Lecture notes on

LINEAR ALGEBRA

PS01CMTH24

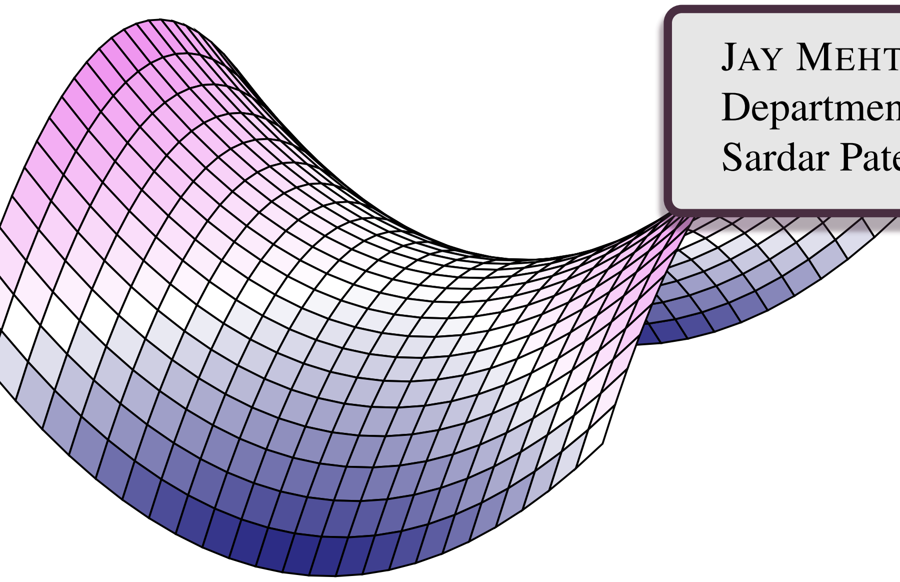
$U_1 \oplus U_2$

$V \cong \hat{\hat{V}}$

$A(V)$

$T(W) \subset W$

$\det(A)$



JAY MEHTA
Department of Mathematics,
Sardar Patel University.

SEMESTER - I
2018-19

Preface and Acknowledgments

These lecture notes are informal notes and is the outcome of the course given for M.Sc. Semester - I at Department of Mathematics, Sardar Patel University. These handouts are aimed to provide a reading material to the students in addition to other references and literature mentioned in the syllabus and to save time in taking notes during the classroom discussion. The notes are tailored for the (Linear Algebra) syllabus of M.Sc. Semester-I of the University and do not cover all the topics of Linear Algebra. Solutions to some of the exercises are not provided as they were given as seminar exercises and discussed in the seminar session during the semester. Students are advised not to rely completely on these notes and are also encouraged to cover the topics discussed during the class and seminar problem sessions.

The notes are prepared from the reference books and the lecture notes of previous years. We have followed the book Topics in Algebra by I. N. Herstein for almost all the topics except for the last section. However, we differ vastly from Herstein in notations and operations of linear maps. There may be a few errors in the notes. I welcome the readers to give their valuable suggestions, comments or point out the corrections, if they find any.

Acknowledgment.

I would like to thank Prof. A. B. Patel and Prof. P. A. Dabhi for their implicit and explicit help in preparing the notes. I also thank my students Asfak, Kashyap and others for pointing out the errors and typos in the notes.

Jay Mehta

Contents

SYLLABUS

PS01CMTH24: Linear Algebra

Unit I: Quick review of vector spaces, examples of sequence and function spaces. Linear spans; linear dependence/independence and basis, Examples of finite dimensional and infinite dimensional vector spaces, quotient space and its dimension. Dual space, dual basis, dimension of the annihilator. Solution of the system of simultaneous linear homogeneous equations.

Unit II: Definitions and examples of algebra, algebra analog of Cayley theorem. Minimal polynomial of a linear transformation. Rank of a linear transformation. Characteristic roots, characteristic vectors and results related to characteristic vectors. Matrix associated with a linear transformation on finite dimensional vector space. Isomorphism between the space of linear transformations and the space of matrices. Similarity of matrices and similarity of linear transformations.

Unit III: Relation of the minimal polynomials of a linear transformation and its induced linear transformation on a quotient space, triangular matrix associated to a linear transformation. Nilpotent linear transformation. Canonical matrix associated to a nilpotent linear transformation. Existence and uniqueness of invariants of a nilpotent linear transformation. Jordan form of a linear transformation.

Unit IV: Trace and its applications, Jocobsons lemma. Transpose of a matrix. Definition of the determinant of a matrix, determinant of a triangular matrix, a matrix with equal rows, a product of matrices. Application of determinant: regularity of a matrix, Cramers Rule to solve system of simultaneous non-homogeneous linear equations. Quadratic forms: diagonalization of a symmetric matrix. Symmetric matrix associated to a quadratic form, classification of quadratics.

Text Books

1. Herstein I. N., Topics in algebra, Wiley Eastern Ltd., New Delhi, (2nd Edition, 1975).
2. Kwak J. H., Hong S., Linear Algebra, (Second edition), Birkhauser.

Reference Books

1. Kumaresan, S., Linear Algebra: A Geometric Approach, Prentice Hall of India, 2000.
2. Simmons G. F., Introduction to Topology and Modern Analysis, McGraw-Hill Co., Tokyo, 1963.
3. Helson, H., Linear Algebra, (Second Edition), Hindustan Book Agency, TRIM-4, 1994.
4. Ramachandra Rao A. and Bhimasankaram P., Linear Algebra (Second Edition), Hindustan Book Agency, TRIM.

CHAPTER 1

VECTOR SPACES

1.1 Vector Spaces

Definition 1.1.1. Let V be a non empty set and F be a field. Suppose there exists two opeartions: (1) **vector addition**, which combines two elements of V denoted by '$+'$ and (2) **scalar multiplication**, which combines an each element of F with an element of V denoted by '$\cdot$' or simply by juxtaposition. Then V, along with the two operations, is called a vector space over the field F if the following properties hold:

(A1) V is closed under addition, $u+v \in V$ for all $u,v \in V$.
(A2) Addition is associative, $u+(v+w) = (u+v)+w$ for all $u,v,w \in V$.
(A3) Addition is commutative, $u+v = v+u$ for all $u,v \in V$.
(A4) There is an element $0 \in V$, called the zero vector, such that $u+0 = u$ for all $u \in V$.
(A5) For each element $u \in V$ there exists an element $-u \in V$, called the inverse of u, such that $u+(-u) = 0$.
(M1) Closure property for scalar multiplication, $\alpha u \in V$ for all $\alpha \in F$ and $u \in V$.
(M2) $\alpha(\beta u) = (\alpha\beta)u$ for all $\alpha,\beta \in F$ and $u \in V$.
(M3) Vector addition is distributive, $\alpha(u+v) = \alpha u+\alpha v$ for all $\alpha \in F$ and $u,v \in V$.
(M4) Distributivity over scalar addition, $(\alpha+\beta)u = \alpha u+\beta u$ for all $\alpha,\beta \in F$ and $u \in V$.
(M5) Unit scalar, $1u = u$ for all $u \in V$.

The elements, no matter what they might really be, of V are called *vectors* and the elements of the field F are called *scalars*. Obeserve that properties (A1)-(A5) shows that $(V,+)$ is an abelian group. Another important observation to make here is that the vector space V is a composite object consisting of a non empty set V along with the two operations defined above and a field F of scalars. The same underlying set V may be a part of a number of distinct vector spaces by considering a different field F (see Example 1.1.3 and Example 1.1.5) or altering the operations (see Example 1.1.3 and Example 1.1.7). When there is no ambiguity or scope of any confusion, we may just state that V is a vector space or else we shall say that V is a vector space over the field F thereby specifying the field.

1.1.1 Examples of vector spaces

In this section we give a variety of examples to be comfortable with the notion of vector space.

Example 1.1.2. The trivial vector space $V = \{0\}$ for any field F. One may define the singleton vector space $V = \{z\}$ for some element z with the following operations:

$$z + z = z \text{ and } \alpha z = z.$$

Here, the zero element is also z since V being singleton. So, z behaves like a zero vector is a singleton vector space.

Example 1.1.3. Question: Is a field F also a vector space over itself? The answer is Yes. In general, the set of n-tuples, $F^{(n)} = \{(v_1, \ldots, v_n) : v_i \in F,\ 1 \le i \le n\}$ is a vector space over the field F with vector addition and scalar multiplication defined as follows:

$$\begin{aligned}(u_1, \ldots, u_n) + (v_1, \ldots, v_n) &:= (u_1 + v_1, \ldots, u_n + v_n);\\ \alpha(u_1, \ldots, u_n) &:= (\alpha u_1, \ldots, \alpha u_n).\end{aligned}$$

In particular, for $n \in \mathbb{N}$, $\mathbb{R}^n$ is a vector over $\mathbb{R}$ of reals and $\mathbb{C}^n$ is a vector space over the field $\mathbb{C}$ of complex numbers.

Example 1.1.4. The set of infinite sequences of elements of a field F is a vector space over the field F. More precisely, $\mathscr{S}(F) = \{(x(n))_{n\in\mathbb{N}} : x(n) \in F\}$ is a vector space over the field F with vector addition and scalar multiplication defined as follows:

$$\begin{aligned}(x(1), x(2), \ldots) + (y(1), y(2), \ldots) &:= (x(1) + y(1), x(2) + y(2), \ldots);\\ \alpha(x(1), x(2), \ldots) &:= (\alpha x(1), \alpha x(2), \ldots).\end{aligned}$$

In particular, $\mathbb{C}^\infty$ and $\mathbb{R}^\infty$ are vector spaces over $\mathbb{C}$ and $\mathbb{R}$ respectively.

Example 1.1.5. Let F be a field and K be a subfield of F. Then F^n is a vector space over the field K with vector addition and scalar multiplication defined as in Example 1.1.3. In particular, $\mathbb{C}^n$ is a vector space over $\mathbb{R}$. Note that this is different from the one defined in Example 1.1.3.

Example 1.1.6. The vector space of matrices: Let F be a field and m,n be positive integers. Then the set $M_{m,n}(F)$ of all $m \times n$ matrices with entries from F is a vector space over F with usual matrix (entrywise) addition and scalar multiplication, i.e.

$$(A+B)_{ij} := A_{ij} + B_{ij} \text{ and } (\alpha A)_{ij} := \alpha A_{ij}.$$

Example 1.1.7. Let $V = \{(x,y) : x,y \in \mathbb{C}\}$ is a vector space over $\mathbb{C}$ with the following opeations:

$$\begin{aligned}\text{Vector addition:}\quad (x_1,y_1)+(x_2,y_2) &:= (x_1+x_2+1, y_1+y_2+1);\\ \text{Scalar multiplication:}\quad \alpha(x,y) &:= (\alpha x+\alpha-1, \alpha y+\alpha-1).\end{aligned}$$

The zero vector here is $\mathbf{0} = (-1,-1)$ and the additive inverse of the element (x,y) is $(-x-2, -y-2)$. Note that the underlying set V here is $\mathbb{C}^2$. However, it is different from the vector space $\mathbb{C}^2$ over $\mathbb{C}$ considered in Example 1.1.3 due to the operations defined differently.

Example 1.1.8. Let F be a field. Then the set $F[x]$ of all polynomials in x over F is a vector space with usual operations (addition of two polynomials and multiplication of a polynomial by an element of F).

Example 1.1.9. The vector space of polynomials, $F_n[x]$: Let $F_n[x]$ be the set of all polynomials of degree less than or equal to n in the variable x with coefficients from the field F (say for example $\mathbb{R}$ or $\mathbb{C}$). Then $F_n[x]$ is a vector space over F with vector addition and scalar multiplication defined as follows:

$$\begin{aligned}(a_0+a_1x+\cdots+a_nx^n)+(b_0+b_1x+\cdots+b_nx^n) &:= (a_0+b_0)+(a_1+b_1)x+\cdots+\\ &\quad (a_n+b_n)x^n;\\ \alpha(a_0+a_1x+\cdots+a_nx^n) &:= (\alpha a_0)+(\alpha a_1)x+\cdots+(\alpha a_n)x^n,\end{aligned}$$

where $\alpha, a_i \in F$, $1 \le i \le n$.

There is a relation between Example 1.1.3 and Example 1.1.9. Every element of $F_{n-1}[x]$ is of the form $a_0+a_1x+\cdots+a_{n-1}x^{n-1}$, where $a_i \in F$, $0 \le i \le n-1$. If we map this element onto an element $(a_0,a_1,\ldots,a_{n-1})$ of $F^{(n)}$, then we have one-one correspondence between $F_{n-1}[x]$ and $F^{(n)}$. Once homomorphism and isomorphism are defined, we can expect that they are isomorphic.

Observe that $F_n[x]$ is a subset of $F[x]$ and it is also a vector space over F under the same operations as of $F[x]$. $F_n[x]$ is said to be a subspace of $F[x]$. Before we give more examples we define subspace of a vector space.

Definition 1.1.10. Let V be a vector space over F. A subset W of V is called a subspace of V if itself forms a vector space over F under the operations of V.

Exercise 1.1.11. Show that W is a subspace of V if and only if for every $w_1,w_2 \in W$ and $\alpha,\beta \in F$, $\alpha w_1+\beta w_2 \in W$, i.e., $w_1+w_2 \in W$ and $\alpha w_1 \in W$ if and only if $\alpha w_1+\beta w_2 \in W$.

