=0}^{r} \lambda_i^0 P_i$ that is, part (a) holds also for $m = 0$).

Example 16: Is matrix $A = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ diagonalisable?

Solution: The characteristic polynomial of $A =$
$f(t) \begin{vmatrix} t & -1 & 0 \\ -1 & t & 0 \\ 0 & 0 & t-1 \end{vmatrix} = (t+1)(t-1)^2.$

Matrix A is diagonalizable though it only has two distinct eigenvalues. This is because there is one linear independent eigenvector corresponding to $\lambda_1 = -1$, but there exist two linearly independent eigenvectors corresponding to $\lambda_2 = 1$.
Therefore, we can form a basis $V_3 \in R$ consisting of the eigenvectors;

$$\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

The matrix $P = \begin{bmatrix} 1 & 1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ is invertible, and $P^{-1}P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

Example 17: Verify the diagonalizability of the matrix $A = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 1 & -1 \\ 0 & 2 & 4 \end{bmatrix}$.

Solution:

The characteristic polynomial of $A = f(t) = \begin{vmatrix} t-2 & -1 & 0 \\ 0 & t-1 & 1 \\ 0 & -2 & t-4 \end{vmatrix} = (t-2)^2(t-3)$

Therefore, the eigenvalues are $\lambda_1 = 2$ and $\lambda_2 = 3$

The eigenvectors corresponding to $\lambda_1 = 2$ are given by $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$

The eigenvectors corresponding to $\lambda_2 = 3$ are given by $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$

Therefore $P = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$ is invertible, and $P^{-1} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$

$\therefore P^{-1}P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

SELF-ASSESSMENT EXERCISE

Obtain the eigenvalues and eigenvectors of the following matrices:

i. $A = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 3 & 2 \\ -1 & 1 & 2 \end{bmatrix}$

ii. $A = \begin{bmatrix} 1 & -4 & -2 \\ 0 & 3 & 1 \\ 1 & 2 & 4 \end{bmatrix}$

iii. $A = \begin{bmatrix} 2 & 0 & 1 \\ -1 & 4 & -1 \\ -1 & 2 & 0 \end{bmatrix}$

Conclusion

We conclude that any $n \times n$matrix A is said to be diagonalizable if A is similar to a diagonal matrix, that is, $P^{-1}AP$ is diagonal for some non-singular$n \times n$ matrix P.

1.4 Summary

We end this unit by summarizing what has been done in it.

- An eigenvalue of a linear transformation $T: V \to V$ (or matrix A) is the same as "characteristic value", "proper value" or "latent root" of T(or A).
- Let$A = [a_{ij}]$be any $n \times n$ matrix, then the characteristic polynomial is defined as $f_A(t) = det(tI - A)$ and characteristic equation of a linear transformation (or matrix) is defined as $f_A(t) = det(tI - A) = 0$.
- The roots of the characteristic polynomials are the eigenvalues of the linear transformation T (or matrix A).
- A scalar $\lambda$is an eigenvalue of a $n \times n$matrix A over F if there exists $X \in V_n(F); X \neq 0$, such that $AX = \lambda X$ are eigenvectors of the matrix A corresponding to the eigenvalueλ.
- A scalarλ is an eigenvalue of a linear transformation T (or matrix A) if and only if it is a root of the characteristic polynomial of T (or A).
- For an eigenvalue$\lambda$of T, the non-zero subspace W_k is called the eigenspace of T associated with the eigenvalue.
- Eigenvectors of a linear transformation (or matrix) corresponding to distinct eigenvalues are linearly independent.
- A linear transformation $T: V \to V$ is diagonalizable if and only if V has a basis consisting of eigenvectors of T.
- A linear transformation (or matrix) is diagonalizable if its eigenvalues are distinct.

So how, in practice, do we "diagonalize" a matrix A, that is, find an invertible matrix P such that $P^{-1}AP = D$ is diagonal? We saw an example of this earlier.

The matrix equation can be rewritten as $AP = PD$, from which we see that the columns of P are the eigenvectors of A.

So, the procedure is:

i. Find the eigenvalues of A,

ii. Find a basis of eigenvectors;

iii. Then let P be the matrix which has the eigenvectors as columns, and D the diagonal matrix whose diagonal entries are the eigenvalues.

1.7 References/Further Readings

Robert A. Beezer (2014). A First Course in Linear Algebra Congruent Press, Gig Harbor, Washington, USA 3(40).

Arbind K Lal Sukant Pati (2018). Linear Algebra through Matrices.

Peter J. Cameron (2008). Notes on Linear Algebra.

K.A. Stroud and Dexter J. Booth (2007). Engineering Mathematics 6th Edition. Palgrave Macmillan.

