DIRECTORATE OF DISTANCE EDUCATION

UNIVERSITY OF NORTH BENGAL

MASTER OF SCIENCES- MATHEMATICS

SEMESTER -III

LINEAR ALGEBRA

DEMATH3OLEC1

BLOCK-2

UNIVERSITY OF NORTH BENGAL

Postal Address:

The Registrar,

University of North Bengal,

Raja Rammohunpur,

P.O.-N.B.U., Dist-Darjeeling,

West Bengal, Pin-734013,

India.

Phone: (O) +91 0353-2776331/2699008

Fax: (0353) 2776313, 2699001

Email: regnbu@sancharnet.in ; regnbu@nbu.ac.in

Wesbsite: www.nbu.ac.in

First Published in 2019

FOREWORD

The Self Learning Material (SLM) is written with the aim of providing simple and organized study content to all the learners. The SLMs are prepared on the framework of being mutually cohesive, internally consistent and structured as per the university's syllabi. It is a humble attempt to give glimpses of the various approaches and dimensions to the topic of study and to kindle the learner's interest to the subject

We have tried to put together information from various sources into this book that has been written in an engaging style with interesting and relevant examples. It introduces you to the insights of subject concepts and theories and presents them in a way that is easy to understand and comprehend.

We always believe in continuous improvement and would periodically update the content in the very interest of the learners. It may be added that despite enormous efforts and coordination, there is every possibility for some omission or inadequacy in few areas or topics, which would definitely be rectified in future.

We hope you enjoy learning from this book and the experience truly enrich your learning and help you to advance in your career and future endeavours.

LINEAR ALGEBRA

BLOCK-1

BLOCK-2

BLOCK-2 LINEAR ALGEBRA

Introduction to Block

Linear algebra is the most applicable area of mathematics. It is one of the fields, that is accepted universally to be the prerequisite to be the in-depth understanding of the machine learning. This field is considered to be the mathematics of data and is especially used in the field of statistics, and used as a tool in Fourier series, computer graphics and so on. It is the study of vector spaces, lines and planes, and some mappings that are required to perform the linear transformations. It includes vectors, matrices and linear functions. It is the study of linear sets of equations and its transformation properties.

In this block we are going to explore the concept of Dual space and linear transformation. Comprehend the inner product space and its applications. Enumerate Quadratic and Bilinear forms. Understand in details about the Jordan Cannonical Form. Comprehend the Annihilating polynomials, diagonal forms, triangular forms. Understand the concepts of Direct Sum Decompositions, Invariant Direct sums & The Primary Decomposition Theorem.

Jordan canonical form is a representation of a linear transformation over a finite-dimensional complex vector space by a particular kind of upper triangular matrix. Every such linear transformation has a unique Jordan canonical form, which has useful properties: it is easy to describe and well-suited for computations. Certain terms like monic polynomial, minimal polynomial as well as annihilating polynomial and characteristic polynomial are clarified in details.

UNIT-8: LINEAR TRANSFORMATION AND DUAL SPACE

STRUCTURE

8.0 OBJECTIVE

Understand the concept of Matrix of Linear Transformation

Comprehend the Similarity of Matrices

Understand the concept of Dual Space

8.1 INTRODUCTION

In this section, we prove that if $\mathbb{V}$ and $\mathbb{W}$ are vector spaces over F with dimensions n and m, respectively, then any $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$ corresponds to a set of $m \times n$ matrices. Before proceeding further, the readers should recall the results on ordered basis.

8.2 MATRIX OF LINEAR TRANSFORMATION

So, let $\mathcal{A}$ = (**v**1, . . . , **v**n) and $\mathcal{B}$ = (**w**1, . . . , **w**m) be ordered bases of $\mathbb{V}$ and $\mathbb{W}$, respectively. Also, let $\mathcal{A}$ = [**v**1, . . . , **v**n] and $\mathcal{B}$ = [**w**1, . . . , **w**m] be the basis matrix of $\mathcal{A}$ and $\mathcal{B}$, respectively.

Then, **v** = $A[v]_{\mathcal{A}}$and **w** = $B[\boldsymbol{w}]_{\mathcal{B}}$, for all **v** $\in\mathbb{V}$ and **w** $\in$W. For $T \in \mathcal{L}(\mathbb{V},\mathbb{W})$ and **v** $\in \mathbb{V}$,

$$B[\boldsymbol{T}(\boldsymbol{v})]_{\mathcal{B}} = T(\boldsymbol{v}) = T(A[\boldsymbol{v}]A) = T(A)[\boldsymbol{v}]A$$
$$= [T(\boldsymbol{v}1) \cdots T(\boldsymbol{v}n)][\boldsymbol{v}]A$$
$$= [B[T(\boldsymbol{v_1})]_{\mathcal{B}} \cdots B[T(\boldsymbol{v_n})]_{\mathcal{B}}][\boldsymbol{v}]_{\mathcal{A}}$$
$$= B[T(\boldsymbol{v_1})]_{\mathcal{B}} \cdots B[T(\boldsymbol{v_n})]_{\mathcal{B}}[\boldsymbol{v}]_{\mathcal{A}}.$$

Therefore, $[\boldsymbol{T}(\boldsymbol{v})]_{\mathcal{B}} = [T(\boldsymbol{v_1})]_{\mathcal{B}} \cdots B[T(\boldsymbol{v_n})]_{\mathcal{B}}[\boldsymbol{v}]_{\mathcal{A}}$. as a vector in $\mathbb{W}$ has a unique expansion
in terms of basis elements. Note that the matrix $[[T(\boldsymbol{v_1})]_{\mathcal{B}} \cdots$
$B[T(\boldsymbol{v_n})]_{\mathcal{B}}]$, denoted $T[\mathcal{A},\mathcal{B}]$,
is an $m \times n$ matrix and is unique with respect to the ordered basis B as the i-th column equals
$[T(\mathbf{v}i)]B$, for $1 \le i \le n$. So, we immediately have the following definition and result.

Definition 8.2.1. [Matrix of a Linear Transformation] Let A = (**v**1, . . . , **v**n) and B =
(**w**1, . . . , **w**m) be ordered bases of V and W, respectively. If $T \in \mathcal{L}(\mathbb{V},\mathbb{W})$ then the matrix
$T[A, B]$ is called the **coordinate matrix** of T or the **matrix of the linear transformation**
T with respect to the basis A and B, respectively.

When there is no mention of bases, we take it to be the standard ordered bases and denote the corresponding matrix by $[T]$.

Note that if **c** is the coordinate vector of an element **v** $\in$V then, $T[A, B]$**c** is the coordinate
vector of T(**v**). That is, the matrix $T[A, B]$ takes coordinate vector of the

domain points to the

coordinate vector of its images.

Theorem 8.2.2. *Let A* = (**v**1, . . . , **v***n*) *and B* = (**w**1, . . . , **w***m*) *be ordered bases of* $\mathbb{V}$ *and* $\mathbb{W}$*, respectively. If* $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$ *then there exists a matrix* $S \in M_{m\times n}(\mathbb{F})$ *with*

$$S = T[\mathcal{A}, \mathcal{B}] = [[T(\boldsymbol{v_1})]_{\mathcal{B}} \cdots B[T(\boldsymbol{v_n})]_{\mathcal{B}}], \text{ and } [T(\mathbf{x})]_B = S[\mathbf{x}]_A,$$

for all $\mathbf{x} \in$ V.

Remark 8.1.3. *Let* $\mathbb{V}$ *and* $\mathbb{W}$ *be vector spaces over* $\mathbb{F}$ *with ordered bases*

*A*1 = (**v**1, . . . , **v***n*)

*and B*1 = (**w**1, . . . , **w***m*)*, respectively. Also, for* $\alpha \in \mathbb{F}$ *with* $\alpha \neq 0$*, let A*2

= $(\alpha\mathbf{v}_1, \ldots, \alpha\mathbf{v}_n)$ *and*

*B*1 = $(\alpha\mathbf{w}_1, \ldots, \alpha\mathbf{w}_m)$ *be another set of ordered bases of* V *and* W*, respectively. Then, for any* $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$

$$T[\mathcal{A}_2, B_2] = \left[[T(\alpha\boldsymbol{v_1})]_{\mathcal{B}_2} \cdots [T(\alpha\boldsymbol{v}1)]_{\mathcal{B}_2}\right]$$
$$= [[T(\boldsymbol{v_n})]_{\mathcal{B}_1} \cdots [T(\boldsymbol{v}1)]_{\mathcal{B}_1}] = T[A_1, B_1].$$

Thus, we see that the same matrix can be the matrix representation of T for two different pairs

of bases.

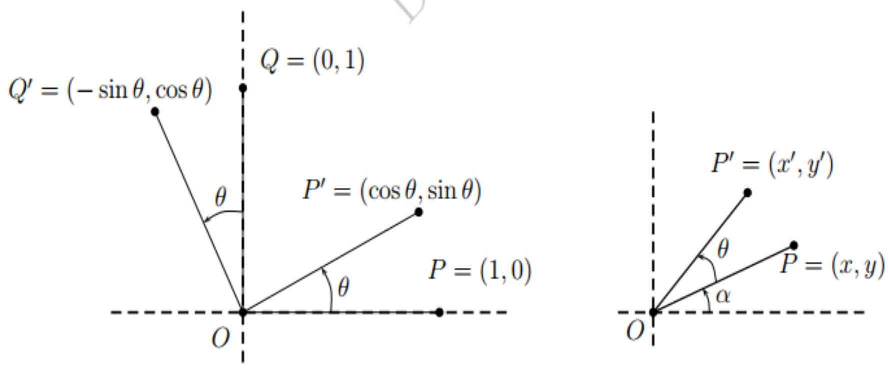


Figure 8.2: Counter-clockwise Rotation by an angle $\boldsymbol{\theta}$

Example: 1. Let $T \in L(\mathbb{R}^2)$ represent a counterclockwise rotation by an angle θ, $0 \leq \theta < 2\pi$. Then, using Figure 8.1, $x = OP \cos \alpha$ and $y = OP \sin \alpha$, verify that

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} OP' \cos(\alpha+\theta) \\ OP' \sin(\alpha+\theta) \end{bmatrix} = \begin{bmatrix} OP\big(\cos\alpha\cos\theta - \sin\alpha\sin\theta\big) \\ OP\big(\sin\alpha\cos\theta + \cos\alpha\sin\theta\big) \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$

Or equivalently, the matrix in the standard ordered basis of $\mathbb{R}^2$ equals

$$[T] = \big[T(\mathbf{e}_1), T(\mathbf{e}_2)\big] = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}. \qquad \text{(A)}$$

2. Let $T \in \mathcal{L}(\mathbb{R}^2)$ with $T((x, y)^T) = (x + y, x - y)^T$.

(a) Then $[T] = \big[[T(\mathbf{e}_1)] \;\; [T(\mathbf{e}_2)]\big] = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$.

(b) On the image space take the ordered basis $\mathcal{B} = \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix}\right)$. Then

$$[T] = \big[[T(\mathbf{e}_1)]_{\mathcal{B}} \;\; [T(\mathbf{e}_2)]_{\mathcal{B}}\big] = \left[\begin{bmatrix} 1 \\ 1 \end{bmatrix}_{\mathcal{B}} \;\; \begin{bmatrix} 1 \\ -1 \end{bmatrix}_{\mathcal{B}}\right] = \begin{bmatrix} 0 & 2 \\ 1 & -1 \end{bmatrix}.$$

(c) In the above, let the ordered basis of the domain space be $\mathcal{A} = \left(\begin{bmatrix} -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \end{bmatrix}\right)$. Then

$$T[\mathcal{A}, \mathcal{B}] = \left[\left[T\begin{bmatrix} -1 \\ 1 \end{bmatrix}\right]_{\mathcal{B}} \;\; \left[T\begin{bmatrix} 3 \\ 1 \end{bmatrix}\right]_{\mathcal{B}}\right] = \left[\begin{bmatrix} 0 \\ -2 \end{bmatrix}_{\mathcal{B}} \;\; \begin{bmatrix} 4 \\ 2 \end{bmatrix}_{\mathcal{B}}\right] = \begin{bmatrix} 2 & 2 \\ -2 & 2 \end{bmatrix}.$$

3. Let $A = (\mathbf{e}1, \mathbf{e}2)$ and $B = (\mathbf{e}1 + \mathbf{e}2, \mathbf{e}1 - \mathbf{e}2)$ be two ordered bases of $\mathbb{R}^2$.
Then Compute $T[A, A]$ and $T[B, B]$, where $T((x, y)^T) = (x + y, x - 2y)^T$

.

Solution: Let $A = \mathrm{Id}_2$ and $B = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$. Then, $A^{-1} = \mathrm{Id}_2$ and $B^{-1} = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$. So,

$$T[\mathcal{A}, \mathcal{A}] = \left[\left[T\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right)\right]_{\mathcal{A}}, \left[T\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right)\right]_{\mathcal{A}}\right] = \left[\begin{bmatrix} 1 \\ 1 \end{bmatrix}_{\mathcal{A}}, \begin{bmatrix} 1 \\ -2 \end{bmatrix}_{\mathcal{A}}\right] = \begin{bmatrix} 1 & 1 \\ 1 & -2 \end{bmatrix} \text{ and}$$

$$T[\mathcal{B}, \mathcal{B}] = \left[\left[T\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right)\right]_{\mathcal{B}}, \left[T\left(\begin{bmatrix} 1 \\ -1 \end{bmatrix}\right)\right]_{\mathcal{B}}\right] = \left[\begin{bmatrix} 2 \\ -1 \end{bmatrix}_{\mathcal{B}}, \begin{bmatrix} 0 \\ 3 \end{bmatrix}_{\mathcal{B}}\right] = \begin{bmatrix} \frac{1}{2} & \frac{3}{2} \\ \frac{3}{2} & -\frac{3}{2} \end{bmatrix}$$

as $\begin{bmatrix} 2 \\ -1 \end{bmatrix}_{\mathcal{B}} = B^{-1}\begin{bmatrix} 2 \\ -1 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 3 \end{bmatrix}_{\mathcal{B}} = B^{-1}\begin{bmatrix} 0 \\ 3 \end{bmatrix}$. Also, verify that $T[\mathcal{B}, \mathcal{B}] = B^{-1}T[\mathcal{A}, \mathcal{A}]B$.

Example [Finding T from $T[A, B]$]

1. Let $\mathbb{V}$and $\mathbb{W}$ be vector spaces over $\mathbb{F}$ with ordered bases A and B, respectively. Suppose
we are given the matrix $S = T[A, B]$. Then determine the corresponding $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$.

Solution: Let B be the basis matrix corresponding to the ordered basis B. Then, using
Equation and Theorem 8.1.2, we see that
$T(\mathbf{v}) = B[T(\mathbf{v})]B = BT[A, B][\mathbf{v}]A = BS[\mathbf{v}]A$.
2. In particular, if V = W = Fn and $A = B$ then we see that
$T(\mathbf{v}) = BSB{-}1\mathbf{v}$.
(B)

8.3 SIMILARITY OF MATRICES

Let $\mathbb{V}$ be a vector space over $\mathbb{F}$ with dim($\mathbb{V}$) = n and ordered basis B. Then any $T \in \mathcal{L}(\mathbb{V})$
corresponds to a matrix in $M_n(\mathbb{F})$. What happens if the ordered basis needs to change? We answer this in this subsection.

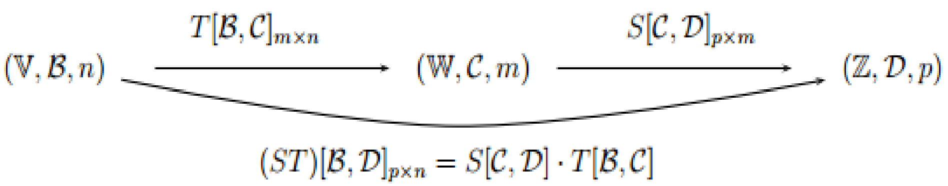


Fig 8.3: Composition of Linear Transformations

Theorem 8.3.1 (Composition of Linear Transformations)**.** *Let* V, W *and* Z *be finite dimensional vector spaces over* F *with ordered bases B, C and D, respectively. Also, let $T \in L$*(V, W)
and $S \in L$(W, Z)*. Then $S \circ T = ST \in \mathcal{L}(\mathbb{V}, \mathbb{Z})$ (see Figure 8.2). Then*

$$(ST)[B, D] = S[C, D] \cdot T[B, C].$$

Proof. Let $B = (\mathbf{u}1, \ldots, \mathbf{u}n)$, $C = (\mathbf{v}1, \ldots, \mathbf{v}m)$ and $D = (\mathbf{w}1, \ldots, \mathbf{w}p)$ be the ordered bases of

V, W and Z, respectively. Then using Theorem 4.3.2, we have

$(ST)[B, D] = [[ST(\mathbf{u}1)]D, \ldots, [ST(\mathbf{u}n)]D] = [[S(T(\mathbf{u}1))]D, \ldots, [S(T(\mathbf{u}n))]D]$

$= [S[C, D]\,[T(\mathbf{u}1)]C, \ldots, S[C, D]\,[T(\mathbf{u}n)]C]$

$= S[C, D]\,[[T(\mathbf{u}1)]C, \ldots, [T(\mathbf{u}n)]C] = S[C, D] \cdot T[B, C].$

Hence, the proof of the theorem is complete.

As an immediate corollary of Theorem 4.4.1 we have the following result.

Theorem 8.3.2 (Inverse of a Linear Transformation)**.** *Let* V *is a vector* space with dim(V) = n.

If T ∈ L(𝕍) is invertible then for any ordered basis B and C of the domain and co-domain,

respectively, one has (T [C, B])−1 = T −1[B, C]. That is, the inverse of the coordinate matrix of

T is the coordinate matrix of the inverse linear transform.

Proof. As T is invertible, TT^{-1} = Id. Thus Theorem 8.2.1 imply

$In = \mathrm{Id}[B, B] = (TT^{-1})[B, B] = T[C, B] \cdot T^{-1}[B, C].$

Hence, by definition of inverse, $T^{-1}[B, C] = (T[C, B])^{-1}$ and the required result follows.

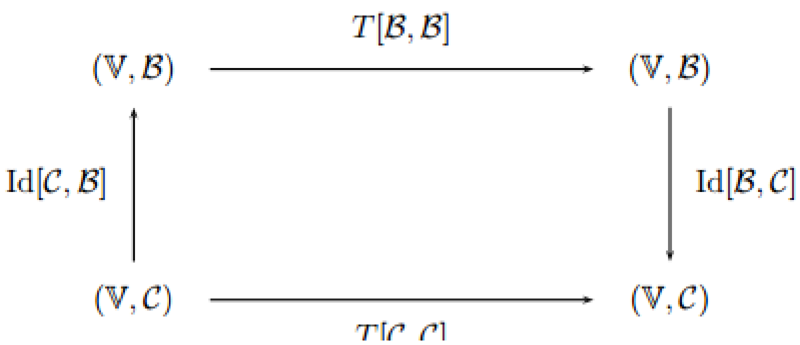


Figure 8.3: *T* [*C*, *C*] = Id[*B*, *C*] · *T* [*B*, *B*] · Id[*C*, *B*] - Similarity of Matrices

Let $\mathbb{V}$ be a finite dimensional vector space. Then, the next result answers the question "what
happens to the matrix $T [B, B]$ if the ordered basis B changes to C?"

Theorem 8.2.3. *Let* $B = (\mathbf{u}1, \ldots, \mathbf{u}n)$ *and* $C = (\mathbf{v}1, \ldots, \mathbf{v}n)$ *be two* ordered bases of V and Id
the identity operator. Then, for any linear operator $T \in \mathcal{L}(\mathbb{V})$

$$T [C, C] = Id[B, C] \cdot T [B, B] \cdot Id[C, B] = (Id[C, B])^{-1} \cdot T [B, B] \cdot Id[C, B]. \qquad \text{(A)}$$

Proof. As Id is an identity operator, $T [B, C]$ as $(\mathrm{Id} \circ T \circ \mathrm{Id})[B, C]$ (see Figure 8.3 for clarity).

Thus, using Theorem 8.3.1, we get

$$T [B, C] = (\mathrm{Id} \circ T \circ \mathrm{Id})[B, C] = \mathrm{Id}[B, C] \cdot T [B, B] \cdot \mathrm{Id}[C, B].$$

Hence, using Theorem 8.2.2, the required result follows.

Let $\mathbb{V}$ be a vector space and let $T \in L(\mathbb{V})$. If $\dim(\mathbb{V}) = n$ then every ordered basis B of V gives an $n \times n$ matrix $T [B, B]$. So, as we change the ordered basis, the coordinate matrix of T changes. Theorem 8.2.1 tells us that all these matrices are related by an invertible matrix.
Thus, we are led to the following definitions.

Definition 8.3.4. [Change of Basis Matrix] Let $\mathbb{V}$ be a vector space with ordered bases B
and C. If $T \in \mathcal{L}(\mathbb{V})$ then, $T [C, C] = \mathrm{Id}[B, C] \cdot T [B, B] \cdot \mathrm{Id}[C, B]$. The matrix $\mathrm{Id}[B, C]$ is called the **change of basis matrix** from B to C.

Definition 8.3.5. [Similar Matrices] Let $X, Y \in \mathrm{M}n(\mathbb{C})$. Then, X and Y are said to be
similar if there exists a non-singular matrix P such that $P^{-1}XP = Y \Leftrightarrow XP = PY$.

Example : Let *B* = 1 + *x*, 1 + 2*x* + *x*2, 2 + *x* and *C* = 1, 1 + *x*, 1 + *x* + *x*2 be ordered

bases of R[*x*; 2]. Then, verify that Id[*B*, *C*]−1 = Id[*C*, *B*], as

$$\mathrm{Id}[\mathcal{C},\mathcal{B}] = [[1]_{\mathcal{B}}, [1+x]_{\mathcal{B}}, [1+x+x^2]_{\mathcal{B}}] = \begin{bmatrix} -1 & 1 & -2 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \text{ and}$$

$$\mathrm{Id}[\mathcal{B},\mathcal{C}] = [[1+x]_{\mathcal{C}}, [1+2x+x^2]_{\mathcal{C}}, [2+x]_{\mathcal{C}}] = \begin{bmatrix} 0 & -1 & 1 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix}.$$

Check Your Progress

1. Explain matrix of linear transformation

2. State Composition of Linear Transformations with proof

3. Define

(a)- Change of Basis

(b) Similar Matrices

8.4 DUAL SPACE*

Definition 8.4.1. [linear Functional] Let 𝕍 be a vector space over F. Then a map $T \in L(V, F)$ is called a **linear functional** on 𝕍.

Definition 8.4.2. [Dual Space] Let 𝕍 be a vector space over F. Then $\mathcal{L}(\mathbb{V}, \mathbb{F})$ is called the

dual space of V and is denoted by V*. The **double dual space** of 𝕍, denoted V**, is the dual

space of V*.

We first give an immediate corollary of Theorem 4.2.17.

Corollary 8.4.3. *Let* $\mathbb{V}$ *and* $\mathbb{W}$ *be vector spaces over* F *with* dim $\mathbb{V}$ = *n* *and* dim $\mathbb{W}$ = *m.*

1. Then $\mathcal{L}(\mathbb{V}, \mathbb{W}) \cong F^{mn}$. Moreover, $\{\mathbf{f}ij | 1 \leq i \leq n, 1 \leq j \leq m\}$ is a basis of $\mathcal{L}(\mathbb{V}, \mathbb{W})$.

2. In particular, if W = F then $\mathcal{L}(\mathbb{V}, \mathbb{W}) = V^{**} = \mathbb{F}^n$. Moreover, if $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a basis of

V then the set $\{\mathbf{f}i | 1 \leq i \leq n\}$ is a basis of V*, where

$$\mathbf{f}_i(\mathbf{v}_k) = \begin{cases} 1, & if\ k = i \\ 0, & k \neq i. \end{cases}$$

The basis $\{\mathbf{f}i | 1 \leq i \leq n\}$ is called the **dual basis** of $\mathbb{F}^n$.

So, we see that $\mathbb{V}$*. can be understood through a basis of $\mathbb{V}$. Thus, one can understand $\mathbb{V}$**.

again via a basis of $\mathbb{V}$*. But, the question arises "can we understand it directly via the vector

space V itself?" We answer this in affirmative by giving a canonical isomorphism from V to $\mathbb{V}$**.

To do so, for each $\mathbf{v} \in \mathbb{V}$, we define a map $L\mathbf{v} : \mathbb{V}^* \to \mathbb{F}$ by $L\mathbf{v}(\mathbf{f}) = \mathbf{f}(\mathbf{v})$, for each $\mathbf{f} \in \mathbb{V}^*$. Then

L**v** is a linear functional as

$$L\mathbf{v}(\alpha\mathbf{f} + \mathbf{g}) = (\alpha\mathbf{f} + \mathbf{g})(\mathbf{v}) = \alpha\mathbf{f}(\mathbf{v}) + \mathbf{g}(\mathbf{v}) = \alpha L\mathbf{v}(\mathbf{f}) + L\mathbf{v}(\mathbf{g}).$$

So, for each $\mathbf{v} \in \mathbb{V}$, we have obtained a linear functional $L\mathbf{v} \in \mathbb{V}^{**}$.Note that, if $\mathbf{v} \neq \mathbf{w}$ then, L

$\mathbf{v} \neq L\mathbf{w}$. Indeed, if $L\mathbf{v} = L\mathbf{w}$ then, $L\mathbf{v}(f) = L\mathbf{w}(f)$, for all $f \in \mathbb{V}^*$. Thus, $f(\mathbf{v})$

= f(**w**), for all f ∈ 𝕍*. That is, f(**v** − **w**) = 0, for each f ∈ 𝕍*. Hence, we get **v** − **w** = **0**, or equivalently, **v** = **w**.

We use the above argument to give the required canonical isomorphism.

Theorem 8.4.4. Let V be a vector space over F. If dim(𝕍) = n then the *canonical map*

T : 𝕍 → 𝕍**.*defined by* *T* (**v**) = *L***v** *is an isomorphism.*

Proof. Note that for each **f** ∈ 𝕍*,

$L_{\alpha\mathbf{v}+\mathbf{u}}(\mathbf{f}) = \mathbf{f}(\alpha\mathbf{v} + \mathbf{u}) = \alpha\mathbf{f}(\mathbf{v}) + \mathbf{f}(\mathbf{u}) = \alpha L_{\mathbf{v}}(\mathbf{f}) + L_{\mathbf{u}}(\mathbf{f}) = (\alpha L_{\mathbf{v}} + L_{\mathbf{u}})(\mathbf{f})$.

Thus, $L_{\alpha\mathbf{v}+\mathbf{u}} = \alpha L_{\mathbf{v}}+L_{\mathbf{u}}$. Hence, $T(\alpha\mathbf{v}+\mathbf{u}) = \alpha T(\mathbf{v})+T(\mathbf{u})$. Thus, *T* is a linear transformation.

For verifying *T* is one-one, assume that *T* (**v**) = *T* (**u**), for some **u**, **v** ∈ 𝕍. Then, $L_{\mathbf{v}} = L_{\mathbf{u}}$. Now,

use the argument just before this theorem to get **v** = **u**. Therefore, *T* is one-one.Thus, *T* gives an inclusion (one-one) map from 𝕍 to 𝕍 **.

Further, applying Corollary 8.3.3.2 to 𝕍 *, gives dim(𝕍 **) = dim(𝕍 *) = *n*. Hence, the required result follows.

Corollary 8.4.5. *Let* 𝕍 *be a vector space of dimension n with basis B* = *{***v**1, . . . , **v***n}.*

1. Then, a basis of 𝕍 **, the double dual of 𝕍, equals D = {L**v**1, . . . , L**v**n}. Thus, for each

T ∈ 𝕍 ** there exists **x** ∈ V such that T (**f**) = **f**(**x**), for all **f** ∈ 𝕍 *. Or equivalently, there

exists **x** ∈ 𝕍 such that T = T**x**.

2. If C = {**f**1, . . . , **f**n} is the dual basis of 𝕍* defined using the basis B then D is indeed the dual basis of 𝕍 ** obtained using the basis C of 𝕍 *. Thus, each basis of 𝕍 * is the dual basis of some basis of V.

Proof. Part 1 is direct as T : 𝕍 → 𝕍 ** was a canonical inclusion map.

For Part 2, we need to

$$L_{\mathbf{v}_i}(\mathbf{f}_j) = \begin{cases} 1, & \text{if } j = i \\ 0, & \text{if } j \neq i \end{cases} \quad \text{or equivalently } \mathbf{f}_j(\mathbf{v}_i) = \begin{cases} 1, & \text{if } j = i \\ 0, & \text{if } j \neq i \end{cases}$$

show that

which indeed holds true using Corollary 8.4.3.2.