Now, we consider some examples which are subspaces of the vector space $\mathscr{S}(K)$ where the field K is either $\mathbb{R}$ or $\mathbb{C}$.

Example 1.1.12. Vector space of summable sequences. Let ℓ^1 be the set of all sequences $(x(n))_{n\in\mathbb{N}}$ over $\mathbb{R}$ (or $\mathbb{C}$) such that $\sum_{n=1}^{\infty}|x(n)|<\infty$. Then ℓ^1 is a vector space over $\mathbb{R}$ (or $\mathbb{C}$).
Hint: Using triangular inequality.

Example 1.1.13. Vector space of pth power summable sequences. Let ℓ^p is the set of all sequences $(x(n))_{n\in\mathbb{N}}$ over $\mathbb{R}$ (or $\mathbb{C}$) such that $\sum_{n=1}^{\infty}|x(n)|^p<\infty$. Then ℓ^p is a vector space over $\mathbb{R}$ (or $\mathbb{C}$).
Hint: $|x(n)+y(n)| \le |x(n)|+|y(n)| \le 2\max\{|x(n)|,|y(n)|\}$. Hence,

$$\begin{aligned}|x(n)+y(n)|^p \le 2^p(\max\{|x(n)|,|y(n)|\})^p &= 2^p\max\{|x(n)|^p,|y(n)|^p\}\\ &\le 2^p(|x(n)|^p+|y(n)|^p).\end{aligned}$$

One can also use Minkowski's inequality given below:

$$\left(\sum_{n=1}^{\infty}|x(n)+y(n)|^p\right)^{\frac{1}{p}} \le \left(\sum_{n=1}^{\infty}|x(n)|^p\right)^{\frac{1}{p}} + \left(\sum_{n=1}^{\infty}|y(n)|^p\right)^{\frac{1}{p}}.$$

Example 1.1.14. The vector space ℓ^∞ of all bounded sequences.

Example 1.1.15. The vector space $\mathbf{c}$ of all convergent sequences.

Note that if $x(n) \to a$ and $y(n) \to a$ then $x(n) + y(n) \to 2a$. The set of all sequences converging to a does not form a vector space unless $a = 0$. Thus, we have the following example:

Example 1.1.16. The vector space $\mathbf{c_0}$ of all convergent sequences whose limit is 0. $\mathbf{c_0}$ is a subspace of $\mathbf{c}$.

Example 1.1.17. Let c_{00} be the set of all sequences with finitely many non-zero terms. Thus, if $(x(n))_{n\in\mathbb{N}} \in c_{00}$ then there exists $k_0 > 0$ such that $x(n) = 0$ for all $n > k_0$, i.e.

$$(x(n))_{n\in\mathbb{N}} = (x(1),\ldots,x(k_0),0,0,\ldots).$$

Then c_{00} is a vector space over the field $K = \mathbb{R}$ or $\mathbb{C}$ under the same operations of $\mathscr{S}(K)$.

Observe that, among all the spaces of sequences we saw so far, we have the following relation:

$$c_{00} \subset \ell^p \subset c_0 \subset c \subset \ell^\infty \subset \mathscr{S}(K), \qquad 1 \le p < \infty$$

where $K = \mathbb{R}$ or $\mathbb{C}$. There is a relation between the spaces c_{00} and $K[x]$, where $K = \mathbb{R}$ or $\mathbb{C}$ (Example 1.1.17 and Example 1.1.8). For any element $(a(n)) = (a(1),\ldots,a(k),0,0,\ldots) \in c_{00}$ we get an element $p(x) = a(0) + a(1)x + \cdots + a(k)x^k \in K[x]$ and vice versa. Thus we have one to one correspondence between c_{00} and $K[x]$ and once we define isomorphism, we can see that they are isomorphic.

Example 1.1.18. Let X be a non-empty set and F be a field. Let $\mathscr{F}(X,F)$ denote the set of all functions $f : X \to F$. Then $\mathscr{F}(X,F)$ is a vector space over F with pointwise addition and scalar multiplication defined as follows:

$$(f+g)(x) := f(x) + g(x)$$
$$(\alpha f)(x) := \alpha f(x).$$

Example 1.1.19. Let X be a non-empty set and the field $K = \mathbb{R}$ or $\mathbb{C}$. Let $\mathscr{B}(X,K)$ denote the set of all bounded functions from X to K, i.e.

$$\mathscr{B}(X,K) = \{f : X \to K : \text{there exists } M > 0 \text{ such that } |f(x)| \le M, \text{ for all } x \in X\}.$$

Then $\mathscr{B}(X,K)$ is a vector space over K under the same operations as in $\mathscr{F}(X,K)$.

Example 1.1.20. Let X be a metric (or topological) space and $K = \mathbb{R}$ or $\mathbb{C}$. Then $C(X,K)$, the set of all continuous functions from X to K, is a vector space over K under the same operations as in $\mathscr{F}(X,K)$.

Example 1.1.21. Is it true that intersection of two subspaces of a vector space is also a subspace? The answer is Yes. Let $C_b(X,K)$ denote the set of all bounded continuous functions from X to K. Then $C_b(X,K)$ is a vector space as

$$C_b(X,K) = C(X,K) \cap \mathscr{B}(X,K).$$

If we assume X to be a compact metric (or topological) space and $f \in C(X,K)$ then $f(X)(\subset \mathbb{R}$ or $\mathbb{C})$ is compact as we know that continuous image of compact set is compact. By Heine-Borel theorem, $f(X)$ is bounded and hence in case of compact space, $C_b(X,K) = C(X,K)$.

Definition 1.1.22. Let $f \in C(X,K)$ (continuity of function is not necessary) for a metric (or topological) space X. Then f is said to be vanishing at infinity if for every $\varepsilon > 0$, there exists a compact subset Y of X such that

$$|f(x)| < \varepsilon \text{ for } x \in Y^c,$$

where $x \in Y^c$ denoted the complement of Y. In other words, for each positive ε, the set $\{x \in X : |f(x)| \geq \varepsilon\}$ is compact. The set of all continuous functions on X vanishing at infinity is denoted by $C_0(X)$.
For example, $X = K = \mathbb{R}$, $f(x) = e^{-|x|}$ vanishes at infinity. Also the function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = \frac{1}{1+x^2}$.

Definition 1.1.23. Let X be a metric (or a toplogical) space and $f \in C(X,K)$ for $K = \mathbb{R}$ or $\mathbb{C}$. The support of f denoted by supp(f) is the set of points in X where f is non-zero, i.e.,

$$\text{supp}(f) := \overline{\{x \in X : f(x) \neq 0\}}.$$

A function $f : X \to K$ is said to have compact support if there exists a compact subset Y of X such that

$$f(x) = 0 \text{ for } x \in Y^c,$$

where $x \in Y^c$ denoted the complement of Y. In other words, the set $\{x \in X : f(x) \neq 0\}$ is compact. The set of all continuous functions on X with compact support is denoted by $C_c(X)$.
For example, $f : \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} 1 - x^2 & \text{if } |x| < 1 \\ 0 & \text{if } |x| \geq 1. \end{cases}$$

Example 1.1.24. $C_0(X)$ and $C_c(X)$ are vector spaces over the field K with the operations of $\mathscr{F}(X,K)$.
Hint: Let $f, g \in C_0(X)$. Given $\varepsilon > 0$ there exists compact subset Y_1 and Y_2 of X such that

$$|f(x)| < \frac{\varepsilon}{2} \text{ for all } x \in Y_1^c$$
$$|g(x)| < \frac{\varepsilon}{2} \text{ for all } x \in Y_2^c$$

Take $Y = Y_1 \cup Y_2$. Then

$$|(f+g)(x)| = |f(x) + g(x)| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon \text{ for all } x \in Y^c.$$

Thus, $f + g \in C_0(X)$. Similarly, one can show that $\alpha f \in C_0(X)$ which concludes that $C_0(X)$ is a vector space. The proof of $C_c(X)$ is similar.

1.1.2 Properties of Vector Space

Now, we see some properties of vector spaces. We have the following lemma. The proof is easy and left as an exercise.

Lemma 1.1.25. *If V is a vector space over F then*

1. *The zero vector $0 \in V$ is unique.*
2. *Additive inverse is unique.*
3. *$\alpha 0 = 0$ for $\alpha \in F$.*
4. *$0v = 0$ for all $v \in V$.*
5. *$(-\alpha)v = -(\alpha v)$ for $\alpha \in F$, $v \in V$. In particular, for $\alpha = 1$ we have $-v = (-1)v$. Thus we get additive inverse from scalar multiplication.*
6. *If $\alpha v = 0$ then either $\alpha = 0$ or $v = 0$.*

Definition 1.1.26. Let U and V be vector spaces over F. A map $T : U \to V$ of U into V is said to be a *homomorphism* if for all $u_!, u_2 \in U$ and $\alpha \in F$

1. $T(u_1 + u_2) = T(u_1) + T(u_2)$;
2. $T(\alpha u_1) = \alpha T(u_1)$.

If a homomorphism $T : U \to V$ is one-one, we call it an *isomorphism*.

A vector space homomorphism is also called a linear map, linear transformation, linear operator or simply an operator. The set of all homomorphisms from U to V is denoted by Hom (U,V).

Definition 1.1.27. Two vector spaces U and V over the same field F are said to be *isomorphic* if there is an onto isomorphism from U to V. In other words, if there is a homomorphism $T : U \to V$ such that T is one-one and onto. We denote it by $U \cong V$.

Definition 1.1.28. Let U and V two vector spaces over the same field F. Let $T : U \to V$ be a homomorphism.

1. The *kernel* (or the null space) of T is defined as $\{u \in U : Tu = 0\}$ where 0 is the identity element of the addition in V. It is denoted by $\ker T$.
2. The range space of T is defined as $\{Tu : u \in U\}$ and is denoted by $R(T)$.

Exercise 1.1.29. Show that $\ker T$ and $R(T)$ are subspaces of U and V respectively.

Definition 1.1.30 (Quotient Space)**.** Let V be a vector space over F and let W be a subspace of V. Then W is a normal subgroup of V. Considering V and W as abelian groups we construct an abelian group V/W called the quotient group under the operation

$$(u+W) + (v+W) = (u+v) + W \quad (u, v \in V).$$

We want to make V/W a vector space over F. For this purpose we define scalar multiplication as follows:

$$\alpha(v+W) = \alpha v + W \quad (\alpha \in F, v+W \in V/W).$$

We must first check that the operation is well-defined, i.e., if $u+W = v+W$ then we must have $\alpha(u+W) = \alpha(v+W)$. Now,

$$\begin{aligned} & u+W = v+W \\ \Rightarrow\ & u-v \in W \\ \Rightarrow\ & \alpha(u-v) \in W \qquad \text{(since } W \text{ is a subspace)} \\ \Rightarrow\ & \alpha u - \alpha v \in W \\ \Rightarrow\ & \alpha u + W = \alpha v + W \end{aligned}$$

This ensures that the scalar multiplication is well-defined. The other vector-space axioms for V/W are easy to verify and left as an exercise. Thus, we have shown that V/W is a vector space called the *quotient space* of V by W. More precisely, we have the following lemma.

Lemma 1.1.31. *If V is a vector space over F and W is a subspace of V, then V/W is a vector space over F under the following operations:*

1. $(v_1+W)+(v_2+W) = (v_1+v_2)+W \quad (v_1+W, v_2+W \in V/W)$.
2. $\alpha(v_1+W) = \alpha v_1 + W \quad (v_1+W \in V/W,\ \alpha \in F)$.

Theorem 1.1.32 (First Homomorphism Theorem)**.** *If T is a homomorphism of U onto V with kernel W, then V is isomorphic to U/W. Conversely, if U is a vector space over F and W a subspace of U, then there is a homomorphism of U onto U/W.*

Proof. Since $W = \ker T$ is a subspace of U, U/W is a vector space over F. Define a map $S : U/W \to V$ as

$$S(u+W) = T(u) \qquad (u+W \in U/W). \tag{1.1}$$

First we show that S is a well-defined map. Let $u, v \in U$ such that

$$\begin{aligned} & u+W = v+W \\ \Rightarrow\ & u-v \in W \\ \Rightarrow\ & T(u-v) = 0 \\ \Rightarrow\ & Tu = Tv \qquad \text{(since } T \text{ is a homomorphism)} \\ \Rightarrow\ & S(u+W) = S(v+W) \end{aligned}$$

Next, we show that S is a homomorphism. For $u+W, v+W \in U/W$,

$$\begin{aligned} S((u+W)+(v+W)) &= S((u+v)+W) \\ &= T(u+v) && \text{(by definition of } S) \\ &= Tu + Tv && \text{(since } T \text{ is a homomorphism)} \\ &= S(u+W) + S(v+W) && \text{(by definition of } S) \end{aligned}$$

and

$$\begin{aligned} S(\alpha(u+W)) &= S(\alpha u + W) \\ &= T(\alpha u) \\ &= \alpha T(u) && (\because T \text{ is a homormophism}) \\ &= \alpha S(u+W). \end{aligned}$$

This shows that S is a homomorphism. Now we show that S is one-one. For this, let $u+W, v+W \in U/W$ such that

$$\begin{aligned} & S(u+W) = S(v+W) \\ \Rightarrow\ & Tu = Tv && \text{(by definition of } S\text{)} \\ \Rightarrow\ & T(u-v) = 0 && \text{(since } T \text{ is a homomorphism)} \\ \Rightarrow\ & u-v \in W (= \ker T) \\ \Rightarrow\ & (u-v)+W = W = 0+W \\ \Rightarrow\ & u+W = v+W \end{aligned}$$

Now, finally it remains to show that S is onto. Let $v \in V$. Since T is onto, there exists $u \in U$ such that $Tu = v$. But by definition of S, $Tu = S(u+W) = v$. This shows that S is onto and hence U/W is isomorphic to V.