UNIT 2 CHARACTERISTIC AND MINIMAL POLYNOMIAL

Unit Structure

2.1 Introduction

This unit is basically a continuation of the previous unit, but the emphasis is on a different aspect of the problem discussed in the previous unit.

Let $T: V \to V$ be a linear transformation on a n-dimensional vector space V over the field F. The two most important polynomials that are associated with T are the characteristic polynomial of T and the minimal polynomial of T.

In this unit, we first show that every square matrix (of linear transformation $T: V \to V$) satisfies its characteristic equation, and use this to compute the inverse of the concerned matrix (or linear transformation), if it exists.

Then we define the minimal polynomial of a square matrix, and discuss the relationship between the characteristic and minimal polynomials. This leads us to a simple way of obtaining the minimal polynomial of a matrix (or linear transformation).

2.2 Learning Outcomes

By the end of this unit, you will be able to:

- State and prove the Cayley-Hamilton theorem;
- Find the inverse of an invertible matrix using this theorem;
- Prove that a scalar$\lambda$is an eigenvalue if and only it is a root of the minimal polynomial;
- Obtain the minimal polynomial of a matrix (or linear transformation) if the characteristic polynomial is known.

2.3 Characteristic and Minimal Polynomials

2.3.1 Cayley-Hamilton Theorem

This section presents the Cayley-Hamilton theorem, which is related to the characteristic equation of a matrix. It is named after the British Mathematicians; Arthur Cayley (1821-1895) and William Hamilton (1805-1865), they were responsible for a lot of work done in the theorem of determinants.

Example 1: Consider the 3×3 matrix $A = \begin{bmatrix} 0 & 1 & 2 \\ -1 & 2 & 1 \\ 0 & 3 & 2 \end{bmatrix}$, then $tI - A = \begin{bmatrix} t & -1 & -2 \\ 1 & t-2 & -1 \\ 0 & -3 & t-2 \end{bmatrix}$

Let C_{ij} be the $(i,j)^{th}$ cofactor of $(tI - A)$, then,

$C_{11} = (t-2)^2 - 3 = t^2 - 4t + 1$
$C_{12} = t - 2$
$C_{13} = -3$
$C_{21} = -1(t-2) - 6 = t + 4$
$C_{22} = t(t-2) = t^2 - 2t$
$C_{23} = 3t$
$C_{31} = 1 + 2(t-2) = 2t - 3$
$C_{32} = -t + 2$
$C_{33} = t(t-2) + 1 = t^2 - 2t + 1$

Therefore, the Matrix of cofactors is $C_{ij} = \begin{bmatrix} t^2-4t+1 & t-2 & -3 \\ t+4 & t^2-2t & 3t \\ 2t-3 & t-2 & t^2-2t+1 \end{bmatrix}$;

Hence $Adj(tI - A) = C_{ij}^T = \begin{bmatrix} t^2-4t+1 & t+4 & 2t-3 \\ t-2 & t^2-2t & t-2 \\ -3 & 3t & t^2-2t+1 \end{bmatrix}$

$$= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} t^2 + \begin{bmatrix} -4 & 1 & 2 \\ 1 & -2 & 1 \\ 0 & 3 & -2 \end{bmatrix} t + \begin{bmatrix} 1 & 4 & -3 \\ -2 & 0 & -2 \\ -3 & 0 & 1 \end{bmatrix}$$

This is a polynomial in t of degree 2, with matrix coefficients.

Similarly, if we consider the $n \times n$ matrix $A = |a_{ij}|$, then $Adj(tI - A)$ is a polynomial of degree $\leq n - 1$, with matrix coefficients.

Let $Adj(tI - A) = B_1 t^{n-1} + B_2 t^{n-2} + \cdots + B_{n-1} t^{n-1} + B_n$ ………………… (1)

where $B_1, B_2, \cdots, B_n$ are $n \times n$ matrices over F.
Now, the characteristic polynomial of A is given by

$$F(t) = f_A(t) = det(tI - A) = |tI - A|$$

$$= \begin{vmatrix} t - a_{11} & -a_{12} & \cdots & \cdots & -a_{1n} \\ -a_{21} & t - a_{22} & \cdots & \cdots & -a_{2n} \\ \vdots & \vdots & \cdots & \cdots & \vdots \\ \vdots & \vdots & \cdots & \cdots & \vdots \\ -a_{n1} & -a_{n2} & \cdots & \cdots & t - a_{nn} \end{vmatrix}, \text{ where } A = [a_{ij}]$$

$$= t^n + c_1 t^{n-1} + c_2 t^{n-2} + \cdots + c_{n-1} t + c_n$$

…………………….. (2)

The coefficients in (1) and (2) shall be used to prove the Cayley-Hamilton theorem.