Let $\mathbb{V}$ be a finite dimensional vector space. Then Corollary 8.3.5 implies that the spaces $\mathbb{V}$ and $\mathbb{V}$ * are naturally dual to each other.

We are now ready to prove the main result of this subsection. To start with, let $\mathbb{V}$ and $\mathbb{W}$ be vector spaces over F. Then, for each $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$, we want to define a map $\hat{T} : \mathbb{W}^* \to \mathbb{V}.^*$

So, if $g \in \mathbb{W}$* then, $\hat{T}(\mathbf{g})$ a linear functional from V to F. So, we need to be evaluate T b($\mathbf{g}$) at

an element of V. Thus, we define T b($\mathbf{g}$)($\mathbf{v}$) = g (T ($\mathbf{v}$)), for all $\mathbf{v} \in$ V.

Now, we note that

$\hat{T} \in \mathcal{L}(\mathbb{W}^* \mathbb{V}^*)$, as for every g, $h \in \mathbb{W}^*$,

$$\hat{T}(\alpha\boldsymbol{g} + \boldsymbol{h})(\boldsymbol{v}) = (\alpha\boldsymbol{g} + \boldsymbol{h})(T(\boldsymbol{v})) = \alpha\boldsymbol{g}(T(\boldsymbol{v})) + \boldsymbol{h}(T(\boldsymbol{v})) = \alpha\hat{T}(\boldsymbol{g}) + \hat{T}(\boldsymbol{h})(\boldsymbol{v}),$$

For all $\mathbf{v} \in \mathbb{V}$ implies that $\hat{T}(\alpha\boldsymbol{g} + \boldsymbol{h}) = \alpha\hat{T}(\boldsymbol{g}) + \hat{T}(\boldsymbol{h})$.

Theorem 8.4.6. *Let* $\mathbb{V}$ *and* $\mathbb{W}$ *be two vector spaces over* $\mathbb{F}$ *with ordered bases* A = ($\mathbf{v}$1, . . . , $\mathbf{v}n$)
and B = ($\mathbf{w}$1, . . . , $\mathbf{w}m$), *respectively. Also, let* A* = ($\mathbf{f}$1, . . . , $\mathbf{f}n$) *and* B* = ($\mathbf{g}$1, . . . , $\mathbf{g}m$) *be the*
corresponding ordered bases of the dual spaces $\mathbb{V}$ * and $\mathbb{W}$ *, respectively. Then, T b[B*, A*] = (T [A, B])T ,the transpose of the coordinate matrix T.

Proof. Note that we need to compute $\hat{\mathrm{T}}$ [B*, A*] = $[[\hat{\mathrm{T}}(\mathbf{g_1})]_{A^*}, \ldots, [\hat{\mathrm{T}}(\mathbf{g_{1m}})]_{A^*}]$ and prove that it equals the transpose of the matrix T [A, B]. So, let

$$T[\mathcal{A},\mathcal{B}] = [[T(\mathbf{v}_1)]_{\mathcal{B}}, \ldots, [T(\mathbf{v}_n)]_{\mathcal{B}}] = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}.$$

Thus, to prove the required result, we need to show that

$$\left[\widehat{T}(\mathbf{g}_j)\right]_{\mathcal{A}^*} = [\mathbf{f}_1, \ldots, \mathbf{f}_n] \begin{bmatrix} a_{j1} \\ a_{j2} \\ \vdots \\ a_{jn} \end{bmatrix} = \sum_{k=1}^{n} a_{jk}\mathbf{f}_k, \text{ for } 1 \le j \le m.$$

(A)

Now, recall that the functionals **f***i*'s and **g***j*'s satisfy

$$\left(\sum_{k=1}^{n} \alpha_k \mathbf{f}_k\right)(\mathbf{v}_t) = \sum_{k=1}^{n} \alpha_k \left(\mathbf{f}_k(\mathbf{v}_t)\right) = \alpha_t,$$

for $1 \le t \le n$ and [**g***j*(**w**1), . . . , **g***j*(**w***m*)] = e_j^T, a row vector with 1 at the *j*-th place and 0, elsewhere. So, let *B* = [**w**1, . . . , **w***m*] and evaluate $\hat{T}$ (**g***j*) at **v***t*'s, the elements of *A*.

$$\begin{aligned} \left(\widehat{T}(\mathbf{g}_j)\right)(\mathbf{v}_t) &= \mathbf{g}_j\left(T(\mathbf{v}_t)\right) = \mathbf{g}_j\left(B\left[T(\mathbf{v}_t)\right]_{\mathcal{B}}\right) = [\mathbf{g}_j(\mathbf{w}_1), \ldots, \mathbf{g}_j(\mathbf{w}_m)]\left[T(\mathbf{v}_t)\right]_{\mathcal{B}} \\ &= \mathbf{e}_j^T \begin{bmatrix} a_{1t} \\ a_{2t} \\ \vdots \\ a_{mt} \end{bmatrix} = a_{jt} = \left(\sum_{k=1}^{n} a_{jk}\mathbf{f}_k\right)(\mathbf{v}_t). \end{aligned}$$

Thus, the linear functional $\hat{T}$ ($\mathbf{g}_j$) and $\sum_{k=1}^{n} a_{jk} f_k$ are equal at **v***t*, for $1 \le t$ $\le n$, the basis vectors
of V. Hence $\hat{T}$ (**g***j*) = $\sum_{k=1}^{n} a_{jk} f_k$ which gives Equation (A).

Remark 8.4.7 . *The above proof of Theorem also shows the following.*

Notes

1. For each T ∈ $\mathcal{L}(\mathbb{V}, \mathbb{W})$ there exists a unique map T b ∈ $\mathcal{L}(\mathbb{V}^*, \mathbb{W}^*)$such that

$\hat{T}$ (**g**)(**v**) = **g** (T (**v**)) , for each **g** ∈ W*.

2. The coordinate matrices T [A, B] and T b[B*, A*] are transpose of each other, where the ordered bases A* of $\mathbb{V}^*$ and B* of $\mathbb{W}^*$ correspond, respectively, to the ordered bases A of $\mathbb{V}$ and B of $\mathbb{W}$.

3. Thus, the results on matrices and its transpose can be re-written in the language a vector space and its dual space.

Check Your Progress

4. Define Dual space –

5. Explain canonical isomorphism.

8.5 LET'S SUM UP

Application of vector space in matrix. Similarity of matrices and dual space concept has been clarified.

8.6 KEYWORDS

1. Evaluated --To **evaluate** an algebraic expression, you have to substitute a number for each variable and perform the arithmetic operations.
2. Canonical - In **mathematics** and computer science, a **canonical**, normal, or standard form of a **mathematical** object is a standard way of presenting that object as a **mathematical** expression.

3. Change of **Coordinates Matrix**. A change of **coordinates matrix**, also called a transition **matrix**, specifies the transformation from one vector basis to another under a change of basis
4. The **domain** is the group of numbers that can be entered into a **function** to create a valid output.

8.7 QUESTION FOR REVIEW

1. 1. Let T ∈ L($\mathbb{R}^2$) represent the reflection about the line y = mx. Find [T].

2. Find the matrix of the linear transformations given below.

1. Let B = $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$ be an ordered basis of R3. Now, define T ∈ L($\mathbb{R}^3$) by T ($\mathbf{x}_1$) = $\mathbf{x}_2$,

T ($\mathbf{x}_2$) = $\mathbf{x}_3$ and T ($\mathbf{x}_3$) = **x**1. Determine T [B, B]. Is T invertible?

3. Define T ∈ L($\mathbb{R}^3$) by T ((x, y, z)T) = (x + y + 2z, x − y − 3z, 2x + 3y + z)T . Let B be the standard ordered basis and C = (1, 1, 1), (1, −1, 1), (1, 1, 2) be another ordered basis of $\mathbb{R}^3$.

Then find

(a) matrices T [B, B] and T [ℂ, ℂ].

(b) the matrix P such that P −1T [B, B] P = T [ℂ, ℂ].

4. Define T : $\mathbb{C}^3 \to \mathbb{C}$ by T ((x, y, z)T) = x. Is it a linear functional?

5. *Let* 𝕍 *be a vector space. Suppose there exists* **v** *∈* 𝕍 *such that* **f**(**v**) = 0, *for all* **f** *∈* 𝕍*. *Then prove that* **v** = **0**.

8.8 SUGGESTED READINGS

1. K. Hauffman and R. Kunz, Linear Algebra, Pearson Education (INDIA), 2003.
2. G. Strang, Linear Algebra And Its Applications, 4th Edition, Brooks/Cole, 2006.
3. S. Lang, Linear Algebra, Springer, 1989.
4. David S. Dummit and Richard M. Foote, Abstract Algebra (3e), John Wiley and Sons.

5. R. Gallian Joseph, Contemporary Abstract Algebra, Narosa Publishing House.

6. Thomas Hungerford, Algebra, Springer GTM.

7. I.N. Herstein, Topics in Abstract Algebra, Wiley Eastern Limited.

8. D.S. Malik, J.M. Mordesen, M.K. Sen, Fundamentals of Abstract Algebra, The McGraw-Hill Companies, Inc.

8.9 ANSWER TO CHECK YOUR PROGRESS

1. [Provide definition and example-- 8.1.1]

2. [Provide statement and proof – 8.2.1]

3. Provide definition (a) – 8.2.4 & (b) – 8.2.5

4. Provide definition –8.3.2

5. Provide the statement of theorem and proof ---8.3.4

UNIT 9: INNER PRODUCT SPACES

STRUCTURE

9.0 OBJECTIVE

Understand the concept and meaning of Inner product space

Enumerate Cauchy Schwartz Inequality

Understand The Concept Of Angle Between Two Vectors And Normed Linear Space

Understand The Background Of GRAM-SCHMIDT ORTHONORMALIZATION PROCESS

9.1 INTRODUCTION

Recall the dot product in $\mathbb{R}^2$ and $\mathbb{R}^3$. Dot product helped us to compute the length of vectors and angle between vectors. This enabled us to rephrase geometrical problems in $\mathbb{R}^2$ and $\mathbb{R}^3$ in the language of vectors. We generalize the idea of dot product to achieve similar goal for a general vector space over $\mathbb{R}$ or $\mathbb{C}$. So, in this chapter $\mathbb{F}$ will denote either $\mathbb{R}$ or $\mathbb{C}$.

9.2 INNER PRODUCT SPACE - DEFINITION AND BASIC PROPERTIES

Definition 9.2.1. [Inner Product] Let $\mathbb{V}$ be a vector space over $\mathbb{F}$. An **inner product** over$\mathbb{V}$, denoted by *h* , *i*, is a map from $\mathbb{V} \times \mathbb{V}$ to F satisfying

1. $\langle a\mathbf{u} + b\mathbf{v}, \mathbf{w}\rangle = a\,\langle \boldsymbol{u},\boldsymbol{w}\rangle + b\,\langle \boldsymbol{v},\boldsymbol{w}\rangle$ for all **u**, **v**, **w** $\in \mathbb{V}$ and $a, b \in \mathbb{F}$,

2. $\langle \boldsymbol{u},\boldsymbol{v}\rangle = \langle \boldsymbol{v},\boldsymbol{u}\rangle$, the complex conjugate of *h***u**, **v***i*, for all **u**, **v** $\in \mathbb{V}$ and

3. $\langle \boldsymbol{u},\boldsymbol{u}\rangle \geq 0$ for all $\mathbf{u} \in \mathbb{V}$. Furthermore, equality holds if and only if $\mathbf{u} = \mathbf{0}$.

Remark 9.2.2. *Using the definition of inner product, we immediately observe that*

1. $\langle \boldsymbol{v},\alpha\boldsymbol{w}\rangle = \overline{\langle \alpha\boldsymbol{w},\boldsymbol{v}\rangle} = \bar{\alpha}\,\overline{\langle \boldsymbol{w},\boldsymbol{v}\rangle} = \bar{\alpha}\,\langle \boldsymbol{v},\boldsymbol{w}\rangle$, *for all* $\alpha \in \mathbb{F}$ *and* **v**, **w** $\in$ V.

2. If $\langle \boldsymbol{u},\boldsymbol{v}\rangle = 0$ *for all* $\mathbf{v} \in \mathbb{V}$ *then in particular* $\langle \boldsymbol{u},\boldsymbol{u}\rangle = 0$. *Hence,* $\mathbf{u} = \mathbf{0}$.

Definition 9.2.3. [Inner Product Space] Let $\mathbb{V}$ be a vector space with an inner product $\langle\ ,\ \rangle$

Then, $(\mathbb{V}, \langle\ ,\ \rangle)$ is called an **inner product space** (in short, IPS).

Example. Examples 1 and 2 that appear below are called the **standard inner product** or the **dot product** on $\mathbb{R}^n$ and $\mathbb{C}^n$, respectively. Whenever an inner product is not clearly mentioned, it will be assumed to be the standard inner product.

1. For $\mathbf{u} = (u1, \ldots, un)^T$, $\mathbf{v} = (v1, \ldots, vn)^T \in \mathbb{R}^n$ define *h***u**, **v***i* $= u1v1 + \cdots + unvn = \mathbf{v}^T\mathbf{u}$.

Then, $\langle\, ,\rangle$ is indeed an inner product and hence $(\mathbb{R}^n, \langle\, ,\rangle\,)$ is an IPS.

2. For $\mathbf{u} = (u1, \ldots, un)*$, $\mathbf{v} = (v1, \ldots, vn)* \in \mathbb{C}^n$ define $\langle \boldsymbol{u}, \boldsymbol{v}\rangle = u1v1 + \cdots + unvn = \mathbf{v}*\mathbf{u}$.

Then, $(\mathbb{C}^n, \langle\, ,\rangle\,)$ is an IPS.

9.3 CAUCHY SCHWARTZ INEQUALITY

As $\langle \boldsymbol{u}, \boldsymbol{u}\,\rangle\ > 0$, for all $\mathbf{u} \neq \mathbf{0}$, we use inner product to define length of a vector.

Definition 9.3.1. [Length / Norm of a Vector] Let $\mathbb{V}$ be a vector space over F. Then, for any vector $\mathbf{u} \in \mathbb{V}$, we define the **length (norm)** of **u**, denoted $||\mathbf{u}|| = \sqrt{\langle \boldsymbol{u}, \boldsymbol{u}\rangle}$the positive square root. A vector of norm 1 is called a **unit vector**. Thus, $\frac{\boldsymbol{u}}{||\mathbf{u}||}$ is called the **unit vector in the direction** of **u**.

Example: 1. Let $\mathbb{V}$ be an IPS and $\mathbf{u} \in \mathbb{V}$. Then, for any scalar α, $||\alpha\mathbf{u}|| = |\alpha|.||u||$

2. Let $\mathbf{u} = (1, -1, 2, -3)^T \in \mathbb{R}^4$. Then, $||\mathbf{u}|| = \sqrt{1 + 1 + 4 + 9} = \sqrt{15}$. Thus, $\frac{1}{\sqrt{15}}\mathbf{u}$ and $-\frac{1}{\sqrt{15}}\mathbf{u}$

are vectors of norm 1. Moreover $\frac{1}{\sqrt{15}}\mathbf{u}$ is a unit vector in the direction of **u**.

3. $||\mathbf{x} + \mathbf{y}||^2 + ||\mathbf{x} - \mathbf{y}||^2 = 2(||x||^2 + ||y||^2)$, for all x^T , $y^T \in \mathbb{R}^n$. This equality is called the Parallelogram Law as in a parallelogram the sum of square of the lengths of the diagonals is equal to twice the sum of squares of the lengths of the sides

4. Apollonius' Identity: *Let the length of the sides of a triangle be a, b, c* $\in \mathbb{R}$ *and that of the median be d* $\in \mathbb{R}$*.*

$$b^2 + c^2 = 2\left(d^2 + \left(\frac{a}{2}\right)^2\right).$$

Notes

Theorem 9.3.2 (Cauchy-Bunyakovskii-Schwartz inequality)**.** *Let* $\mathbb{V}$ *be an inner product space over* $\mathbb{F}$*. Then, for any* **u**, **v** $\in \mathbb{V}$

$$|\langle \boldsymbol{u}, \boldsymbol{v}\rangle| \leq ||\boldsymbol{u}||\ ||\boldsymbol{v}||. \qquad \text{(A)}$$

Moreover, equality holds in Inequality (A) if and only if **u** and **v** are linearly dependent.

Furthermore, if $\mathbf{u} \neq \mathbf{0}$ then

$$\mathbf{v} = \left\langle \mathbf{v}, \frac{\mathbf{u}}{\|\mathbf{u}\|} \right\rangle \frac{\mathbf{u}}{\|\mathbf{u}\|}.$$

Proof. If $\mathbf{u} = \mathbf{0}$ then Inequality (A) holds. Hence, let $\mathbf{u} \leq \mathbf{0}$. Then, by Definition 9.1.1.3, $\langle \lambda\boldsymbol{u} + \boldsymbol{v}, \lambda\boldsymbol{u} + \boldsymbol{v}\rangle \geq 0$ for all $\lambda \in \mathbb{F}$ and $\mathbf{v} \in \mathbb{V}$. In particular, for

$$\lambda = -\frac{\langle \mathbf{v}, \mathbf{u}\rangle}{\|\mathbf{u}\|^2},$$

$$\begin{aligned} 0 &\leq \langle \lambda\mathbf{u} + \mathbf{v}, \lambda\mathbf{u} + \mathbf{v}\rangle = \lambda\bar{\lambda}\|\mathbf{u}\|^2 + \lambda\langle \mathbf{u}, \mathbf{v}\rangle + \bar{\lambda}\langle \mathbf{v}, \mathbf{u}\rangle + \|\mathbf{v}\|^2 \\ &= \frac{\langle \mathbf{v}, \mathbf{u}\rangle}{\|\mathbf{u}\|^2}\frac{\overline{\langle \mathbf{v}, \mathbf{u}\rangle}}{\|\mathbf{u}\|^2}\|\mathbf{u}\|^2 - \frac{\langle \mathbf{v}, \mathbf{u}\rangle}{\|\mathbf{u}\|^2}\langle \mathbf{u}, \mathbf{v}\rangle - \frac{\overline{\langle \mathbf{v}, \mathbf{u}\rangle}}{\|\mathbf{u}\|^2}\langle \mathbf{v}, \mathbf{u}\rangle + \|\mathbf{v}\|^2 = \|\mathbf{v}\|^2 - \frac{|\langle \mathbf{v}, \mathbf{u}\rangle|^2}{\|\mathbf{u}\|^2}. \end{aligned}$$

Or, in other words $|\langle \boldsymbol{v}, \boldsymbol{u}\rangle|^2 \leq ||\mathbf{u}||^2||\mathbf{v}||^2$ and the proof of the inequality is over. Now, note that equality holds in Inequality (A) if and only if $||\lambda\mathbf{u} + \mathbf{v}, \lambda\mathbf{u} + \mathbf{v}|| = 0$, or equivalently, $\lambda\mathbf{u} + \mathbf{v} = \mathbf{0}$. Hence, **u** and **v** are linearly dependent. Moreover,

$$0 = \langle \mathbf{0}, \mathbf{u}\rangle = \langle \lambda\mathbf{u} + \mathbf{v}, \mathbf{u}\rangle = \lambda\langle \mathbf{u}, \mathbf{u}\rangle + \langle \mathbf{v}, \mathbf{u}\rangle$$

implies that $\mathbf{v} = -\lambda\mathbf{u} = -\frac{\langle \mathbf{v}, \mathbf{u}\rangle}{\|\mathbf{u}\|^2}\mathbf{u} = \left\langle \mathbf{v}, \frac{\mathbf{u}}{\|\mathbf{u}\|}\right\rangle \frac{\mathbf{u}}{\|\mathbf{u}\|}.$

Corollary 9.3.3. *Let* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$*. Then,*

$$\left(\sum_{i=1}^{n} \mathbf{x}_i\mathbf{y}_i\right)^2 \leq \left(\sum_{i=1}^{n} \mathbf{x}_i^2\right)\left(\sum_{i=1}^{n} \mathbf{y}_i^2\right).$$

Check your progress

1. Define Inner Product and Inner Product Space

2. Explain length or norm of vector

9.4 ANGLE BETWEEN TWO VECTORS

Let $\mathbb{V}$ be a real vector space. Then, for **u**, **v** $\in \mathbb{V}$, the Cauchy-Schwartz inequality implies that

$-1 \leq \frac{\langle u,v \rangle}{||u|| ||v||} \leq 1$. We use this together with the properties of the cosine function to define the angle between two vectors in an inner product space.

Definition 9.4.1. [Angle between Vectors] Let $\mathbb{V}$ be a real vector space. If $\theta \in [0, \pi]$ is the angle between **u**, **v** $\in \mathbb{V} \setminus \{\mathbf{0}\}$ then we define

$$\cos\theta = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{u}\| \, \|\mathbf{v}\|}.$$

Example :1. Take $(1, 0)^T$, $(1, 1)^T \in \mathbb{R}^2$. Then, $\cos\theta = \frac{1}{\sqrt{2}}$. So $\theta = \pi/4$.

2. Take $(1, 1, 0)^T$, $(1, 1, 1)^T \in \mathbb{R}^3$. Then, angle between them, say $\beta = \cos^{-1} 2/\sqrt{6}$

3. Angle depends on the IP. Take $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = 2\mathbf{x}1\mathbf{y}1 + \mathbf{x}1\mathbf{y}2 + \mathbf{x}2\mathbf{y}1 + \mathbf{x}2\mathbf{y}2$ on $\mathbb{R}^2$. Then, angle between $(1, 0)^T$, $(1, 1)^T \in \mathbb{R}^2$ equals $\cos^{-1}\frac{3}{\sqrt{10}}$.

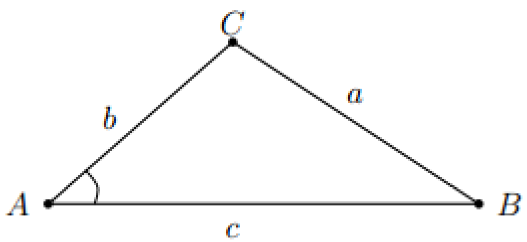


Figure A: Triangle with vertices *A*, *B* and *C*

We will now prove that if *A*, *B* and *C* are the vertices of a triangle (see Figure A) and *a*, *b* and *c*, respectively, are the lengths of the corresponding sides then $\cos(A) = \frac{b^2+c^2-a^2}{2bc}$. This in turn implies that the angle between vectors has been rightly defined.

Lemma 9.4.2 *Let A, B and C be the vertices of a triangle (see Figure 5.1) with corresponding side lengths a, b and c, respectively, in a real inner product space* $\mathbb{V}$ *then*

$$\cos(A) = \frac{b^2 + c^2 - a^2}{2bc}$$

Proof. Let **0**, **u** and **v** be the coordinates of the vertices *A*, *B* and *C*, respectively, of the triangle *ABC*. Then, $\overrightarrow{AB} = \mathbf{u}$, $\overrightarrow{AC} = \mathbf{v}$ and $\overrightarrow{BC} = \mathbf{v} - \mathbf{u}$. Thus, we need to prove that

$$\cos(A) = \frac{\|\mathbf{v}\|^2 + \|\mathbf{u}\|^2 - \|\mathbf{v} - \mathbf{u}\|^2}{2\|\mathbf{v}\|\|\mathbf{u}\|} \Leftrightarrow \|\mathbf{v}\|^2 + \|\mathbf{u}\|^2 - \|\mathbf{v} - \mathbf{u}\|^2 = 2\,\|\mathbf{v}\|\,\|\mathbf{u}\|\,\cos(A).$$

Now, by definition $||\mathbf{v}-\mathbf{u}||^2 = ||\mathbf{v}||^2+||\mathbf{u}||^2-2\,\langle \boldsymbol{u}, \boldsymbol{v}\rangle$and hence

$||\mathbf{v}||^2+||\mathbf{u}||^2-||\mathbf{v}-\mathbf{u}||^2 = 2\,\langle \boldsymbol{u}, \boldsymbol{v}\rangle$

As $\langle \boldsymbol{v}, \boldsymbol{u}\rangle = ||\mathbf{v}||\;||\mathbf{u}||\cos(A)$, the required result follows.

Definition 9.4.3. [Orthogonality / Perpendicularity] Let $\mathbb{V}$ be an inner product space over $\mathbb{R}$. Then,

1. the vectors **u**, **v** $\in \mathbb{V}$ are called **orthogonal/perpendicular** if $\langle \boldsymbol{u}, \boldsymbol{v}\rangle = 0$.

2. Let $S \subseteq \mathbb{V}$. Then, the **orthogonal complement** of *S* in $\mathbb{V}$, denoted $S^\perp$, equals

$$S^\perp = \{\mathbf{v} \in \;:\; \langle \boldsymbol{v}, \boldsymbol{w}\rangle = = 0, \text{ for all } \mathbf{w} \in S\}.$$

Example 1. **0** is orthogonal to every vector as $\langle \mathbf{0}, \boldsymbol{x}\rangle = = 0$ for all $\mathbf{x} \in \mathbb{V}$.

2. If $\mathbb{V}$ is a vector space over $\mathbb{R}$ or $\mathbb{C}$ then **0** is the only vector that is orthogonal to itself.

3.Let $\mathbf{u} = (1, 2)^T$. What is $\mathbf{u}^\perp$ in $\mathbb{R}^2$?

Solution: $\{(x, y)^T \in \mathbb{R}^2 \mid x + 2y = 0\}$.

Is this Null(**u**)?

Note that $(2, -1)^T$ is a basis of $\mathbf{u}^\perp$ and for any vector $\mathbf{x} \in \mathbb{R}^2$,

$$\mathbf{x} = \langle \mathbf{x}, \mathbf{u} \rangle \frac{\mathbf{u}}{\|\mathbf{u}\|^2} + \left(\mathbf{x} - \langle \mathbf{x}, \mathbf{u} \rangle \frac{\mathbf{u}}{\|\mathbf{u}\|^2} \right) = \frac{x_1 + 2x_2}{5}(1,2)^T + \frac{2x_1 - x_2}{5}(2,-1)^T$$

Is a decomposition of **x** into two vectors, one parallel to **u** and the other parallel to $\mathbf{u}^\perp$?

4. Fix $\mathbf{u} = (1, 1, 1, 1)^T$, $\mathbf{v} = (1, 1, -1, 0)^T \in \mathbb{R}^4$. Determine $\mathbf{z}, \mathbf{w} \in \mathbb{R}^4$ such that $\mathbf{u} = \mathbf{z} + \mathbf{w}$ with the condition that **z** is parallel to **v** and **w** is orthogonal to **v**.

Solution: As **z** is parallel to **v**, $\mathbf{z} = k\mathbf{v} = (k, k, -k, 0)^T$, for some $k \in \mathbb{R}$. Since **w** is orthogonal to **v** the vector $\mathbf{w} = (a, b, c, d)^T$ satisfies $a + b - c = 0$. Thus, $c = a + b$ and

$$(1, 1, 1, 1)^T = \mathbf{u} = \mathbf{z} + \mathbf{w} = (k, k, -k, 0)^T + (a, b, a + b, d)^T .$$

Comparing the corresponding coordinates, gives the linear system $d = 1$, $a + k = 1$, $b + k = 1$ and $a + b - k = 1$ in the variables a, b, d and k. Thus, solving for a, b, d and k gives $\mathbf{z} = 1/3\ (1, 1, -1, 0)^T$ and $\mathbf{w} = 1/3\ (2, 2, 4, 3)^T$.

9.5 NORMED LINEAR SPACE

To proceed further, recall that a vector space over $\mathbb{R}$ or $\mathbb{C}$ was a linear space.

Definition 9.5.1 . [Normed Linear Space] Let $\mathbb{V}$ be a linear space.

1. Then, a **norm** on $\mathbb{V}$ is a function $f(\mathbf{x}) = ||\mathbf{x}||$from $\mathbb{V}$ to R such that

(a) $||\mathbf{x}|| \geq 0$ for all $\mathbf{x} \in \mathbb{V}$ and if $||\mathbf{x}|| = 0$ then $\mathbf{x} = \mathbf{0}$.

(b) $||\alpha\mathbf{x}|| = |\ \alpha\ |\ ||\mathbf{x}||$ for all $\alpha \in \mathbb{F}$ and $\mathbf{x} \in \mathbb{V}$.

(c) $||\mathbf{x} + \mathbf{y}|| \leq ||\mathbf{x}|| + ||\mathbf{y}||$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{V}$ (triangle inequality).