Conversely, for a vector space U and its subspace W define a homomorphism $\phi : U \to U/W$ by $\phi(u) = u+W$ for $u \in U$. Then clearly ϕ is an onto homomorphism. □

Note: The homomorphism ϕ defined above is called the *canonical* or the natural homomorphism of U onto U/W.

Definition 1.1.33 (Internal Direct Sum)**.** Let V be a vector space over F and let $U_1, \ldots, U_n$ be subspaces of V. V is said to be the *internal direct sum* of $U_1, \ldots, U_n$ if every element $v \in V$ can be written uniquely as $v = u_1 + u_2 + \cdots + u_n$, where $u_i \in U_i$, $i = 1, 2, \ldots, n$. We write $V = U_1 \oplus U_2 \oplus \cdots \oplus U_n$.

Definition 1.1.34 (External Direct Sum)**.** Let $V_1, V_2, \ldots, V_n$ be vector spaces over F. Consider the set V of all ordered n-tuples $(v_1, v_2, \ldots, v_n)$ where $v_i \in V_i$, $i = 1, 2, \ldots, n$. We say that two elements $(v_1, v_2, \ldots, v_n)$ and $(v_1', v_2', \ldots, v_n')$ are equal if and only if for each i, $v_i = v_i'$. For $v = (v_1, v_2, \ldots, v_n)$, $v' = (v_1', v_2', \ldots, v_n') \in V$ and $\alpha \in F$, we define

$$v + v' := (v_1 + v_1', v_2 + v_2', \ldots, v_n + v_n')$$

and

$$\alpha(v_1, v_2, \ldots, v_n) := (\alpha v_1, \alpha v_2, \ldots, \alpha v_n).$$

Then V is a vector space over F with the operations defined above, called the *external direct sum* of vector spaces $V_1, V_2, \ldots, V_n$ and denoted by $V = V_1 \times V_2 \times \cdots \times V_n$.

Theorem 1.1.35. *If V is the internal direct sum of $U_1, \ldots, U_n$ then V is isomorphic to the external direct sum of $U_1, \ldots, U_n$.*

Proof. Since V is internal direct sum of $U_1, \ldots, U_n$, we write $V = U_1 \oplus \cdots \oplus U_n$. Then given $u \in V$, u can be uniquely written as $u = u_1 + \cdots + u_n$, $u_i \in U_i$, $1 \le i \le n$. Define a map $T : V \to U_1 \times U_2 \times \cdots \times U_n$ by

$$Tu = T(u_1 + u_2 + \cdots + u_n) = (u_1, u_2, \ldots, u_n).$$

Clearly, the map is well-defined. First we show that T is a homomorphism. Let $u, v \in V$. Then u and v can be uniquely written as $u = u_1 + \cdots + u_n$ and $v = v_1 + \cdots + v_n$. Now,

$$T(u+v) = T((u_1+v_1) \cdots + (u_n+v_n))$$

$$= (u_1 + v_1, \dots, u_n + v_n)$$
$$= (u_1, \dots, u_n) + (v_1, \dots, v_n) = Tu + Tv.$$

Also

$$\begin{aligned} T(\alpha u) = T(\alpha(u_1 + u_2 + \cdots + u_n)) &= T(\alpha u_1 + \alpha u_2 + \cdots + \alpha u_n) \\ &= (\alpha u_1, \alpha u_2, \dots, \alpha u_n) \\ &= \alpha(u_1, u_2, \dots, u_n) = \alpha T(u). \end{aligned}$$

Now, we show that T is one-one. Let $u, v \in V$ such that $Tu = Tv$. Then

$$\begin{aligned} & T(u_1 + \cdots + u_n) = T(v_1 + \cdots + v_n) \\ \Rightarrow\ & (u_1, \dots, u_n) = (v_1, \dots, v_n) \\ \Rightarrow\ & u_i = v_i \text{ for all } i = 1, 2, \dots, n \\ \Rightarrow\ & u_1 + \cdots + u_n = v_1 + \cdots + v_n \\ \Rightarrow\ & u = v. \end{aligned}$$

Clearly, T is onto because for $(u_1, \dots, u_n) \in U_1 \times \cdots \times U_n$, $u = u_1 + \cdots + u_n \in V$ such that $Tu = T(u_1 + \cdots + u_n) = (u_1, \dots, u_n)$. Thus T is an isomorphism of V onto $U_1 \times \cdots \times U_n$. □

Since internal direct sum and external direct sum are isomorphic as vector spaces, we will now refer to it as merely direct sum without specifying internal or external.

1.2 Linear Independence and Bases

Definition 1.2.1. Let V be a vector space over F and $v_1, v_2, \dots, v_n \in V$. Then an element of the form $\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n$, where $\alpha_i \in F, i = 1, 2, \dots, n$ is called a *linear combination* (over F) of $v_1, v_2, \dots, v_n$.

Definition 1.2.2. Let V be a vector space over F and S be a non empty subset of V. Then the set of all linear combinations of finite sets of elements of S, denoted by $L(S)$, is called the *linear span* or the span of S, i.e.,

$$L(S) = \{\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n : v_1, v_2, \dots, v_n \in S,\ \alpha_1, \alpha_2, \dots, \alpha_n \in F,\ n \in \mathbb{N}\}.$$

Lemma 1.2.3. *$L(S)$ is a subspace of V.*

Proof. Let $v = \lambda_1 s_1 + \ldots + \lambda_n s_n$ and $w = \mu_1 t_1 + \ldots + \mu_m t_m$ be two elements of $L(S)$ where $\lambda_i, \mu_i \in F$ and $s_i, t_i \in S$. Thus, for $\alpha, \beta \in F$,

$$\begin{aligned} \alpha v + \beta w &= \alpha(\lambda_1 s_1 + \ldots + \lambda_n s_n) + \beta(\mu_1 t_1 + \ldots + \mu_m t_m) \\ &= (\alpha\lambda_1)s_1 + \ldots + (\alpha\lambda_n)s_n + (\beta\mu_1)t_1 + \ldots + (\beta\mu_m)t_m \\ &\in L(S) \end{aligned}$$

Thus, $L(S)$ is a subpace of V. □

Now, we state the following lemma which gives the properties of linear span. The proof is straightforward and easy and left as an exercise.

Lemma 1.2.4. *Let S and T be two non-empty subsets of V. Then*

1. $S \subset T$ *implies* $L(S) \subset L(T)$.
2. $L(S \cup T) = L(S) + L(T)$.
3. $L(L(S)) = L(S)$.
4. S *is a subspace of* V *if and only if* $L(S) = S$.

Definition 1.2.5. Let V be a vector space over F. Then V is said to be *finite dimensional* (over F) if there is a finite set S such that $L(S) = V$.

For example, $F^{(n)}$ is finite dimensional over F for if S is a subset of $F^{(n)}$ consisting of n vectors such that $S = \{(1,0,\dots,0),(0,1,0,\dots,0),\dots,(0,0,\dots,0,1)\}$ then $L(S) = F^{(n)}$.

At this stage, we have just defined only what is a finite dimensional vector space. We will define what is dimension of a vector space later.

Definition 1.2.6. Let V be a vector space over F and $v_1,\dots,v_n \in V$. We say that they are *linearly dependent* over F if there exists $\alpha_1,\dots,\alpha_n \in F$, not all of them 0, such that $\alpha_1 v_1 + \cdots + \alpha_n v_n = 0$.

$v_1,\dots,v_n$ are said to be *linearly independent* if they are not linearly dependent, i.e., if $\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n = 0$ then $\alpha_i = 0, \forall i$. Note that if $v_1,\dots,v_n$ are linearly independent then none of them can be zero, for if $v_1 = 0$ (say), then for any $\alpha \neq 0$ in F we have $\alpha v_1 + 0v_2 + \cdots + 0v_n = 0$.

Definition 1.2.7. A non-empty subset M of a vector space V is called linearly independent over F if every finite subset of M is linearly independent over F.

Example 1.2.8. Clearly, in $\mathbb{R}^3$, $(1,0,0),(0,1,0)$ and $(0,0,1)$ are linearly independent over $\mathbb{R}$. Verify that $p_1(x) = x+1$, $p_2(x) = 3x^2+x+3$ and $p_3(x) = 3x^2+3x+5$ are three linearly dependent elements of $\mathbb{R}_3[x]$ over $\mathbb{R}$.

Remark 1.2.9. Observe that the notion of linear dependence is considered over the field F. Hence it depends not only on the given vectors but also on the field. For example, the field of complex numbers $\mathbb{C}$ can be considered as a vector space over $\mathbb{R}$ and also over $\mathbb{C}$. The elements $v_1 = 1$ and $v_2 = i$ are linearly independent over $\mathbb{R}$ as there are no non-zero real numbers α_1 and α_2 such that $(\alpha_1 \times 1) + (\alpha_2 \times i) = 0$. However, when considered over $\mathbb{C}$, they are linearly dependent since $iv_1 + (-1)v_2 = 0$.

1.2.1 Properties of linear independence and span

Let $x = (-1,1,0)$, $y = (1,-1,1)$ and $z = (2,-2,3)$ denote three elements of $\mathbb{R}^3$. Let $u = (-4,4,0) \in \mathbb{R}^3$. Then clearly $u \in L(\{x,y,z\})$ since u can be written in as $u = 5x+3y-z$. Note that u can also be written as $u = 6x+6y-2z$. Thus in this case, the expression of an element in the span of x,y,z is not unique. Now consider another set $\{e_1,e_2,e_3\}$ where $e_1 = (1,0,0)$, $e_2 = (0,1,0)$ and $e_3 = (0,0,1)$ in $\mathbb{R}^3$. Clearly u also belongs to its span since u

can be written as $u = -e_1 + 4e_2 + 0e_3$. As in the previous case, is it possible to find another representation of u in $L(\{e_1, e_2, e_3\})$? **No?** It seems that such expression is unique in this case. What is the difference! One difference is that x, y, z are linearly dependent ($z = x + 3y$) while we know that e_1, e_2, e_3 are L.I. This tempts us to ask the following question:

Question: If $v_1, \ldots, v_n \in V$ are linearly independent then is it true that every element in their span has a unique representation (over F)?

The answer to this question is affirmative. More precisely, we have the following lemma:

Lemma 1.2.10. *If $v_1, \ldots, v_n \in V$ are linearly independent, then every element in their linear span can be uniquely expressed in the form $\lambda_1 v_1 + \cdots + \lambda_n v_n$ with $\lambda_i \in F$.*

Proof. By definition of linear span, every element in $L(\{v_1, \ldots, v_n\})$ is of the form $\lambda_1 v_1 + \cdots + \lambda_n v_n$. Suppose if possible $v \in L(\{v_1, \ldots, v_n\})$ can be expressed in two different ways, say, $v = \lambda_1 v_1 + \cdots + \lambda_n v_n$ and $v = \mu_1 v_1 + \cdots + \mu_n v_n$. Then

$$\lambda_1 v_1 + \cdots + \lambda_n v_n = v = \mu_1 v_1 + \cdots + \mu_n v_n.$$

This implies,

$$(\lambda_1 - \mu_1)v_1 + (\lambda_2 - \mu_2)v_2 + \cdots + (\lambda_n - \mu_n)v_n = 0.$$

Since, $v_1, v_2, \ldots, , v_n$ are given to be linearly independent, we conclude that $\lambda_i - \mu_i = 0$ for all i and hence every element $v \in L(\{v_1, \ldots, v_n\})$ has a unique representation. □

Question 1.2.11. Is the converse true? That is if every element $v \in L(\{v_1, \ldots, v_n\})$ has a unique representation given in above lemma, then can we say that $v_1, v_2, \ldots, v_n$ are linearly independent?