Theorem 1 (Cayley-Hamilton):
Let $F(t) = t^n + c_1 t^{n-1} + c_2 t^{n-2} + \cdots + c_{n-1} t + c_n$ be the characteristic polynomial of an $n \times n$ matrix A, then $F(A) = A^n + c_1 A^{n-1} + c_2 A^{n-2} + \cdots + c_{n-1} A + c_n I = 0$
(Note that 0 denotes the $n \times n$ zero matrix, and $I = I_n$)

$$\begin{aligned} (tI - A)Adj(tI - A) &= Adj(tI - A) \\ &= det(tI - A)I \\ &= f(t)I \end{aligned}$$

……..………….. (3)

Equate equations (1) and (3) above becomes

$$(tI - A)(B_1 t^{n-1} + B_2 t^{n-2} + \cdots + B_{n-1} t^{n-1} + B_n) = F(t)I$$
$$= It^n + c_1 It^{n-1} + c_2 It^{n-2} + \cdots + c_{n-1} It + c_n I$$

Substituting the value of $F(t)$ to get

$$= f(t)$$

Now, comparing constant term and the coefficients of $t, t^2, \cdots, t^n$ on both sides we have

$$\begin{array}{ccccc} & - & AB_n & = & c_n I \\ B_n & - & AB_{n-1} & = & c_{n-1} I \\ B_{n-1} & - & AB_{n-2} & = & c_{n-2} I \\ \vdots & \vdots & \vdots & = & \vdots \\ \vdots & \vdots & \vdots & = & \vdots \\ B_3 & - & AB_2 & = & c_2 I \\ B_2 & - & AB_1 & = & c_1 I \\ & & B_1 & = & I \end{array}$$

Pre-multiplying the first equation by I, the second by A, the third by A^2 that last by A^n, and adding all these equations, we obtain

$$0 = c_n I + c_{n-1} A + c_{n-2} A^2 + \cdots + c_2 A^{n-2} + c_1 A^{n-1} + A^n = F(A)$$

Thus, the Cayley-Hamilton theorem is proved.
This theorem can also be stated as

"Every square matrix satisfies its characteristic polynomial".

Remark 1: You may be tempted to give the following 'quick' proof of Theorem 1:

$$F(t) = det(tI - A)$$
$$F(A) = det(AI - A) = det(A - A) = det(0) = 0$$

This proof is false. Why? Well, the left-hand side of the above equation, $F(A)$ is an $n \times n$ matrix while the right-hand side is the scalar 0, being the value of $det(0)$.

Now, as usual, we give the analogue of Theorem 1 for linear transformations:

Theorem 2 (Cayley-Hamilton): Let T be a linear transformation on a finite- dimensional vector space V. If $f(t)$ is the characteristic polynomial of T, then $F(T) = 0$

Proof: Let $dimV = n$ and let $B = \{v_1, v_2, \cdots, v_n\}$ be a basis of V.
In Unit 1, we have observed that
$F(t)$ = the characteristic polynomial of T
= the characteristic polynomial of the matrix $[T]_B$
Let $[T]_B = A$
If $F(t) = t^n + c_1t^{n-1} + c_2t^{n-2} + \cdots + c_{n-1}t + c_n$ then, by Theorem 1,

$$F(A) = A^n + c_1A^{n-1} + c_2A^{n-2} + \cdots + c_{n-1}A + c_nI = 0$$

Now, in Theorem 2 of Unit 3we proved that $[T]_B$is a vector space isomorphism. Thus,

$$[F(T)]_B = [T^n + c_1T^{n-1} + c_2T^{n-2} + \cdots + c_{n-1}T + c_nI]_B$$
$$= [T^n]_B + c_1[T^{n-1}]_B + c_2[T^{n-2}]_B + \cdots + c_{n-1}[T]_B + c_n[I]_B$$
$$= A^n + c_1A^{n-1} + c_2A^{n-2} + \cdots + c_{n-1}A + c_nI$$
$$= F(A) = 0$$

Again, using the one-one property of $[T]_B$, this implies that $F(T) = 0$.
Thus, Theorem 2 is true.

Example 2: Verify the Cayley-Hamilton theorem for $A = \begin{pmatrix} 3 & 2 \\ -1 & 0 \end{pmatrix}$

Solution: The characteristic polynomial of A is $\begin{vmatrix} 3 & 2 \\ -1 & 0 \end{vmatrix} = t^2 - 3t + 2$

Let's verify that $A^2 - 3A + 2I = 0$

Now, $A^2 = \begin{pmatrix} 3 & 2 \\ -1 & 0 \end{pmatrix}\begin{pmatrix} 3 & 2 \\ -1 & 0 \end{pmatrix} = \begin{pmatrix} 7 & 6 \\ -3 & -2 \end{pmatrix}$

$\therefore A^2 - 3A + 2I = \begin{pmatrix} 7 & 6 \\ -3 & -2 \end{pmatrix} - 3\begin{pmatrix} 3 & 2 \\ -1 & 0 \end{pmatrix} + 2\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$

Therefore, the Cayley-Hamilton theorem is true in this case.