2. A linear space with a norm on it is called a **normed linear space** (NLS).

Theorem 9.5.2. *Let* $\mathbb{V}$ *be a normed linear space and* $\mathbf{x}, \mathbf{y} \in \mathbb{V}$ *Then,*

$$||\mathbf{x}|| - ||\mathbf{y}|| \leq ||\mathbf{x} - \mathbf{y}||$$

Proof.

As $||\mathbf{x}|| = ||\mathbf{x} - \mathbf{y} + \mathbf{y}|| \leq ||\mathbf{x} - \mathbf{y}|| + ||\mathbf{y}||$ one has $||\mathbf{x}|| - ||\mathbf{y}|| \leq ||\mathbf{x} - \mathbf{y}||$.

Similarly, one obtains

$$||\mathbf{y}|| - ||\mathbf{x}|| \leq ||\mathbf{y} - \mathbf{x}|| = ||\mathbf{x} - \mathbf{y}||$$

Combining the two, the required result follows.

Example: Let $\mathbb{V}$ be an IPS. Is it true that $f(\mathbf{x}) = \sqrt{\langle x, x\rangle}$ is a norm?

Solution: Yes. The readers should verify the first two conditions. For the third condition, recalling the Cauchy-Schwartz inequality, we get

$$\begin{aligned} f(\mathbf{x}+\mathbf{y})^2 &= \langle \mathbf{x}+\mathbf{y}, \mathbf{x}+\mathbf{y}\rangle = \langle \mathbf{x}, \mathbf{x}\rangle + \langle \mathbf{x}, \mathbf{y}\rangle + \langle \mathbf{y}, \mathbf{x}\rangle + \langle \mathbf{y}, \mathbf{y}\rangle \\ &\leq \|\mathbf{x}\|^2 + \|\mathbf{x}\|\|\mathbf{y}\| + \|\mathbf{x}\|\|\mathbf{y}\| + \|\mathbf{y}\|^2 = (f(\mathbf{x}) + f(\mathbf{y}))^2. \end{aligned}$$

Thus, $||\mathbf{x}|| = \sqrt{\langle x, x\rangle}$ is a norm, called the norm **induced** by the inner product *h·, ·i*.

Theorem 9.5.3 . *Let* $||\cdot||$ *be a norm on a* NLS $\mathbb{V}$. *Then, k · k is induced by some inner product if and only if k · k satisfies the* parallelogram law*:* $||\mathbf{x} + \mathbf{y}||^2 + ||\mathbf{x} - \mathbf{y}||^2 = 2||\mathbf{x}||^2 + 2||\mathbf{y}||^2$.

Example

1. For $\mathbf{x} = (x1, x2)^T \in \mathbb{R}^2$, we define $k\mathbf{x}k1 = |\mathbf{x}1| + |\mathbf{x}2|$. Verify that $||\mathbf{x}||_1$ is indeed a norm. But, for $\mathbf{x} = \mathbf{e}_1$ and $\mathbf{y} = \mathbf{e}_2$, $2(||\mathbf{x}||^2 + ||\mathbf{y}||^2) = 4$ whereas

$||\mathbf{x} + \mathbf{y}||^2 + ||\mathbf{x} - \mathbf{y}||^2 = ||(1, 1)||^2 + ||(1, -1)||^2 = (|1| + |1|)^2 + (|1| + |-1|)^2 = 8.$

So, the parallelogram law fails. Thus, $||\mathbf{x}||_1$ is not induced by any inner product in $\mathbb{R}^2$.

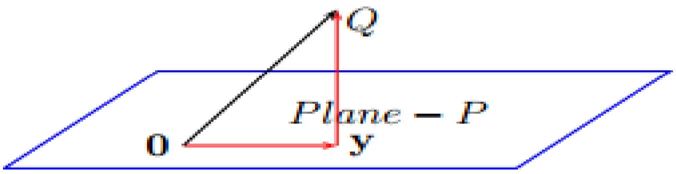
9.6 GRAM-SCHMIDT ORTHONORMALIZATION PROCESS

Definition 9..61. Let $\mathbb{V}$ be an IPS. Then, a non-empty set $S = \{\mathbf{v}1, \ldots, \mathbf{v}n\} \subseteq \mathbb{V}$ is called an **orthogonal set** if $\mathbf{v}i$ and $\mathbf{v}j$ are **mutually orthogonal**, for $1 \leq i \neq j \leq n$, *i.e.*,

$$\langle \boldsymbol{u_i}\, \boldsymbol{u_j} \rangle = 0, \text{ for } 1 \leq i < j \leq n.$$

Further, if $||\mathbf{v}_i|| = 1$, for $1 \leq i \leq n$, Then S is called an **orthonormal set**. If S is also a basis of V then S is called an **orthonormal basis** of V.

Example. Which point on the plane P is closest to the point, say Q?



Solution: Let **y** be the foot of the perpendicular from Q on P . Thus, by Pythagoras Theorem, this point is unique. So, the question arises: how do we find **y**?

Note that $\overrightarrow{yQ}$ gives a normal vector of the plane P . Hence, $\vec{Q} = \mathbf{y} + \overrightarrow{yQ}$.

So, need to decompose , $\vec{Q}$ into two vectors such that one of them lies on the plane P and the other is orthogonal to the plane.

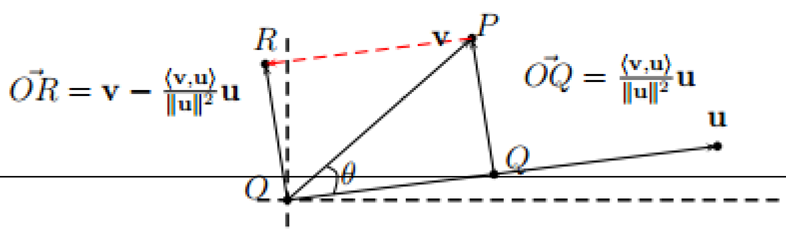


Figure B: Decomposition of vector **v**

Thus, we see that given **u**, **v** $\in$ V \ {**0**}, we need to find two vectors, say **y** and **z**, such that **y** is parallel to **u** and **z** is perpendicular to **u**. Thus, **y** = **u** cos(θ) and **z** = **u** sin(θ), where θ is the angle between **u** and **v**.

We do this as follows (see Figure B). Let $\hat{\boldsymbol{u}} = \frac{\boldsymbol{u}}{||\boldsymbol{u}||}$ be the unit vector in the direction of **u**. Then, using trigonometry, $os(\theta) = \frac{||\vec{OQ}||}{||\vec{OP}||}$. Hence|| $\vec{\boldsymbol{O}Q}$|| = || $\vec{\boldsymbol{O}P}$|| cos(θ). Now using Definition of angle between vectors

$$\|\vec{OQ}\| = \|\mathbf{v}\| \left|\frac{\langle \mathbf{v},\mathbf{u}\rangle}{\|\mathbf{v}\|\,\|\mathbf{u}\|}\right| = \left|\frac{\langle \mathbf{v},\mathbf{u}\rangle}{\|\mathbf{u}\|}\right|,$$

where the absolute value is taken as the length/norm is a positive quantity. Thus,

$$\vec{OQ} = \|\vec{OQ}\|\ \hat{\mathbf{u}} = \left\langle \mathbf{v}, \frac{\mathbf{u}}{\|\mathbf{u}\|}\right\rangle \frac{\mathbf{u}}{\|\mathbf{u}\|}.$$

Hence

$$\mathbf{y} = \vec{OQ} = \left\langle \mathbf{v}, \frac{\mathbf{u}}{\|\mathbf{u}\|}\right\rangle \frac{\mathbf{u}}{\|\mathbf{u}\|} \text{ and } \mathbf{z} = \mathbf{v} - \left\langle \mathbf{v}, \frac{\mathbf{u}}{\|\mathbf{u}\|}\right\rangle \frac{\mathbf{u}}{\|\mathbf{u}\|}.$$

In literature, the vector **y** = $\vec{\boldsymbol{O}Q}$ is called the **orthogonal projection** of **v** on **u**, denoted $\text{Proj}_{\mathbf{u}}(\mathbf{v})$. Thus

$$\text{Proj}_{\mathbf{u}}(\mathbf{v}) = \left\langle \mathbf{v}, \frac{\mathbf{u}}{\|\mathbf{u}\|}\right\rangle \frac{\mathbf{u}}{\|\mathbf{u}\|} \text{ and } \|\text{Proj}_{\mathbf{u}}(\mathbf{v})\| = \|\vec{OQ}\| = \left|\frac{\langle \mathbf{v},\mathbf{u}\rangle}{\|\mathbf{u}\|}\right|. \tag{C}$$

Moreover, the distance of **u** from the point P equals

$$\|\vec{OR}\| = \|\vec{PQ}\| = \left\|\mathbf{v} - \left\langle \mathbf{v}, \frac{\mathbf{u}}{\|\mathbf{u}\|}\right\rangle \frac{\mathbf{u}}{\|\mathbf{u}\|}\right\|.$$

Example: Determine the projection of $\mathbf{v} = (1, 1, 1, 1)^T$ on $\mathbf{u} = (1, 1, -1, 0)^T$.

Solution: By Equation (C), we have

$$\mathrm{Proj}_{\mathbf{v}}(\mathbf{u}) = \langle \mathbf{v}, \mathbf{u} \rangle \frac{\mathbf{u}}{\|\mathbf{u}\|^2} = \frac{1}{3}(1, 1, -1, 0)^T$$

And $\mathbf{w} = (1, 1, 1, 1)T - \mathrm{Proj}\mathbf{v}(\mathbf{u}) = 1\ 3(2, 2, 4, 3)T$ is orthogonal to $\mathbf{u}$.

Example: Let $\mathbf{u} = (1, 1, 1, 1)T$, $\mathbf{v} = (1, 1, -1, 0)T$, $\mathbf{w} = (1, 1, 0, -1)T \in \mathbb{R}^4$. Write $\mathbf{v} = \mathbf{v}1 + \mathbf{v}2$, where $\mathbf{v}1$ is parallel to $\mathbf{u}$ and $\mathbf{v}2$ is orthogonal to $\mathbf{u}$. Also, write $\mathbf{w} = \mathbf{w}1 + \mathbf{w}2 + \mathbf{w}3$ such that

$\mathbf{w}1$ is parallel to $\mathbf{u}$, $\mathbf{w}2$ is parallel to $\mathbf{v}2$ and $\mathbf{w}3$ is orthogonal to both $\mathbf{u}$ and $\mathbf{v}2$.

Solution: Note that

(a) $\mathbf{v}_1 = \mathrm{Proj}_{\mathbf{u}}(\mathbf{v}) = \langle \mathbf{v}, \mathbf{u} \rangle \frac{\mathbf{u}}{\|\mathbf{u}\|^2} = \frac{1}{4}\mathbf{u} = \frac{1}{4}(1, 1, 1, 1)^T$ is parallel to $\mathbf{u}$.

(b) $\mathbf{v}_2 = \mathbf{v} - \frac{1}{4}\mathbf{u} = \frac{1}{4}(3, 3, -5, -1)^T$ is orthogonal to $\mathbf{u}$.

Note that $\mathrm{Proj}_{\mathbf{u}}(\mathbf{w})$ is parallel to $\mathbf{u}$ and $\mathrm{Proj}_{\mathbf{v}_2}(\mathbf{w})$ is parallel to $\mathbf{v}_2$. Hence, we have

(a) $\mathbf{w}_1 = \mathrm{Proj}_{\mathbf{u}}(\mathbf{w}) = \langle \mathbf{w}, \mathbf{u} \rangle \frac{\mathbf{u}}{\|\mathbf{u}\|^2} = \frac{1}{4}\mathbf{u} = \frac{1}{4}(1, 1, 1, 1)^T$ is parallel to $\mathbf{u}$,

(b) $\mathbf{w}_2 = \mathrm{Proj}_{\mathbf{v}_2}(\mathbf{w}) = \langle \mathbf{w}, \mathbf{v}_2 \rangle \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|^2} = \frac{7}{44}(3, 3, -5, -1)^T$ is parallel to $\mathbf{v}_2$ and

(c) $\mathbf{w}_3 = \mathbf{w} - \mathbf{w}_1 - \mathbf{w}_2 = \frac{3}{11}(1, 1, 2, -4)^T$ is orthogonal to both $\mathbf{u}$ and $\mathbf{v}_2$.

Theorem 9.6.2 *Let $S = \{\mathbf{u}1, \ldots, \mathbf{u}n\}$ be an orthonormal subset of an* IPS $\mathbb{V}(\mathbb{F})$

1. Then, S is a linearly independent subset of $\mathbb{V}$.
2. Suppose $\mathbf{v} \in LS(S)$ with $\mathbf{v} = \sum_{i=1}^{n} \alpha_i \mathbf{u}_i$, for some αi's in $\mathbb{F}$. Then,

(a) $\alpha_i = \langle \mathbf{v}, \mathbf{u}_i \rangle$.

(b) $\|\mathbf{v}\|^2 = \|\sum_{i=1}^{n} \alpha_i \mathbf{u}_i\|^2 = \sum_{i=1}^{n} |\alpha_i|^2$.

3. Let **z** ∈ $\mathbb{V}$ and **w** = $\sum_{i=1}^{n}\langle z, \mathbf{u_i}\rangle\mathbf{u_i}$. Then, **z** = **w** + (**z** − **w**) with ⟨**z** − **w**, **w**⟩= 0, i.e., **z** − **w** ∈ LS(S)⊥. Further, $\|\mathbf{z}\|^2 = \|\mathbf{w}\|^2 + \|\mathbf{z} - \mathbf{w}\|^2 \geq \|\mathbf{w}\|^2$.

4. Let dim($\mathbb{V}$) = n. Then, $\langle \mathbf{v}, \mathbf{u_i}\rangle = 0$ for all i = 1, 2, . . . , n if and only if **v** = **0**.

Proof. Part 1: Consider the linear system c1**u**1 + · · · + cn**u**n = **0** in the

$$0 = \langle \mathbf{0}, \mathbf{u}_i\rangle = \langle c_1\mathbf{u}_1 + \cdots + c_n\mathbf{u}_n, \mathbf{u}_i\rangle = \sum_{j=1}^{n} c_j\langle \mathbf{u}_j, \mathbf{u}_i\rangle = c_i\langle \mathbf{u}_i, \mathbf{u}_i\rangle = c_i.$$

variables *c*1, . . . , *cn*. As

$\langle \mathbf{0}, \boldsymbol{u}\rangle = 0$ and $\langle \boldsymbol{u_j}, \boldsymbol{u_i}\rangle = 0$, for all $j \neq i$, we have

Hence, $c_i = 0$, for $1 \leq i \leq n$. Thus, the above linear system has only the trivial solution. So,

the set *S* is linearly independent.

Part 2: Note that

$$\langle \mathbf{v}, \mathbf{u}_i\rangle = \langle \sum_{j=1}^{n} \alpha_j\mathbf{u}_j, \mathbf{u}_i\rangle = \sum_{j=1}^{n} \alpha_j\langle \mathbf{u}_j, \mathbf{u}_i\rangle = \alpha_i\langle \mathbf{u}_i, \mathbf{u}_i\rangle = \alpha_i.$$

This completes the first sub-part. For the second sub-part, we have

$$\begin{aligned}\|\sum_{i=1}^{n}\alpha_i\mathbf{u}_i\|^2 &= \left\langle \sum_{i=1}^{n}\alpha_i\mathbf{u}_i, \sum_{i=1}^{n}\alpha_i\mathbf{u}_i\right\rangle = \sum_{i=1}^{n}\alpha_i\left\langle \mathbf{u}_i, \sum_{j=1}^{n}\alpha_j\mathbf{u}_j\right\rangle \\ &= \sum_{i=1}^{n}\alpha_i\sum_{j=1}^{n}\overline{\alpha_j}\langle \mathbf{u}_i, \mathbf{u}_j\rangle = \sum_{i=1}^{n}\alpha_i\overline{\alpha_i}\langle \mathbf{u}_i, \mathbf{u}_i\rangle = \sum_{i=1}^{n}|\alpha_i|^2.\end{aligned}$$

Part 3: Note that for $1 \leq i \leq n$,

$$\begin{aligned}\langle \mathbf{z} - \mathbf{w}, \mathbf{u}_i\rangle &= \langle \mathbf{z}, \mathbf{u}_i\rangle - \langle \mathbf{w}, \mathbf{u}_i\rangle = \langle \mathbf{z}, \mathbf{u}_i\rangle - \left\langle \sum_{j=1}^{n}\langle \mathbf{z}, \mathbf{u}_j\rangle\mathbf{u}_j, \mathbf{u}_i\right\rangle \\ &= \langle \mathbf{z}, \mathbf{u}_i\rangle - \sum_{j=1}^{n}\langle \mathbf{z}, \mathbf{u}_j\rangle\langle \mathbf{u}_j, \mathbf{u}_i\rangle = \langle \mathbf{z}, \mathbf{u}_i\rangle - \langle \mathbf{z}, \mathbf{u}_i\rangle = 0.\end{aligned}$$

So, **z** − **w** ∈LS(*S*)⊥. Hence, $\langle z - w, w\rangle = 0$ as **w** ∈LS(*S*). Further, $||\mathbf{z}||^2 = ||\mathbf{w} + (\mathbf{z} - \mathbf{w})||^2 = ||\mathbf{w}||^2 + ||\mathbf{z} - \mathbf{w}||^2 \geq ||\mathbf{w}||^2$.

Part 4: Follows directly using Part 2b as {**u**1,..., **u***n*} is a basis of 𝕍.

Theorem 9.6.3 . *Let* 𝕍 *be a finite dimensional* ips *with an orthonormal basis* {**v**1, · · · , **v***n*}.
Then, for each **x**, **y** ∈𝕍,

$$\langle \mathbf{x}, \mathbf{y}\rangle = \sum_{i=1}^{n} \langle \mathbf{x}, \mathbf{v}_i\rangle \overline{\langle \mathbf{y}, \mathbf{v}_i\rangle}.$$

Furthermore, if **x** = **y** *then* k**x**k2 = $\sum_{i=1}^{n} |\langle x, v_i\rangle|^2$ *(generalizing the* **Pythagoras Theorem***).*
As a corollary to Theorem 9.5.2, we have the following result.

Theorem 9.5.4 (Bessel's Inequality)**.** *Let* V *be an* ips *with* {**v**1, · · · , **v***n*} *as an orthogonal set.*
Then,

$$\sum_{k=1}^{n} \frac{|\langle \mathbf{z}, \mathbf{v}_k\rangle|^2}{\|\mathbf{v}_k\|^2} \leq \|\mathbf{z}\|^2,$$ for each **z** ∈ V.

Equality holds if and only if $\mathbf{z} = \sum_{k=1}^{n} \frac{\langle \mathbf{z}, \mathbf{v}_k\rangle}{\|\mathbf{v}_k\|^2} \mathbf{v}_k.$

Proof. For 1 ≤ k ≤ n, define $u_k = \frac{v_k}{||v_k||}$ and use Theorem 9.5.2.4 to get the required result.

Check your proress

3. Explain normed linear space.

4. Explain Orthonormal projection

9.7 LET'S SUM UP

In this chapter we got clarity about the inner product space, Cauchy-Schwartz inequality, orthogonality, orthonormal basis and orthogonal projection

9.8 KEYWORDS

1. Vertices - The common endpoint of two or more rays or line segments
2. Coordinates - A set of values that show an exact position. On graphs it is usually a pair of numbers: the first number shows the distance along, and the second number shows the distance up or down
3. Linearly Independent. A set of vectors is maximally **linearly independent** if including any other vector in the vector space would make it **linearly dependent**
4. Parallel lines are two lines that are always the same distance apart and never touch.

9.9 QUESTION FOR REVIEW

1. *Let* $\mathbf{u} = (-1, 1, 2, 3, 7)^T \in \mathbb{C}^5$. *Find all* $\alpha \in \mathbb{C}$ *such that* $||\alpha\mathbf{u}|| = 1$.

2. Consider $\mathbb{R}^3$ *with the standard inner product. Find --* $S^{\perp}$ *for* $S = \{(1, 1, 1)^T, (0, 1, -1)^T\}$ *and* $S = LS((1, 1, 1)^T, (0, 1, -1)^T)$.

3. Let $A \in M_n(\mathbb{C})$ *satisfy* $||A\mathbf{x}|| \leq ||\mathbf{x}||$ *for all* $\mathbf{x} \in \mathbb{C}_n$. *Then, prove that if* $\alpha \in \mathbb{C}$ *with* $|\alpha| > 1$ *then* $A - \alpha I$ *is invertible.*

4. Prove Bessel's Inequality

9.10 SUGGESTED READINGS

1. K. Hauffman and R. Kunz, Linear Algebra, Pearson Education (INDIA), 2003.

2. G. Strang, Linear Algebra And Its Applications, 4th Edition, Brooks/Cole, 2006.

3. S. Lang, Linear Algebra, Springer, 1989.

4. David S. Dummit and Richard M. Foote, Abstract Algebra (3e), John Wiley and Sons.

5. R. Gallian Joseph, Contemporary Abstract Algebra, Narosa Publishing House.

6. Thomas Hungerford, Algebra, Springer GTM.

7. I.N. Herstein, Topics in Abstract Algebra, Wiley Eastern Limited.

8. D.S. Malik, J.M. Mordesen, M.K. Sen, Fundamentals of Abstract Algebra, the McGraw-Hill Companies, Inc.

9.11 ANSWER TO CHECK YOUR PROGRESS

1. [Provide definition and example-- 9.1.1 & 9.1.3]

2. Provide related definitions – 9.2.1

3. Provide definition and explanation – 9.4.1

4. Explain example of 9.5.1

UNIT 10: QUADRATIC FORMS

STRUCTURE

10.0 OBJECTIVE

Understand the concept of Quadratic forms & reduction of Quadratic forms, Understand the canonical forms , Understand the Sylvester law of inertia.

10.1 INTRODUCTION

A lot of applications of mathematics involve dealing with quadratic forms: you meet them in statistics (analysis of variance) and mechanics (energy of rotating bodies), among other places. In this section we begin the study of quadratic forms.

10.2 QUADRATIC FORMS

For almost everything in the remainder of this chapter, we assume that *the characteristic of the field* $\mathbb{K}$ *is not equal to* 2.

This means that $2 \neq 0$ in $\mathbb{K}$, so that the element $1/2$ exists in $\mathbb{K}$. Of our list of "standard" fields, this only excludes $\mathbb{F}_2$, the integers mod 2. (For example, in $\mathbb{F}_5$, we have $1/2 = 3$.)

A quadratic form as a function which, when written out in coordinates, is a polynomial in which every term has total degree 2 in the variables. For example, is a quadratic form in three variables.

$$q(x,y,z) = x^2 + 4xy + 2xz - 3y^2 - 2yz - z^2$$

Definition 10.2.1 A *quadratic form* in n variables $x_1,\ldots,x_n$ over a field $\mathbb{K}$ is a polynomial

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j$$

in the variables in which every term has degree two (that is, is a multiple of $x_i x_j$ for some i, j).

In the above representation of a quadratic form, we see that if $i \neq j$, then the term in $x_i x_j$ comes twice, so that the coefficient of $x_i x_j$ is $a_{ij} + a_{ji}$. We are free to choose any two values for a_{ij} and a_{ji} as long as they have the right sum; but we will always make the choice so that the two values are equal. That is, to obtain a term $c\, x_i x_j$, we take $a_{ij} = a_{ji} = c/2$. (This is why we require that the characteristic of the field is not 2.)

Any quadratic form is thus represented by a *symmetric* matrix A with (i, j) entry a_{ij} (that is, a matrix satisfying $A = A$>). *This is the third job of matrices in linear algebra:* Symmetric matrices represent quadratic forms.

We think of a quadratic form as defined above as being a function from the vector space $\mathbb{K}^n$ to the field $\mathbb{K}$. It is clear from the definition that

$$q(x_1,\ldots,x_n) = v^\top A v, \text{ where } v = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

Now if we change the basis for V , we obtain a different representation for the same function q. The effect of a change of basis is a linear substitution $v = Pv'$ on the variables, where P is the transition matrix between the bases. Thus we have so we have the following:

$$v^T Av = (Pv')^\top (P v') = (v')^T (P^T AP) v',$$

Proposition 10.2.2 *A basis change with transition matrix P replaces the symmetric matrix A representing a quadratic form by the matrix* $P^T AP$.

Definition 10.2.3 Two symmetric matrices A, A' over a field $\mathbb{K}$ are *congruent* if $A' = P^T AP$ for some invertible matrix P.

Proposition 10.2.4 *Two symmetric matrices are congruent if and only if they represent the same quadratic form with respect to different bases.*

10.3 REDUCTION OF QUADRATIC FORMS

Even if we cannot find a canonical form for quadratic forms, we can simplify them very greatly.

Theorem 10.3.1 Let q be a quadratic form in n variables x1,...,xn, over a field $\mathbb{K}$ whose characteristic is not 2. Then by a suitable linear substitution to new variables y1,...,yn, we can obtain *for some* $c_1,\dots,c_n \in \mathbb{K}$.

$$q = c_1 y_1^2 + c_2 y_2^2 + \cdots + c_n y_n^2$$

Proof : Our proof is by induction on n. We call a quadratic form which is written as in the conclusion of the theorem *diagonal*. A form in one variable is certainly diagonal, so the induction starts. Now assume that the theorem is true for forms in $n-1$ variables. Take

$$q(x_1,\dots,x_n) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} x_i x_j,$$

where $a_{ij} = a_{ji}$ for $i \neq j$.

Case 1: Assume that $a_{ii} \neq 0$ for some i. By a permutation of the variables (which is certainly a linear substitution), we can assume that $a_{11} \neq 0$. Let

$$y_1 = x_1 + \sum_{i=2}^{n} (a_{1i}/a_{11})x_i.$$

Then we have

$$a_{11}y_1^2 = a_{11}x_1^2 + 2\sum_{i=2}^{n} a_{1i}x_1x_i + q'(x_2,\ldots,x_n),$$

where q' is a quadratic form in $x_2,\ldots,x_n$. That is, all the terms involving x_1 in q have been incorporated into $a_{11}y_1^2$. So we have

$$q(x_1,\ldots,x_n) = a_{11}y_1^2 + q''(x_2,\ldots,x_n),$$

where q'' is the part of q not containing x_1 minus q'. By induction, there is a change of variable so that and so we are done (taking $c_1 = a_{11}$).

$$q''(x_2,\ldots,x_n) = \sum_{i=2}^{n} c_i y_i^2,$$

Case 2: All a_{ii} are zero, but $a_{ij} \neq 0$ for some $i \neq j$. Now

$$x_{ij} = \tfrac{1}{4}\left((x_i + x_j)^2 - (x_i - x_j)^2\right),$$

Case 3: All a_{ij} are zero. Now q is the zero form, and there is nothing to prove: take $c1 = \cdots = cn = 0$.

Example : Consider the quadratic form $q(x,y,z) = x^2 + 2xy + 4xz + y^2 + 4z^2$. We have

$$(x+y+2z)^2 = x^2 + 2xy + 4xz + y^2 + 4z^2 + 4yz,$$

$$\begin{aligned} q &= (x+y+2z)^2 - 4yz \\ &= (x+y+2z)^2 - (y+z)^2 + (y-z)^2 \\ &= u^2 + v^2 - w^2, \end{aligned}$$

where $u = x + y + 2z$, $v = y - z$, $w = y + z$. Otherwise said, the matrix representing the quadratic form, namely is congruent to the matrix

$$A = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 1 & 0 \\ 2 & 0 & 4 \end{bmatrix}$$

$$A' = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}.$$

Can you find an invertible matrix P such that $P{>}AP = A'$?

Thus any quadratic form can be reduced to the diagonal shape $\alpha_1 x_1^2 + \cdots + \alpha n x_n^2$ by a linear substitution. But this is still not a "canonical form for congruence".

For example, if $y_1 = x_1/c$, then $\alpha_1 x_1^2 = (\alpha 1c2)\, y_1^2$. In other words, we can multiply any αi by any factor which is a perfect square in $\mathbb{K}$. Over the complex numbers C, every element has a square root. Suppose that $\alpha_1,\ldots,\alpha_r \neq 0$, and $\alpha_{r+1} = \cdots = \alpha_n = 0$. Putting

$$y_i = \begin{cases} (\sqrt{\alpha_i})x_i & \text{for } 1 \le i \le r, \\ x_i & \text{for } r+1 \le i \le n, \end{cases}$$

we have $q = y_1^2. + \cdots + y_r^2..$

We will see later that r is an "invariant" of q: however we do the reduction, we arrive at the same value of r. Over the real numbers R, things are not much worse. Since any positive real number has a square root, we may suppose that $\alpha_1,\ldots,\alpha_s > 0$, $\alpha_{s+1},\ldots,\alpha_{s+t} < 0$, and $\alpha_{s+t+1},\ldots,\alpha_n = 0$. Now putting

$$y_i = \begin{cases} (\sqrt{\alpha_i})x_i & \text{for } 1 \le i \le s, \\ (\sqrt{-\alpha_i})x_i & \text{for } s+1 \le i \le s+t, \\ x_i & \text{for } s+t+1 \le i \le n, \end{cases}$$

We get $q = x_1^2 + \cdots + x + s2 - xs\ 2{+}1 - \cdots - xs\ 2{+}t$.