Theorem 1.2.12. *Let V be a vector space over F, $v_1, v_2, \ldots, v_n \in V$ and $v_1 \neq 0$. Then either they are linearly independent or there exists a $k \leq n$ such that v_k is a linear combination of the preceding ones, $v_1, v_2, \ldots, v_{k-1}$.*

Proof. If $v_1, v_2, \ldots, v_n$ are all linearly independent then we are done. Suppose that $v_1, \ldots, v_n$ are not linearly independent. Then obviously $n \geq 2$ as $\{v_1\}$ is linearly independent. Since $v_1, \ldots, v_n$ are linearly dependent, there exist $\alpha_1, \ldots, \alpha_n \in F$, not all zero, such that $\alpha_1 v_1 + \cdots + \alpha_n v_n = 0$. Let k be the largest integer such that $\alpha_k \neq 0$, i.e., $\alpha_{k+1} = \alpha_{k+2} = \cdots = \alpha_n = 0$. This $k \geq 2$. If $k = 1$, then $\alpha_1 v_1 = 0$. Since $\alpha_1 \neq 0$, $v_1 = 0$. This is not possible. So, we get $\alpha_1 v_1 + \cdots + \alpha_k v_k = 0$. Since $\alpha_k \neq 0$, we have

$$\begin{aligned} v_k &= \alpha_k^{-1}(-\alpha_1 v_1 - \alpha_2 v_2 - \cdots - \alpha_{k-1} v_{k-1}) \\ &= (-\alpha_k^{-1}\alpha_1)v_1 + (-\alpha_k^{-1}\alpha_2)v_2 + \cdots + (-\alpha_k^{-1}\alpha_{k-1})v_{k-1}. \end{aligned}$$

Thus, v_k is a linear combination of its predecessors. □

Corollary 1.2.13. *Let V be a vector space over F and $v_1, \ldots, v_n \in V$ such that $W = L(\{v_1, \ldots, v_n\})$. If $v_1, \ldots, v_k$ are linearly independent then we can find a linearly independent subset $\{v_1, \ldots, v_k, v_{i_1}, \ldots, v_{i_r}\}$ of $\{v_1, \ldots, v_n\}$ whose linear span is also W.*

Proof. If $v_1,\dots,v_n$ are linearly independent, then we are done.
Assume that $v_1,\dots,v_n$ are linearly dependent. Then by the above theorem there is a v_j such that v_j is a linear combination of $v_1,\dots,v_{j-1}$, i.e.,

$$v_j = \beta_1 v_1 + \cdots + \beta_{j-1} v_{j-1}.$$

Since, $v_1,\dots,v_k$ are given to be linearly independent, $j>k$.
Claim: $L(\{v_1,\dots,v_{j-1},v_{j+1},\dots,v_n\}) = W$.
Since $\{v_1,\dots,v_{j-1},v_{j+1},\dots,v_n\} \subset \{v_1,\dots,v_n\}$, we have

$$L(\{v_1,\dots,v_{j-1},v_{j+1},\dots,v_n\}) \subset L(\{v_1,\dots,v_n\}) = W.$$

Now, let $w \in W = L(\{v_1,\dots,v_n\})$. Then there exists $\alpha_1,\dots,\alpha_n \in F$ such that $w = \alpha_1 v_1 + \cdots + \alpha_n v_n$. Then, we have

$$\begin{aligned} w &= \alpha_1 v_1 + \cdots + \alpha_{j-1} v_{j-1} + \alpha_j v_j + \alpha_{j+1} v_{j+1} + \cdots + \alpha_n v_n \\ &= \alpha_1 v_1 + \cdots + \alpha_{j-1} v_{j-1} + \alpha_j(\beta_1 v_1 + \cdots + \beta_{j-1} v_{j-1}) + \alpha_{j+1} v_{j+1} + \cdots + \alpha_n v_n \\ &= (\alpha_1 + \alpha_j\beta_1) v_1 + (\alpha_2 + \alpha_j\beta_2) v_2 + \cdots + (\alpha_{j-1} + \alpha_j\beta_{j-1}) v_{j-1} + \alpha_{j+1} v_{j+1} + \cdots + \alpha_n v_n \end{aligned}$$

Therefore, $w \in L(\{v_1,\dots,v_{j-1},v_{j+1},\dots,v_n\})$. Hence, $W = L(\{v_1,\dots,v_{j-1},v_{j+1},\dots,v_n\})$. If $v_1,\dots,v_{j-1},v_{j+1},\dots,v_n$ are linearly independent then we are done. If not, then continuing this procedure, we get a subset $\{v_1,\dots,v_k,v_{i_1},\dots,v_{i_r}\}$ of $\{v_1,\dots,v_n\}$ such that it is linearly independent and $W = L(\{v_1,\dots,v_k,v_{i_1},\dots,v_{i_r}\})$. □

Definition 1.2.14. Let V be a vector space over F. A subset S of V is called a *basis* of V if S is linearly independent (that is every finite subset of S is linearly independent) and $L(S) = V$.

Corollary 1.2.15. *Let V be a finite-dimensional vector space over F and $v_1,\dots,v_n \in V$ such that $L(\{v_1,v_2,\dots,v_n\}) = V$. Then there exists a subset $\{u_1,\dots,u_m\}$ of $\{v_1,\dots,v_n\}$ such that $\{u_1,\dots,u_m\}$ is a basis of V.*

Lemma 1.2.16. *Let V be a vector space and $\{v_1,v_2,\dots,v_n\}$ be a basis of V. If $\{w_1,\dots,w_m\}$ in V are linearly independent then $m \le n$.*

Proof. Since $\{v_1,v_2,\dots,v_n\}$ is a basis of V, every element in V can be written as a linear combination of $v_1,v_2,\dots,v_n$. In particular, $w_m \in V$ can be written as a linear combination of $v_1,v_2,\dots,v_n$. Therefore the set $\{w_m,v_1,v_2,\dots,v_n\}$ is linearly dependent and clearly

$$L(\{w_m,v_1,v_2,\dots,v_n\}) = V.$$

Therefore, we can find a proper subset $\{w_m,v_{i_1},\dots,v_{i_r}\}$ of $\{w_m,v_1,v_2,\dots,v_n\}$ which is linearly independent and which spans V, i.e., $\{w_m,v_{i_1},\dots,v_{i_r}\}$ forms a basis of V. Thus, we have inserted one w at the cost of at least on v from our set. Therefore, $r \le n-1$.

Now, w_{m-1} can be written as a linear combination of the new basis $\{w_m, v_{i_1}, \ldots, v_{i_r}\}$. Hence, the set $\{w_{w-1}, w_m, v_{i_1}, \ldots, v_{i_r}\}$ is linearly dependent. Repeating the above procedure we can find new basis of V of the form

$$\{w_{m-1}, w_m, v_{j_1}, \ldots, v_{j_s}\} \qquad (s \leq n-2).$$

Continuing this way, we eventually come down to a stage where the basis of V is of the form

$$B = \{w_2, w_3, \ldots, w_m, v_\alpha, v_\beta, \ldots\}.$$

Now, we write w_1 as a linear combination of the new basis B of V given above. Since $w_1, w_2, \ldots, w_m$ are linearly independent, w_1 cannot be written as a linear combination of $w_2, \ldots, w_m$. Hence, the basis B must contain some v.

To obtain the basis B, we have inserted $m-1$ w's and each time removed at least one v, and still there is some v left in B. This implies $m-1 \leq n-1$ and so $m \leq n$. □

Corollary 1.2.15 indicates that number of elements in a basis is less than or equal to number of elements in a span. By Lemma 1.2.16, we can say that cardinality of any linearly independent set is less than or equal to cardinality of a basis. Combining them, we have the following relation:

$$\text{Card(L.I. set)} \leq \text{Card(Basis)} \leq \text{Card(Span)}.$$

Corollary 1.2.17. *If V is finite-dimensional vector space over F then any two bases of V have the same number of elements.*

Proof. Let $\{v_1, \ldots, v_n\}$ and $\{w_1, \ldots, w_m\}$ be two bases of V over F. In particular, $w_1, \ldots, w_m \in V$ are linearly independent over F and considering $\{v_1 \ldots, v_n\}$ as basis of V, by Lemma 1.2.16, we have $m \leq n$.

Now, $v_1 \ldots, v_n \in V$ are linearly independent over F and considering $\{w_1 \ldots, w_n\}$ as a basis of V, by the above lemma, we have $n \leq m$. Hence, $m = n$. □

Exercise 1.2.18. If T is an isomorphism of V onto W, prove that T maps a basis of V onto a basis of W.

Solution. The problem is given as a seminar exercise. □

Corollary 1.2.19. *$F^{(n)}$ and $F^{(m)}$ are isomorphic if and only if $n = m$.*

Proof. $F^{(n)}$ has a basis consisting of n vectors, $(1,0,\ldots,0), (0,1,0,\ldots,0), \ldots, (0,0,\ldots,0,1)$. $F^{(m)}$ has a similar basis consisting of m such elements. By above exercise, an isomorphism maps basis onto a basis. Hence, by Corollary 1.2.17, $m = n$. □

Exercise 1.2.20. If V is a finite dimensional vector space over F then V is isomorphic to $F^{(n)}$ for some n.

Solution. V is given to be a finite dimensional vector space over F. Let $\{v_1, v_2, \ldots, v_n\}$ be a basis of V. Then every element $v \in V$ has a unique representation of the form

$$\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n \qquad (\alpha_i \in F,\ i = 1, 2, \ldots, n).$$

We define a map $T : V \to F^{(n)}$ by $T(v) = T(\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n) = (\alpha_1, \alpha_2, \ldots, \alpha_n)$. It is easy to see that T is an isomorphism of V onto $F^{(n)}$. Thus, $V \cong F^{(n)}$. □

Definition 1.2.21. Let V be a finite dimensional vector space over F. The dimension of V over F is the number of elements in any basis of V over F. It is usually denoted by $\dim V$ and sometimes by $\dim_F V$ (to stress that dimension of V is over the field F).

Exercise 1.2.22. Let $\mathscr{V} = \{V : V \text{ is a vector space over } F\}$. Define a relation '$\sim$' on $\mathscr{V}$ as follows: for $U, V \in \mathscr{V}$, $U \sim V$ if U is isomorphic to V. Show that '$\sim$' is an equivalence relation, i.e., isomorphism of vector spaces is an equivalence relation.

Corollary 1.2.23. *Let V and W be two finite dimensional vector spaces over F such that their dimensions are same, i.e.,* $\dim V = \dim W$. *Then V and W are isomorphic.*

Proof. If $\dim V = \dim W = n$. Then by Exercise 1.2.20, V is isomorphic to $F^{(n)}$ and W is isomorphic to $F^{(n)}$. Hence, by the transitivity property of the above exercise, V and W are isomorphic. □

By Exercise 1.2.18 and Corollary 1.2.23, we have the following result:

Lemma 1.2.24. *If V and W are two finite dimensional vector spaces over F then V and W are isomorphic to each other if and only if* $\dim V = \dim W$.

We have already seen that given a spanning set we can find its subset which forms a basis. The following lemma show that given a linearly independent set we can extend it to a basis.

Lemma 1.2.25. *Let V be a finite-dimensional vector space over F such that $u_1, \ldots, u_m \in V$ are linearly independent. Then we can find vectors $u_{m+1}, \ldots, u_{m+r}$ in V such that $u_1, \ldots, u_m, u_{m+1}, \ldots, u_{m+r}$ is a basis of V.*

Proof. Since V is a finite dimensional vector space it has a basis, say $\{v_1, \ldots, v_n\}$. Consider the set $B = \{u_1, \ldots, u_m, v_1, \ldots, v_n\}$. Clearly the set B spans V. Then by Corollary 1.2.13, there is a linearly independent subset $\{u_1, \ldots, u_m, v_{i_1}, \ldots, v_{i_r}\}$ of B such that it spans V. We just write $v_{i_1} = u_{m+1}, \ldots, v_{i_r} = u_{m+r}$. □

Lemma 1.2.26. *Let V be finite dimensional vector space over F and W be a subspace of V. Then W is finite dimensional and* $\dim W \leq \dim V$. *Also,* $\dim V/W = \dim V - \dim W$.

Proof. Let $\dim V = n$. Then by Lemma 1.2.16, any $n+1$ elements in V are linearly dependent. In particular, any $n+1$ elements in W are linearly dependent. Let $\{w_1,\dots,w_m\}$ be a maximal linearly independent set in W, then $m \leq n$. We want to show that $\dim W = m$. Let $w \in W$ then set $\{w, w_1,\dots,w_m\}$ is linearly dependent and hence

$$\alpha w + \alpha_1 w_1 + \cdots + \alpha_m w_m = 0$$

such that not all the scalars α_i's are zero. If $\alpha = 0$, then we have

$$\alpha_1 w_1 + \cdots + \alpha_m w_m = 0.$$

Since $w_1,\dots,w_m$ are linearly independent, $\alpha_i = 0$ for all $i = 1,2,\dots,m$ which implies $\{w, w_1,\dots,w_m\}$ is a linearly independent set. This is contradiction since $\{w_1,\dots,w_m\}$ is the largest linearly independent set in W. Hence, $\alpha \neq 0$ and so we can write

$$w = (-\alpha^{-1}\alpha_1)w_1 + \cdots + (-\alpha^{-1}\alpha_m)w_m.$$

This means $w_1,\dots,w_m$ spans W and since it is a L.I. set, it is basis of W. Thus it follows that $\dim W \leq \dim V$.

Now, $w_1,\dots,w_m$ is a basis of W and so it is a linearly independent set in V. Then by Lemma 1.2.25, it can be extended to a basis, $\{w_1,\dots,w_m,v_1,\dots,v_r\}$, of V where $\dim V = m+r$ and $\dim W = m$.