Example 3: Verify the Cayley-Hamilton theorem for A, where $A = \begin{pmatrix} 0 & 1 & 0 \\ 3 & 0 & 1 \\ 1 & -2 & -1 \end{pmatrix}$

Solution: The characteristic polynomial of A is $\begin{vmatrix} t & -1 & 0 \\ 3 & t & -1 \\ 1 & 2 & t+1 \end{vmatrix} = t^3 + t^2 - t - 4$

Let's verify that $A^3 + A^2 - A - 4I = 0$

Now, $A^2 = \begin{pmatrix} 0 & 1 & 0 \\ 3 & 0 & 1 \\ 1 & -2 & -1 \end{pmatrix}\begin{pmatrix} 0 & 1 & 0 \\ 3 & 0 & 1 \\ 1 & -2 & -1 \end{pmatrix} = \begin{pmatrix} 3 & 0 & 1 \\ 1 & 1 & -1 \\ -7 & 3 & -1 \end{pmatrix}$

Also, $A^3 = A^2A = \begin{pmatrix} 3 & 0 & 1 \\ 1 & 1 & -1 \\ -7 & 3 & -1 \end{pmatrix}\begin{pmatrix} 0 & 1 & 0 \\ 3 & 0 & 1 \\ 1 & -2 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 1 & -1 \\ 2 & 3 & 2 \\ 8 & -5 & 4 \end{pmatrix}$

$$A^3 + A^2 - A - 4I$$

$$= \begin{pmatrix} 1 & 1 & -1 \\ 2 & 3 & 2 \\ 8 & -5 & 4 \end{pmatrix} + \begin{pmatrix} 3 & 0 & 1 \\ 1 & 1 & -1 \\ -7 & 3 & -1 \end{pmatrix} - \begin{pmatrix} 0 & 1 & 0 \\ 3 & 0 & 1 \\ 1 & -2 & -1 \end{pmatrix} - 4\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Hence the Cayley-Hamilton theorem is also true in this case.

Exercise: Verify the Cayley-Hamilton theorem for the following matrices:

i) $\begin{pmatrix} 7 & 6 & 0 \\ 2 & 3 & 0 \\ -2 & -2 & 1 \end{pmatrix}$ ii) $\begin{pmatrix} 1 & 0 & 1 \\ 0 & 3 & 1 \\ 3 & 3 & 4 \end{pmatrix}$

We shall now use Theorem 1 to prove a result that gives us a method for obtaining the inverse of an invertible matrix.

Theorem 3: $F(t) = t^n + c_1t^{n-1} + c_2t^{n-2} + \cdots + c_{n-1}t + c_n$ be the characteristic polynomial of an $n \times n$ matrix A. Then A^{-1} exists if $c_n \neq 0$ and, in this case,

$$A^{-1} = -c^{-1}(A^{n-1} + c_1A^{n-2} + c_2A^{n-3} + \cdots + c_{n-1})$$

Proof: By Theorem 1,

$F(A) = A^n + c_1A^{n-1} + c_2A^{n-2} + \cdots + c_{n-1}A + c_nI = 0$

This implies that

$A(A^{n-1} + c_1A^{n-2} + c_2A^{n-3} + \cdots + c_{n-1}) = -c_nI$ or

$(A^{n-1} + c_1A^{n-2} + c_2A^{n-3} + \cdots + c_{n-1})A = -c_nI$

$\Rightarrow -c_n^{-1}(A^{n-1} + c_1A^{n-2} + c_2A^{n-3} + \cdots + c_{n-1}) = IA^{-1}$

$\Rightarrow -c_n^{-1}(A^{n-1} + c_1A^{n-2} + c_2A^{n-3} + \cdots + c_{n-1}) = A^{-1}$

Thus, A is invertible.

Example 4: Is $A = \begin{pmatrix} 2 & 1 & 1 \\ -1 & 2 & -1 \\ -1 & 1 & 3 \end{pmatrix}$ invertible? If so, find A^{-1}.