Again, we will see later that s and t don't depend on how we do the reduction.

[This is the theorem known as *Sylvester's Law of Inertia*.]

10.3.2 Quadratic and bilinear forms

The formal definition of a quadratic form looks a bit different from the version we gave earlier, though it amounts to the same thing. First we define a bilinear form.

Definition 10.3.3 (a) Let $b : V \times V \to K$ be a function of two variables from V with values in K. We say that b is a *bilinear form* if it is a linear function of each variable when the other is kept constant: that is,

$b(v,w1 + w2) = b(v,w1) + b(v,w2)$, $b(v,cw) = cb(v,w)$,

With two similar equations involving the first variable. A bilinear form b is *symmetric* if $b(v,w) = b(w,v)$ for all $v,w \in V$.

(b) Let $q : V \to K$ be a function. We say that q is a *quadratic form* if

– $q(cv) = c^2q(v)$ for all $c \in K$, $v \in V$;

– the function b defined by

$b(v,w) = \frac{1}{2}(q(v+w) - q(v) - q(w))$

is a bilinear form on V.

Remarks The bilinear form in the second part is symmetric; and the division by 2 in the definition is permissible because of our assumption that the characteristic of K is not 2.

If we think of the prototype of a quadratic form as being the function x^2, then the first equation says $(cx)^2 = c^2x^2$, while the second has the form

$$\frac{1}{2}((x+y)^2 - x^2 - y^2) = xy,$$

and xy is the prototype of a bilinear form: it is a linear function of x when y is constant, and *vice versa*.

Note that the formula $b(x,y) = \frac{1}{2}(q(x+y) - q(x) - q(y))$ (which is known as the *polarization formula*) says that the bilinear form is determined by the quadratic term.

Conversely, if we know the symmetric bilinear form b, then we have

$$2q(v) = 4q(v) - 2q(v) = q(v+v) - q(v) - q(v) = 2b(v,v),$$

so that $q(v) = b(v,v)$, and we see that the quadratic form is determined by the symmetric bilinear form. So these are equivalent objects.

If b is a symmetric bilinear form on V and $B = (v1,...,vn)$ is a basis for V, then we can represent b by the $n\times n$ matrix A whose (i, j) entry is $a_{ij} = b(v_i,v_j)$.

Note that A is a symmetric matrix. It is easy to see that this is the same as the matrix representing the quadratic form.

Here is a third way of thinking about a quadratic form. Let V^* be the dual space of V, and let $\alpha : V \to V^*$ be a linear map. Then for $v \in V$, we have $\alpha(v) \in V^*$, and so $\alpha(v)(w)$ is an element of K. The function

$$b(v,w) = \alpha(v)(w)$$

is a bilinear form on V . If $\alpha(v)(w) = \alpha(w)(v)$ for all $v,w \in V$, then this bilinear form is symmetric. Conversely, a symmetric bilinear form b gives rise to a linearbmap $\alpha : V \to V^*$ satisfying $\alpha(v)(w) = \alpha(w)(v)$, by the rule that $\alpha(v)$ is the linear map $w \to b(v,w)$.

Now given $\alpha : V \to V^*$, choose a basis B for V , and let B^* be the dual basis for V^* Then α is represented by a matrix A relative to the bases B and B^*

Proposition 10.3.4 The following objects are equivalent on a vector space over a field whose characteristic is not 2:

(a) a quadratic form on V ;

(b) a symmetric bilinear form on V ;

(c) a linear map $\alpha : V \to V^*$ satisfying $\alpha(v)(w) = \alpha(w)(v)$ for all $v,w \in V$ .Moreover, if corresponding objects of these three types are represented *by matrices as described above, then we get the same matrix A in each* case. Also, a change of basis in V with transition matrix P replaces A by $P^\top AP$.

Proof : Only the last part needs proof. We have seen it for a quadratic form, and the argument for a bilinear form is the same. So suppose that $\alpha : V \to V^*$, and we change from B to B' in V with transition matrix P. We saw that the transition matrix between the dual bases in V^* is $(P^\top)^{-1}$. Now go back to the discussion

of linear maps between different vector spaces in Chapter 4. If $\alpha : V \to W$ and

we change bases in V and W with transition matrices P and Q, then the matrix A representing α is changed to $Q^{-1}AP$. Apply this with $Q = P^\top)-1$, so that $Q^{-1} = P^\top$,and we see that the new matrix is $P^\top AP$, as required.

Check your progress

1. Explain Quadartic form

2. What do you understand by bilinear form, enumerate it.

10.4 CANONICAL FORMS FOR COMPLEX AND REAL FORMS

Finally, in this section, we return to quadratic forms (or symmetric matrices) over the real and complex numbers, and find canonical forms under congruence. Recall that two symmetric matrices A and A' are congruent if $A' = P^{\top}AP$ for some invertible matrix P; as we have seen, this is the same as saying that the represent the same quadratic form relative to different bases.

Theorem 10.4.1 *Any n× n complex symmetric matrix A is congruent to a matrix of the form*

$$\begin{bmatrix} I_r & O \\ O & O \end{bmatrix}$$

for some r. Moreover, r = rank(A), and so A is congruent to two matrices of this form then they both have the same value of r.

Proof We already saw that A is congruent to a matrix of this form. Moreover, if P is invertible, then so is $P^{\top}$, and so as claimed

$$r = \text{rank}(P^{\top}AP) = \text{rank}(A)$$

The next result is Sylvester's Law of Inertia.

Theorem 10.4.2 Any n× n real symmetric matrix A is congruent to a matrix of the form

$$\begin{bmatrix} I_s & O & O \\ O & -I_t & O \\ O & O & O \end{bmatrix}$$

for some s,t. Moreover, if A is congruent to two matrices of this form, then they have the same values of s and of t.

Proof Again we have seen that A is congruent to a matrix of this form. Arguing as in the complex case, we see that $s + t = \text{rank}(A)$, and so any

$$q = y_1^2 + \cdots + y_s^2 - y_{s+1}^2 - \cdots - y_{s+t}^2 = z_1^2 + \cdots + z_{s'}^2 - z_{s'+1}^2 - \cdots - z_{s+t}^2.$$

two matrices of this form congruent to *A* have the same values of $s+t$. Suppose that two different reductions give the values s,t and s',t' respectively, with $s+t = s' + t' = n$. Suppose for a contradiction that $s < s'$. Now let q be the quadratic form represented by *A*. Then we are told that there are linear functions $y1,\dots,yn$ and $z1,\dots,zn$ of the original variables $x1,\dots,xn$ of q such that

Now consider the equations

$y_1 = 0,\dots,\qquad y_s = 0,\qquad z_{s'+1} = 0,\dots\qquad z_n = 0$

regarded as linear equations in the original variables $x1,\dots,xn$. The number of equations is $s+(n - s') = n-(s' - s) < n$. According to a lemma from much earlier in the course (we used it in the proof of the Exchange Lemma!), the equations have a non-zero solution. That is, there are values of $x1,\dots,xn$, not all zero, such that the variables $y1,\dots,ys$ and $z_{s'+1},\dots,z_n$ are all zero.

Since $y_1 = \cdots = ys = 0$, we have for these values

$$q = -y_{s+1}^2 - \cdots - y_n^2 \leq 0.$$

But since z

$s'+1 = \cdots = zn = 0$, we also have

$$q = z_1^2 + \cdots + z_{s'}^2 > 0.$$

But this is a contradiction. So we cannot have $s < s'$. Similarly we cannot have $s0 < s$ either. So we must have $s = s'$, as required to be proved.

We saw that $s+t$ is the rank of *A*. The number $s-t$ is known as the *signature* of *A*. Of course, both the rank and the signature are independent of how we reduce the matrix (or quadratic form); and if we know the rank and signature, we can easily recover s and t.

Let q be a quadratic form in n variables represented by the real symmetric matrix *A*. Let q (or *A*) have rank $s+t$ and signature $s-t$, that is, have s positive and t negative terms in its diagonal form. We say that q (or *A*) is

• ***positive definite*** if $s = n$ (and $t = 0$), that is, if $q(v) \geq 0$ for all v, with equality only if $v = 0$;

• ***positive semidefinite*** if $t = 0$, that is, if $q(v) \geq 0$ for all v;

• ***negative definite*** if $t = n$ (and $s = 0$), that is, if $q(v) \leq 0$ for all v, with equality only if $v = 0$;

• ***negative semi-definite*** if $s = 0$, that is, if $q(v) \leq 0$ for all v;

• ***indefinite*** if $s > 0$ and $t > 0$, that is, if $q(v)$ takes both positive and negative values.

Lemma 10.4.3. Let $A \in \mathbb{M}_n(\mathbb{C})$. Then A is Hermitian if and only if at least one of the following
statements hold:

1. S^*AS is Hermitian for all $S \in \mathbb{M}_n$.
2. A is normal and has real eigenvalues.
3. $\mathbf{x}^*A\mathbf{x} \in \mathbb{R}$ for all $\mathbf{x} \in \mathbb{C}_n$.

Proof. Let $S \in \mathbb{M}_n$, $(S^*AS)^* = S^*A^*S = S^*AS$. Thus S^*AS is Hermitian.
Suppose $A = A^*$. Then, A is clearly normal as $AA^* = A^2 = {}^*A$. Further, if $(\lambda, \mathbf{x})$ is an
eigenpair then $\lambda\mathbf{x}^*\mathbf{x} = \mathbf{x}^*A\mathbf{x} \in \mathrm{R}$ implies $\lambda \in \mathrm{R}$.
For the last part, note that $\mathbf{x}^*A\mathbf{x} \in \mathrm{C}$. Thus $\overline{\mathbf{x}^*A\mathbf{x}} = (\mathbf{x}^*A\mathbf{x})^* = \mathbf{x}^*A^*\mathbf{x} = \mathbf{x}^*A\mathbf{x}$, we get $\mathrm{Im}(\mathbf{x}^*A\mathbf{x}) = 0$. Thus, $\mathbf{x}^*A\mathbf{x} \in \mathbb{R}$.

If S^*AS is Hermitian for all $S \in \mathbb{M}_n$. then taking $S = In$ gives A is Hermitian.
If A is normal then $A = U^* \operatorname{diag}(\lambda 1, \ldots, \lambda n)U$ for some unitary matrix U. Since $\lambda i \in \mathrm{R}$,
$A^* = (U^* \operatorname{diag}(\lambda 1, \ldots, \lambda n)U)^* = U^* \operatorname{diag}(\lambda 1, \ldots, \lambda n)U = U^* \operatorname{diag}(\lambda 1, \ldots, \lambda n)U = A$.
So, A is Hermitian.
If $\mathbf{x}^*A\mathbf{x} \in \mathrm{R}$ for all $\mathbf{x} \in \mathbb{C}_n$. then $aii = \mathbf{e}^*i\, A\mathbf{e}i \in \mathrm{R}$. Also,
$a_{ii} + a^{jj} + a_{ij} + a_{ji} = (\mathbf{e}i + \mathbf{e}j)^*A(\mathbf{e}i + \mathbf{e}j) \in \mathrm{R}$.
So,
$\mathrm{Im}(aij) = -\mathrm{Im}(aji)$. Similarly, $a_{ii} + a_{jj} + ia_{ij} - ia_{ji} = (\mathbf{e}i + i\mathbf{e}j)^*A(\mathbf{e}i + i\mathbf{e}j) \in$ Rimplies that $\mathrm{Re}(a_{ij}) = \mathrm{Re}(a_{ji})$. Thus, $A = A^*$.

Remark 10.3.4 Let $A \in \mathbb{M}_n(\mathbb{R})$. Then the condition $\mathbf{x}^*A\mathbf{x} \in \mathbb{R}$ in Definition 6.3.9 is always true and hence doesn't put any restriction on the matrix A. So, in Definition 6.3.9, we assume that $A^T = A$, i.e., A is a symmetric matrix.

Example:

1. Let $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ or $A = \begin{bmatrix} 3 & 1+i \\ 1-i & 4 \end{bmatrix}$. Then, A is positive definite.

Theorem 10.4.5 Let A ∈ $\mathbb{M}_n(\mathbb{C})$. Then, the following statements are

2. Let $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ or $A = \begin{bmatrix} \sqrt{2} & 1+i \\ 1-i & \sqrt{2} \end{bmatrix}$. Then, A is positive semi-definite but not positive definite.

equivalent.

1. A is positive semi-definite.
2. $A^* = A$ and each eigenvalue of A is non-negative.
3. $A = B^* B$ for some $B \in \mathbb{M}_n(\mathbb{C})$.

Proof. $1 \Rightarrow 2$: Let A be positive semi-definite. Then, by Lemma 10.3.3 A is Hermitian. If

$(\alpha, \mathbf{v})$ is an eigen-pair of A then $\alpha||\mathbf{v}||^2 = \mathbf{v}^*A\mathbf{v} \geq 0$. So, $\alpha \geq 0$.

$2 \Rightarrow 3$: Let $\sigma(A) = \{\alpha 1, \ldots, \alpha n\}$. Then, by spectral theorem, there exists a unitary matrix U such that $U^*AU = D$ with $D = \text{diag}(\alpha 1, \ldots, \alpha n)$. As $\alpha i \geq 0$, for $1 \leq i \leq n$ define

$D^{1/2} = \text{diag}(\sqrt{\alpha_1}, \ldots, \sqrt{\alpha_n},)$. Then, $A = U D^{1/2}\ [D^{1/2} U^*] = B^* B$.

$3 \Rightarrow 1$: Let $A = B^* B$. Then, for $\mathbf{x} \in \mathbb{C}_n$, $\mathbf{x}^*A\mathbf{x} = \mathbf{x}^*B^*B\mathbf{x} = ||B\mathbf{x}||^2 \geq 0$. Thus, the required result follows.

A similar argument gives the next result and hence the proof is omitted.

Theorem 10. 4.6. Let A ∈ $\mathbb{M}_n(\mathbb{C})$.Then, the following statements are equivalent.

1. A is positive definite.
2. $A^* = A$ and each eigenvalue of A is positive.
3. $A = B^*B$ for a non-singular matrix $B \in \mathbb{M}_n(\mathbb{C})$.

Remark 10.4.7. Let A ∈ $\mathbb{M}_n(\mathbb{C})$.be a Hermitian matrix with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. Then, there exists a unitary matrix $U = [\mathbf{u}1, \mathbf{u}2, \ldots, \mathbf{u}n]$ and a diagonal matrix $D = \text{diag}(\lambda 1, \lambda 2, \ldots, \lambda n)$ such that $A = UDU^*$. Now, for $1 \leq i \leq n$, define $\alpha i = \max\{\lambda i, 0\}$ and $\beta i = \min\{\lambda i, 0\}$.

Then

1. for D1 = diag(α1, α2, . . . , αn), the matrix A1 = UD1U* is positive semi-definite.

2. for D2 = diag(β1, β2, . . . , βn), the matrix A2 = UD2U* is positive semi-definite.

3. $A = A_1 - A_2$. The matrix A1 is generally called the positive semi-definite part of A.

Definition 10.4.8. [Multilinear Function] Let 𝕍 be a vector space over 𝔽. Then,

1. for a fixed $m \in \mathbb{N}$, a function $f : \mathbb{V}^m \to \mathbb{F}$ is called an *m*-**multilinear** function if *f* is linear in each component. That is,

$f(\mathbf{v}1, \ldots, \mathbf{v}i-1, (\mathbf{v}i + \alpha\mathbf{u}), \mathbf{v}i+1 \ldots, \mathbf{v}m) = f(\mathbf{v}1, \ldots, \mathbf{v}i-1, \mathbf{v}i, \mathbf{v}i+1 \ldots, \mathbf{v}m) + \alpha f(\mathbf{v}1, \ldots, \mathbf{v}i-1, \mathbf{u}, \mathbf{v}i+1 \ldots, \mathbf{v}m)$

for $\alpha \in \mathbb{F}$, $\mathbf{u} \in \mathbb{V}$ and $\mathbf{v}i \in \mathbb{V}$, for $1 \leq i \leq m$.

2. An *m*-multilinear form is also called an *m*-**form**.

3. A 2-form is called a **bilinear form**.

Definition 10.4.9. [Sesquilinear, Hermitian and Quadratic Forms]
Let $A = [aij] \in \mathbb{M}_n(\mathbb{C})$.
be a Hermitian matrix and let $\mathbf{x}, \mathbf{y} \in \mathbb{C}_n$. Then, a **sesquilinear form** in $\mathbf{x}, \mathbf{y} \in \mathbb{C}_n$ is defined as $H(\mathbf{x}, \mathbf{y}) = \mathbf{y}{*}A\mathbf{x}$. In particular, $H(\mathbf{x}, \mathbf{x})$, denoted $H(\mathbf{x})$, is called a **Hermitian form**. In case $A \in Mn(\mathbb{R})$, $H(\mathbf{x})$ is called a **quadratic** form.

Remark 10.4.10. Observe that

1. if A = In then the bilinear/sesquilinear form reduces to the standard inner product.

2. H(**x**, **y**) is 'linear' in the first component and 'conjugate linear' in the second component.

3. The quadratic form H(**x**) is a real number. Hence, for α ∈ R, the equation H(**x**) = α, represents a conic in $\mathbb{R}^n$.

10.5 SYLVESTER'S LAW OF INERTIA

The main idea of this section is to express $H(\mathbf{x})$ as sum or difference of squares. Since $H(\mathbf{x})$ is a quadratic in **x**, replacing **x** by $c\mathbf{x}$, for $c \in \mathbb{C}$, just gives a multiplication factor by $|c|^2$. Hence,

one needs to study only the normalized vectors. Let us consider Example 6.1.1 again. There we see that

$$\mathbf{x}^T A\mathbf{x} = 3\frac{(x+y)^2}{2} - \frac{(x-y)^2}{2} = (x+2y)^2 - 3y^2, \text{ and} \quad (1)$$

$$\mathbf{x}^T B\mathbf{x} = 5\frac{(x+2y)^2}{5} + 10\frac{(2x-y)^2}{5} = (3x - \frac{2y}{3})^2 + \frac{50y^2}{9}. \quad (2)$$

Note that both the expressions in Equation (1) is the difference of two non-negative terms.

Whereas, both the expressions in Equation (2) consists of sum of two non-negative terms. Is this just a coincidence?

In general, let $A \in \mathbb{M}_n(\mathbb{C})$ be a Hermitian matrix. Then, $\sigma(A) = \{\alpha 1, \ldots, \alpha n\} \subseteq \mathbb{R}$ and there exists a unitary matrix U such that $U^*AU = D = \text{diag}(\alpha 1, \ldots, \alpha n)$.

Let $\mathbf{x} = U\mathbf{z}$. Then, $k\mathbf{x}k = 1$ and U is unitary implies that $||\mathbf{z}|| = 1$. If $\mathbf{z} = (\mathbf{z}1, \ldots, \mathbf{z}n)^*$ then

$$H(\mathbf{x}) = \mathbf{z}^*U^*AU\mathbf{z} = \mathbf{z}^*D\mathbf{z} = \sum_{i=1}^{n} \alpha_i|\mathbf{z}_i|^2 = \sum_{i=1}^{p} |\sqrt{\alpha_i}\, \mathbf{z}_i|^2 - \sum_{i=p+1}^{r} \left|\sqrt{|\alpha_i|}\, \mathbf{z}_i\right|^2, \quad (3)$$

where $\alpha 1, \ldots, \alpha p > 0$, $\alpha p+1, \ldots, \alpha r < 0$ and $\alpha r+1, \ldots, \alpha n = 0$. Thus, we see that the possible values of $H(\mathbf{x})$ seem to depend only on the eigenvalues of A. Since U is an invertible matrix, the components $\mathbf{z}i$'s of $\mathbf{z} = U^{-1}\mathbf{x} = U^*\mathbf{x}$ are commonly known as the **linearly independent linear forms**. Note that each $\mathbf{z}i$ is a linear expression in the components of **x**. Also, note that in Equation (6.3.3), p corresponds to the number of positive eigenvalues and $r - p$ to the number of negative eigenvalues.

For a better understanding, we define the following numbers.

Definition 10.5.1. [Inertia and Signature of a Matrix] Let $A \in \mathbb{M}_n(\mathbb{C})$ be a Hermitian matrix. The **inertia** of A, denoted $i(A)$, is the triplet $(i+(A), i-(A), i'(A))$, where $i+(A)$ is the number of positive eigenvalues of

A, $i-(A)$ is the number of negative eigenvalues of A and $i'(A)$ is the nullity of A. The difference $i+(A) - i-(A)$ is called the **signature** of A.

Lemma 10.5.2 . [Sylvester's Law of Inertia] Let $A \in \mathbb{M}_n(\mathbb{C})$ be a Hermitian matrix and let

$\mathbf{x} \in \mathbb{C}_n$. Then, every Hermitian form $H(\mathbf{x}) = \mathbf{x}^* A\mathbf{x}$, in n variables can be written as

$$H(\mathbf{x}) = |\mathbf{y}_1|^2 + \cdots + |\mathbf{y}_p|^2 - |\mathbf{y}_{p+1}|^2 - \cdots - |\mathbf{y}_r|^2$$

where

$\mathbf{y}1, \ldots, \mathbf{y}r$ are linearly independent linear forms in the components of $\mathbf{x}$ and the integers p and r satisfying $0 \le p \le r \le n$, depend only on A.

Proof. Equation (3) implies that $H(\mathbf{x})$ has the required form. We only need to show that p and r are uniquely determined by A. Hence, let us assume on the contrary that there exist p, q, r, s ∈ N with p > q such that

$$\begin{aligned} H(\mathbf{x}) &= |\mathbf{y}_1|^2 + \cdots + |\mathbf{y}_p|^2 - |\mathbf{y}_{p+1}|^2 - \cdots - |\mathbf{y}_r|^2 \quad (4) \\ &= |\mathbf{z}_1|^2 + \cdots + |\mathbf{z}_q|^2 - |\mathbf{z}_{q+1}|^2 - \cdots - |\mathbf{z}_s|^2, \quad (5) \end{aligned}$$

where $\mathbf{y} = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = M\mathbf{x}$, $\mathbf{z} = \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} = N\mathbf{x}$ with $Y1 = \begin{bmatrix} y_1 \\ \vdots \\ y_p \end{bmatrix}$ and $Z1 = \begin{bmatrix} z_1 \\ \vdots \\ z_p \end{bmatrix}$ for some invertible matrices M and N. Now the invertibility of M and N implies $\mathbf{z} = B\mathbf{y}$, for some invertible matrix B.

Decompose

$$\begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} = \begin{bmatrix} B_1 & B_2 \\ B_3 & B_4 \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}.$$

where $B1$ is a $q \times p$ matrix. Then

$$B = \begin{bmatrix} B_1 & B_2 \\ B_3 & B_4 \end{bmatrix}$$

As $p > q$, the homogeneous linear system $B_1 Y_1 = \mathbf{0}$ has a nontrivial solution, say $\tilde{Y}_1 = \begin{bmatrix} \tilde{y}_1 \\ \vdots \\ \tilde{y}_p \end{bmatrix}$ and consider $\tilde{y} = \begin{bmatrix} Y_1 \\ 0 \end{bmatrix}$ .Then for this choice of **y** e, $Z1$ = **0** and thus, using Equations (4) and (5), we have

$$H(\tilde{\mathbf{y}}) = |\tilde{y_1}|^2 + |\tilde{y_2}|^2 + \cdots + |\tilde{y_p}|^2 - 0 = 0 - (|z_{q+1}|^2 + \cdots + |z_s|^2).$$

Now, this can hold only if $\widetilde{Y}_1$= **0**, a contradiction to Y f 1 being a non-trivial solution. Hence

$p = q$. Similarly, the case $r > s$ can be resolved. This completes the proof of the lemma.

Remark 10.5.3. Since A is Hermitian, Rank(A) equals the number of nonzero eigenvalues. Hence, Rank(A) = r. The number r is called the **rank** and the number r − 2p is called the **inertial degree** of the Hermitian form H(**x**).

Check your progress

3. What is Hermitian and state its necessary condition?

---

4. Define Signature and inertia of the matrix

10.6 LET’S SUM UP

We understand the quadratic forms, its representation and its reduction. We have understood the relation between quadratic form and bilinear form. Comprehended the Slyvester law of Inertia

10.7 KEYWORDS

1. Symmetric matrices – a **symmetric matrix** is a square **matrix** that is equal to its transpose.
2. Unitary matrix – a **matrix** that has an inverse and a transpose whose corresponding elements are pairs of conjugate complex numbers.
3. Restriction – the **restriction** of a function is a new function, denoted or , obtained by choosing a smaller domain A for the original function

4. Permissible – that may be **permitted** : **allowable**

10.8 QUESTION FOR REVIEW

1. Let $A \in \mathbb{M}n(\mathbb{C})$ be a Hermitian matrix. If the signature and the rank of A is known then prove that one can find out the inertia of A.

2. Explain reduction of Quadratic forms

3. Explain Slyvester law of Inertia

10.9 SUGGESTED READINGS

i. K. Hauffman and R. Kunz, Linear Algebra, Pearson Education (INDIA), 2003.

ii. G. Strang, Linear Algebra And Its Applications, 4th Edition, Brooks/Cole, 2006.

iii. S. Lang, Linear Algebra, Springer, 1989.

iv. David S. Dummit and Richard M. Foote, Abstract Algebra (3e), John Wiley and Sons.

v. R. Gallian Joseph, Contemporary Abstract Algebra, Narosa Publishing House.

vi. Thomas Hungerford, Algebra, Springer GTM.

vii. I.N. Herstein, Topics in Abstract Algebra, Wiley Eastern Limited.

viii. D.S. Malik, J.M. Mordesen, M.K. Sen, Fundamentals of Abstract Algebra, The McGraw-Hill Companies, Inc.

10.10 ANSWER TO CHECK YOUR PROGRESS

1. Provide definition and example—10.1.1

2. Provide explanation and definition – 10.2.3

3. Provide definition and explanation – 10.3.3

4. Provide definition – 10.4.1

UNIT 11: JORDAN CANONICAL FORM

STRUCTURE

11.0 OBJECTIVE

Understand the generalized Schur's theorem and its application

Comprehend the JORDAN CANONICAL FORM THEOREM

ENUMERATE TE CONCEPT OF minimal polynomial

11.1 INTRODUCTION

Jordan canonical form is a representation of a linear transformation over a finite-dimensional complex vector space by a particular kind of upper triangular matrix. Every such linear

transformation has a unique Jordan canonical form, which has useful properties: it is easy to describe and well-suited for computations.
Less abstractly, one can speak of the Jordan canonical form of a square matrix; every square matrix is similar to a unique matrix in Jordan canonical form, since similar matrices correspond to representations of the same linear transformation with respect to different bases, by the change of basis theorem.
Jordan canonical form can be thought of as a generalization of diagonalizability to arbitrary linear transformations (or matrices); indeed, the Jordan canonical form of a diagonalizable linear transformation (or a diagonalizable matrix) is a diagonal matrix.

11.2 GENERALIZED SCHUR'S THEOREM

We start this chapter with the following theorem which generalizes the Schur Upper triangularization theorem.

Theorem 11.2.1. [Generalized Schur's Theorem] Let A ∈ Mn(C). Suppose $\lambda_1, \dots, \lambda_k$ are the distinct eigenvalues of A with multiplicities m1, . . . , mk, respectively. Then, there exists a non-singular matrix W such that

$$W^{-1}AW = \bigoplus_{i=1}^{k} T_i, \ \ where, \ T_i \in \mathbb{M}_{m_i}(\mathbb{C}), \ \ for \ 1 \leq i \leq k$$

and T_i's are upper triangular matrices with constant diagonal λ_i. If A has real entries with real eigenvalues then W can be chosen to have real entries.