If we show that $\{v_1+W,\dots,v_r+W\}$ is a basis of V/W, then we are done as $\dim V/W = r = (m+r)-m = \dim V - \dim W$. First we show that it spans V/W. Let $v+W \in V/W$ for some $v \in V$. Now, since $\{w_1,\dots,w_m,v_1,\dots,v_r\}$ is a basis of V, every $v \in V$ can be written as

$$v = \alpha_1 w_1 + \ldots + \alpha_m w_m + \beta_1 v_1 + \ldots + \beta_r v_r.$$

Then

$$\begin{aligned} v+W &= \alpha_1(w_1+W) + \ldots + \alpha_m(w_m+W) + \beta_1(v_1+W) + \ldots + \beta_r(v_r+W) \\ &= \beta_1(v_1+W) + \ldots + \beta_r(v_r+W) \qquad (\text{as } w_i \in W \Rightarrow w_i + W = 0) \end{aligned}$$

Thus, $\{v_1+W,\dots,v_r+W\}$ spans V/W. Now, we show that it is a linearly independent set. Let $\gamma_1(v_1+W) + \cdots + \gamma_r(v_r+W) = 0$. Then $\gamma_1 v_1 + \cdots + \gamma_r v_r \in W$ and so it can be written as a linear combination of elements of the basis of W as follows:

$$\begin{aligned} \gamma_1 v_1 + \cdots + \gamma_r v_r &= \lambda_1 w_1 + \cdots + \lambda_m w_m \\ \Rightarrow \gamma_1 v_1 + \cdots + \gamma_r v_r - \lambda_1 w_1 + \cdots + \lambda_m w_m &= 0. \end{aligned}$$

Since $\{w_1,\dots,w_m,v_1,\dots,v_r\}$ is a basis of V, above equation implies

$$\gamma_1 = \cdots = \gamma_r = \lambda_1 = \cdots = \lambda_m = 0.$$

Thus, $\{v_1+W,\dots,v_r+W\}$ is a basis of V/W with r elements. Hence,

$$\dim V/W = r = \dim V - m = \dim V - \dim W.$$

□

Exercise 1.2.27. If A and B are subspaces of a vector space V then prove that $(A+B)/B$ is isomorphic to $A/(A\cap B)$, where

$$A+B = \{v \in V : v = a+b,\ a \in A,\ b \in B\}.$$

Solution. Given as a seminar exercise. This result is known as the *Second Isomorphism Theorem.* □

Corollary 1.2.28. *If A and B are are finite dimensional subspaces of a vector space V, then $(A+B)$ is finite dimensional and we have*

$$\dim(A+B) = \dim A + \dim B - \dim(A\cap B).$$

Proof. By Exercise 1.2.27, we have

$$\frac{(A+B)}{B} \cong \frac{A}{(A\cap B)}.$$

By Exercise 1.2.18, we know that if two vectors spaces are isomorphic then their dimensions are same. Hence,

$$\dim \frac{(A+B)}{B} = \dim \frac{A}{(A\cap B)}.$$

Then by Lemma 1.2.26, we have

$$\begin{aligned} &\dim(A+B) - \dim B = \dim A - \dim(A\cap B) \\ \Rightarrow\ &\dim(A+B) = \dim A + \dim B - \dim(A\cap B). \end{aligned}$$

□

1.3 Dual Spaces

Let V and W be two vector spaces over F. We have defined $\mathrm{Hom}(V,W)$ to be the set of all vector space homomorphisms from V into W. We intend to make it a vector space over F. For $S,T \in \mathrm{Hom}(V,W)$, we define $S+T$ as

$$(S+T)(v) := S(v)+T(v)$$

for all $v \in V$. Now, we check that $S+T$ is a homomorphism. If $v_1, v_2 \in V$ and $\alpha \in F$, then

$$\begin{aligned} (S+T)(v_1+v_2) &= S(v_1+v_2)+T(v_1+v_2) && \text{(by definition of } S+T) \\ &= (S(v_1)+S(v_2))+(T(v_1)+T(v_2)) && \text{(since } S,T \text{ are homomorphisms)} \\ &= (S(v_1)+T(v_1))+(S(v_2)+T(v_2)) && \\ &= (S+T)(v_1)+(S+T)(v_2) && \text{(by definition of } S+T) \end{aligned}$$

and

$$\begin{aligned} (S+T)(\alpha v_1) &= S(\alpha v_1)+T(\alpha v_1) && \text{(by definition of } S+T) \\ &= \alpha S(v_1)+\alpha T(v_1) && \text{(since } S,T \text{ are homomorphisms)} \\ &= \alpha(S+T)(v_1) && \text{(by definition of } S+T). \end{aligned}$$

Thus, $(S+T)$ is a homomorphism of V into W i.e., $S+T \in \mathrm{Hom}(V,W)$. The zero homomorphism $0 \in \mathrm{Hom}(V,W)$ is defined by $0(v) = 0$ for all $v \in V$ and we have $S+0 = S$ for all

$S \in \mathrm{Hom}(V,W)$. Also, for any $S \in \mathrm{Hom}(V,W)$, let $(-S)$ be defined by $(-S)(v) = -(S(v))$. Thus, it is evident that $\mathrm{Hom}(V,W)$ is an abelian group under the addition defined above.

Now, we define scalar multiplication on $\mathrm{Hom}(V,W)$ to make it into a vector space over F. For $\lambda \in F$ and $S \in \mathrm{Hom}(V,W)$. We define λS by

$$(\lambda S)(v) := \lambda(S(v)) \qquad (\text{ for all } v \in V).$$

One can check that λS defined above is in $\mathrm{Hom}(V,W)$. The other properties of vector space can easily verified. Thus, we have the following result:

Lemma 1.3.1. *Let V and W be vector spaces over F. Then* $\mathrm{Hom}(V,W)$ *is a vector space over F under the operations defined above.*

Proof. Seminar exercise (partly discussed in class). □

Exercise 1.3.2. If $S,T \in \mathrm{Hom}(V,W)$ and $\{v_1,\dots,v_m\}$ be a basis of V over F. If $S(v_i) = T(v_i)$ for all $i = 1,\dots,m$, then prove that $S = T$. In other words, if two homomorphisms agree on the basis then they must be same.

Solution. Seminar exercise. □

When V and W are finite dimensional vector spaces, the following theorem relates the dimension of $\mathrm{Hom}(V,W)$ to that of V and W.

Theorem 1.3.3. *Let V and W be vector spaces over F of dimensions m and n respectively. Then* $\mathrm{Hom}(V,W)$ *is of dimension mn over F.*

Proof. We prove the theorem by finding a basis of $\mathrm{Hom}(V,W)$ containing mn elements. Let $\{v_1,\dots,v_m\}$ be a basis of V over F and $\{w_1,\dots,w_n\}$ be a basis of W over F. If $v \in V$ then $v = \lambda_1 v_1 + \cdots + \lambda_m v_m$ where $\lambda_1,\dots,\lambda_m$ are uniquely determined elements of F.

For $1 \le i \le n$ and $1 \le j \le m$, define $T_{ij} : V \to W$ by $T_{ij}(v) = \lambda_j w_i$ for $v \in V$. Observe that on basis elements T_{ij} is defined as

$$T_{ij}(v_k) = \begin{cases} w_i & \text{if } j = k \\ 0 & \text{if } j \neq k. \end{cases}$$

First we show that T_{ij} is a homomorphism. For this, let $u,v \in V$ such that $u = \alpha_1 v_1 + \cdots + \alpha_m v_m$ and $v = \beta_1 v_1 + \cdots + \beta_m v_m$ and $\gamma,\mu \in F$. Then

$$\begin{aligned} T_{ij}(\gamma u + \mu v) &= T_{ij}(\gamma(\alpha_1 v_1 + \cdots + \alpha_m v_m) + \mu(\beta_1 v_1 + \cdots + \beta_m v_m)) \\ &= T_{ij}((\gamma\alpha_1 + \mu\beta_1)v_1 + \cdots + (\gamma\alpha_m + \mu\beta_m)v_m) \\ &= (\gamma\alpha_j + \mu\beta_j)w_i \qquad (\text{by definition of } T_{ij}) \\ &= \gamma\alpha_j w_i + \mu\beta_j w_i \\ &= \gamma T_{ij}(u) + \mu T_{ij}(v). \end{aligned}$$

Thus, $T_{ij} \in \mathrm{Hom}(V,W)$. Let $B = \{T_{ij} : 1 \le i \le n,\ 1 \le j \le m\}$. Then B consists of mn elements of $\mathrm{Hom}(V,W)$. If we show that B is a basis of $\mathrm{Hom}(V,W)$ over F then we are done.

For this, first we shall show that B spans $\mathrm{Hom}(V,W)$. Let $S \in \mathrm{Hom}(V,W)$. Then we have to show that $S = L(B)$, i.e., S is a linear combination of T_{ij}'s. For a basis element v_1 of V, $S(v_1) \in W$. So, $S(v_1)$ can be written as a linear combination of elements in the basis of W. Let $S(v_1) = \alpha_{11}w_1 + \alpha_{21}w_2 + \cdots + \alpha_{n1}w_n$. In fact, for all $j = 1,2,\ldots,m$,

$$S(v_j) = \alpha_{1j}w_1 + \alpha_{2j}w_2 + \cdots + \alpha_{nj}w_n. \tag{1.2}$$

Now, consider

$$\begin{aligned} S_0 &= \sum_{\substack{1\le i\le n\\ 1\le j\le m}} \alpha_{ij}T_{ij} \\ &= \alpha_{11}T_{11} + \alpha_{21}T_{21} + \cdots + \alpha_{n1}T_{n1} + \alpha_{12}T_{12} + \alpha_{22}T_{22} + \cdots + \alpha_{n2}T_{n2} + \cdots + \\ &\quad + \alpha_{1m}T_{1m} + \alpha_{2m}T_{2m} + \cdots + \alpha_{nm}T_{nm} \end{aligned}$$

Note that $S_0 \in L(B)$. If we show that $S = S_0$ then we have $S \in L(B)$ and we are done. By (above) Exercise 1.3.2 it suffices to show that $S_0(v_k) = S(v_k)$ for all basis elements v_k of V. Now, we compute the value of S_0 at v_k.

$$\begin{aligned} S_0(v_k) &= \left(\sum_{\substack{1\le i\le n\\ 1\le j\le m}} \alpha_{ij}T_{ij} \right)(v_k) \\ &= \alpha_{11}T_{11}(v_k) + \cdots + \alpha_{n1}T_{n1}(v_k) + \alpha_{12}T_{12}(v_k) + \cdots + \alpha_{n2}T_{n2}(v_k) + \cdots + \\ &\quad + \alpha_{1m}T_{1m}(v_k) + \cdots + \alpha_{nm}T_{nm}(v_k) \\ &= \sum_{i=1}^{n} \alpha_{ik}w_i \qquad \text{(by definition of } T_{ij}\text{)} \\ &= \alpha_{1k}w_1 + \alpha_{2k}w_2 + \cdots + \alpha_{nk}w_n \\ &= S(v_k) \qquad \text{(by equation (1.2))}. \end{aligned}$$

Thus, $S(v_k) = S_0(v_k)$ for all $k = 1,2,\ldots,m$ and hence $S \in L(B)$.

Now, we show that the set B is linearly independent. For $\beta_{ij} \in F$, let

$$\sum_{\substack{1\le i\le n\\ 1\le j\le m}} \beta_{ij}T_{ij} = 0.$$

Then we have to show that $\beta_{ij} = 0$ for all $i = 1,2,\ldots,n$ and $j = 1,2,\ldots,m$. Applying this to the basis element v_k of V, we have

$$\begin{aligned} &\left(\sum_{\substack{1\le i\le n\\ 1\le j\le m}} \beta_{ij}T_{ij} \right)(v_k) = 0 \\ \Rightarrow\ &\sum_{i=1}^{n} \beta_{ik}w_i = 0 \qquad \text{(by definition of } T_{ij}\text{)} \\ \Rightarrow\ &\beta_{1k}w_1 + \beta_{2k}w_2 + \cdots + \beta_{nk}w_k = 0. \end{aligned}$$

Since, $\{w_1, w_2, \ldots, w_n\}$ is a basis of W, we have $\beta_{ik} = 0$ for all $i = 1,2,\ldots,n$. This is true for all $k = 1,2,\ldots,m$. Hence, $\beta_{ij} = 0 \quad \forall\, i,j$. This shows that B is a linearly independent set and hence a basis of $\mathrm{Hom}(V,W)$. Hence, $\dim(\mathrm{Hom}(V,W)) = mn$. □

Remark 1.3.4. As a consequence of above theorem, if $V \neq \{0\}$ and $W \neq \{0\}$ are finite dimensional vector spaces of dimension m and n respectively then $m \geq 1$ and $n \geq 1$ and hence $\dim(\text{Hom}(V,W)) = mn \geq 1$. This means $\text{Hom}(V,W)$ does not just contain only one trivial homomorphism 0, i.e. $\text{Hom}(V,W) \neq 0$.

Corollary 1.3.5. *If* $\dim_F V = m$ *then* $\dim_F \text{Hom}(V,V) = m^2$.

Proof. In the theorem put $W = V$. So $m = n$ and hence $mn = m^2$. □

Corollary 1.3.6. *If* $\dim_F V = m$ *then* $\dim_F \text{Hom}(V,F) = m$.

Proof. We know that $F^{(n)}$ is a vector space over F of dimension n. Here, F is a vector space of dimension is 1 over F. Put $W = F$ in the above theorem, then we have $\dim_F \text{Hom}(V,F) = m$. □

Definition 1.3.7 (Dual space)**.** Let V be a vector space over a field F. A homomorphism from V to F is also called a *linear functional*. The collection of all linear functionals on V is denoted by $\widehat{V} = \text{Hom}(V,F)$ is a vector space over F and called the *dual space* or simply the *dual* of V over F. Thus,

$$\widehat{V} = \{f : V \to F : f \text{ is a homomorphism}\}.$$

Remark 1.3.8. By Corollary 1.3.6, if V is finite dimensional then $\dim V = \dim \widehat{V}$. Then we can say that V is isomorphic to its dual $\widehat{V}$. However, this is true only when V is finite dimensional. If V is not finite dimensional then **no** such isomorphism exists.