Solution: The characteristic polynomial of A,

$$F(t) = \begin{vmatrix} t-2 & -1 & -1 \\ 1 & t-2 & 1 \\ 1 & -1 & t-3 \end{vmatrix} = t^3 - 7t^2 + 19t - 19$$

Since the constant term of $F(t) = -19 \neq 0$, hence A is invertible.
Now, by Theorem 1, $A^3 - 7A^2 + 19A - 19I = 0$

$\Rightarrow \left(\frac{1}{19}\right) A(A^2 - 7A + 19I) = I$

$\therefore A^{-1} = \left(\frac{1}{19}\right)(A^2 - 7A + 19I)$

Now, $A^2 = \begin{pmatrix} 2 & 1 & 1 \\ -1 & 2 & -1 \\ -1 & 1 & 3 \end{pmatrix}\begin{pmatrix} 2 & 1 & 1 \\ -1 & 2 & -1 \\ -1 & 1 & 3 \end{pmatrix} = \begin{pmatrix} 2 & 5 & 4 \\ -3 & 2 & -6 \\ -6 & 4 & 7 \end{pmatrix}$

Hence $A^{-1} = \left(\frac{1}{19}\right)\begin{pmatrix} 7 & -2 & -3 \\ 4 & 7 & 1 \\ 1 & -3 & 5 \end{pmatrix}$

To make sure that there has been no error in calculation, multiply this matrix by A, you should get I.
Now try the following exercise.
Obtain A^{-1} wherever possible for the matrices:

i) $\begin{pmatrix} 7 & 6 & 0 \\ 2 & 3 & 0 \\ -2 & -2 & 1 \end{pmatrix}$

ii) $\begin{pmatrix} 1 & 0 & 1 \\ 0 & 3 & 1 \\ 3 & 3 & 4 \end{pmatrix}$

2.3.2 Minimal Polynomial

In Module 1 Unit 4, we defined the minimal polynomial of a linear transformation $T: V \to V$. We said that it is the **monic polynomial of least degree** with coefficients in *F*, which is satisfied by *T*. But we weren't able to give a method of obtaining the minimal polynomial of *T*.
In this section, we will show that the minimal polynomial divides the characteristic polynomial. Moreover, the roots of the minimal polynomial are the same as those of the characteristic polynomial. Since it is easy to obtain the characteristic polynomial of *T*, these facts will give us a simply way of finding the minimal polynomial of *T*.

Let us first recall some properties of the minimal polynomial **(MP)** of T that we gave in unit 4. Let $p(t)$ be the minimal polynomial of *T*, then
MP1: $p(t)$ is a monic polynomial with coefficients in *F*.

MP2: If $q(t)$ is a non-zero polynomial over F such that $deg\ q < \ deg\ p$, then $q(t)\ \neq\ 0$.

MP3: If, for some polynomial g (over F, $g(T) = 0$, then $p(t)|\ g(t)$). That is, there exists a polynomial $h(t)$ over F such that $g(t)\ =\ p(t)h(t)$. We will now obtain the first link in the relationship between minimal polynomial and the characteristic polynomial

MP4: The minimal polynomial $p(t)$of a linear transformation divides its characteristic polynomial$f(t)$.

Proof: Let the characteristic polynomial and the minimal polynomial of *T* be $f(t)$and $p(t)$, respectively. By Theorem 2, $f(t) = 0$.
So, by **MP4**, $p(t)$ divides $f(t)$, as desired.

Before going on to show the full relationship between the minimal and characteristic polynomials, we state (but won't prove!) two theorems that will be used again and again, in this course as well as other courses.

Theorem 4: (Division algorithm for polynomials): Let *f* and g be two polynomials in *t* with coefficients in a field *F* such that *f* ≠ *0*. Then

i. there exist polynomials in r with coefficients in F such that $g = fq + r$, where r = *0* or $deg\ r < deg\ f$, and
ii. if we also have $gfq_1 + r_1$,with $r_1 = 0$ or $deg\ r_1 < deg\ f$, then $q = q_1$ and $r = r_1$

An immediate corollary follows.

Corollary: If g is a polynomial over *F* with $\lambda \neq 0$ as a root, then $g(t) = (t - \lambda)q(t)$, for some polynomial *q* over *F*.

Proof: By the division algorithm, taking $f = (t - \lambda)$ to have

$$g(t) = (t - \lambda)q(t) + r(t) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(1)$$

With $r = 0$ or $deg\ r < deg\ (t - \lambda) = 1$
If $deg\ r < 1$, then *r* is a constant.
Putting $t = \lambda$ in (1) gives $g(\lambda) = r(\lambda) = r$ since *r* is a constant.
But $g(\lambda) = 0$, since λ is root of g, then $r = 0$.
Thus, the only possibility is $r = 0$.
Hence, $g(t) = (t - \lambda)q(t)$

And now we come to a very important result that you may be using often, without realizing coefficients has at least one root in *C*.
In other words, this theorem says that any polynomial
$f(t) = \alpha_{n-1}t^{n-1} + \alpha_{n-2}t^{n-2} + \cdots + \alpha_1 t + \alpha_0$ (where $\alpha_0, \cdots, \alpha_n \in C, \alpha_n \neq 0, n \geq 1$) has at least one root in *C*.