Proof. By Schur Upper Triangularization, there exists a unitary matrix U such that U*AU = T, an upper triangular matrix with diag(T) = (λ1, . . . , λ1, . . . , λk, . . . , λk). Now, for any upper triangular matrix B, a real number α and i < j, consider the matrix F(B, i, j, α) = Eij(−α)BEij(α). Then, for 1 ≤ k, l ≤ n,

$$(F(B,i,j,\alpha))_{k\ell} = \begin{cases} B_{ij} - \alpha B_{jj} + \alpha B_{ii}, & \text{whenever } k = i, \ell = j \\ B_{i\ell} - \alpha B_{j\ell}, & \text{whenever } \ell \neq j \\ B_{kj} + \alpha B_{ki}, & \text{whenever } k \neq i \\ B_{k\ell}, & \text{otherwise.} \end{cases} \qquad (1)$$

Notes

Now, using Equation (1), the diagonal entries of *F*(*T, i, j, α*) and *T* are equal and

$$(F(T,i,j,\alpha))_{ij} = \begin{cases} T_{ij}, & \text{whenever } T_{jj} = T_{ii} \\ 0, & \text{whenever } T_{jj} \neq T_{ii} \text{ and } \alpha = \frac{T_{ij}}{T_{jj}-T_{ii}}. \end{cases}$$

Thus, if we denote the matrix F(T, i, j, α) by T1 then (F(T1, i − 1, j, α))$_{i-1,j}$ = 0, for some choice of α, whenever $(T_1)_{i-1,i-1} \neq$ Tjj. Moreover, this operation also preserves the 0 created by F(T, i, j, α) at (i, j)-th place. Similarly, F(T1, i, j + 1, α) preserves the 0 created by F(T, i, j, α) at (i, j)-th place. So, we can successively apply the following sequence of operations to get T → F(T, m_1, m_1+1, α) = T_1 → F(T1, m_1−1, m_1+1, β) → · · · → F(T_{m_1-1}, 1, m_1+1, γ) = Tm_1 where α, β, . . . , γ are appropriately chosen and Tm1[:, m1 + 1] = $\lambda_2 \mathbf{e}_{\mathbf{m_1}}$+1. Thus, observe that the above operation can be applied for different choices of i and j with i < j to get the required result.

Definition 11.2.2. [Jordan Block and Jordan Matrix]

1. Let *λ* ∈ ℂ and *k* be a positive integer. Then, by the **Jordan block** *Jk*(*λ*) ∈ $\mathbb{M}_k$(ℂ), we
understand the matrix

$$\begin{bmatrix} \lambda & 1 & & \\ & \ddots & \ddots & \\ & & \lambda & 1 \\ & & & \lambda \end{bmatrix}$$

2. A **Jordan matrix** is a direct sum of Jordan blocks. That is, if *A* is a Jordan matrix
having *r* blocks then there exist positive integers *ki*'s and complex numbers *λi*'s (not
necessarily distinct), for 1 ≤ *i* ≤ *r* such that

$$A = J_{k_1}(\lambda_1) \oplus \cdots \oplus J_{k_r}(\lambda r).$$

1. $J_1(0)$ = [0] is the only Jordan matrix of size 1.

2. $J_1(0) \oplus J_1(0) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ and $J_2(0) = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ are Jordan matrices of size 2.

3. Even though , $J_1(0) \oplus J_2(0)$ and $J_2(0) \oplus J_1(0)$ are two Jordan matrices of size 3, we do

not differentiate between them as they are similar

4. $J_1(0) \oplus J_1(0) \oplus J_1(0) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ $J_2(0) \oplus J_1(0) = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ and

$J_3(0) = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$

Remark 11.2.3. [Jordan blocks] *Fix a positive integer k. Then,*

1. $J_k(\lambda)$ *is an upper triangular matrix with* λ *as an eigenvalue.*

2. $J_k(\lambda) = \lambda I_k + J_k(0)$.

3. Alg.Mul $\lambda(J_k(\lambda)) = k$.

4. The matrix $J_k(0)$ *satisfies the following properties.*

5. Thus, using Remark 11.1.3.4d Geo.Mul $\lambda(J_k(\lambda)) = 1$

(a) $\text{Rank}((J_k(0)^i) = k - i, \text{ for } 1 \le i \le k.$

(b) $J_k^T(0)J_k(0) = \begin{bmatrix} 0 & 0 \\ 0 & I_{k-1} \end{bmatrix}.$

(c) $J_k(0)^p = 0 \text{ whenever } p \ge k.$

(d) $J_k(0)\mathbf{e}_i = \mathbf{e}_{i-1} \text{ for } i = 2, \ldots, k.$

(e) $\left(I - J_k^T(0)J_k(0)\right)\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{0} \end{bmatrix} = \langle \mathbf{x}, \mathbf{e}_1 \rangle \mathbf{e}_1.$

Definition 11.2.4. [Index of an Eigenvalue] Let J be a Jordan matrix containing $Jt(\lambda)$, for some positive integer t and some complex number λ. Then, the smallest value of k for which $\text{Rank}((J - \lambda I)^k)$ stops decreasing is the order of the largest Jordan block $J_k(\lambda)$ in J. This number k is called the **index of the eigenvalue** λ.

Lemma 11.2.5. *Let A* $\in \mathbb{M}_n(\mathbb{C})$ be strictly upper triangular. Then, A is similar to a direct sum of Jordan blocks. That is, there exists a non-singular matrix S and integers n1 ≥ . . . ≥ nm ≥ 1

such that

$$A = S^{-1}(J_{n_1}(0) \oplus \cdots \oplus J_{n_m}(0))S.$$

If A ∈ $\mathbb{M}_n(\mathbb{R})$ then S can be chosen to have real entries.

Proof. We will prove the result by induction on n. For n = 1, the statement is trivial. So, let the result be true for matrices of size $\leq n-1$ and let $A \in \mathbb{M}_n(\mathbb{C})$ be strictly upper triangular.

Then, $A = \begin{bmatrix} 0 & a^T \\ 0 & A_1 \end{bmatrix}$. By induction hypothesis there exists an invertible matrix $S1$ such that

$$A_1 = S_1^{-1}\Big(J_{n_1}(0) \oplus \cdots \oplus J_{n_m}(0)\Big)S_1 \text{ with } \sum_{i=1}^{m} n_i = n-1.$$

Thus,

$$\begin{bmatrix} 1 & 0 \\ 0 & S_1^{-1} \end{bmatrix} A \begin{bmatrix} 1 & 0 \\ 0 & S_1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & S_1^{-1} \end{bmatrix} \begin{bmatrix} 0 & \mathbf{a}^T \\ 0 & A_1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & S_1 \end{bmatrix} = \begin{bmatrix} 0 & \mathbf{a}^T S_1 \\ 0 & S^{-1}A_1 S_1 \end{bmatrix} = \begin{bmatrix} 0 & \mathbf{a}_1^T & \mathbf{a}_2^T \\ 0 & J_{n_1}(0) & 0 \\ 0 & 0 & J \end{bmatrix},$$

where $S_1^{-1}\ (J_{n_1}(0)\ \oplus \cdots \oplus\ J_{n_m}(0))\ \ S_1 = J_{n_1}\ (0) \oplus J$ and $\mathbf{a}^T\, S_1 =$ $[\boldsymbol{a}_1^T \boldsymbol{a}_2^T]$. Now, writing J_{n_1} to mean J_{n_1} (0) and using Remark 11.1.3.4e, we have

$$\begin{bmatrix} 1 & -\mathbf{a}_1^T J_{n_1}^T & 0 \\ 0 & I_{n_1} & 0 \\ 0 & 0 & I \end{bmatrix} \begin{bmatrix} 0 & \mathbf{a}_1^T & \mathbf{a}_2^T \\ 0 & J_{n_1} & 0 \\ 0 & 0 & J \end{bmatrix} \begin{bmatrix} 1 & \mathbf{a}_1^T J_{n_1}^T & 0 \\ 0 & I_{n_1} & 0 \\ 0 & 0 & I \end{bmatrix} = \begin{bmatrix} 0 & \langle \mathbf{a}_1, \mathbf{e}_1 \rangle \mathbf{e}_1^T & \mathbf{a}_2^T \\ 0 & J_{n_1} & 0 \\ 0 & 0 & J \end{bmatrix}$$

So, we now need to consider two cases depending on whether $[\mathbf{a}1, \mathbf{e}1] = 0$ or $[\mathbf{a}1, \mathbf{e}1] \neq 0$. In the first case, A is similar to $\begin{bmatrix} 0 & 0 & a_2^T \\ 0 & J_{n_1} & 0 \\ 0 & 0 & J \end{bmatrix}$. This in turn is similar to $\begin{bmatrix} J_{n_1} & 0 & 0 \\ 0 & 0 & a_2^T \\ 0 & 0 & J \end{bmatrix}$ by permuting the first row and column. At this stage, one can apply induction and if necessary do a block permutation, in order to keep the block sizes in decreasing order.

So, let us now assume that $[\mathbf{a}1, \mathbf{e}1] \neq 0$. Then, writing $\alpha = [\mathbf{a}1, \mathbf{e}1]$, we have

$$\begin{bmatrix} I & \mathbf{e}_{i+1}\mathbf{a}_2^T J^{i-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} J_{n_1+1} & \mathbf{e}_i \mathbf{a}_2^T J^{i-1} \\ 0 & J \end{bmatrix} \begin{bmatrix} I & -\mathbf{e}_{i+1}\mathbf{a}_2^T J^{i-1} \\ 0 & I \end{bmatrix} = \begin{bmatrix} J_{n_1+1} & \mathbf{e}_{i+1}\mathbf{a}_2^T J^{i} \\ 0 & J \end{bmatrix}, \text{ for } i \geq 1.$$

Now, using Remark 11.1.3.4c, verify that

$$\begin{bmatrix} \frac{1}{\alpha} & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & \frac{1}{\alpha}I \end{bmatrix} \begin{bmatrix} 0 & \alpha \mathbf{e}_1^T & \mathbf{a}_2^T \\ 0 & J_{n_1} & 0 \\ 0 & 0 & J \end{bmatrix} \begin{bmatrix} \alpha & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & \alpha I \end{bmatrix} = \begin{bmatrix} 0 & \mathbf{e}_1^T & \mathbf{a}_2^T \\ 0 & J_{n_1} & 0 \\ 0 & 0 & J \end{bmatrix} \equiv \begin{bmatrix} J_{n_1+1} & \mathbf{e}_1\mathbf{a}_2^T \\ 0 & J \end{bmatrix}.$$

Hence, for $p = n - n_1 - 1$, we have

$$\begin{bmatrix} I & \mathbf{e}_{p+1}\mathbf{a}_2^T J^{p-1} \\ 0 & I \end{bmatrix} \cdots \begin{bmatrix} I & \mathbf{e}_2\mathbf{a}^T \\ 0 & I \end{bmatrix} \begin{bmatrix} J_{n_1+1} & \mathbf{e}_1\mathbf{a}^T \\ 0 & J \end{bmatrix} \begin{bmatrix} I & -\mathbf{e}_2\mathbf{a}^T \\ 0 & I \end{bmatrix} \cdots \begin{bmatrix} I & -\mathbf{e}_{p+1}\mathbf{a}^T J^{p-1} \\ 0 & I \end{bmatrix} = \begin{bmatrix} J_{n_1+1} & 0 \\ 0 & J \end{bmatrix}$$

If necessary, we need to do a block permutation, in order to keep the block sizes in decreasing order. Hence, the required result follows

Corollary 11.2.6. *A* $\in$ M*n*(C). Then, A is similar to J, a Jordan matrix.

Proof. Let λ1, . . . , λk be the distinct eigenvalues of A with algebraic multiplicities m1, . . . , mk.

By Theorem 11.1.1, there exists a non-singular matrix S such that S^{-1} AS $= \bigoplus_{i=1}^{k} T_i$, where Ti is an upper triangular with diagonal (λi, . . . , λi).

Thus Ti $- \lambda_i I_{m_i}$ is a strictly upper triangular matrix. Thus, by Theorem 11.1.5, there exist a non-singular matrix Si such that

$$S_i^{-1}(T_i - \lambda_i I_{m_i})S_i = J(0)$$

a Jordan matrix with 0 on the diagonal and the size of the Jordan blocks decreases as we move down the diagonal. So, $S_i^{-1}T_i\,S_i = J(\lambda i)$ is a Jordan matrix with λi on diagonal and the size of the Jordan blocks decreases as we move down the diagonal.

Now, take $W = \bigoplus_{i=1}^{k} Si$. Then, verify that $W^{-1}AW$ is a Jordan matrix.

Let $A \in \mathbb{M}_n(\mathbb{C})$. Suppose $\lambda \in \sigma(A)$ and J is a Jordan matrix that is similar to A. Then, for each fixed i, $1 \le i \le n$, by `$i(\lambda)$, we denote the number of Jordan blocks $Jk(\lambda)$ in J for which $k \ge i$.

Remark 11.2.7 . *Let* $A \in \mathbb{M}_n(\mathbb{C})$. *Suppose* $\lambda \in \sigma(A)$ *and J is a Jordan matrix that is similar to A. Then, for* $l \le k \le n$,

$$l_k(\lambda) = \mathrm{Rank}(A - \lambda I)^{k-1} - \mathrm{Rank}(A - \lambda I)^k.$$

Proof. We need to consider only the Jordan blocks $J_k(\lambda)$, for different values of k. Hence, without loss of generality, let us assume that $J = \bigoplus_{i=1}^{n} a_i J_i(\lambda)$, where a_i's are non-negative integers and J contains exactly a_i copies of the Jordan block $J_i(\lambda)$, for $1 \leq i \leq n$.

We observe the following:

1. $n = \sum_{i \geq 1} i a_i$.
2. $\text{Rank}(J - \lambda I) = \sum_{i \geq 2} (i-1) a_i$.
3. $\text{Rank}((J - \lambda I)^2) = \sum_{i \geq 3} (i-2) a_i$.
4. In general, for $1 \leq k \leq n$, $\text{Rank}((J - \lambda I)^k) = \sum_{i \geq k+1} (i-k) a_i$.

Thus, writing l_i in place of $l_i(\lambda)$,we get

$$\begin{aligned}
\ell_1 &= \sum_{i\geq 1} a_i = \sum_{i \geq 1} i a_i - \sum_{i \geq 2} (i-1) a_i = n - \text{Rank}(J - \lambda I), \\
\ell_2 &= \sum_{i\geq 2} a_i = \sum_{i \geq 2} (i-1) a_i - \sum_{i \geq 3} (i-2) a_i = \text{Rank}(J - \lambda I) - \text{Rank}((J - \lambda I)^2), \\
&\vdots \\
\ell_k &= \sum_{i\geq k} a_i = \sum_{i \geq k} (i-(k-1)) a_i - \sum_{i \geq k+1} (i-k) a_i = \text{Rank}((J - \lambda I)^{k-1}) - \text{Rank}((J - \lambda I)^k).
\end{aligned}$$

Now, the required result follows as rank is invariant under similarity operation and the matrices J and A are similar.

Lemma 11.2.8. [Similar Jordan Matrices] Let J and J0 be two similar Jordan matrices of size n. Then, J is a block permutation of J0.

Proof. For 1 ≤ i ≤ n, let `i and `0i be, respectively, the number of Jordan blocks of J and J0 of size at least i corresponding to λ. Since J and J0 are similar, the matrices (J − λI)i and (J0 − λI)i are similar for all i, 1 ≤ i ≤ n. Therefore, their ranks are equal for all i ≥ 1 and hence, `i = `0i for all i ≥ 1. Thus the required result follows.

Check your progress

1. Define the following

a. **Jordan Block**

b. Jordan Matrix

---

2. What do you understand by Similar Jordan Matrices.

11.3 JORDAN CANONICAL FORM THEOREM

Theorem 11.3.1. [Jordan Canonical Form Theorem] *Let A* $\in \mathbb{M}_n(\mathbb{C})$. Then, A is similar to a Jordan matrix J, which is unique up to permutation of Jordan blocks. If A $\in \mathbb{M}_n(\mathbb{R})$ and has real eigenvalues then the similarity transformation matrix S may be chosen to have real entries.

This matrix J is called the the Jordan canonical form of A, denoted Jordan CF(A)

Example: Let us use the idea from Lemma 11.1.7 to find the Jordan Canonical Form of the following matrices.

$$\text{Let } A = J_4(0)^2 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Solution: Note that $l_1 = 4 - \text{Rank}(A - 0I) = 2$. So, there are two Jordan blocks. Also, $l_2 = \text{Rank}(A - 0I) - \text{Rank}((A - 0I)2) = 2$. So, there are at least 2 Jordan blocks of size 2. As there are exactly two Jordan blocks, both the blocks must have size 2. Hence, Jordan CF(A) = $J2(0) \oplus J2(0)$

2.

$$\text{Let } A_1 = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Solution: Let $B = A1 - I$. Then, $l_1 = 4 - \text{Rank}(B) = 1$. So, B has exactly one Jordan block and hence $A1$ is similar to j4(1)

3.

$$A_2 = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Solution: Let $C = A2 - I$. Then, $l_1 = 4 - \text{Rank}(C) = 2$. So, C has exactly two Jordan blocks. Also, $l_2 = \text{Rank}(C) - \text{Rank}(C2) = 1$ and $l_3 = \text{Rank}(C2) - \text{Rank}(C3) = 1$. So, there is at least 1 Jordan blocks of size 3.

Thus, we see that there are two Jordan blocks and one of them is of size 3. Also, the size of the matrix is 4. Thus, $A2$ is similar to $J3(1) \oplus J1(1)$.

4. Let $A = J4(1)^2 \oplus A1 \oplus A2$, where $A1$ and $A2$ are given in the previous exercises.

Solution: One can directly get the answer from the previous exercises as the matrix A is already in the block diagonal form. But, we compute it again for better understanding.

Let $B = A - I$. Then,

$l_1 = 16 - \text{Rank}(B) = 5$,

$l_2 = \text{Rank}(B) - \text{Rank}(B2) = 11 - 7 = 4$,

$l_3 = \text{Rank}(B2) - \text{Rank}(B3) = 7 - 3 = 4$ and

$l_4 = \text{Rank}(B3) - \text{Rank}(B4) = 3 - 0 = 3$.

Hence, $J4(1)$ appears thrice (as $l_4 = 3$ and $l_5 = 0$), $J3(1)$ also appears once (as $l_3 - l_4 = 1$),

$J2$ (1) does not appear as (as $l_2 - l_3 = 0$) and $J1(1)$ appears once (as $l_1 - l_2 = 1$). Thus, the required result follows.

Theorem 11.3.2. [A is similar to A^T] *Let $A \in \mathbb{M}_n(\mathbb{C})$. Then, A is similar to A^T.*

Proof. Let

$$K_n = \begin{bmatrix} & & 1 \\ & \cdot^{\cdot^{\cdot}} & \\ 1 & & \end{bmatrix}.$$

Then, observe that $K^{-1} = K$ and $KJ_n(a)K = Jn(a)^T$, as the (i, j)-th entry of A goes to $(n - i + 1, n - j + 1)$-th position in KAK. Hence

$$\left[\bigoplus K_{n_i}\right]\left[\bigoplus J_{n_i}(\lambda_i)\right]\left[\bigoplus K_{n_i}\right] = \left[\bigoplus J_{n_i}(\lambda_i)\right]^T.$$

Thus, J is similar to J^T . But, A is similar to J and hence A^T is similar to J^T and finally we get A is similar to A^T . Therefore, the required result follows.

11.4 MINIMAL POLYNOMIAL

We start this section with the following definition. Recall that a polynomial $p(x) = a_0 + a_1x +$
$\cdots + a_nx^n$ with $a_n = 1$ is called a monic polynomial.

Definition 11.4.1. [Companion Matrix] Let $P(t) = t^n + a_{n-1}t^{n-1} + \cdots +$ a_0 be a monic
polynomial in t of degree n. Then, the $n \times n$ matrix

$$A = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & 0 & \cdots & 0 & -a_2 \\ 0 & 0 & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & -a_{n-2} \\ 0 & 0 & 0 & & 1 & -a_{n-1} \end{bmatrix};$$

denoted $A(n : a_0, . . . , a_{n-1})$ or Companion(P), is called the **companion matrix** of $P(t)$.

Definition 11.4.2. [Annihilating Polynomial] Let $A \in \mathbb{M}_n(\mathbb{C})$. Then, the polynomial $P(t)$ is said to **annihilate** (destroy) A if $P(A) = \mathbf{0}$.

Let $P(x)$ be the characteristic polynomial of A. Then, by the Cayley-Hamilton Theorem, $P(A) = \mathbf{0}$. So, if $f(x) = P(x)g(x)$, for any multiple of $g(x)$, then $f(A) = P(A)g(A) = \mathbf{0}g(A) = \mathbf{0}$. Thus, there are infinitely many polynomials which annihilate A. In this section, we will concentrate on a monic polynomial of least positive degree that annihilates A.

Definition 11.4.3. [Minimal polynomial] Let $A \in \mathbb{M}_n(\mathbb{C})$. Then, the **minimal polynomial**
of A, denoted $m_A(x)$, is a monic polynomial of least positive degree satisfying $m_A(A) = \mathbf{0}$.

Theorem 11.3.4. *Let A be the companion matrix of the monic polynomial* $P(t) = t^n + a_{n-1} t^{n-1} + \cdots + a_0$. *Then,* $P(t)$ *is both the characteristic and the minimal polynomial of A.*

Proof. Expanding $\det(tI_n - \text{Companion}(P))$ along the first row, we have

$$\det(tI_n - A(n : a_0,\ldots,a_{n-1})) = t\det(t I_{n-1} - A(n-1 : a_1,\ldots,a_{n-1})) + (-1)^{n+1} a_0(-1)^{n-1}$$

$$= t^2 \det(tI_{n-2} - A(n-2 : a_2,\ldots,a_{n-1})) + a_0 + a_1 t$$

$$\vdots$$

$$= P(t).$$

Thus, $P(t)$ is the characteristic polynomial of A and hence $P(A) = \mathbf{0}$. We will now show that $P(t)$ is the minimal polynomial of A. To do so, we first observe that $A\mathbf{e}_1 = \mathbf{e}_2,\ldots,A\mathbf{e}_{n-1} = \mathbf{e}_n$. That is,

$$A^k\mathbf{e}_1 = \mathbf{e}_{k+1}, \qquad \text{for } 1 \le k \le n-1. \qquad (1)$$

Now, Suppose we have a monic polynomial $Q(t) = t^m + b_{m-1} t^{m-1} + \cdots + b_0$, with $m < n$, such that $Q(A) = \mathbf{0}$. Then, using Equation (1), we get

$\mathbf{0} = Q(A)\mathbf{e}_1 = A^m\mathbf{e}_1 + b_{m-1} A^{m-1}\mathbf{e}_1 + \cdots + b_0 I\mathbf{e}_1 = \mathbf{e}_{m+1} + b_{m-1}\mathbf{e}_m + \cdots + b_0\mathbf{e}_1$, a contradiction to the linear independence of $\{\mathbf{e}_1,\ldots,\mathbf{e}_{m+1}\} \subseteq \{\mathbf{e}_1,\ldots,\mathbf{e}_n\}$.

The next result gives us the existence of such a polynomial for every matrix A. To do so, recall that the well-ordering principle implies that if S is a subset of natural numbers then it contains a least element.

Lemma 11.4.5. [Existence of the Minimal Polynomial] Let A ∈ Mn(C). Then, there exists a unique monic polynomial m(x) of minimum (positive) degree such that m(A) = **0**. Further, if f(x) is any polynomial with f(A) = **0** then m(x) divides f(x).

Proof. Let P (x) be the characteristic polynomial of A. Then, deg(P (x)) = n and by the Cayley-Hamilton Theorem, $P(A) = \mathbf{0}$. So, consider the set

$$S = \{\deg(f(x)) : f(x) \text{ is a nonzero polynomial, } f(A) = \mathbf{0}\}.$$

Then, S is a non-empty subset of N as $n \in S$. Thus, by well-ordering principle there exists a smallest positive integer, say M, and a corresponding polynomial, say $m(x)$, such that

$\deg(m(x)) = M$, $m(A) = \mathbf{0}$.

Also, without loss of generality, we can assume that $m(x)$ is monic and unique (nonuniqueness will lead to a polynomial of smaller degree in S).

Now, suppose there is a polynomial $f(x)$ such that $f(A) = \mathbf{0}$. Then, by division algorithm, there exist polynomials $q(x)$ and $r(x)$ such that $f(x) = m(x)q(x) + r(x)$, where either $r(x)$ is identically the zero polynomial of $\deg(r(x)) < M = \deg(m(x))$. As

$$\mathbf{0} = f(A) = m(A)q(A) + r(A) = \mathbf{0}q(A) + r(A) = r(A),$$

we get $r(A) = \mathbf{0}$. But, $m(x)$ was the least degree polynomial with $m(A) = \mathbf{0}$ and hence $r(x)$ is the zero polynomial. That is, $m(x)$ divides $f(x)$.

As an immediate corollary, we have the following result.

Corollary 11.4.6. [Minimal polynomial divides the Characteristic Polynomial] Let $m_A(x)$ and $P_A(x)$ be, respectively, the minimal and the characteristic polynomials of $A \in \mathbb{M}_n(\mathbb{C})$.

1. Then, $m_A(x)$ divides $P_A(x)$.
2. Further, if λ is an eigenvalue of A then $m_A(\lambda) = 0$.

Proof. The first part following directly from Lemma 11.3.5. For the second part, let $(\lambda, \mathbf{x})$ be an

eigen-pair. Then, $f(A)\mathbf{x} = f(\lambda)\mathbf{x}$, for any polynomial of f, implies that

$$m_A(\lambda)\mathbf{x} = m_A(A)\mathbf{x} = \mathbf{0}\mathbf{x} = \mathbf{0}.$$

But, $\mathbf{x} \neq \mathbf{0}$ and hence $m_A(\lambda) = 0$. Thus, the required result follows. We also have the following result.

Lemma 11.4.7. Let A and B be two similar matrices. Then, they have the same minimal polynomial.

Proof. Since A and B are similar, there exists an invertible matrix S such that $A = S^{-1} BS$. Hence, $f(A) = F(S^{-1} BS) = S^{-1} f(B)S$, for any polynomial f. Hence, $mA(A) = \mathbf{0}$ if and only if $m_A(B) = \mathbf{0}$ and thus the required result follows.

Theorem 11.3.8. *Let A* $\in \mathbb{M}_n(\mathbb{C})$ *and let λ1, . . . , λk be the distinct eigenvalues of A. If ni is the size of the largest Jordan block for λi in J =* Jordan CFA *then*

$$m_A(x) = \prod_{i=1}^{k}(x - \lambda_i)^{n_i}.$$

Proof. Using 11.3.6, we see that

$$m_A(x) = \prod_{i=1}^{k}(x-\lambda_i)^{\alpha_i},$$

for some αi's with $1 \leq \alpha i \leq$ Alg.MUL$\lambda_i(A)$. As $m_A(A) = \mathbf{0}$, using Lemma 11.3.7 we have

$$m_A(J) = \prod_{i=1}^{k}(J - \lambda_i I)^{\alpha_i} = \mathbf{0}.$$

But, observe that for the Jordan block $J_{n_i}(\lambda_i)$, one has

1. $(J_{n_i}(\lambda_i) - \lambda_i I)^{\alpha i} = \mathbf{0}$ if and only if ${}^{\alpha i} \geq n_i$, and

2. $(J_{n_m}(\lambda_m) - \lambda_i I)^{\alpha i}$ is invertible, for all $m \neq i$.

Thus

$$\prod_{i=1}^{k}(J - \lambda_i I)^{n_i} = \mathbf{0} \text{ and } \prod_{i=1}^{k}(x - \lambda_i)^{n_i} \text{ divides } \prod_{i=1}^{k}(x - \lambda_i)^{\alpha_i} = m_A(x) \text{ and } \prod_{i=1}^{k}(x - \lambda_i)^{n_i}$$

is a monic polynomial, the result follows. As an immediate consequence, we also have the following result which corresponds to the converse of the above theorem.

Theorem 11.4.9. Let A ∈ $\mathbb{M}_n(\mathbb{C})$ and let λ1, . . . , λk be the distinct eigenvalues of A. If the minimal polynomial of A equals $\prod_{i=1}^{k}(x - \lambda_i)^{ni}$ then n_i is the size of the largest Jordan block for λi in J = Jordan CFA.

Proof. It directly follows from Theorem 11.3.8.

We now give equivalent conditions for a square matrix to be diagonalizable.

Theorem 11.4.10. Let A ∈ $\mathbb{M}_n(\mathbb{C})$ Then, the following statements are equivalent.

1. A is diagonalizable.

2. Every zero of mA(x) has multiplicity 1.

3. Whenever mA(α) = 0, for some α, then $\frac{d}{dx} m_A(x)|_{x=\alpha} \neq 0$.

Proof. Part 1 ⇒ Part 2. If A is diagonalizable, then each Jordan block in J = Jordan CFA has size 1. Hence, by Theorem 11.3.8, $m_A(x) = \prod_{i=1}^{k} (x - \lambda i)$, where λi's are the distinct eigenvalues of A.