Definition 1.3.9. Let V be a finite dimensional vector space over F and let $\{v_1, v_2, \ldots, v_n\}$ be a basis of V over F. Let $\hat{v}_i$ be an element of $\widehat{V}$ defined as $\hat{v}_i(\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n) = \alpha_i$, i.e.,

$$\hat{v}_i(v_j) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases}$$

Then $\hat{v}_i$ are same as T_{ij} in the previous theorem with $W = F$ here which is one dimensional over F. Thus, $v_1, \hat{\ldots}, \hat{v}_n$ forms a basis of $\widehat{V}$. This basis is called the *dual basis* of $v_1, \ldots, v_n$.

Lemma 1.3.10. *Let V be a finite dimensional vector space over F and $v \in V$, $v \neq 0$. Then there is an element $f \in \widehat{V}$ such that $f(v) \neq 0$.*

Proof. If $v \neq 0$ then $\{v\}$ is linearly independent. Therefore by Lemma 1.2.25, it can be extended to a basis $\{v = v_1, v_2, \ldots, v_n\}$ of V. As in above definition, define $f : V \to F$ by $f(\alpha v + \alpha_2 v_2 + \cdots + \alpha_n v_n) = \alpha$. Then $f \in \widehat{V}$ and $f(v) = 1 \neq 0$. Hence, the lemma. □

The above lemma is true for infinite dimensional vector spaces also but here we are concerned only for finite dimensional vector space V.

1.3.1 Second dual

Let V be a vector space over F and $v_0 \in V$. For every $f \in \widehat{V}$ denote the map $f \mapsto f(v_0)$ by T_{v_0}. That is $T_{v_0}(f) = f(v_0)$ for $f \in \widehat{V}$. Thus, T_{v_0} is a map from $\widehat{V}$ to F. Furthermore, we check that T_{v_0} is a homomorphism. For any $f, g \in \widehat{V}$ and $\lambda \in F$,

$$\begin{aligned} T_{v_0}(f+g) &= (f+g)(v_0) = f(v_0) + g(v_0) = T_{v_0}(f) + T_{v_0}(g); \\ T_{v_0}(\lambda f) &= (\lambda f)(v_0) = \lambda(f(v_0)) = \lambda T_{v_0}(f). \end{aligned}$$

Thus, $T_{v_0} : \widehat{V} \to F$ is a homomorphism, i.e., it is an element of dual space of $\widehat{V}$. We say that T_{v_0} is an element of second dual of V defined as follows:

Definition 1.3.11. Let V be a vector space over F and $\widehat{V}$ denote the dual space of V. Then the set of all homomorphisms from $\widehat{V}$ to F is a vector space over F called the *second dual* of V and denote by $\widehat{\widehat{V}}$ i.e.,

$$\widehat{\widehat{V}} = \{T : \widehat{V} \to F : T \text{ is a homomorphism}\}.$$

As seen above for every $v_0 \in V$ we get an element $T_{v_0} \in \widehat{\widehat{V}}$. We can thus define a map $v_0 \mapsto T_{v_0}$ from V to $\widehat{\widehat{V}}$ and we expect it to be isomorphism. More precisely, we have the following theorem:

Theorem 1.3.12. *Let V be a vector space over F. Then V is isomorphic to a subspace of $\widehat{\widehat{V}}$, i.e., there is an isomorphism of V into $\widehat{\widehat{V}}$. If V is finite dimensional then $V \cong \widehat{\widehat{V}}$ (V is isomorphic to $\widehat{\widehat{V}}$).*

Proof. Define a map $\psi : V \to \widehat{\widehat{V}}$ by $\psi(v) = T_v$ where T_v is defined as above, $T_v(f) = f(v)$ for $f \in \widehat{V}$. First we show that ψ is a homomorphism. For this we have to show that for any $v, w \in V$ and $\lambda \in F$,

$$\psi(v+w) = \psi(v) + \psi(w) \quad \text{and} \quad \psi(\lambda v) = \lambda \psi(v).$$

That is we have to show that

$$T_{v+w} = T_v + T_w \quad \text{and} \quad T_{\lambda v} = \lambda T_v.$$

Now,

$$\begin{aligned} T_{v+w}(f) &= f(v+w) && \text{(by definition of } T) \\ &= f(v) + f(w) && \text{(since } f \text{ is a homomorphism of } V \text{ into } F) \\ &= T_v(f) + T_w(f) && \text{(by definition of } T) \\ &= (T_v + T_w)(f). \end{aligned}$$

Also,

$$T_{\lambda v}(f) = f(\lambda v) \qquad \text{(by definition of } T)$$

$$\begin{aligned} &= \lambda f(v) && (\text{since } f \text{ is a homomorphism of } V \text{ into } F) \\ &= \lambda(T_v(f)) && (\text{by definition of } T). \\ &= (\lambda T_v)(f) \end{aligned}$$

Thus, $\psi : V \to \widehat{\widehat{V}}$ is a homomorphism. Now we show that ψ is one-one. To show this, we shall show that if $\psi(v) = 0$ then $v = 0$. Now,

$$\begin{aligned} \psi(v) = 0 &\Rightarrow T_v = 0 \\ &\Rightarrow T_v(f) = 0 \text{ for all } f \in \widehat{V} \\ &\Rightarrow f(v) = 0 \text{ for all } f \in \widehat{V}. \end{aligned}$$

Thus, $\psi(v) = 0$ implies $f(v) = 0$ for all $f \in \widehat{V}$. If $v \neq 0$ then by Lemma 1.3.10, there exists at least one $f \in \widehat{V}$ such that $f(v) \neq 0$. But here we have $f(v) = 0$ for all $f \in \widehat{V}$. Hence v must be 0 which concludes that ψ is one-one.

When V is finite dimensional we know that

$$\dim(V) = \dim(\widehat{V}) = \dim(\widehat{\widehat{V}}).$$

Since, ψ is an isomorphism, by equality of dimensions, we conclude that ψ is onto. Hence, V is isomorphic to its second dual $\widehat{\widehat{V}}$. □

The above theorem holds even when V is infinite dimensional. However, in case of infinite dimensional vector space V, the isomorphism ψ is not onto.

Definition 1.3.13. Let V be a vector space and W be a subspace of V. The *annihilator* of W is denoted by W^0 and defined as

$$W^0 = \{f \in \widehat{V} \mid f(w) = 0 \text{ for all } w \in W\}.$$

Exercise 1.3.14. If V is a vector space and U, W are subspaces of V then prove that:

1. W^0 is a subspace of $\widehat{V}$.
2. If $U \subset W$, then $W^0 \subset U^0$.

Question 1.3.15. Is the converse of (2) in above true? That is, given $W^0 \subset U^0$ can we say that $U \subset W$?

Definition 1.3.16. Let V be a vector space and W be a subspace of V. If $f \in \widehat{V}$, i.e., $f : V \to F$ be a linear functional, then the restriction of f to W, $\tilde{f}$ or $f|_W$ is a map from W to F and defined as

$$\tilde{f} = f|_W : W \to F \quad \text{such that} \quad \tilde{f}(w) = f(w), \text{ for all } w \in W.$$

For $w_1, w_2 \in W$ and $\alpha, \beta \in F$,

$$\begin{aligned} \tilde{f}(\alpha w_1 + \beta w_2) &= f|_W(\alpha w_1 + \beta w_2) \\ &= f(\alpha w_1 + \beta w_2) && (\text{by definition of } \tilde{f}) \end{aligned}$$

$$\begin{aligned} &= \alpha f(w_1) + \beta f(w_2) \quad \text{(since } f \text{ is a homomorphism)} \\ &= \alpha f\big|_W(w_1) + \beta f\big|_W(w_2) \\ &= \alpha \tilde{f}(w_1) + \beta \tilde{f}(w_2) \end{aligned}$$

Thus, $\tilde{f} : W \to F$ is a homomorphism. Hence, if $f \in \widehat{V}$ then $\tilde{f} = f\big|_W \in \widehat{W}$.

Theorem 1.3.17. *Let V be a finite dimensional vector space over F and W be a subspace of V. Then $\widehat{W}$ is isomorphic to $\widehat{V}/W^0$ and* $\dim W^0 = \dim V - \dim W$.

Proof. Define a map $\phi : \widehat{V} \to \widehat{W}$ by $\phi(f) = \tilde{f} = f\big|_W$ for all $f \in \widehat{V}$. If we show that ϕ is an onto homomorphism with $\ker\phi = W^0$, then by Theorem 1.1.32 (First Homomorphism Theorem), we have $\widehat{W} \cong \widehat{V}/W^0$ and we are done.
First we show that ϕ is a homomorphism: For $f_1, f_2 \in \widehat{V}$ and $\alpha \in F$,

$$\begin{aligned} \phi(f_1 + f_2)(w) &= (f_1 + f_2)\big|_W(w) \\ &= (f_1 + f_2)(w) \\ &= f_1(w) + f_2(w) \\ &= f_1\big|_W(w) + f_2\big|_W(w) \\ &= (\phi(f_1) + \phi(f_2))(w) \end{aligned}$$

Thus, $\phi(f_1 + f_2)(w) = (\phi(f_1) + \phi(f_2))(w)$ for all $w \in W$. This implies, $\phi(f_1 + f_2) = \phi(f_1) + \phi(f_2)$. Also,

$$\begin{aligned} \phi(\alpha f_1)(w) &= (\alpha f_1)\big|_W(w) \\ &= \alpha f_1(w) \\ &= \alpha(f_1\big|_W)(w) = \alpha\phi(w). \end{aligned}$$

Thus, $\phi(\alpha f_1) = \alpha\phi(f_1)$. Hence, ϕ is a homomorphism.
Now, we show that $\ker\phi = W^0$: By definition,

$$\begin{aligned} \ker\phi &= \{f \in \widehat{V} : \phi(f) = 0\} \\ &= \{f \in \widehat{V} : \tilde{f} = f\big|_W = 0\} \qquad \text{(where } \tilde{f} \in \widehat{W}\text{)} \\ &= \{f \in \widehat{V} : f(w) = 0 \text{ for all } w \in W\} \\ &= W^0. \end{aligned}$$

Next, we show that ϕ is onto. Let $h \in \widehat{W}$. Then we want to find $f \in \widehat{V}$ such that $\phi(f) = \tilde{f} = h$ i.e., $f\big|_W = h$. Let $\{w_1, \ldots, w_m\}$ be a basis of W. Then by Lemma 1.2.25, it can be extended to a basis $\{w_1, \ldots, w_m, v_1, \ldots, v_r\}$ of V. Let $W_1 = L(\{v_1, \ldots, v_r\})$. Then $V = W \oplus W_1$, i.e., every $v \in V$ can be uniquely written as $v = w + w_1$, where $w \in W$ and $w_1 \in W_1$. For $h \in \widehat{W}$, define a function $f : V \to F$ by $f(v) = f(w + w_1) = h(w)$. Then by definition of f, clearly $f\big|_W = h$, i.e., $\phi(f) = h$. We have to just check that $f \in \widehat{V}$. Let $v, v' \in V$ and $\alpha, \beta \in F$. Then $v = w + w_1$ and $v' = w' + w_1'$ where $w, w' \in W$, $w_1, w_1' \in W_1$ and

$$f(\alpha v + \beta v') = f\big(\alpha(w + w_1) + \beta(w' + w_1')\big)$$

$$
\begin{aligned}
&= f\big((\alpha w+\beta w')+(\alpha w_1+\beta w_1')\big) && \\
&= h(\alpha w+\beta w') && \text{(by definition of } f) \\
&= \alpha h(w)+\beta h(w') && \text{(since } h \text{ is a linear functional on } W) \\
&= \alpha f(v)+\beta f(v') && \text{(by definition of } f).
\end{aligned}
$$

Thus, f is a linear functional on V, i.e., $f \in \widehat{V}$ and by the definition of $\phi(f) = \tilde{f} = h$. Hence, ϕ is onto. We have proved that $\widehat{W} \cong \widehat{V}/W^0$. Hence, by Lemma 1.2.26, we have $\dim \widehat{W} = \dim \widehat{V} - \dim W^0$. Since, V is given to be finite dimensional, we know that $\dim \widehat{V} = \dim V$ and hence,

$$\dim W^0 = \dim V - \dim W.$$

□

From the technique used to prove that ϕ is onto in the above theorem, we can deduce the following result:

Exercise 1.3.18. If V is finite dimensional and W is a subspace of V prove that there is a subspace W_1 of V such that $V = W \bigoplus W_1$.

Solution. Seminar exercise. □

Example 1.3.19. Consider a subspace $W = \{(x,y,z) : x+2y+z = 0\}$ of $\mathbb{R}^3$. Find W^0 and state its dimension.