Remark 2: In the theorem, if λ_1 is a root of $f(t) = 0$, then by theorem 1,

$f(t) = (t - \lambda_1)f_1(t);$ here, $deg\ f_1 = n - 1$

If $f_1(t)$ is not constant, then the equation $f_1(t) = 0$ has a root $\lambda_2 \in C$, and

$f_1(t) = (t - \lambda_2)f_2(t).$

Consequently, $f(t) = (t - \lambda_1)(t - \lambda_2)f_2(t);$ here $deg\ f_2 = n - 2.$

Using the fundamental theorem repeatedly, we obtain

$F(t) = \alpha_n(t - \lambda_1)(t - \lambda_2)\cdots(t - \lambda_n);$ for some $\lambda_1, \lambda_2, \cdots, \lambda_n \in C$, which are not necessarily distinct. (This process has to stop after *n* steps since $deg\ f = n$)

Thus, all the roots of $f(t) \in C$ and these are *n* in number. They may not all be distinct.

Suppose $\lambda_1, \lambda_2, \cdots, \lambda_k$ are the distinct roots, and they are repeated $m_1, m_2, \cdots, m_k$ times, respectively.

Then, $m_1 + m_2 + \cdots + m_k = n$ and

$f(t) = \alpha_n(t - \lambda_1)^{m_1}(t - \lambda_2)^{m_2}\cdots(t - \lambda_n)^{m_k}.$

For example, the polynomial equation $t^3 - it^2 + t - i$ has no real roots, but it has two distinct complex roots, namely, $t = -1$ and $t = \sqrt{-i}.$

So, we write $t^3 - it^2 + t - i.$

Here i is repeated twice and $-\ i$ only occurs once.

We can similarly show that any polynomial $f(t)$ over *r* can be written as a product of linear polynomials and quadratic polynomials.

For example, the real polynomial, $t^3 - 1 = (t - 1)(t^2 + t + 1).$

Now we go to show the second and final link that relates the minimal and characteristic polynomials of $T: V \to V$, where V is a vector space over F.

Let $p(t)$ be the minimal polynomial of *T*. We shall show that a scalar λ is a root of $p(t)$.

The proof will utilize the following remark.

Remark 3: If λ is an eigenvalue of *T*, then $Tx - \lambda x$ for some $x \in V, x \neq 0$

But $Tx - \lambda x \Rightarrow T^2x = T(Tx) = T(T\lambda x) = \lambda^2 x$

By induction, it is easy to see that $T^k x = \lambda^k x;\ \forall\ k.$

Now, if $g(t) = a_n t^n + a_{n-1}t^{n-1} + a_{n-2}t^{n-2} + \cdots + a_1 t + a_0$ is a polynomial over F, then, $g(T) = a_n T^n + a_{n-1}T^{n-1} + a_{n-2}T^{n-2} + \cdots + a_1 T + a_0 I$

This means that

$$\begin{aligned} g(T)x &= a_n T^n x + a_{n-1}T^{n-1}x + \cdots + a_1 Tx + a_0 x \\ &= a_n \lambda^n x + a_{n-1}\lambda^{n-1}x + \cdots + a_1 \lambda x + a_0 x \\ &= g(\lambda)x \end{aligned}$$

Thus, λ is an eigenvalue of $T \Rightarrow g(\lambda)$ is an eigenvalue of $g(T)$.

Theorem 5: Let T be a linear transformation on a finite-dimensional vector V over the field F. Then $\lambda \in F$ is an eigenvalue of T if and only if λ is a root of the minimal **polynomial of *T* have the same roots.**

Proof: Let p be the minimal polynomial of T and let $\lambda \in F$. Suppose λ is an eigenvalue of T, then $Tx - \lambda x$ for some $0 \neq x \in V$.
Also, by Remark 3, $p(T)x = 0$, but $p(T) = 0$.
Thus, $0 = p(\lambda) = 0$, that is, λ is a root of $p(t)$.
Conversely, let λ be a root of $p(\lambda) = 0$ and by Theorem 5,
$p(t) = (t - \lambda)q(t),\ deg\ q < deg\ p, q \neq 0$.
By the property MP3, $\exists\ v \in V$ such that $q(T)\ v \neq 0$.
Let $x = q(T)v \neq 0$, then, $(T - \lambda I)x = (t - \lambda I)q(T)v = p(T)v = 0$
$\Rightarrow Tx - \lambda x = 0 \Rightarrow Tx = \lambda x.$
Hence, λ is an eigenvalue of T.
So, λ is an eigenvalue of T iff λ is a root of the minimal polynomial of *T*.