Part 2 ⇒ Part 3. Let $m_A(x) = \prod_{i=1}^{k} (x - \lambda i)$, where λi's are the distinct eigenvalues of A. Then, $mA(x) = 0$ if and only if $x = \lambda i$, for some i, $1 \leq i \leq k$. In that case, it is easy to verify that $\frac{d}{dx} m_A(x) \neq 0$, for each λi.

Part 3 ⇒ Part 1. Suppose that for each α satisfying $m_A(\alpha) = 0$, one has $\frac{d}{dx} m_A(\alpha) \neq 0$.

Then, it follows that each zero of $m_A(x)$ has multiplicity 1. Also, using Corollary 11.3.6, each zero of $m_A(x)$ is an eigenvalue of A and hence by Theorem 7.2.8, the size of each Jordan block is 1. Thus, A is diagonalizable.

We now have the following remarks and observations.

Remark 11.4.11. *1.* Let f(x) be a monic polynomial and A = Companion(f) be the companion matrix of f. Then, by Theorem 11.3.4) f(A) = **0** and no monic polynomial of smaller degree annihilates A. Thus PA(x) = mA(x) = f(x), where PA(x) is the characteristic polynomial and mA(x), the minimal polynomial of A.

2. Let A ∈ $\mathbb{M}$n($\mathbb{C}$). Then, A is similar to Companion(f), for some monic polynomial f if and only if $m_A(x) = f(x)$.

Proof. Let B = Companion (f). Then, using Lemma 11.3. 7, we see that $m_A(x) = m_B(x)$. But, by Remark 11.3.11.1, we get $m_B(x) = f(x)$ and hence the required result follows.

Conversely, assume that mA(x) = f(x). But, by Remark 11.3.11.1, $m_B(x) = f(x) = P_B(x)$, the characteristic polynomial of B. Since mA(x) = mB(x), the matrices A and B have the same largest Jordan blocks for each eigenvalue λ. As PB = mB, we know that for each λ, there is only one Jordan block in Jordan CFB. Thus, Jordan CFA = Jordan CFB and hence A is similar to Companion (f).

Check your progress

3. Define Jordan canonical form

4. Explain --Minimal polynomial divides the Characteristic Polynomial

11.5 LET'S SUM UP

The Jordan canonical form is convenient for computations. In particular, matrix powers and exponentials are straightforward to compute once the Jordan canonical form is known.

11.6 KEYWORDS

1. Linear Substitution--- The method of solving "by **substitution**" works by solving one of the equations (you choose which one) for one of the variables (you choose which one), and then plugging this back into the other equation, "**substituting**" for the chosen variable and solving for the other.

2. Canonical form – In **mathematics** and computer science, a **canonical**, normal, or standard **form** of a **mathematical** object is a standard way of presenting that object as a **mathematical** expression

3. well-ordering principle - The **well-ordering principle** is a property of the positive integers which is equivalent to the statement of the principle of mathematical induction.
4. Polynomial – is an expression consisting of variables (also called indeterminates) and coefficients, that involves only the operations of addition, subtraction, multiplication, and non-negative integer exponents of variables
5. Monic -- a **monic** polynomial is a single-variable polynomial (that is, a univariate polynomial) in which the leading coefficient (the nonzero coefficient of highest degree) is equal to 1.

11.7 QUESTION FOR REVIEW

1. Fix a positive integer k and a complex number λ. Then, prove that (a) Rank(Jk(λ) − λIk) = k − 1.

2. Let J be a Jordan matrix that contains ` Jordan blocks for λ. Then, prove that

(a) Rank(J − λI) = n − `.

3. *Convert* $\begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$ *to* $J3(0)$ *and* $\begin{bmatrix} 0 & 1 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ *to* $J2(0) \oplus J1(0)$.

4. Find A^{10}. Can you find a formula for A^k for any positive integer k, where

$$A = \begin{pmatrix} 5 & -1 \\ 9 & -1 \end{pmatrix}$$

11.8 SUGGESTED READINGS

- K. Hauffman and R. Kunz, Linear Algebra, Pearson Education (INDIA), 2003.
- G. Strang, Linear Algebra And Its Applications, 4th Edition, Brooks/Cole, 2006.
- S. Lang, Linear Algebra, Springer, 1989.
- David S. Dummit and Richard M. Foote, Abstract Algebra (3e), John Wiley and Sons.
- R. Gallian Joseph, Contemporary Abstract Algebra, Narosa Publishing House.
- Thomas Hungerford, Algebra, Springer GTM.
- I.N. Herstein, Topics in Abstract Algebra, Wiley Eastern Limited.
- D.S. Malik, J.M. Mordesen, M.K. Sen, Fundamentals of Abstract Algebra, The McGraw-Hill Companies, Inc.

11.9 ANSWER TO CHECK YOUR PROGRESS

1. Provide definition and example—11.1.2

2. Provide statement of theorem and proof – 11.1.8

3. Provide definition and explanation – 11.2.1

4. Provide statement of corollary and prrof– 11.3.6

UNIT 12: BILINEAR FORMS

STRUCTURE

12.0 OBJECTIVE

Understand the basic concept involved in Bilinear forms

Understand the Matrix Representation of Bilinear Forms

UNDERSTAND Change of Base and positive definite

Comprehend the Geometry Associated To A Positive Form and Bilinear forms over a complex vector space

12.1 INTRODUCTION

In mathematics, a **bilinear form** on a vector space V is a bilinear map $V \times V \to K$, where K is the field of scalars. In other words, a bilinear form is a function $B : V \times V \to K$ that is linear in each argument separately On a complex vector space, a bilinear form takes values in the complex numbers. In fact, a bilinear form can take values in any vector space, since the axioms make sense as long as vector addition and scalar multiplication are defined.

12.2 CONCEPTS

12.2.1Dot Product Definition : Let $X, Y \in \mathbb{R}^n$ then we define

$(X \cdot Y) = X^tY = x_1y_1 + x_2y_2 + \cdots + x_ny_n$.

The important features of the dot product are: Bilinearity

$((X_1 + X_2) \cdot Y) = (X_1 \cdot Y) + (X_2 \cdot Y)$

$(X \cdot (Y_1 + Y_2)) = (X \cdot Y_1) + (X \cdot Y_2)$

$(cX \cdot Y) = c(X \cdot Y) = (X \cdot cY)$

Symmetry $(X \cdot Y) = (Y \cdot X)$

Positivity $X \neq 0 \Rightarrow (X \cdot X) > 0$

Notice that bilinearity says that if we fix one element of the dot product then $(- \cdot Y) : \mathbb{R}^n \to \mathbb{R}$is a linear transformation. And it is this property which we will focus on first.

12.2.2 Bilinear Forms. Let V be a vector space over F. We define a bilinear form to be a function $f : V \times V \to F$ such that

$(\forall v_1, v_2, w \in V) f(v_1 + v_2, w) = f(v_1, w) + f(v_2, w)$

$(\forall v, w_1, w_2 \in V) f(v, w_1 + w_2) = f(v, w_1) + f(v, w_2)$

$(\forall v, w \in V, c \in F) f(cv, w) = cf(v, w) = f(v, cw)$

We will often use the notation $\langle v, w \rangle$ for $f(v, w)$. A typical example of a bilinear form is the dot product on $\mathbb{R}^n$. We shall usually write $\langle x, y \rangle$instead of f(x, y) for simplicity and we shall also identify each 1 × 1 matrix with its unique entry.

12.2.3 Symmetric Bilinear Form Definition: We say a bilinear form $\langle , \rangle$ is Symmetric if

$$(\forall v, w) \langle v, w \rangle = \langle w, v \rangle$$

12.2.4 Skew-Symmetric Bilinear Form Definition:

We say a bilinear form $\langle , \rangle$ is Skew-Symmetric if

$$(\forall v)\ \langle v, v \rangle = 0$$

Lemma 12.2.5 *Let V be a vector space over a field F of characteristic $\neq$ 2. Let $\langle , \rangle$ be a bilinear form on V . Then $\langle , \rangle$ is skew-symmetric if and only if*

$$(\forall v, w \in V)\ \langle v, w \rangle = -\langle w. v\rangle$$

Proof. $\Rightarrow$: Well we then know that

$$0\langle v + w. v + w \rangle = \langle v, v\rangle + \langle w, w\rangle + \langle v, w\rangle + \langle w, v\rangle$$
$$= \langle v. w\rangle + \langle w, v\rangle$$

So we have 0 = $\langle v, w\rangle + \langle w, v\rangle$ and hence $\langle v, w\rangle = -\langle w, v\rangle$

$\Leftarrow$: Well we then know that $\langle v, v\rangle = -\langle v, v\rangle$and so 2 $\langle v, v\rangle$= 0. So either 2 = 0 or $\langle v, v\rangle$= 0. But we are assuming that the characteristic isn't 2 and so we must have $\langle v, v\rangle$= 0.

It is for this last part that we need the characteristic isn't 2. In the case of characteristic 2 we see that the condition that $\langle v, w \rangle = -hw, vi$ is the same as saying that $\langle , \rangle$ is symmetric because $a = -a$ for all a. But, as we will see there are in general skew symmetric matrices over fields of characteristic 2 which are not symmetric.

Hence the above definition is the right one.

12.3 MATRIX REPRESENTATION OF BILINEAR FORMS

The most common examples of bilinear forms are those which act on the space F^n of column vectors as follows.

Let A be an $n \times n$ matrix. Then.

$$\langle X, Y\rangle = X^tAY$$

notice that this is a 1×1 matrix. The first thing we need to check is that this is in fact a bilinear form.

Lemma 12.3.1. Let V an n-dimensional vector space over F and let X, Y ∈ V be represented as column vectors relative to some basis. Further let A be an n × n matrix in F. Then hX, Y i = XtAY is a bilinear form.

Proof. We need to check the following.

• $< (X_1+X_2), Y >= (X_1+X_2)^t AY = (X_1^t+X_2^t)AY = (X_1^t\ AY)+(X_2^t\ AY) = \langle X_1, Y\rangle + \langle X_2, Y\rangle$

•$\langle X, (Y1+Y2)\rangle = X^t A(Y1+Y2) = (X^t AY1) + (X^t AY2) = \langle X, Y1\rangle + \langle X, Y2\rangle$

• $cX, Y = (cX^t AY) = (X^t AcY) = c\,\langle X, Y\rangle = \langle X, cY\rangle$

And so in fact *h, i* is a linear transform. Now, given a finite dimension vector space we want to show that any given bilinear form is of the above form.

Bilinear forms have matrixes

Lemma 12.3.2. Let ⟨, ⟩be a bilinear form on V , a finite dimensional vector space and lets let B = {b1,··· , bn} be a basis for V . Then there is a matrix A such that ⟨X, Y⟩= = X^tAY where X, Y are considered column vectors relative to the basis B.

Proof. We want to show that there is a matrix *A* such that

$$(\forall\{x_i, y_i : i \in n\} \subseteq F)\langle x_1, \cdots, x_n\rangle^t A\,\langle y_1, \cdots, y_n\rangle = \langle \Sigma_{\ i\in n} x_i b_i, \Sigma_{\ i\in n} y_i b_i\rangle$$

Well we know that any matrix we come up with will correspond to a bilinear form. So in particular if we can come up with a matrix which agrees with our bilinear form on the basis elements then the bilinear form associated to the matrix must be the one we want.

Specifically what we need is $A = (a_{i,j})$ where $a_{i,j} = \langle vi, vj\rangle$. Then by bilinearity *XtAY* = *hX, Y i* for all vectors *X, Y* .

Definition 12.3.3 . We say that $A = (\langle bi, bj\rangle)$ is the Matrix Associated to the Bilinear form ⟨, ⟩ relative to the basis *{b1,··· , bn}.*

Check your rporgress

1. Define Symmetric Bilinear Form & Skew Symmetric Bilinear Form

2. What do you understand by Bilinear forms have matrixes?

12.4 CHANGE OF BASE

One of the most important questions regarding these matrixes is what happens when we change bases. This leads us to the following theorem.

Theorem 12.4.1. Let A be the matrix associated to a bilinear form $\langle , \rangle$with respect to a basis. Then the matrixes which represent the same form with respect to different basis are those of the form

$$QAQ^t \text{ for some } Q \in GL_n(F).$$

Proof. Let P be the element of GLn(F) which represents the linear transformation which changes the base. So we have

$X^* = PX$, $Y^* = PY$ and hX, Y i = $\langle X^*, Y^* \rangle$ (as X, X^* and Y, Y^* are just different representation of the same vectors)

We then know that

$$\langle X, Y \rangle = X^t AY = (P^{-1} X^*)^t A(P^{-1} X Y^*) = (X^*)^t (P^{-1})^t AP^{-1} Y^*$$

But we also know that

$$\langle X, Y \rangle = \langle X^*, Y^* \rangle = (X^*)^t A^* Y^*$$

for A* the matrix representing the bilinear for relative to the new basis. Hence letting $Q = (P^{-1})^t$ we must have

$$A^* = QAQ^t.$$

12.4.2 Dot product matrix:

Now lets consider what happens to the dot product if we change basis. Recall that

$$(X \cdot Y) = X^t Y$$

And so we have that the matrix associated to the standard dot product is just the identity matrix.

12.4.3 Orthogonal.

Recall that a matrix is said to be orthogonal if

$P^t P = I$ or $P^{-1} = P^t$.

Lemma12.4.3.1 . *If you change base relative to an orthogonal change of base then the dot product is preserved.*

Proof. Let P be the orthogonal change of base. Well I is the matrix associated with the identity so by previous theorems this means that the matrix associated with the dot product under the new basis is

$(P^{-1})^t I (P^{-1}) = (P^t)^t I P^t = PP^t = I$

So changing the basis by an orthogonal matrix preserves the dot product. Similarly we have

Lemma 12.4.3.2 The matrixes which represent the dot product are those of the form PP^t for $P \in GLn(R)$.

Proof. By previous theorem.

Recall the three conditions on the dot product which were important. First off was Bilinearity. But this isn't a helpful as we know that X^tAY is bilinear for every A. The next is symmetry. This is in fact useful.

12.4.4 Definition Symmetric Matrix. We say a matrix is symmetric if $A = A^t$.

Lemma 12.4.4.1 A bilinear form is symmetric if and only if the matrix associated to it is symmetric.

Proof. Symmetry is equivalent to

$$\langle X, Y \rangle = X^tAY = Y^tAX = \langle Y, X \rangle$$

but we have $(Y^tAX) = (Y^tAX)^t = X^tA^tY$ because the transpose of a 1×1 matrix is itself.

Hence being a symmetric bilinear for is equivalent to

$$(\forall X, Y) X^tAY = X^tA^tY$$

and this is equivalent to $A = A^t$.

The third condition is that $(X \cdot X) > 0$ if $X \neq 0$ (Positivity).

12.5 POSITIVE DEFINITE:

Definition 12.5.1 : We call a bilinear form $\langle \, , \rangle$ on V Positive Definite if

$$(\forall v \in V, v \neq 0) \; \langle v, v \rangle > 0$$

12.5.2 Orthonormal basis: Given a bilinear form $\langle \, , \rangle$ on a vector space V we say that two vectors $v, w \in V$ are orthogonal ($v \perp w$) if $\langle v, w \rangle = 0$

Notes

Let $B = hv_1, \cdots, vn_i$ be a basis for V. We then say that B is an orthonormal basis if

$$(\forall i \neq j)\langle v_i, v_j\rangle = 0$$
$$(\forall i)\langle v_i, v_i\rangle = 1$$

Lemma 12.5.3. If B is an orthonormal basis for V with respect to $\langle\, ,\rangle$ then the matrix associated to $\langle\, ,\rangle$ relative to B is the identity. Proof. Immediate. Now we are going to show that for any positive definite bilinear form an orthonormal basis exists.

Orthonormal basis always exist for symmetric p

Theorem 12.5.4. Let $\langle\, ,\rangle$ be a positive definite symmetric bilinear form on a finite dimensional vector space V . Then there is an orthonormal basis for V

Proof. The method we are going to use is called the Gram-Schmidt procedure We are going to start with a basis $B = (v_1,\cdots, v_n)$

Step 1: The first step will be to normalize v1. Now we know that

$$\langle cv1, cv1\rangle = c^2\langle v1, v1\rangle$$

But, because we know that that $\langle\, ,\rangle$ is positive definite,

$$\langle v1, v1\rangle > 0$$

and so we know phv1, v1i is a real number and hence if we let

$$w1 = \sqrt{\langle v1, v1\rangle}\, v1$$

then we see that $\langle w1, w1\rangle > = 1$.

Step 2a: Now we want to look for a linear combination of $v2$, $w1$ which is orthogonal to $w1$. The value is

$$w = v2 - \langle v2, w1\rangle w1$$

because

$$\langle w, w1\rangle = \langle v2, w1\rangle - \langle v2, w1\rangle\langle w1, w1\rangle = 0$$

Step 2b: Then normalize w and call that vector $w2$. Further ($w1$, $w2$, $v3$, . . . , vn) is a basis for V. ···

Step k a: Suppose we have defined orthonormal vectors ($w1, \ldots, wk-1$) and that ($w1, \ldots, wk-1, vk, \ldots, vn$) is a basis.

Then we want to look for a linear combination of vk, $w1, \ldots, wk-1$ which is orthogonal to wi for all $i < k$. The value is

$w = v_k - \langle v_k, w_1\rangle w_1 + \cdots + \langle v_k, w_{k-1}\rangle w_{k-1}$

because

$\langle w, w_i\rangle = \langle v_k, w_i\rangle - \langle v_k, w_1\rangle \langle w_1, w_i\rangle + \cdots + \langle v_k, w_{k-1}\rangle \langle w_{k-1}, w_i\rangle$

But $\langle w_j, w_i\rangle = 0$ if $i \neq j$ and $\langle w_j, w_j\rangle = 1$ so we have

$\langle w, w_i\rangle = \langle v_k, w_i\rangle - \langle v_k, w_i\rangle \langle w_i, w_i\rangle = 0$

Step kb: Then normalize w and call that vector w_k. Further $v_k \in Span(w_1, \ldots, w_k, v_{k+1}, \ldots, v_n)$ and so $(w_1, \ldots, w_k, v_{k+1}, \ldots, v_n)$ is a basis. Hence after iterating this process n times we see that $(w_1, \ldots, w_n)$ is an orthonormal basis for V. We then have the following theorem.

12.5.5 . *The following are equivalent for a real $n \times n$ matrix*

(1) A represents the dot product.

(2) There is an invertible matrix $P \in GL_n(\mathbb{R})$ such that $A = P^tP$

(3) A is symmetric and positive definite.

Proof. We have already shown that (1) and (2) are equivalent. Further, the fact that (1) → (3) is by virtue of the fact that the dot product satisfies positivity and symmetry. So all that is left is to show that (3) → (2). Well the first thing to notice is that if A is positive definite then so is the form $\langle X, Y\rangle = X^tAY$. So in particular there is an orthonormal basis B' with respect to $\langle\, ,\rangle$. Now we know then that with respect to the basis B' the matrix associated to $\langle\, ,\rangle$ is I (because B' is orthonormal). But at the same time we know that if P is the matrix associated to the change of base from B' to the standard basis of $\mathbb{R}^n$ then $A = P^tA'P = P^tP$ and so A satisfies (2).

12.6 GEOMETRY ASSOCIATED TO A POSITIVE FORM

Suppose we have a bilinear form $\langle\, ,\rangle$ on a real vector space. Then, it is possible to define the length of a vector as follows.

Euclidian space Definition 12.6.1. Let $v \in V$ and let $\langle\,,\rangle$ be a positivedefinite bilinear form. Then we can define

$$|v| = \sqrt{(\langle v, v\rangle)}$$

We often call a real vector space with a length a Euclidian space

Lemma 12.6.2. *Let $\langle\,,\rangle$ be a positive definite bilinear form on a real vector space V. Then we have*

(Schwarz Inequality) $|\langle v, w\rangle| \leq |v| \cdot |w|$

(Triangle Inequality) $|v + w| \leq |v| + |w|$

Now given a subspace of W we have that **Restriction.**

Lemma 12.6.3. Let V be a vector space and $\langle\,,\rangle$ a bilinear form on V . Then if W ⊆ V is a subspace then $\langle\,,\rangle$ restricts to a bilinear form on h,i. Further if h,i is positive definite or symmetric on V then $\langle\,,\rangle$ is positive definite or symmetric on W

12.6.4 Inner product: An inner product on a real vector space V is a bilinear form which is both positive definite and symmetric.

12.6.5Angles and length: Suppose that h ,i is an inner product on a real vector space V . Then one may define the length of a vector v ∈ V by setting

$$||\boldsymbol{v}|| = \sqrt{\langle \boldsymbol{v}, \boldsymbol{v}\rangle}$$

and the angle θ between two vectors v, w ∈ V by setting

$$\cos\theta = \frac{\langle \boldsymbol{v}, \boldsymbol{w}\rangle}{||\boldsymbol{v}|| \cdot ||\boldsymbol{w}||}.$$

These formulas are known to hold for the inner product on $\mathbb{R}^n$.

$$\boldsymbol{w}_2 = \boldsymbol{v}_2 - \frac{\langle \boldsymbol{v}_2, \boldsymbol{w}_1\rangle}{\langle \boldsymbol{w}_1, \boldsymbol{w}_1\rangle}\boldsymbol{w}_1.$$

Then w1, w2 are orthogonal and have the same span as v1, v2. Proceeding by induction, suppose w1, w2, . . . , wk are orthogonal and have the same span as v1, v2, . . . , vk. Once we then define

$$\boldsymbol{w}_{k+1} = \boldsymbol{v}_{k+1} - \sum_{i=1}^{k} \frac{\langle \boldsymbol{v}_{k+1}, \boldsymbol{w}_i\rangle}{\langle \boldsymbol{w}_i, \boldsymbol{w}_i\rangle}\boldsymbol{w}_i,$$

we end up with vectors w1, w2, . . . , wk+1 which are orthogonal and have the same span as the original vectors v1, v2, . . . , vk+1. Using the formula from the last step repeatedly, one may thus obtain an orthogonal basis w1, w2, . . . , wn for the vector space V.

Example: We find an orthogonal basis of R3, starting with the basis

$$\boldsymbol{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \quad \boldsymbol{v}_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad \boldsymbol{v}_3 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

We define the first vector by w1 = v1 and the second vector by

$$\boldsymbol{w}_2 = \boldsymbol{v}_2 - \frac{\langle \boldsymbol{v}_2, \boldsymbol{w}_1 \rangle}{\langle \boldsymbol{w}_1, \boldsymbol{w}_1 \rangle} \boldsymbol{w}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \frac{2}{2} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}.$$

Then w1, w2 are orthogonal and we may define the third vector by

$$\begin{aligned} \boldsymbol{w}_3 &= \boldsymbol{v}_3 - \frac{\langle \boldsymbol{v}_3, \boldsymbol{w}_1 \rangle}{\langle \boldsymbol{w}_1, \boldsymbol{w}_1 \rangle} \boldsymbol{w}_1 - \frac{\langle \boldsymbol{v}_3, \boldsymbol{w}_2 \rangle}{\langle \boldsymbol{w}_2, \boldsymbol{w}_2 \rangle} \boldsymbol{w}_2 \\ &= \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} - \frac{4}{2} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} - \frac{2}{1} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}. \end{aligned}$$

12.7 BILINEAR FORMS OVER A COMPLEX VECTOR SPACE

Bilinear forms are defined on a complex vector space in the same way that they are defined on a real vector space. However, one needs to conjugate one of the variables to ensure positivity of the dot product. The complex transpose of a matrix is denoted by $A^* = \overline{A^t}$ and it is also known as the adjoint of A. One has $x^*x \geq 0$ for all $x \in \mathbb{C}^n$.

Notes

Bilinear forms on $\mathbb{R}^n$	Bilinear forms on $\mathbb{C}^n$
Linear in the first variable	Conjugate linear in the first variable
$\langle u+v, w\rangle = \langle u, w\rangle + \langle v, w\rangle$	$\langle u+v, w\rangle = \langle u, w\rangle + \langle v, w\rangle$
$\langle \lambda u, v\rangle = \lambda \langle u, v\rangle$	$\langle \lambda u, v\rangle = \overline{\lambda} \langle u, v\rangle$
Linear in the second variable	Linear in the second variable
$\langle x, y\rangle = x^t A y$ for some A	$\langle x, y\rangle = x^* A y$ for some A
Symmetric, if $A^t = A$	Hermitian, if $A^* = A$
Symmetric, if $a_{ij} = a_{ji}$	Hermitian, if $a_{ij} = \overline{a}_{ji}$

Check your progress

3. What is symmetric bilinear form?

4. Explain Orthonormal basis always exist for symmetric p

12.8 LET'S SUM UP

These theories also have interesting **applications** in classical coding theory. As mentioned above, the main difference lies in the structure of the underlying association schemes, which makes the analysis considerably more difficult in the case of symmetric **bilinear forms**.

12.9 KEYWORDS

1. Vector Space - A **vector space** (also called a **linear space**) is a collection of objects called **vectors**, which may be added together and multiplied ("scaled") by numbers, called *scalars*.

2. **Transpose matrix** - In linear algebra, the **transpose** of a matrix is an operator which flips a matrix over its diagonal, that is it switches the row and column indices of the matrix by producing another matrix
3. **Complex Matrix** : A square complex matrix whose transpose is equal to the matrix with every entry replaced by its complex conjugate (denoted here with an overline) is called a Hermitian matrix (equivalent to the matrix being equal to its conjugate transpose);
4. **Span -** In linear algebra, the linear **span** (also called the linear hull or just **span**) of a set S of **vectors** in a vector space is the smallest linear subspace that contains the set.

12.10 QUESTION FOR REVIEW

1. Prove that the sum of two bilinear forms is a bilinear form

2. Prove that the product of scalar and bilinear form is a bilinear form.

3. Explain Bilinear forms over a complex vector space

12.11 SUGGESTED READINGS

- K. Hauffman and R. Kunz, Linear Algebra, Pearson Education (INDIA), 2003.
- G. Strang, Linear Algebra And Its Applications, 4th Edition, Brooks/Cole, 2006.
- S. Lang, Linear Algebra, Springer, 1989.
- David S. Dummit and Richard M. Foote, Abstract Algebra (3e), John Wiley and Sons.
- R. Gallian Joseph, Contemporary Abstract Algebra, Narosa Publishing House.
- Thomas Hungerford, Algebra, Springer GTM.
- I.N. Herstein, Topics in Abstract Algebra, Wiley Eastern Limited.
- D.S. Malik, J.M. Mordesen, M.K. Sen, Fundamentals of Abstract Algebra, The McGraw-Hill Companies, Inc.

12.12 ANSWER TO CHECK YOUR PROGRESS

1. Provide definition and example—12.1.3 & 12.1.4

2. Provide statement of lemma and proof – 12.2.2

3. Provide statement of theorem and proof – 12.3.4.1

4. Provide statement of theorem and proof– 12.4.4

UNIT 13: ANNIHILATING POLYNOMIALS & DECOMPOSITION - I

STRUCTURE

13.0 OBJECTIVE

- Know about the polynomials over the field F, the degree of polynomial, monic polynomial, annihilating polynomials as well as minimal polynomials.
- Understand that the linear operator is annihilated by its characteristic polynomial.
- Understand that we consider all monic polynomials with coefficients in F and the degree
  of the minimal polynomial is the least positive integer such that a linear relation is
  obtained annihilated.

- Know the structure of the triangular form of a matrix of a linear operator T on a space V
  over the field F.
- Understand that we can diagonalize two or more commuting matrices simultaneously.
- Know that the matrix of a linear operator T commutes with that of a polynomial of a linear operator T.

13.1 INTRODUCTION

In this unit we investigate more properties of a linear transformation. We define certain terms like monic polynomial, minimal polynomial as well as annihilating polynomial and characteristic polynomial. It is seen that the theorem of Cayley-Hamilton in this unit helps us in narrowing down the reach for the minimal polynomials of various operators.In this unit we are again exploring the properties of a linear operator on the spaceV over the field F. In an upper triangular or lower triangular matrix the elements in the diagonal are the characteristic values.

13.2 CONCEPTS

13.2.1 Polynomial Over F: Let F(x) be the subspace of Fn spanned by vectors 1, x, x2..... An element of F(x) is called a polynomial over F.

13.2.2 Degree of a Polynomial: F(x) consists of all (finite) linear combinations of x and its powers. If *f* is a non-zero polynomial of the form

$$f = f_0x^0 + f_1x + f_2x^2 + \cdots + f_nx^n$$

such that $f_n \neq 0$ and $n \geq 0$ and $f_n = 0$ for all integers k > n; this integer is obviously unique and is called the degree of f.
The scalars $f_0, f_1, f_2, \ldots, f_n$ are sometimes called the coefficients of *f* in the field F.