Solution. By definition of annihilator,

$$W^0 = \{f_{(\alpha,\beta,\gamma)} \in \hat{\mathbb{R}}^3 : f(w) = 0,\ w \in W\}.$$

Note that

$$
\begin{aligned}
W &= \{(x,y,-x-2y) : x,y \in \mathbb{R}\} \\
&= \{x(1,0,-1)+y(0,1,-2) : x,y \in \mathbb{R}\}
\end{aligned}
$$

and hence $\dim W = 2$. Then by above theorem,

$$\dim W^0 = \dim \mathbb{R}^3 - \dim W = 3-2 = 1.$$

Now, let $f_{(\alpha,\beta,\gamma)} \in W^0$. Then

$$f_{(\alpha,\beta,\gamma)}(1,0,-1) = f_{(\alpha,\beta,\gamma)}(0,1,-2) = 0.$$

This implies, $\alpha-\gamma = 0$ and $\beta - 2\gamma = 0$. Therefore, $\alpha = \gamma \quad$ and $\quad \beta = 2\gamma$. Thus,

$$f_{(\alpha,\beta,\gamma)} = f_{(\gamma,2\gamma,\gamma)}.$$

So, $W^0 = \{f_{(\gamma,2\gamma,\gamma)} \in \hat{\mathbb{R}}^3 : \gamma \in \mathbb{R}\}$. □

Exercise 1.3.20. If F is the field of real numbers, then find W^0 where

(a) W is spanned by $(1,2,3)$ and $(0,4,-1)$.
(b) W is spanned by $(0,0,1,-1)$, $(2,1,1,0)$ and $(2,1,1,-1)$.

Solution. Seminar exercise. □

Now, observe that $W \subset V$ and $W^{00} \subset \widehat{\widehat{V}}$ are subspaces of V and its second dual respectively. As such they are not comparable. However, as seen before, we have an identification of V to $\widehat{\widehat{V}}$ by the isomorphism $\psi : V \to \widehat{\widehat{V}}$ defined as $v_0 \mapsto T_{v_0}$. In this sense, we have the following corollary:

Corollary 1.3.21. $W^{00} = W$.

Proof. First we show that $W \subset W^{00}$. For this, let $w \in W$. Then we have to show that $\psi(w) = T_w \in W^{00}$. Observe that for any $f \in W^0$ (i.e., $f(w) = 0$ for all $w \in W$),

$$T_w(f) = f(w) = 0.$$

Thus, $T_w \in \widehat{\widehat{V}}$ such that $T_w(f) = 0$ for all $f \in W^0$. Then by definition of annihilator $T_w \in W^{00}$ and hence $W \subset W^{00}$.

Now, considering W^0 as a subspace of $\widehat{V}$ by Theorem 1.3.17 (above), we have

$$\widehat{W^0} \cong \widehat{\widehat{V}}/W^{00}$$

and hence

$$\begin{aligned} & \dim(\widehat{W^0}) = \dim(\widehat{\widehat{V}}) - \dim(W^{00}) \\ \Rightarrow\ & \dim(W^0) = \dim(V) - \dim(W^{00}) \\ \Rightarrow\ & \dim(V) - \dim(W) = \dim(V) - \dim(W^{00}) \qquad \text{(by previous theorem)} \\ \Rightarrow\ & \dim(W) = \dim(W^{00}). \end{aligned}$$

We have $W \subset W^{00}$ and $\dim(W) = \dim(W^{00})$, it follows that $W = W^{00}$. □

1.3.2 An application to the system of Linear equations

Theorem 1.3.17 has an application to the study of systems of *linear homogeneous equations*. Consider the system of m equations in n unknowns

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= 0, \\ a_{21}x_1 + \cdots + a_{2n}x_n &= 0, \\ \vdots \qquad\quad & \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= 0, \end{aligned}$$

where $a_{ij} \in F$. Let U be the subspace of $F^{(n)}$ spanned by the m vectors $(a_{11}, \ldots, a_{1n})$, $(a_{21}, \ldots, a_{2n}), \ldots, (a_{m1}, \ldots, a_{mn})$. Then the dimension of U is called the rank of the above system. The following theorem gives the number of linearly independent solutions to such a system of rank r:

Theorem 1.3.22. *If the system of homogeneous linear equations:*

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= 0, \\ a_{21}x_1 + \cdots + a_{2n}x_n &= 0, \\ \vdots \qquad & \qquad \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= 0, \end{aligned}$$

where $a_{ij} \in F$ is of rank r, then there are $n-r$ linearly independent solutions in $F^{(n)}$.

Proof. Let $U = L(\{(a_{11}, \ldots, a_{1n}), (a_{21}, \ldots, a_{2n}), \ldots, (a_{m1}, \ldots, a_{mn})\})$ be the subspace of $F^{(n)}$. Since the system is given to be of rank r, we have $\dim U = r$.

Let $v_1 = (1, 0, \ldots, 0)$, $v_2 = (0, 1, 0, \ldots, 0), \ldots, v_n = (0, \ldots, 0, 1)$ be standard basis of $F^{(n)}$ and $\{\hat{v_1}, \hat{v_2}, \ldots, \hat{v_n}\}$ be its corresponding dual basis in $\hat{F}^{(n)}$. Then every $f \in \hat{F}^{(n)}$ can be written as $f = x_1\hat{v_1} + x_2\hat{v_2} + \cdots + x_n\hat{v_n}$, where $x_i \in F$. Now,

$$\begin{aligned} f \in U^0 &\Leftrightarrow f(a_{i1}, a_{i2}, \ldots, a_{in}) = 0 \\ &\Leftrightarrow f(a_{i1}v_1 + a_{i2}v_2 + \cdots + a_{in}v_n) = 0 \\ &\Leftrightarrow (x_1\hat{v_1} + x_2\hat{v_2} + \cdots + x_n\hat{v_n})(a_{i1}v_1 + a_{i2}v_2 + \cdots + a_{in}v_n) = 0 \\ &\Leftrightarrow x_1a_{i1} + x_2a_{i2} + \cdots + x_na_{in} = 0 \end{aligned}$$

since by definition of dual basis $\hat{v_i}(v_j) = 0$ for $i \neq j$ and $\hat{v_i}(v_i) = 1$.
Thus, if $f \in U^0$ then the equations in the system are satisfied. Conversely, if $(x_1, \ldots, x_n)$ is a solution then there is an element $x_1\hat{v_1} + x_2\hat{v_2} + \cdots + x_n\hat{v_n}$ in U^0. Hence, we conclude that the number of linearly independent solutions of the above system is same as $\dim U^0$. By Theorem 1.3.17,

$$\dim U^0 = \dim F^{(n)} - \dim U = n - r.$$

□

CHAPTER 2

LINEAR TRANSFORMATIONS

We have seen that $\mathrm{Hom}(V,W)$ is a vector space over the field F given that V and W are vector spaces over F. Furthermore, if V and W are finite dimensional then we know that

$$\dim(\mathrm{Hom}(V,W)) = \dim(V)\times\dim(W).$$

In the last chapter, we considered its special case when $W = F$, i.e., $\mathrm{Hom}(V,F) = \widehat{V}$ with $\dim(\mathrm{Hom}(V,F)) = \dim V$ if V is finite dimensional over F.

In this chapter, we will concentrate on the case where $W = V$, i.e. $\mathrm{Hom}(V,V)$, where V is considered to be a finite-dimensional vector space over F. Thus, $\mathrm{Hom}(V,V)$, the set of all homomorphisms of V into itself, is a vector space over F under the following operations: for $T_1,T_2 \in \mathrm{Hom}(V,V)$ and $\alpha \in F$,

$$(T_1+T_2)(v) = T_1(v)+T_2(v)$$
$$(\alpha T_1)(v) = \alpha(T_1(v)).$$

Also, if V is finite dimensional over F then

$$\dim(\mathrm{Hom}(V,V)) = (\dim V)^2.$$

2.1 The Algebra of Linear Transformations

Let V be a vector space over a field F. Let $T_1,T_2 \in \mathrm{Hom}(V,V)$. Since $T_2(v)\in V$ for any $v\in V$, $T_1(T_2(v))$ makes sense. So, we define the product of two linear transformations T_1 and T_2 by

$$(T_1T_2)(v) = T_1(T_2(v)) \qquad \text{for any } v\in V.$$

We have to check that $T_1T_2 \in \mathrm{Hom}(V,V)$, i.e., T_1T_2 is a homomorphism from V to V. For any $u,v\in V$ and $\alpha,\beta\in F$

$$(T_1T_2)(\alpha u+\beta v) = T_1(T_2(\alpha u+\beta v))$$

$$
\begin{aligned}
&= T_1(\alpha T_2(u) + \beta T_2(v)) && (\text{since } T_2 \in \mathrm{Hom}(V,V)) \\
&= \alpha T_1(T_2(u)) + \beta T_1(T_2(v)) && (\text{since } T_1 \in \mathrm{Hom}(V,V)) \\
&= \alpha (T_1T_2)(u) + \beta (T_1T_2)(v).
\end{aligned}
$$

As an easy exercise one can also check the following properties of the product in $\mathrm{Hom}(V,V)$:

1. $T_1(T_2+T_3) = T_1T_2 + T_1T_3$;
2. $(T_2+T_3)T_1 = T_2T_1 + T_3T_1$;
3. $T_1(T_2T_3) = (T_1T_2)T_3$;
4. $\alpha(T_1T_2) = (\alpha T_1)T_2 = T_1(\alpha T_2)$;

for all $T_1, T_2, T_3 \in \mathrm{Hom}(V,V)$ and all $\alpha \in F$.

Properties 1, 2 and 3 makes $\mathrm{Hom}(V,V)$ an associative ring. Property 4 involves both scalar multiplication and product in $\mathrm{Hom}(V,V)$ and hence connects its character as a vector space and an associative ring. In addition to these, $\mathrm{Hom}(V,V)$ has a unit element with respect to multiplication. Let $I \in \mathrm{Hom}(V,V)$ defined by $I(v) = v$ for all $v \in V$. Then I is the unit element as for any $T \in \mathrm{Hom}(V,V)$, one has $IT = TI = T$. This makes $\mathrm{Hom}(V,V)$ a ring with unit element.

Thus, $\mathrm{Hom}(V,V)$ is an associative ring as well as a vector space over F. We call such an algebraic structure as an *algebra over F*, formally defined as follows:

Definition 2.1.1. An associative ring A is called an *algebra over F* if A is a vector space over the field F such that for all $a, b \in A$ and all $\alpha \in F$, $\alpha(ab) = (\alpha a)b = a(\alpha b)$.

Examples 2.1.2. We consider some examples of algebras (vector spaces which are also associative rings):

1. As seen above, $\mathrm{Hom}(V,V)$ is an algebra over F. From now and in what follows, we use the notation $A(V)$ for $\mathrm{Hom}(V,V)$. Thus, $A(V)$ is an algebra (over F) of all homomorphisms (linear transformations) of V into V. Sometimes we shall also denote it by $A_F(V)$ to emphasis the role of F.
2. The vector space $F^{(n)}$ is an algebra over F where the product of any two elements $(a_1, a_2, \ldots, a_n), (b_1, b_2, \ldots, b_n) \in F^{(n)}$ is defined as follows:
$$(a_1, a_2, \ldots, a_n)(b_1, b_2, \ldots, b_n) = (a_1b_1, a_2b_2, \ldots, a_nb_n).$$
3. The vector space $F[x]$ of all polynomials with coefficients from a field F becomes an algebra with the product defined as product of two polynomials.

Remark 2.1.3. The vector space $F_n[x]$, i.e. the space of all polynomials with coefficients in a field F and of degree at most n, fails to be an algebra as it is not closed under multiplication. $x^n \times x^n = x^{2n} \notin F_n[x]$.

Theorem 2.1.4. *If $\mathscr{A}$ is an algebra over F with unit element then $\mathscr{A}$ is isomorphic to a subalgebra of $A(V)$ for some vector space V over F.*

Proof. Since $\mathscr{A}$ is an algebra over F it is also a vector space over F. We shall use $V = \mathscr{A}$ to prove the theorem i.e., we show that $\mathscr{A}$ is isomorphic to a subalgebra of $A(\mathscr{A})$.

For $a \in \mathscr{A}$, let $T_a : \mathscr{A} \to \mathscr{A}$ by $T_a(v) = av$, $(v \in \mathscr{A})$. We now show that T_a is a linear transformation on $\mathscr{A}$. For $v_1, v_2 \in \mathscr{A}$, and $\alpha \in F$,

$$T_a(v_1 + v_2) = a(v_1 + v_2) = av_1 + av_2 = T_a(v_1) + T_a(v_2).$$
$$T_a(\alpha v_1) = a(\alpha v_1) = \alpha(av_1) = \alpha T_a(v_1).$$

Thus, $T_a \in A(\mathscr{A})$.

Now, define a map $\psi : \mathscr{A} \to A(\mathscr{A})$ by $\psi(a) = T_a$ for every $a \in \mathscr{A}$. We prove the theorem by showing that ψ is an isomorphism of $\mathscr{A}$ into $A(\mathscr{A})$. First we show that ψ is an (algebra) homorphism. For this we have to show that for any $a, b \in \mathscr{A}$ and $\alpha, \beta \in F$,

$$\psi(\alpha a + \beta b) = \alpha \psi(a) + \beta \psi(b), \qquad \text{(for vector space homomorphism)}$$
$$\text{i.e.} \qquad T_{\alpha a + \beta b} = \alpha T_a + \beta T_b.$$

and

$$\psi(ab) = \psi(a)\psi(b), \qquad \text{(for ring homomorphism)}$$
$$\text{i.e., } T_{ab} = T_a T_b.$$

Now, for every $v \in \mathscr{A}$,

$$T_{\alpha a + \beta b}(v) = (\alpha a + \beta b)v = \alpha(av) + \beta(bv) = \alpha T_a(v) + \beta T_b(v) = (\alpha T_a + \beta T_b)(v).$$

Also,

$$T_{ab}(v) = (ab)v = a(bv) = T_a(T_b(v)) = (T_a T_b)(v).$$

Thus, ψ is an algebra homomorphism of $\mathscr{A}$ into $A(\mathscr{A})$.