In Unit 1, we have already observed that $\lambda$is an eigenvalue of *T* if and only if λ is a root of the characteristic polynomial of *T*.
Hence, we have shown that both the minimal and characteristic polynomials of *T* have the same roots, namely, the eigenvalues of *T*.

Caution: ***Though the roots of the characteristic polynomial and the minimal polynomial coincide, the two polynomials are not the same, in general.***

For example, if the characteristic polynomial $T: R^4 \rightarrow R^4$ is $(t+1)^2(t-2)^2$, then the minimal polynomial could be $(t+1)(t-1)$, or $(t+1)^2(t-2)$, or $(t+2)(t-2)^2$, or even $(t+1)^2(t-2)^2$, depending on which of these polynomials is satisfied by *T*.
In general, let $(t-\lambda_1)^{n_1}(t-\lambda_2)^{n_2}\cdots(t-\lambda_n)^{n_r}$ be the characteristic polynomial of a linear transformation *T*, where $deg\ f = n$; $(n_1 + n_2 + \cdots + n_r = n)$, and $\lambda_1, \lambda_2, \cdots, \lambda_r \in C$ are distinct. Then the minimal polynomial $p(t)$ is given by
$p(t) = (t-\lambda_1)^{m_1}(t-\lambda_2)^{m_2}\cdots(t-\lambda_n)^{m_r}$, where $1 \leq m_i \leq n_i$ for $i = 1,2, \dots\dots, r$.
In case *T* has *n* distinct eigenvalues, then $f(t) = (t-\lambda_1)(t-\lambda_2)\cdots(t-\lambda_n)$ and therefore, $p(t) = (t-\lambda_1)(t-\lambda_2)\cdots(t-\lambda_n) = f(t)$.

Definition: The minimal polynomial of a matrix *A* over *f* is the monic polynomial $p(t)$ such that

i. $p(A) = 0$, and
ii. If $q(t)$ is a non-zero polynomial over *F* such that $deg\ q < deg\ p$, then $q(A) \neq 0$.

Two theorems which are analogues to Theorems3and 5shall be stated. Their proofs are also similar to those of Theorems3 and 5.

Theorem 6: The minimal polynomial and the characteristic polynomial of a linear transformation (or matrix) have the same roots.

Theorem 7: The roots of the minimal polynomial and characteristic polynomial of a matrix are the same, and the roots are the eigenvalues of the matrix.
Let us use these theorems now.

Example 5: Obtain the minimal polynomial of $A = \begin{bmatrix} 5 & -6 & -6 \\ -1 & 4 & 2 \\ 3 & -6 & -4 \end{bmatrix}$

The characteristic polynomial of A is $f(t) = \begin{pmatrix} t-5 & 6 & 6 \\ 1 & t-4 & -2 \\ -3 & 6 & t+4 \end{pmatrix} =$
$(t-1)(t-2)^2$
Therefore, the minimal polynomial $p(t)$ is either $(t-1)(t-2)$or $(t-1)(t-2)^2$
Since$(A-I)(A-2I) = \begin{bmatrix} 4 & -6 & -6 \\ -1 & 3 & 2 \\ 3 & -6 & -5 \end{bmatrix} \begin{bmatrix} 3 & -6 & -6 \\ -1 & 2 & 2 \\ 3 & -6 & -6 \end{bmatrix} =$
$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$
$\Rightarrow p(t) = (t-1)(t-2)$

Example 6: Find the minimal polynomial of $A = \begin{bmatrix} 3 & 1 & -1 \\ 2 & 2 & -1 \\ 2 & 2 & 0 \end{bmatrix}$

The characteristic polynomial of A is

$$f(t) = \begin{pmatrix} t-3 & -1 & 1 \\ -2 & t-2 & 1 \\ -2 & -2 & t \end{pmatrix} = (t-1)(t-2)^2$$

Again, as in the example (3) above, the minimal polynomial $p(t)$ is either $(t-1)(t-2)$ or $(t-1)(t-2)^2$ but in this case,

$$(A-I)(A-2I) = \begin{bmatrix} 2 & 1 & -1 \\ 2 & 1 & -1 \\ 2 & 2 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 & -1 \\ 2 & 0 & -1 \\ 2 & 2 & -2 \end{bmatrix} = \begin{bmatrix} 2 & 0 & -1 \\ 2 & 0 & -1 \\ 4 & 0 & -2 \end{bmatrix}$$

$$\neq \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$\Rightarrow p(t) \neq (t-1)(t-2)$.
Thus, $p(t) = (t-1)(t-2)^2$.