13.2.3 Monic Polynomial: A polynomial *f* (x) over a field F is called monic polynomial if the coefficient of highest degree term in it is unity i.e $f_n = 0$

13.2.4 Annihilating Polynomials: Let A be n n matrix over a field F and f(x) be a polynomial over F. Then if *f*(A) = 0. Then we say that the polynomial *f*(x) annihilates the matrix A.

13.3 ANNIHILATING POLYNOMIALS

It is important to know the class of polynomials that Annihilate T. Suppose T is a linear operator on V, a vector space over the field F. If p is a polynomial over F, then p(T) is again a linear operator on V. If q is another polynomial over F, then

(p + q) (T) = p(T) + q(T)

(pq) (T) = p (T) q (T)

Therefore, the collection of polynomials p which annihilate T, in the sense that

$$p(T) = 0,$$

is an ideal in the polynomial algebra F[x]. It may be the zero ideal, i.e., it may be that T is not annihilated by any non-zero polynomial. But, that cannot happen if the space V is finite dimensional.

Suppose T is a linear operator on the n-dimensional space V. Look at the first (n2 + 1) powers of T:

$$I, T, T^2, T^{n2}$$

This is a sequence of $n^2 + 1$ operators in L(V, V), the space of linear operators on V. The space L(V, V,) has dimension n^2. Therefore, that sequence of $n^2 + 1$ operators must be linearly dependent. i.e., we have

$$c_0 I + c_1 T + \cdots + c_{n2} T^{n2} = 0$$

for some scalars c_i not all zero. So, the ideal of polynomials which annihilate T contains a nonzero polynomial of degree n^2 or less.

Definition 13.3.1 . Let T be a linear operator on a finite-dimensional vector space V over the field F. The minimal polynomial for T is the (unique) monic generator of the ideal of polynomials over F which annihilate T.

The name 'minimal polynomial' stems from the fact that generator of a polynomial ideal is characterized by being the monic polynomial of minimum degree in the ideal. That means that the minimal polynomial p for the linear operator T is uniquely determined by these three properties:

1. p is a monic polynomial over the scalar field F.
2. p(T) = 0
3. No polynomial over F which annihilates T has smaller degree than p has.

If A an $n \times n$ matrix over F, we define the minimal polynomial for A in an analogous way, as the unique monic generator of the ideal of all polynomials over F which annihilate A. If the operator T is represented in some ordered basis by the matrix A, then T and A have the same minimal polynomial. That is because f(T) is represented in the basis by the matrix f(A) so that f(T) = 0 if and only if f(A) = 0.

From the last remark about operators and matrices it follows that similar matrices have the Notes same minimal polynomial. That fact is also clear from the definitions because

$$f(P^{-1}AP) = P^{-1}f(A)P$$

for every polynomial f. There is another basic remark which we should make about minimal polynomials of matrices.

Suppose that A is an n n matrix with entries in the field F. Suppose that F1 is a field which contains F as a subfield. (For example, A might be a matrix with rational entries, while F1 is the field of real numbers. Or, A might be a matrix with real entries, while F1 is the field of complex numbers.) We may regard A either as an n n matrix over F or as an n n matrix over F1. On the surface, it might appear that we obtain two different minimal polynomials for A. Fortunately that is not the case; and we must see why. What is the definition of the minimal polynomial for A, regarded as an n n matrix over the field F? We consider all monic polynomials with coefficients in F which annihilate A, and we choose the one of least degree. If f is a monic polynomial over F:

$$f = x^k + \sum_{j=0}^{k-1} a_j x^i \qquad \dots (1)$$

then f(A) = 0 merely says that we have a linear relation between the powers of A:

$$A^k + a_{k-1}A^{k-1} + \cdots + a_1 A + a_0 I = 0 \qquad \dots (2)$$

The degree of the minimal polynomial is the least positive integer k such that there is a linear relation of the form (2) between the powers I, A, . ⋯ A^k Furthermore, by the uniqueness of the minimal polynomial, there is for that k one and only one relation of the form (2); i.e., once the minimal k is determined, there are unique scalars a a 0 1 , , ⋯ k in F such that (2) holds. They are the coefficients of minimal polynomial. Now (for each k) we have in (2) a system of n^2 linear equations for the 'unknowns'a_0 , , . ⋯a_{k-1} Since the entries of A lie in F, the coefficients of the system of equations (2) are in F. Therefore, if the system has a solution with a a 0 1 , , ⋯ k in F1 it has a solution with a a 0 1 , , ⋯ k in F. It should now be clear that the two minimal polynomials are the same. What do we know thus far about the minimal polynomial for a linearoperator on an n-dimensional space? Only that its degree does not exceed n2. That turns out to be a rather poor estimate, since the degree cannot exceed n. We shall prove shortly that the operator is annihilated by its characteristic polynomial. First, let us observe a more elementary fact.

Theorem 13.3.2: Let T be a linear operator on an n-dimensional vector space V [or, let A be an n n matrix]. The characteristic and minimal polynomials for T [for A] have the same roots, except for multiplicities.

Proof. Let p be the minimal polynomial for T. Let c be a scalar. What we want to show is that p(c) = 0 if and only if c is a characteristic value of T. First, suppose p(c) = 0. Then p = (x – c)q

where q is a polynomial. Since deg q < deg p, the definition of the minimal polynomial p tells us that q (T) ≠ 0. Choose a vector β such that q(T) β ≠ 0. Let α = q(T) β. Then

$$0 = p(T)\beta$$
$$= (T - cI)q(T)\beta$$
$$= (T - cI)\alpha$$

and thus, c is a characteristic value of T. Now, suppose that c is a characteristic value of T, say $T\alpha = c\alpha$ with 0. So

$$p(T)\alpha = p(c)\alpha \ .$$

Since $p(T) = 0$ and $\alpha \neq 0$, we have p(c) 0. Let T be a diagonalizable linear operator and let $c_1, \cdots c_k$ be the distinct characteristic values of T. Then it is easy to see that the minimal polynomial for T is the polynomial.

$$p = (x - c_1) \cdots (x - c_k).$$

If α is a characteristic vector, then one of the operators $T - c_1I, \cdots , T - c_kI$ sends α into 0. Therefore

$$(T - c_1I), \cdots , (T - c_kI)\ \alpha = 0$$

for every characteristic vector α. There is a basis for the underlying space which consists of characteristic vectors of T; hence

$$p(T) = (T - c_1I), \cdots , (T - c_kI) = 0$$

What we have concluded is this. If T is a diagonalizable linear operator, then the minimal polynomial for T is a product of distinct linear factors. As we shall soon see, that property characterizes diagonalizable operators.

Check your progress

1. Define degree of Polynomial & Monic Polynomial

2. State the properties of Minimal Polynomial

13.4 CAYLEY-HAMILTON THEOREM

13.4.1 Theorem: Let T be a linear operator on a finite dimensional vector space V. If *f* is the characteristic polynomial for T, then $f(T) = 0$;

in other words, the minimal polynomial divides the characteristic polynomial for T.

Proof: The proof, although short, may be difficult to understand. Aside from brevity, it has the virtue of providing an illuminating and far from trivial application of the general theory of determinants. Let K be the commutative ring with identity consisting of all polynomials in T. Of course, K is actually a commutative algebra with identity over the scalar field. Choose an ordered basis $\{\alpha_1, \ldots, \alpha_n\}$ for V, and let A be the matrix which represents T in the given basis. Then

$$T\alpha_i = \sum_{i=j}^{n} A_{ji}\alpha_j, \qquad 1 \le j \le n$$

These equations may be written in the equivalent form

$$\sum_{j=1}^{n} (\delta_{ij}T - A_{ji}I)\alpha_j = 0, \qquad 1 \le i \le n.$$

Let B denote the element of $K^{n\times n}$ with entries

$$B_{ij} = \delta_{ij}T - A_{ji}I.$$

When n = 2

$$B = \begin{bmatrix} T - A_{11}I & -A_{21}I \\ -A_{12}I & T - A_{22}I \end{bmatrix}$$

and

$$\begin{aligned} \det B &= (T - A_{11}I)(T - A_{22}I) - A_{12}A_{21}I \\ &= T^2 - (A_{11} + A_{22})T + (A_{11}A_{12} - A_{12}A_{21})I \\ &= f(T) \end{aligned}$$

since f is the determinant of the matrix xI – A whose entries are the polynomials ( ) .

$$(xI - A)_{ij} = \delta_{ij}x - A_{ji}.$$

Notes

We wish to show that f(T) = 0. In order that f(T) be the zero operator, it is necessary and sufficient that $(\det B)\alpha_k = 0$ for k = 1,⋯ , n. By the definition of B, the vectors $\alpha_1, \cdots \alpha_n$ satisfy the equations

$$\sum_{j=1}^{n} B_{ij}\alpha_j = 0, \quad 1 \le i \le n. \qquad \ldots (3)$$

When n = 2, it is suggestive to write (3) in the form

$$\begin{bmatrix} T - A_{11}I & -A_{21}I \\ -A_{12}I & T - A_{22}I \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

In this case, the classical adjoint, adj B is the matrix

$$\tilde{B} = \begin{bmatrix} T - A_{22}I & A_{21}I \\ A_{12}I & T - A_{11}I \end{bmatrix}$$

And

$$\tilde{B}B = \begin{bmatrix} \det B & 0 \\ 0 & \det B \end{bmatrix}.$$

Hence, we have

$$(\det B)\begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} = (\tilde{B}B)\begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}$$

$$= \tilde{B}\left(B\begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}\right)$$

In the general case, let B ˜ = adj B. Then by (3)

$$\sum_{j=1}^{n} \tilde{B}_{ki} B_{ij} \alpha_j = 0$$

For each pair k, i, and summing on i, we have

$$0 = \sum_{i=1}^{n} \sum_{j=1}^{n} \tilde{B}_{ki} B_{ij} \alpha_i$$

$$= \sum_{j=1}^{n} \left(\sum_{i=1}^{n} \tilde{B}_{ki} B_{ij} \right) \alpha_j .$$

Now B ˜ B = (det B)I, so that

$$\sum_{i=1}^{n} \tilde{B}_{ki} B_{ij} = \delta_{kj} \det B.$$

Therefore

$$0 = \sum_{j=1}^{n} \delta_{ki} (\det B) \alpha_j$$

$$= (\det B)_{\alpha k,} \qquad 1 \leq k \leq n.$$

The Cayley-Hamilton theorem is useful to us at this point primarily because it narrows down the search for the minimal polynomials of various operators. If we know the matrix A which represents T in some ordered basis, then we can compute the characteristic polynomial f. We know that the minimal polynomial p divides f and that the two polynomials have the same roots. There is no method for computing precisely the roots of a polynomial (unless its degree is small); however, if f factors

$$f = (x - c_1)^{d1} \cdots (x - c_k)^{dk}, C_{11} \cdots, C_k, \text{distinct}, d_i \geq 1 \qquad .... \text{ (4)}$$

$$p = (x - c_1)^{r1} \cdots (x - c_k)^{rk}, \qquad 1 \leq r_j \leq d_j \qquad .... \text{ (5)}$$

That is all we can say in general. If f is the polynomial (4) and has degree n, then for every polynomial p as in (5) we can find an n n matrix which has f as its characteristic polynomial and p as its minimal polynomial. We shall not prove this now. But, we want to emphasize the fact that the knowledge that the characteristic polynomial has the form (4) tells us that the minimal polynomial has the form (5), and it tells us nothing else about p.

Example: Let A be the 4 × 4 (rational) matrix

$$A = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}$$

The powers of A are easy to compute

$$A^2 = \begin{bmatrix} 2 & 0 & 2 & 0 \\ 0 & 2 & 0 & 2 \\ 2 & 0 & 2 & 0 \\ 0 & 2 & 0 & 2 \end{bmatrix}$$

$$A^3 = \begin{bmatrix} 0 & 4 & 0 & 4 \\ 4 & 0 & 4 & 0 \\ 0 & 4 & 0 & 4 \\ 4 & 0 & 4 & 0 \end{bmatrix}$$

Thus $A^3 = 4A$, i.e., if $p = x^3 – 4x = x(x + 2) (x – 2)$, then $p(A) = 0$. The minimal polynomial for A must divide p. That minimal polynomial is obviously not of degree 1, since that would mean that A was a scalar multiple of the identity. Hence, the candidates for the minimal polynomial are: p, $x(x + 2)$, $x(x – 2)$, $x^2 – 4$. The three quadratic polynomials can be eliminated because it is obvious at a glance that $A^2 \neq - 2A$, $A^2 \neq 2A$, $A^2 \neq 4I$. Therefore p is the minimal polynomial for A. In particular 0, 2, and – 2 are the characteristic values of A. One of the factors x, x – 2, x + 2 must be repeated twice in the characteristic polynomial. Evidently, rank (A) = 2. Consequently there is a two-dimensional space of characteristic vectors associated with the characteristic value 0. From Theorem 2, it should now be clear that the characteristic polynomial is $x^2 (x^2 – 4)$ and that A is similar over the field of rational numbers to the matrix

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & -2 \end{bmatrix}.$$

Example: Verify Cayley-Hamilton's theorem for the linear transformationT represented by the matrix A.

$$A = \begin{bmatrix} 0 & 0 & 1 \\ 3 & 1 & 0 \\ -2 & 1 & 4 \end{bmatrix}$$

Solution: The characteristic polynomial is given by

$$|A - x\,I| = \begin{bmatrix} 0-x & 0 & 1 \\ 3 & 1-x & 0 \\ -2 & 1 & 4-x \end{bmatrix}$$

$$= -x[(1-x)(4-x)] + (3+2-2x)$$

$$= -x(4-5x+x^2)+5-2x$$

$$= -x^3+5x^2-6x+5=0$$

$$f(x) = x^3-5x^2+6x-5=0$$

Now

$$A^2 = \begin{bmatrix} 0 & 0 & 1 \\ 3 & 1 & 0 \\ -2 & 1 & 4 \end{bmatrix}\begin{bmatrix} 0 & 0 & 1 \\ 3 & 1 & 0 \\ -2 & 1 & 4 \end{bmatrix} = \begin{bmatrix} -2 & 1 & 4 \\ 3 & 1 & 3 \\ -5 & 5 & 14 \end{bmatrix}$$

$$A^3 = \begin{bmatrix} -2 & 1 & 4 \\ 3 & 1 & 3 \\ -5 & 5 & 14 \end{bmatrix}\begin{bmatrix} 0 & 0 & 1 \\ 3 & 1 & 0 \\ -2 & 1 & 4 \end{bmatrix} = \begin{bmatrix} -5 & 5 & 14 \\ -3 & 4 & 15 \\ -13 & 19 & 51 \end{bmatrix}$$

$$f(A) = A^3 - 5A^2 + 6A - 5I$$

$$= \begin{bmatrix} -5 & 5 & 14 \\ -3 & 4 & 15 \\ -13 & 19 & 51 \end{bmatrix} - \begin{bmatrix} -10 & 5 & 20 \\ 15 & 5 & 15 \\ -25 & 25 & 14 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 6 \\ 18 & 6 & 0 \\ -12 & 6 & 24 \end{bmatrix} - \begin{bmatrix} 5 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 5 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = 0$$

where 0 being null matrix. So f(A) = 0

$$A = \begin{bmatrix} 2 & -1 & 1 \\ -2 & 2 & -1 \\ 1 & -1 & 2 \end{bmatrix}$$

13.5 SIMULTANEOUS TRIANGULATION AND SIMULTANEOUS DIAGONALIZATION

Let V be a finite-dimensional space and let $\mathcal{F}$ be a family of linear operators on V. We ask when we can simultaneously triangulate or diagonalize the operators in $\mathcal{F}$, i.e., find one basis B such that all of the matrices [T]B, T in $\mathcal{F}$, are triangular (or diagonal). In the case of diagonalization, it is necessary that $\mathcal{F}$ be a commuting family of operators: UT = TU for all T, U in $\mathcal{F}$. That follows from the fact that all diagonal matrices commute. Of course, it is also necessary that each operator in $\mathcal{F}$ be a diagonalizable operator. In order to simultaneously triangulate, each operator in $\mathcal{F}$ must be triangulable. It is not necessary that $\mathcal{F}$ be a commuting family; however that condition is sufficient for simultaneous triangulation (if each T can be individually triangulated).

The subspace W is invariant under (the family of operators) $\mathcal{F}$ if W is invariant under each operator in $\mathcal{F}$.

13.5.1 Lemma: Let $\mathcal{F}$ be a commuting family of triangulable linear operator on V. Let W be a proper subspace of V which is invariant under $\mathcal{F}$. There exists a vector α in V such that

(a) α is not in W;
(b) for each T in $\mathcal{F}$, the vector $T\alpha$ is in the subspace spanned by α and W.

Proof: It is no loss of generality to assume that $\mathcal{F}$ contains only a finite number of operators, because of this observation. Let {T1,...,Tn) be a maximal linearly independent subset of $\mathcal{F}$, i.e., a basis for the subspace spanned by $\mathcal{F}$. If α is a vector such that (b) holds for each Ti, then (b) will hold for every operator which is a linear combination of T1,..., Tr. We can find a vector β1 (not in W) and a scalar c1 such that $(T_1 - c_1I)\beta_1$ is in W. Let V_1 be the collection of all vectors β in V such that $(T_1 - c_1I)\beta$ is in W. Then V_1 is a subspace of V which is properly larger than W. Furthermore, V_1 is invariant under $\mathcal{F}$, for this reason. If T commutes with T_1, then

$$(T_1 - c_1I)(T\beta) = T(T_1 - c_1I)\beta$$

If β is in V1, then (T1 – c1I)β is in W. Since W is invariant under each T in $\mathcal{F}$, we have $T(T_1 - c_1I)\beta$ in W, i.e., Tβ in V1, for all β in V1 and all T in $\mathcal{F}$. Now W is a proper subspace of V1. Let U2 be the linear operator on V1 obtained by restricting T2 to the subspace V1. The minimal polynomial for U2 divides the minimal polynomial for T2. We obtain a vector β2 in V1 (not in W) and a scalar c2 such that $(T_2 - c_2I)\ \beta_2$ is in W. Note that

(a) β_2 is not in W;
(b) $(T_1 - c_1I)\beta_2$ is in W;
(c) $(T_2 - c_2I)\beta_2$ is in W.

Let V 2 be the set of all vectors β in V1 such that $(T_2 - c_2I)\beta$ is in W. Then V2 is invariant under $\mathcal{F}$. Apply the restriction of T3 to V2. If we continue in this way, we shall reach a vector $\alpha = \beta_r$ (not in W) such that (Tj – cjI)α is in W, j = 1,..., r.

Theorem 13.5.2: Let V be a finite-dimensional vector space over the field F. Let $\mathcal{F}$ be a commuting family of triangulable linear operators on V. There exists an ordered basis for V such that every operator in Φ is represented by a triangular matrix in that basis.

Corollary 13.5.3: Let $\mathcal{F}$ be a commuting family of n × n matrices over an algebraically closed field F. There exists a non-singular n × n matrix P with entries in F such that $P^{-1}AP$ is upper-triangular, for every matrix A in $\mathcal{F}$.

Theorem 13.5.4: Let F be a commuting family of diagonalizable linear operators on the finitedimensional vector space V. There exists an ordered basis for V such that every operator in Φ is represented in that basis by a diagonal matrix.

Proof: We could prove this theorem by adapting the lemma before Theorem 1 to the diagonalizable case. However, at this point it is easier to proceed by induction on the dimension of V.

If dim V = 1, there is nothing to prove. Assume the theorem for vector spaces of dimension less than n, and let V be an n-dimensional space. Choose any T in $\mathcal{F}$ which is not a scalar multiple of the identity. Let c1,..., ck be the distinct characteristic values of T, and (for each i) let Wi be the null space of T – c_iI. Fix an index i. Then Wi is invariant under every operator which commutes
with T.

Let $\mathcal{F}$i be the family of linear operators on Wi obtained by restricting the operators in $\mathcal{F}$ to the (invariant) subspace Wi. Each operator in $\mathcal{F}$i is diagonalizable, because its minimal polynomial divides the minimal polynomial for the corresponding operator in $\mathcal{F}$. Since dim Wi < dim V, the operators in $\mathcal{F}$i can be simultaneously diagonalized. In other words, Wi has a basis Bi which consists of vectors which are simultaneously characteristic vectors for every operator in $\mathcal{F}$i.

Since T is diagonalizable, B = (B1,..., Bk) is a basis for V. That is the basis we seek.

If we consider finite dimensional vector space V over a complex field F, then there is a basis such that the matrix of the linear operator T is diagonal. This is due to the key fact that every complex polynomial of positive degree has a root. This tells us that every linear operator has at least one eigenvector.

From the theorem above we now have that every complex n × n matrix A is similar to an upper triangular matrix i.e. there is a matrix P, such that P–1 AP is upper triangular.

Equally we also state that for a linear operator T on a finite dimensional

$$A' = \left[\begin{array}{c|c} \lambda & * \\ \hline O & D \end{array}\right]$$

complex vector space V, there is a basis of V such that the matrix of T with respect to that basis is upper triangular.

Let V contain an eigenvector of A, call it v1. Let be its eigen value. We extend (v1) to a Basis = (v1, v2, …, vn) for V. There will be a matrix P for which the new matrix A = P^{-1} A P has the block form.

where D is an (n – 1) × (n – 1) matrix, is a 1 × 1 matrix of the restriction of T to W (v1). Here O denotes n – 1 zeros below in the first column. By induction on n, we may assume that there exists a matrix Q such that Q^{-1} D Q is upper triangular. If we denote Q1 by the relation

$$Q_1 = \left[\begin{array}{c|c} 1 & O \\ \hline O & Q \end{array}\right]$$

then

$$A'' = Q_1^{-1} A' Q_1 = \left[\begin{array}{c|c} \lambda & * \\ \hline O & Q^{-1}DQ \end{array}\right]$$

is the upper triangular and thus

$$A'' = (P\,Q_1)^{-1}\,A\,(P\,Q_1).$$

Knowing one vector v corresponding to the characteristic value we can find a linear operator P and then Q1 to find A" .

Check your progress

3. State the Cayley-HamiltonTheorem

---

4. Explain Simultaneous Triangulation and Simultaneous Diagonalization

---

13.6 LET'S SUM UP

In this unit certain terms related to linear operator T are defined, i.e., the monic polynomial,bannihilating polynomials, minimal polynomials as well as characteristic polynomials. With the help of Cayley-Hamilton theorem it becomes easier to search for the minimal polynomials of various operators. In this unit we are dealing with matrices that commute with each other. In a triangular matrix the main diagonal has the entries of the characteristic values and it has not zero entries in the upper part of the diagonal only or non-zero entries in the lower of the main diagonal. If two or more matrices commute then we can diagonalize them simultaneously.

13.7 KEYWORDS

1. Annihilating Polynomial: Annihilating polynomial f(x) over the field F is such that for a matrix
A of n n matrix over the field f(A) = 0, then we say that the polynomial annihilates the matrix.
If a linear operator T is represented by the matrix then f(T) = 0 gives us the annihilating polynomial
for the linear operator T.
2. Monic Polynomial: The monic polynomial is a polynomial f(x) whose coefficient of the highest
degree in it is unity.
3. Diagonalizable: Each operator in $\mathcal{F}$ i is diagonalizable, because its minimal polynomial divides
the minimal polynomial for the corresponding operator in $\mathcal{F}$.
4. Ordered Basis: There exists an ordered basis for V such that every operator in $\mathcal{F}$ is represented
by a triangular matrix in that basis.

13.8 QUESTION FOR REVIEW

1. Let A be the following 3 3 matrix over F;

$$A = \begin{bmatrix} 2 & -1 & 1 \\ -2 & 2 & -1 \\ 1 & -1 & 2 \end{bmatrix}$$

Find the characteristic polynomial for A and also the minimal polynomial for A.

2. Let A be the following 3 3 matrix over F;

$$A = \begin{bmatrix} 1 & 3 & 7 \\ 4 & 2 & 3 \\ 1 & 2 & 1 \end{bmatrix}$$

Find the characteristic polynomial for A and also find the minimal polynomial for A.

3. Find an invertible real matrix P such that P–1AP and P–1BP are both diagonal, where A and B are the real matrices

$$A = \begin{bmatrix} 1 & 2 \\ 0 & 2 \end{bmatrix}, \qquad B = \begin{bmatrix} 3 & -8 \\ 0 & -1 \end{bmatrix}$$

4.Let $\mathcal{F}$ be a commuting family of 3 × 3 complex matrices. How many linearly independent matrices can $\mathcal{F}$ contain? What about the n × n case?

13.9 SUGGESTED READINGS

- K. Hauffman and R. Kunz, Linear Algebra, Pearson Education (INDIA), 2003.
- G. Strang, Linear Algebra And Its Applications, 4th Edition, Brooks/Cole, 2006.
- S. Lang, Linear Algebra, Springer, 1989.

- David S. Dummit and Richard M. Foote, Abstract Algebra (3e), John Wiley and Sons.
- R. Gallian Joseph, Contemporary Abstract Algebra, Narosa Publishing House.
- Thomas Hungerford, Algebra, Springer GTM.
- I.N. Herstein, Topics in Abstract Algebra, Wiley Eastern Limited.
- D.S. Malik, J.M. Mordesen, M.K. Sen, Fundamentals of Abstract Algebra, The McGraw-Hill Companies, Inc.

13.10 ANSWER TO CHECK YOUR PROGRESS

1. Provide definition 13.1.2 & 13.1.3

2. Provide 3 properties of minimal polynomial – 13.2.1

3. Provide statement of theorem – 13.3.1

4. Provide statement of theorem and proof related to concept– 13.4.

UNIT 14: ANNIHILATING POLYNOMIALS & DECOMPOSITION - II

STRUCTURE

14.0 OBJECTIVE

- Understand the meanings of invariant subspaces as well as decomposition of a vector

space into the invariant direct sums of the independent subspaces.

Know the projection operators and their properties

See that there is less emphasis is on matrices and more attention is given to subspaces.

- See that the vector space V is decomposed as a direct sum of the invariant subspaces under some linear operator T.
- Understand that the linear operator induces a linear operator Ti on each invariant subspaces Wi by restriction.
- Know that if αi is the vector in the invariant subspace Wi then the vector αi in the finite

vector space V is obtained as a linear combinations of its projections αi in the subspace Wi.

- considering a linear operator T on a finite dimensional space the minimal polynomial for the linear operator is a product of a number of irreducible monic polynomials $p_i^{r_i}$ over the field F where ri are positive integers.
- Know that this structure of the minimal polynomial helps in decomposing the space V as
the direct sum of the invariant subspaces Wi.
- Understand that the general linear operatorT induces a linear operator Ti on Wi by restriction and the minimal polynomial for Ti is the irreducible $p_i^{r_i}$

14.1 INTRODUCTION

This unit and the next units are slightly more complicated than the other previous units.The ideas of invariant subspaces and their relations with the vector space V is obtained. The ideas of projection operators and their properties are introduced. These ideas will help in expressing the given linear operator T in terms of the direct sums of the operators T1j TK as seen in the next unit. Here the vector space is decomposed as the direct sum of the invariant subspaces Wi. The linear operator induces a linear operator Ti for each invariant subspaces Wi. The method of finding the projection operators and their properties is discussed. The direct sum decomposition of the vector space V for a linear operator T in terms of invariant subspaces. The general linear operator T induces a linear operator Ti on the invariant subspace, the minimal polynomial of Ti is the $p_i^{r_i}$ This structure of the induced linear operator helps in introducing the projection operators Ei. These projections associated with the primary decomposition of T, then are polynomials in T, and they commute each will an operator that commutes with T.

14.2 DIRECT-SUM DECOMPOSITIONS

14.1.1 Definition: Let W1,..., Wk be subspaces of the vector space V. We say that W1,..., Wkare independent if

$$\alpha_1 + ... + \alpha_k = 0, \qquad \alpha_i \text{ in } Wi$$

implies that each α_i is 0. For k = 2, the meaning of independence is {0} intersection, i.e., W_1 and W_2 are independent if and only if W1 ⋂ W2 = {0}. If k > 2, the independence of W1,..., Wk says much more than W1 ⋂ ... ⋂ Wk = {0}. It says that each Wj intersects the sum of the other subspaces Wi only in the zero vector.

The significance of independence is this. Let W = W1 + ... + Wk be the subspace spanned by W1,...,Wk. Each vector α in W can be expressed as a sum

$$\alpha = \alpha_1 + ... + \alpha_k, \qquad \alpha_i \text{ in } Wi.$$

If W1,..., Wk are independent, then that expression for α is unique; for if

$$\alpha = \beta_1 + ... + \beta_k, \qquad \beta_i \text{ in } W_i$$

then $0 = (\alpha_1 - \beta_1) + ... + (\alpha_k - \beta_k)$, hence $\alpha_i - \beta_i = 0$, i = 1,..., k. Thus, when W_1,..., W_k are independent, we can operate with the vectors in W as k-tuples (α1,..., αk), αi in Wi, in the same way as we operate with vectors in R^k as k-tuples of numbers.

14.2.2 Lemma: Let V be a finite-dimensional vector space. Let W1,..., Wk be subspaces of V and let W = W 1 + ... + Wk. The following are equivalent.

(a) W1,..., Wk are independent.

(b) For each j, 2 ≤ j ≤ k, we have

$$Wj \cap (W1 + ... + Wj-1) = \{0\}$$

(c) If Bi is an ordered basis for Wi, 1 ≤ i ≤ k, then the sequence B = (B1,..., Bk) is an ordered basis for W.

Proof: Assume (a). Let α be a vector in the intersection Wj ⋂ (W1 + ... + Wj–1). Then there are vectors α_1,..., α_{j-1} with α_i in W_i such that α = α1 + ... + α_{j-1}. Since

$$\alpha_1 + ... + \alpha_{j-1} + (-\alpha) + 0 + ... + 0 = 0$$

and since W1, ..., Wk are independent, it must be that $\alpha_1 = \alpha_2 = ... = \alpha_{j-1} = \alpha = 0$.

Now, let us observe that (b) implies (a). Suppose

$$0 = \alpha = \alpha_1 + ... + \alpha_k, \qquad \alpha_i \text{ in } Wi$$

Let j be the largest integer i such that $\alpha i \neq 0$. Then

$$0 = \alpha = \alpha_1 + ... + \alpha_j, \qquad \alpha j \neq 0.$$

Thus $\alpha_j = -\alpha_1 - ... - \alpha_{j-1}$ is a non-zero vector in $Wj \cap (W1 + ... + Wj-1)$.

Now that we know (a) and (b) are the same, let us see why (a) is equivalent to (c). Assume (a).

Let Bi be basis for Wi, $1 \leq i \leq k$, and let B = (B1,..., Bk). Any linear relation between the vectors in B will have the form

$$\beta 1 + ... + \beta k = 0$$

where βi is some linear combination of the vectors in Bi. Since W1,.., Wk are independent, each βi is 0. Since each Bi is independent, the relation we have between the vectors in B is the trivial relation. If any (and hence all) of the conditions of the last lemma hold, we say that the sum W = W1 + ... + Wk is direct or that W is the direct sum of W1,..., Wk and we write

$W = W1 \oplus \cdots \oplus Wk$

In the literature, the reader may find this direct sum referred to as an independent sum or the interior direct sum of W1,..., Wk.

Example 1: Let V be a finite-dimensional vector space over the field F and let $\{\alpha 1,..., \alpha n\}$ be any basis for V. If Wi is the one-dimensional subspace spanned by αi, then $V = W1 \oplus \cdots \oplus Wn$.

Example 2: Let n be a positive integer and F a subfield of the complex numbers, and let V be the space of all $n \times n$ matrices over F. Let W1 be the subspace of all symmetric matrices, i.e., matrices A such that $A^t = A$. Let W2 be the subspace of all skew-symmetric matrices, i.e., matrices A such that $A^t = -A$. Then $V = W1 \oplus W2$. If A is any matrix in V, the unique expression for A as a sum of matrices, one in W1 and the other in W2, is

$$A = A_1 + A_2$$

$$A_1 = \frac{1}{2}(A + A^t)$$

$$A_2 = \frac{1}{2}(A - A^t)$$

Example 3: Let T be any linear operator on a finite-dimensional space V. Let c1,.., ck be the
distinct characteristic values of T, and let Wi be the space of characteristic vectors associated withthe characteristic value c_i. Then W1,..., Wk are independent. In particular, if T is diagonalizable, then V = W1 ⊕ ⋯ ⊕ Wk.

14.2.3 Definition: If V is a vector space, a projection of V is a linear operator E on V such that $E^2 = E$. Suppose that E is a projection. Let R be the range of E and let N be the null space of E.

1. The vector β is in the range R if and only if Eβ = β. If β = Eα, then Eβ = E2α = Eα = β.

Conversely, if β = Eβ, then (of course) β is in the range of E.

2. V = R ⊕ N.

3. The unique expression for α as a sum of vectors in R and N is α = Eα + (α – Eα).

From (1), (2), (3) it is easy to see the following. If R and N are subspaces of V such that V = R ⊕ N, there is one and only one projection operator E which has range R and null space N. That operator is called the projection on R along N.

Any projection E is (trivially) diagonalizable. If {α1,..., αr} is a basis for R and {αr+1,..., αn} a basis for N, then the basis B = {α1,..., αn} diagonalizes E.

$$[E]_B = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}$$

where I is the r × r identity matrix. That should help explain some of the terminology connected with projections. The reader should look at various cases in the plane R^2 (or 3-space, R^3), to convince himself that the projection on R along N sends each vector into R by projecting it parallel to N.

Projections can be used to describe direct-sum decompositions of the space V. For, suppose V = W1 ⊕ ⋯ ⊕ Wk. For each j we shall define an operator Ej on V. Let α be in V, say α = α1 + ⋯+ αk with αi in Wi.

Define Ejα = αj. Then Ej is a well-defined rule. It is easy to see that Ej is linear, that the range of Ej is Wj, and that E2 j = Ej. The null space of Ej is the subspace

$$(W_1 + \cdots + W_{j-1} + W_{j+1} + \cdots + W_k)$$

for, the statement that E jα = 0 simply means αj = 0, i.e., that α is actually a sum of vectors from the spaces Wi with i ≠ j. In terms of the projection Ej we have

$$\alpha = E_1\alpha + \cdots + E_k\alpha$$

for each α in V. What (1) says is that

$$I = E1 + \cdots + Ek$$

Note also that if i ≠ j, then EiEj = 0, because the range of Ej is the subspace Wj which is contained in the null space of Ei. We shall now summarize our findings and state and prove a converse.

Theorem 14.2.4: If V = W1 ⊕ ⋯ ⊕ Wk, then there exist k linear operators E1,..., Ek on V such that

(i) each Ei is a projection ($E_1^2 = E_i$)

(ii) $E_iE_j = 0$, if i ≠ j;

(iii) I = E1 + ⋯ + Ek;

(iv) the range of Ei is Wi.

Conversely, if E1,..., Ek are k linear operators on V which satisfy conditions (i), (ii) and (iii), and if we let Wi be the range of Ei, then

$$V = Wi \oplus \cdots \oplus Wk.$$

Proof: We have only to prove the converse statement. Suppose E1,..., Ek are linear operators on V which satisfy the first three conditions, and let Wi be the range of Ei. Then certainly

$$V = W1 + \cdots + Wk;$$

for, by condition (iii) we have

$$\alpha = E1\alpha + \cdots + Ek\alpha$$

for each α in V, and $E_i\alpha$ is in Wi. This expression for α is unique, because if

$$\alpha = \alpha_1 + \cdots + \alpha_k$$

with α_i in Wi, say $\alpha_i = E_i\beta_i$, then using (i) and (ii) we have

$$E_j\alpha = \sum_{i=1}^{k} E_j\alpha_i$$

$$= \sum_{i=1}^{k} E_jE_i\beta_i$$

$$= E_j^2\beta_j$$

$$= E_j\beta_j$$

$$= \alpha_j$$

This shows that V is the direct sum of the W

Check your progress

1. What is direct sum decomposition?

---

2. Define Projection

14.3 INVARIANT DIRECT SUMS

Theorem 14.3.1: Let T be a linear operator on the space V, and W1,..., Wk and E1,..., Ek .Then a necessary and sufficient condition that each subspace Wi be invariant under T is that T commutes with each of the projections Ei, i.e.,

TEi = EiT, i = 1,..., k

Proof: Suppose T commutes with each Ei. Let α be in Wj. Then $E_j\alpha = \alpha$, and

Tα = T(Ejα)

= Ej(Tα)

which shows that Tα is in the range of Ej, i.e., that Wj is invariant under T. Assume now that each Wi is invariant under T. We shall show that $TE_j = E_jT$. Let α be any vector in V. Then

α = E1α + ... + Ekα

$$T\alpha = TE1\alpha + ... + TEk\alpha$$

Since Eiα is in Wi, which is invariant under T, we must have T(Eiα) = Eiβi for some vector βi. Then

$$E_jTE_i\alpha = E_jE_i\beta_i$$

$$= \begin{cases} 0, & \text{if } i \neq j \\ E_j\beta_j, & \text{if } i = j \end{cases}$$

$$E_jT\alpha = E_jTE_1\alpha + ... + E_jTE_k\alpha$$

$$= E_j\beta_j$$

$$= TEj\alpha$$

This holds for each α in V, so EjT = TEj.

We shall now describe a diagonalizable operator T in the language of invariant direct sum decompositions (projections which commute with T). This will be a great help to us in understanding some deeper decomposition theorems later. The description which we are about to give is rather complicated, in comparison to the matrix formulation or to the simple statement that the characteristic vectors of T span the underlying space. But, we should bear in mind that this is our first glimpse at a very effective method, by means of which various problems concerned with subspaces, bases, matrices, and the like can be reduced to algebraic calculations with linear operators. With a little experience, the efficiency and elegance of this method of reasoning should become apparent.

Theorem 14.3.2: Let T be a linear operator on a finite-dimensional space V. If T is diagonalizable and if c1,..., ck are the distinct characteristic values of T, then there exist linear operators E1,..., Ek on V such that

(i) $T = c_1E_1 + ... + c_kE_k$;

(ii) $I = 'E_1 + ... + E_k$;

(iii) $E_iE_j = 0$, $i \neq j$;

(iv) $E_1^2 = E_i$ (Ei is a projection);

(v) the range of Ei is the characteristic space for T associated with ci.

Conversely, if there exist k distinct scalars c1,..., ck and k non-zero linear operators E1,..., Ek which satisfy conditions (i), (ii), and (iii), then T is diagonalizable, c1,..., ck are the distinct characteristic values of T, and conditions (iv) and (v) are satisfied also.

Proof: Suppose that T is diagonalizable, with distinct characteristic values c1,..., ck. Let Wi be the space of characteristic vectors associated with the characteristic value ci. As we have seen, $V = W1 \oplus \ldots \oplus Wk$

Let E1,...,Ek be the projections associated with this decomposition. Then (ii), (iii), (iv) and (v) are satisfied. To verify (i), proceed as follows. For each α in V,

$$\alpha = E_1\alpha + \ldots + E_k\alpha$$

and so

$$T\alpha = TE_1\alpha + \ldots + TE_k\alpha$$
$$= c_1E_1\alpha + \ldots + c_kE_k\alpha$$

In other words, T = c1E1 + ... + ckEk.

Now suppose that we are given a linear operator T along with distinct scalars ci and non-zero operators Ei which satisfy (i), (ii) and (iii). Since EiEj = 0 when $i \neq j$, we multiply both sides of I = E1 + ... + Ek by E_i and obtain immediately $E_1^2 = E_i$. Multiplying $T = c_1E_1 + \ldots + c_kE_k$ by Ei, we then have $TE_i = c_iE_i$, which shows that any vector in the range of Ei is in the null space of (T –ciI). Since we have assumed that $Ei \neq 0$, this proves that there is a non-zero vector in the null space of (T – ciI), i.e., that ci is a characteristic value of T. Furthermore, the ci are all of the characteristic values of T; for, if c is any scalar, then

$$T - cI = (c_1 - c)E_1 + \ldots + (c_k - c)E_k$$

so if $(T - cI)\alpha = 0$, we must have $(c_i - c)E_i\alpha = 0$. If α is not the zero vector, then $Ei\, \alpha \neq 0$ for some i, so that for this i we have $c_i - c = 0$.

Certainly T is diagonalizable, since we have shown that every non-zero vector in the range of Ei is a characteristic vector of T, and the fact that $I = E_1 + \ldots + E_k$ shows that these characteristic vectors span V. All that remains to be demonstrated is that the null space of $(T - c_iI)$ is exactly the range of Ei. But this is clear, because if $T\alpha = c_i\alpha$, then

$$\sum_{j=1}^{k}(c_j - c_i)E_j\alpha = 0$$ Hence

$$(cj - ci)Ej\alpha = 0 \qquad \text{for each } j$$

and then

$$Ej\alpha = 0 \qquad j \neq I$$

Since $\alpha = E1\alpha + ... + Ek\alpha$, and $Ej\alpha = 0$ for $j \neq i$, we have $\alpha = Ei\alpha$, which proves that α is in the range of Ei. One part for a diagonalizable operator T, the scalars c1,..., ck and the operators E1,..., Ek are uniquely determined by conditions (i), (ii), (iii), the fact that the ci are distinct, and the fact that the Ei are non-zero. One of the pleasant features of the decomposition $T = c_1E_1 + ... + c_kE_k$ is that if g is any polynomial over the field F, then

$$g(T) = g(c1)E1 + ... + g(ck)Ek.$$

To see how it is proved one need only compute T^r for each positive integer r. For example,

$$T^2 = \sum_{i=1}^{k} c_i E_i \sum_{j=1}^{k} c_j E_j$$

$$= \sum_{i=1}^{k} \sum_{j=1}^{k} c_i c_j E_i E_j$$

$$= \sum_{i=1}^{k} c_i^2 E_i^2$$

$$= \sum_{i=1}^{k} c_i^2 E_i$$

g(A) is simply the diagonal matrix with diagonal entries $g(A_{11})$, ..., $g(A_{nn})$. We should like in particular to note what happens when one applies the Lagrange polynomials corresponding to the scalars c1,..., ck:

$$p_j = \prod_{i \neq j} \frac{(x - c_i)}{(c_j - c_i)}$$

We have $pj(ci) = \delta ij$, which means that

$$p_j(T) = \sum_{i=1}^{k} \delta_{ij} E_i$$

$$= E_j$$

Thus the projections Ej not only commute with T but are polynomials in T. Such calculations with polynomials in T can be used to give an alternative proof of Theorem 2 of unit 14, which characterized diagonalizable operators in terms of their minimal polynomials. The proof is entirely independent of our earlier proof.

If T is diagonalizable, $T = c_1E_1 + ... + c_kE_k$, then

$$g(T) = g(c1)E1 + ... + g(ck)Ek$$

for every polynomial g. Thus g(T) = 0 if and only if g(ci) = 0 for each i. In particular, the minimal polynomial for T is

$$p = (x - c_1) ... (x - c_k)$$

Now suppose T is a linear operator with minimal polynomial p = (x – c1) ... (x – ck), where c1,..., ck are distinct elements of the scalar field. We form the Lagrange polynomials

$$p_j = \prod_{i \neq j} \frac{(x - c_i)}{(c_j - c_i)}$$

So that pj(ci) = δij and for any polynomial g of degree less than or equal to (k – 1) we have

$$g = g(c_1)p_1 + ... + g(c_k)p_k$$

Taking g to be the scalar polynomial 1 and then the polynomial x, we have

$$\left.\begin{aligned} 1 &= p_1 + \cdots + p_k \\ x &= c_1p_1 + \cdots + c_kp_k \end{aligned}\right\} \quad ...(2)$$

You will note that the application to x may not be valid because k may be 1. But if k = 1, T is a scalar multiple of the identity and hence diagonalizable). Now let Ej = pj(T). From (2) we have

$$\left.\begin{aligned} I &= E_1 + \cdots + E_k \\ T &= c_1E_1 + \cdots + c_kE_k \end{aligned}\right\} \quad ...(3)$$

Observe that if $i \neq j$, then pi pj is divisible by the minimal polynomial p, because pi pj contains every (x – cr) as a factor. Thus

$$EiEj = 0, i \neq j \quad ...(4)$$

We must note one further thing, namely, that $Ei \neq 0$ for each i. This is because p is the minimal polynomial for T and so we cannot have $p_i(T) = 0$ since p_i has degree less than the degree of p. This last comment, together with (3), (4), and the fact that the c_i are distinct enables us to apply Theorem 2 to conclude that T is diagonalizable.

14.4 THE PRIMARY DECOMPOSITION THEOREM

Theorem 14.41 (Primary Decomposition Theorem): Let T be a linear operator on the finite dimensional vector space V over the field F. Let p be the minimal polynomial for T,

$$p = p_1^{r_1} \cdots p_k^{r}$$

where the pi are distinct irreducible monic polynomials over F and the r_i are positive integers. Let W_i be the null space of $p_i(T)^{r_i}$, i = 1,..., k. Then

(i) V = W1 ⊕ ... ⊕ Wk;

(ii) each Wi is invariant under T;

(iii) if Ti is the operator induced on Wi by T, then the minimal polynomial for Ti is $p_1^{r_1}$.

Proof: The idea of the proof is this. If the direct-sum decomposition (i) is valid, how can we get

hold of the projections E1,..., Ek associated with the decomposition? The projection Ei will be the identity on Wi and zero on the other Wj.

We shall find a polynomial hi such that hi(T) is the identity on Wi and is zero on the other Wj, and so that h1(T) + ... + hk(T) = I, etc.

For each i, let

$$f_i = \frac{p}{p_i^{r_i}} = \prod_{j \neq i} p_j^{r_j} .$$

Since p1,..., pk are distinct prime polynomials, the polynomials f1,..., fk are relatively prime. Thus there are polynomials g1,..., gk such that

$$\sum_{i=1}^{n} f_i g_i = 1$$

Note also that if i ≠ j, then fi fj is divisible by the polynomial p, because fi fj contains each pm rm as a factor. We shall show that the polynomials $h_i = f_i g_i$ behave in the manner described in the firstparagraph of the proof.

Let Ei = hi(T) = $f_i(T)g_i(T)$. Since $h_1 + ... + h_k = 1$ and p divides fi fj for i ≠ j, we have

$$E1 + ... + Ek = I$$

$$EiEj = 0, \qquad \text{if } i \neq j$$

Thus the Ei are projections which correspond to some direct sum decomposition of the space V.

We wish to show that the range of Ei is exactly the subspace Wi. It is clear that each vector in the range of Ei is in Wi, for if α is in the range of Ei, then α = Eiα and so

$$\begin{aligned} p_i(T)^{r_i}\alpha &= p_i(T)^{r_i}E_i\alpha \\ &= p_i(T)^{r_i}f_i(T)g_i(T)\alpha \\ &= 0 \end{aligned}$$

because $p^{r_i} f_i g_i$ is divisible by the minimal polynomial p. Conversely, suppose that α is in the null space of $p_i(T)^{r_i}$ If j ≠ i, then $f_j g_j$ is divisible by piri and so $f_j(T)g_j(T)\alpha = 0$, i.e., $E_j\alpha = 0$ for j ≠ i. But then it is immediate that $E_i\alpha = \alpha$, i.e., that α is in the range of Ei. This completes the proof of statement (i).

It is certainly clear that the subspaces Wi are invariant under T. If Ti is the operator induced on Wi by T, then evidently $p_i(T)^{r_i} = 0$, because by definition pi(T)ri is 0 on the subspace Wi. This shows that the minimal polynomial for Ti divides $p_i^{r_i}$. Conversely, let g be any polynomial such that g(Ti) = 0. Then g(T)fi(T) = 0. Thus gfi is divisible by the minimal polynomial p of T, i.e., $p_i^{r_i} f_i$ divides g f_i. It is easily seen that piri divides g. Hence the minimal polynomial for Ti is $p_i^{r_i}$.

Corollary 14.4.2: If E1,..., Ek are the projections associated with the primary decomposition of T, then each Ei is a polynomial in T, and accordingly if a linear operator U commutes with T then U commutes with each of the Ei, i.e., each subspace Wi is invariant under U.

In the notation of the proof of Theorem 1, let us take a look at the special case in which the minimal polynomial for T is a product of first degree polynomials, i.e., the case in which each pi

is of the form pi = x – ci. Now the range of Ei is the null space Wi of (T – ciI)ri. Let us put $D = c_1E_1 + ... + c_kE_k$. By Theorem 2 of unit 17, D is a diagonalizable operator which we shall call the

diagonalizable part of T. Let us look at the operator N = T – D. Now

Now

$$T = TE_1 + ... + TE_k$$

$$D = c_1E_1 + ... + c_kE_k$$

$$N = (T - c_1I)E_1 + ... + (T - c_kI)E_k$$

The reader should be familiar enough with projections by now so that he sees that

$$N^2 = (T - c_1I)^2E_1 + ... + (T - c_kI)^2E_k$$

and in general that

$$N^r = (T - c_1I)^rE_1 + ... + (T - c_kI)^rE_k$$

When $r \geq r_i$ for each i, we shall have Nr = 0, because the operator (T – ciI)r will then be 0 on the range of Ei.

Definition14.4.3: Let N be a linear operator on the vector space V. We say that N is nilpotent if there is some positive integer r such that Nr = 0.

Theorem 14.4.4: Let T be a linear operator on the finite-dimensional vector space V over the field F. Suppose that the minimal polynomial forT decomposes over F into a product of linear polynomials. Then there is a diagonalizable operator D on V and a nilpotent operator N on V such that

(i) T = D + N,

(ii) DN = ND

The diagonalizable operator D and the nilpotent operator N are uniquely determined by (i) and

(ii) and each of them is a polynomial in T.

Proof: We have just observed that we can write T = D + N where D is diagonalizable and N is nilpotent, and where D and N not only commute but are polynomials in T. Now suppose that we also have T = D' + N' where D' is diagonalizable, N' is nilpotent, and D'N' = N'D'. We shall prove that D = D' and N = N'.

Since D' and N' commute with one another and T = D' + N', we see that D' and N' commute with T. Thus D' and N' commute with any polynomial in T; hence they commute with D and with N.

Now we have

$$D + N = D' + N'$$

or

$$D - D' = N' - N$$

and all four of these operators commute with one another. Since D and D' are both diagonalizable and they commute, they are simultaneously diagonalizable, and D – D' is diagonalizable. Since N and N' are both nilpotent and they commute, the operator (N' – N) is nilpotent; for, using the fact that N and N' commute

$$(N' - N)^r = \sum_{j=0}^{r} \binom{r}{j} (N')^{r-j}(-N)^j$$

and so when r is sufficiently large every term in this expression for (N' – N)r will be 0. (Actually, a nilpotent operator on an n-dimensional space must have its nth power 0; if we take r = 2n above, that will be large enough. It then follows that r = n is large enough, but this is not obvious from the above expression.) Now D – D' is a diagonalizable operator which is also nilpotent. Such an operator is obviously the zero operator; for since it is nilpotent, the minimal polynomial for this operator is of the form xr for some r ≤ m; but then since the operator is diagonalizable, the minimal polynomial cannot have a repeated root; hence r = 1 and the minimal polynomial is simply x, which says the operator is 0. Thus we see that D = D' and N = N'.

Corollary 14..5 : Let V be a finite-dimensional vector space over an algebraically closed field F, e.g., the field of complex numbers. Then every linear operator T on V can be written as the sum of a diagonalizable operator D and a nilpotent operator N which commute. These operators D and N' are unique and each is a polynomial in T. From these results, one sees that the study of linear operators on vector spaces over an algebraically closed field is essentially reduced to the study of nilpotent operators. For vector spaces over non-algebraically closed fields, we still need to find some substitute for characteristic values and vectors. It is a very interesting fact that these two problems can be handled simultaneously and this is what we shall do in the next units.

In concluding this section, we should like to give examples, which illustrate some of the ideas of the primary decomposition theorem. We have chosen to give it at the end of the section since it deals with differential equations and thus is not purely linear algebra

Example: Prove that the matrix A

$$A = \begin{bmatrix} 1 & 1 & 1 \\ -1 & -1 & -1 \\ 1 & 1 & 0 \end{bmatrix}$$

is nilpotent. Find its index of nilpotency.

Proof:

$$A^2 = \begin{bmatrix} 1 & 1 & 1 \\ -1 & -1 & -1 \\ 1 & 1 & 0 \end{bmatrix}\begin{bmatrix} 1 & 1 & 1 \\ -1 & -1 & -1 \\ 1 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ -1 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$A^3 = \begin{bmatrix} 1 & 1 & 0 \\ -1 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} 1 & 1 & 1 \\ -1 & -1 & -1 \\ 1 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

So $A^3 = 0$. Hence A is nilpotent of the index of nilpotence 3. Notice that $A^2 \neq 0$. (matrix) Also the characteristic polynomial of A is $p(x) = x^3$.

Check your progress

3. Explain invariant Sum

---

4. What is nilpotent?

14.5 LET'S SUM UP

In this unit the importance is given to the ideas of invariant subspaces of a vector space V
for a linear operator T. The vector space V is decomposed into a set of linear invariant subspaces. The sum of the bases vectors of the invariant subspaces defines the basis vectors of the vector space V.
The primary decomposition theorem is based on the fact that the minimal polynomial of
the linear operator is the product of the irreducible. This helps in finding the projection operates which are polynomials in T. The direct decomposition of the vector space V in terms of the invariant subspaces helps in inducing linear operators Ti on these subspaces Wi.
The induced operator Ti on Wi by T has the minimal polynomial as well as due to the factorisation of the minimal polynomial of T.

14. 6 KEYWORDS

5. Skew-symmetric Matrices: Skew-symmetric matrices, i.e., matrices A such that At = –A
6. Subspaces: These subspaces will be taken as independent subspaces of the vector space V and after finding the independent basis of each independent subspace the ordered basis of the whole space will be constructed.
7. Projection Operator: The projection operator E has the property that E2 = E so its characteristic
values can be equal to 0 and unit.

8. Restriction: When the finite space V is decomposed into the direct sum of the invariant subspaces
the linear operator induces a linear operator by the process known as restriction.
9. The Lagrange Polynomials: Help us to find the projection operators for any linear operator T in
terms of the matrix representing T and its characteristic values.
10. Invariant Sub-spaces: If a vector α in V is such that α and Tα are in the subspace W of V then W
is invariant subspace of V over the field F.
11. Nilpotent Transformation: A nilpotent transformation N on the vector space V represented by
a matrix A is such that AK = 0 for some integer K and AK–1 ≠ 0. Here K is the index of nilpotency.
12. Projection Operators: The projection operator Ei acting on the vector αi gives Eαi = αi for the
subspace Wi and gives zero for other. Also E E i i 2 = and EiEj = 0 for i ≠ j

14.7 QUESTION FOR REVIEW

1. Let V be a finite dimensional vector space and W1 is any subspace of V. Prove that there is a subspace W_2 of V such that V = W1 ⊕ W2.

2. Let V be a finite dimensional vector space and let W1,... WK be subspaces of V such that $V = W_1 + W_2 + ... + W_k$ and dim V = dim W_1 + ... + W_K. Prove that $V = W_1 \oplus W_2 \oplus ... \oplus W_k$.

3. Let T be the diagonalizable linear operator on R3 which is represented by the matrix

$$A = \begin{bmatrix} 5 & -6 & -6 \\ -1 & 4 & 2 \\ 3 & -6 & -4 \end{bmatrix}$$

use the Lagrange polynomials to write the representing matrix A in the form A = E1 + 2E2, E

1 + E2 = I, E1E2 = 0. Where I is a unit matrix and 0 is zero matrix.

4. Show that the linear operator T on R^3 represented by the matrix

is nilpotent.

$$A = \begin{bmatrix} 1 & -1 & 1 \\ 2 & -2 & 2 \\ -1 & 1 & -1 \end{bmatrix}$$

14.8 SUGGESTED READINGS

- K. Hauffman and R. Kunz, Linear Algebra, Pearson Education (INDIA), 2003.
- G. Strang, Linear Algebra And Its Applications, 4th Edition, Brooks/Cole, 2006.
- S. Lang, Linear Algebra, Springer, 1989.
- David S. Dummit and Richard M. Foote, Abstract Algebra (3e), John Wiley and Sons.
- R. Gallian Joseph, Contemporary Abstract Algebra, Narosa Publishing House.
- Thomas Hungerford, Algebra, Springer GTM.
- I.N. Herstein, Topics in Abstract Algebra, Wiley Eastern Limited.
- D.S. Malik, J.M. Mordesen, M.K. Sen, Fundamentals of Abstract Algebra, The McGraw-Hill Companies, Inc.

14.9 ANSWER TO CHECK YOUR PROGRESS

1. Provide definition & example-- 14.1.1

2. Provide definition and explanation – 14.1.3

3. Provide statement of theorem & proof – 14.2.1

4. Provide definition-- 14.3.3.