Now, let T be a linear transformation for V to V, and B be a basis of V. Let $A = [T]_B$, if $g(t)$ is any polynomial with coefficients in *f*, then $g(T) = 0$ if and only if $g(A) = 0$.

Thus, the minimal polynomial of T is the same as the minimal of A.
For example, if $T: R^3 \longrightarrow R^3$ is a linear operator which is represented with respect to the standard basis, by the matrix in Example 3, then its minimal polynomial is $(t-1)(t-2)$.

Example 7: What can the minimal polynomial of $T: R^4 \longrightarrow R^4$ be if the characteristic polynomial of $[T]_B$ is
i. $(t-1)(t^3+1)$.
ii. $(t^2+1)^2$

Solution: i) We know that $(t-1)(t^3+1) = (t-1)(t+1)(t^2 - \mathrm{t} + 1)$.
This has 4 distinct complex roots, of which only 1 and -1 are real.
Since all the roots are distinct this polynomial is also the minimal polynomial of T.
ii) $(t^2+1)^2$ has no real roots. It has 2 repeated complex roots, i and $-i$.
Now, the minimal polynomial must be a real polynomial that divides the characteristic polynomial. Therefore, it can be (t^2+1) or $(t^2+1)^2$.
This example shows that if the minimal polynomial is a real polynomial, then it need not be a product of linear polynomials only.
Of course, over *C* it will always be a product of linear polynomials.

Solutions/Answers to Exercises

a) $$f_A(t) = \begin{vmatrix} t & -1 & 0 & -1 \\ -1 & t & -1 & 0 \\ 0 & -1 & t & -1 \\ -1 & 0 & -1 & t \end{vmatrix} = t^2(t-2)(t+2)$$

Therefore, the minimal polynomial can be $t(t-2)(t+2)$ or $t^2(t-2)(t+2)$
Now $A(A-2I)(A+2I) = 0$
$\therefore$ $t(t-2)(t+2)$ is the minimal polynomial of A

b) $$A = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}$$

$$f_A(t) = \begin{vmatrix} t-1 & -1 & 0 \\ 0 & t-1 & -1 \\ -1 & 0 & t-1 \end{vmatrix} = t^3 - t^2 - t$$

This has 3 distinct roots: 0, $\frac{1+i\sqrt{5}}{2}$, $\frac{1-i\sqrt{5}}{2}$
Therefore, the minimal polynomial is the same as $f_A(t)$.

E5) sum of its diagonal elements = 0
sum of eigenvalues $= 0 - 2 + 2 = 0$
$T_r(A) = -(coeff \text{ of } t^3 \text{in } f_A(t)) = 0$
$\therefore T_r(A)$ = sum of its diagonal elements of *A*
= sum of its eigenvalues of *A*

SELF-ASSESSMENT EXERCISE(S)

Try the following exercises now.
Find the minimal polynomial of

a) $A = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix}$

b) $T: R \longrightarrow R^3 / \ (x + y, \ \ y + z, \ z + x)$

The next exercise involves the concept of the trace of a matrix.

If $A = [a_{ij}] \in M_n(F)$, then the trace of A, denoted by $T_r(A)$ is – (coefficient of t^{n-1}in $f_A(t)$).

Let $A = [a_{ij}] \in M_n(F)$. For the matrix A given in Exercise 4, show that
$T_r(A)$ = (sum of its eigenvalues)
= (sum of its diagonal elements)

Conclusion

We end the unit by concluding that if the minimal polynomial is a real polynomial, then it need not be a product of linear polynomials only and that the roots of the minimal polynomial and characteristic polynomial of a matrix are the same, and the roots are the eigenvalues of the matrix.

2.4 Summary

In this unit we have covered the following points.

1) The proof of the Cayley-Hamilton theorem, which says that every square matrix (or linear transformation$T: V \rightarrow V$) satisfies its characteristic equation.
2) The use of the Cayley-Hamilton theorem to find the inverse of a matrix.
3) The definition of the minimal polynomial of a matrix.
4) The proof of the fact that the minimal polynomial and the characteristic polynomial of a linear transformation (or matrix) have the same roots. These roots are precisely the eigenvalues of the concerned linear transformation (or matrix).
5) A method for obtaining the minimal polynomial of a linear transformation (or matrix).

2.5 References/Further Readings

Arbind K Lal Sukant Pati (2018). Linear Algebra through Matrices.

Peter J. Cameron (2008). Notes on Linear Algebra.

Robert A. Beezer (2014). A First Course in Linear Algebra. Congruent Press Gig Harbor, Washington, USA 3(40).