Now, we show that ψ is one-one. For this we shall show that $\ker \psi = \{0\}$.

$$\begin{aligned} \ker \psi &= \{a \in \mathscr{A} : \psi(a) = 0\} \\ &= \{a \in \mathscr{A} : T_a = 0\} \\ &= \{a \in \mathscr{A} : T_a(v) = 0 \text{ for all } v \in \mathscr{A}\}. \end{aligned}$$

Thus, we have $a \in \ker \psi$ if $T_a(v) = 0$ for all $v \in \mathscr{A}$. Since $\mathscr{A}$ is an algebra there is a unit element $e \in \mathscr{A}$. In particular, $T_a(e) = 0$ which implies $ae = a = 0$. Hence, kernel of ψ consists of only one element 0. This proves that ψ is a one-one map and hence it is an isomorphism. □

Definition 2.1.5. Let $\mathscr{A}$ be an algebra over F with unit element e and let $p(x) = \alpha_0 + \alpha_1 x + \cdots + \alpha_n x^n$ be a polynomial in $F[x]$. Then for any $a \in \mathscr{A}$, by $p(a)$ we mean the element $\alpha_0 e + \alpha_1 a + \cdots + \alpha_n a^n$ in $\mathscr{A}$. We say that *a satisfies* $p(x)$ if $p(a) = 0$.

Lemma 2.1.6. *Let $\mathscr{A}$ be an algebra, with unit element, over F. Suppose that dimension of $\mathscr{A}$ is m over F. Then every element in $\mathscr{A}$ satisfies some nontrivial polynomial in $F[x]$ of degree at most m.*

Proof. Let e be the unit element of $\mathscr{A}$ and let $a \in \mathscr{A}$. Consider $m+1$ elements $e, a, a^2, \ldots, a^m$ in $\mathscr{A}$. Since, $\dim \mathscr{A} = m$, these $m+1$ elements must be linearly dependent over F. This implies, there exists $\alpha_0, \alpha_1, \ldots, \alpha_m$ in F, not all zero, such that

$$\alpha_0 e + \alpha_1 a + \cdots + \alpha_m a^m = 0.$$

This means that a satisfies the nontrivial polynomial $q(x) = \alpha_0 + \alpha_1 x + \cdots + \alpha_m x^m$ of degree at most m in $F[x]$. □

Theorem 2.1.7. *If V is an n-dimensional vector space over F, then given any element $T \in A(V)$, there exists a nontrivial polynomial $q(x) \in F[x]$ of degree at most n^2 such that $q(T) = 0$.*

Proof. Here $\dim V = n$ over F. Then $A(V)$ is an algebra and $\dim A(V) = n^2$ over F. By previous lemma, every element $T \in A(V)$ satisfies a polynomial $q(x) \in F[x]$ of degree at most n^2, i.e., $q(T) = 0$. □

Definition 2.1.8. Let V be a finite dimensional vector space over F. A nontrivial polynomial $p(x)$ in $F[x]$ of lowest degree such that $p(T) = 0$ is called a *minimal polynomial* for T over F.

If T satisfies another polynomial $h(x)$ then $p(x)|h(x)$.

Exercise 2.1.9. Let V be a finite dimensional vector space over F and $T \in A(V)$. Then $p(x) \in F[x]$ is a minimal polynomial for T if and only if for any other polynomial, say $h(x)$, satisfied by T, we have $p(x)|h(x)$.

Solution. Hint: Use division algorithm. Given as a seminar exercise. □

Definition 2.1.10. An element $T \in A(V)$ is said to be *right invertible* if there exists an $S \in A(V)$ such that $TS = 1$. Here, 1 denotes the unit element of $A(V)$, i.e. $1(v) = v$ for all $v \in V$.

Similarly, an element $T \in A(V)$ is said to be *left invertible* if there exists a $U \in A(V)$ such that $UT = 1$.

Exercise 2.1.11. If $T \in A(V)$ is both left invertible and right invertible then prove that the left inverse and the right inverse must be equal and that the inverse is unique.

Definition 2.1.12. An element $T \in A(V)$ is said to be *invertible* or *regular* if it is both left invertible and right invertible; that is, if there is an element $S \in A(V)$ such that $ST = TS = 1$. We write S as T^{-1}.

An element in $A(V)$ is said to be *singular* if it is not regular, i.e., if it is not invertible.

It may be possible that an element in $A(V)$ is left invertible but not right invertible or vice-versa. We consider one such instance in the following example:

Example 2.1.13. Consider the vector space $V = \mathbb{R}[x]$, the set of all polynomials in x with real coefficients, over the field $\mathbb{R}$. For any $q(X) \in V$, $q(x) = a_0 + a_1x + \cdots + a_nx^n$ define elements $S, T \in A(V)$ by

$$S(q(x)) = \frac{d}{dx}q(x)$$
$$T(q(x)) = \int_1^x q(x)dx.$$

Then

$$\begin{aligned}
ST(q(x)) &= S(T(q(x)) \\
&= S\left(a_0\int_1^x 1dx + a_1\int_1^x xdx + \cdots + a_n\int_1^x x^n dx\right) \\
&= S\left((a_0x - a_0) + \left(a_1\frac{x^2}{2} - \frac{a_1}{2}\right) + \cdots + \left(a_n\frac{x^{n+1}}{n+1} - \frac{a_n}{n+1}\right)\right) \\
&= a_0 + a_1x + \cdots + a_nx^n = q(x).
\end{aligned}$$

Thus, $ST(q(x)) = q(x)$ for all $q(x) \in V$ which means $ST = 1$. However, $TS \neq 1$ as

$$\begin{aligned}
TS(q(x)) &= T(a_1 + 2a_2x + \cdots + na_nx^{n-1}) \\
&= a_1\int_1^x 1dx + 2a_2\int_1^x xdx + \cdots + na_n\int_1^x x^{n-1}dx \\
&= (a_1x - a_1) + (a_2x^2 - a_2) + \cdots + (a_nx^n - a_n) \\
&\neq q(x).
\end{aligned}$$

Thus, T is left invertible but not right invertible. In other words, we can say that S is right invertible but not left invertible.

Remark 2.1.14. Notice that in above example $V = \mathbb{R}[x]$ is an infinite dimensional vector space. However, it is not possible that a linear transformation only left invertible or only right invertible in case of finite dimensional vector space. We shall show that, for finite-dimensional vector space V, an element in $A(V)$ which is left invertible or right invertible is invertible.

Theorem 2.1.15. *If V is finite dimensional over F, then $T \in A(V)$ is invertible if and only if the constant term in the minimal polynomial for T is not* 0.

Proof. Let $p(x) = \alpha_0 + \alpha_1x + \cdots + \alpha_kx^k \in F[x]$, $\alpha_k \neq 0$ be the minimal polynomial for T over F. First assume that the constant term in $p(x)$ is non-zero i.e., $\alpha_0 \neq 0$. Since, T satisfies $p(x)$, we have $P(T) = 0$. This implies,

$$\begin{aligned}
& \alpha_kT^k + \alpha_{k-1}T^{k-1} + \cdots + \alpha_1T + \alpha_0I = 0 \\
\Rightarrow\quad & T(\alpha_kT^{k-1} + \alpha_{k-1}T^{k-2} + \cdots + \alpha_1) = -\alpha_0I \\
\Rightarrow\quad & T\left(-\tfrac{1}{\alpha_0}(\alpha_kT^{k-1} + \alpha_{k-1}T^{k-2} + \cdots + \alpha_1)\right) = I
\end{aligned}$$

Similarly,

$$\left(-\frac{1}{\alpha_0}(\alpha_kT^{k-1} + \alpha_{k-1}T^{k-2} + \cdots + \alpha_1)\right)T = I$$

Therefore,

$$S = -\frac{1}{\alpha_0}(\alpha_k T^{k-1} + \cdots + \alpha_1 I)$$

acts as an inverse of T and hence T is invertible.

Conversely, assume that T is invertible then we have to show that $\alpha_0 \neq 0$. Suppose if possible, $\alpha_0 = 0$. Then $p(T) = 0$ implies

$$\alpha_1 T + \alpha_2 T^2 + \cdots + \alpha_k T^k = T(\alpha_1 I + \alpha_2 T^2 + \cdots + \alpha_k T^{k-1}) = 0.$$

Since, T is invertible, multiplying by T^{-1} to the left on both sides of the above relation, we get $\alpha_1 I + \alpha_2 T + \cdots + \alpha_k T^{k-1} = 0$. This means T satisfies the polynomial $q(x) = \alpha_1 + \alpha_2 x + \cdots + \alpha_k x^{k-1}$ with $\deg(q(x)) = k-1 < k = \deg(p(x))$. This is contradiction as $p(x)$ is the minimal polynomial of T. So our assumption that $\alpha_0 = 0$ must be false and hence $\alpha_0 \neq 0$. □

Corollary 2.1.16. *Let V be a finite dimensional vector space over F. If $T \in A(V)$ is invertible then T^{-1} is a polynomial expression in T over F.*

Proof. Since T is invertible, by above theorem, $\alpha_0 I + \alpha_1 T + \cdots + \alpha_k T^k =$ with $\alpha_0 \neq 0$. Then, as seen in the theorem, $TT^{-1} = I$ where

$$T^{-1} = -\frac{1}{\alpha_0}(\alpha_1 I + \alpha_2 T + \cdots + \alpha_k T^{k-1}) \qquad (\alpha_i \in F).$$

Clearly, T^{-1} is a polynomial expression in T over F. □

Corollary 2.1.17. *Let V be a finite dimensional vector space over F. If $T \in A(V)$ is singular, then there exists an $S \neq 0$ in $A(V)$ such that $ST = TS = 0$.*

Proof. Since T is singular (not invertible), by above theorem, the constant term in the minimal polynomial for T must be zero. Then the minimal polynomial for T will be of the form $p(x) = \alpha_1 x + \alpha_2 x^2 + \cdots + \alpha_k x^k$. Hence, $p(T) = 0$ implies

$$T(\alpha_1 I + \alpha_2 T + \cdots \alpha_k T^{k-1}) = (\alpha_1 I + \alpha_2 T + \cdots \alpha_k T^{k-1})T = 0.$$

If we take $S = \alpha_1 I + \alpha_2 T + \cdots \alpha_k T^{k-1}$ then $S \neq 0$ as $\alpha_1 + \alpha_2 x + \cdots \alpha_k x^{k-1}$ is of lower degree than $p(x)$. Hence, we found $S \in A(V)$ and $S \neq 0$ such that $ST = TS = 0$. □

Corollary 2.1.18. *Let V be a finite dimensional vector space over F. If $T \in A(V)$ is right invertible (or left invertible) then T is invertible.*

Proof. Suppose T is right invertible. That is there exists a $U \in A(V)$ such that $TU = I$. Let if possible, T is not invertible (singular). Then by above corollary, there exists an $S \in A(V)$, $S \neq 0$ such that $ST = TS = 0$. Then, we have

$$0 = ST = (ST)U = S(TU) = S1 = S \neq 0,$$

which is a contradiction. So, T must be regular. □

Theorem 2.1.19. *Let V be a finite dimensional vector space over F. Then $T \in A(V)$ is singular if and only if there exists a $v \in V$, $v \neq 0$ such that $T(v) = 0$.*

*In other words, T is **regular** if and only if it is **one-one** (or $\ker T = \{0\}$).*

Proof. First, let us assume that T is singular. We know that, T is singular if and only if there exists an $S \neq 0$ in $A(V)$ such that $ST = TS = 0$. Since $S \neq 0$, there is an element $w \in V$ such that $S(w) \neq 0$. Let $v = S(w)$. Then $T(v) = T(S(w)) = (TS)(w) = 0$ (since $TS = 0$). This proves the first part.

Conversely, assume that $T(v) = 0$ for some $v \neq 0$ in V. We have to show that T is singular. Suppose, on the contrary, that T is regular. Then since $T(v) = 0$, multiplying both sides by T^{-1}, we have $T^{-1}(T(v)) = T^{-1}0 = 0$ and hence $v = 0$. This is a contradiction since we have assumed $v \neq 0$. Thus, T must be singular. □

Definition 2.1.20. If $T \in A(V)$, then the *range* of T is denoted by $T(V)$ and defined by

$$T(V) = \{T(v) \mid v \in V\}.$$

Clearly, the range of T is a subspace of V. Note that the range of T is all of V, i.e., $T(V) = V$ if and only if T is onto.

Theorem 2.1.21. *If V is a finite dimensional vector space over F, then $T \in A(V)$ is **regular** if and only if T maps V **onto** V.*

Proof. First we consider that T is regular and show that T is onto. Since T is regular, for any $v \in V$ (co-domain), we can write it as $v = T(T^{-1}(v))$, where $T^{-1}(v) \in V$ (domain). Thus $T(V) = V$.

Conversely given that T is onto we have to show that T is regular. Suppose, if possible, T is singular. Then there exist a $v_1 \neq 0$ in V such that $T(v_1) = 0$. Since $v_1 \neq 0$, the singleton set $\{v_1\}$ is linearly independent. We extend it to a basis $\{v_1, v_2, \dots, v_n\}$ of V. Then $\{T(v_1), T(v_2), \dots, T(v_n)\}$ is a basis of $T(V)$ and so every element in $T(V)$ is spanned by the elements $w_1, w_2, \dots, w_n$ where $w_i = T(v_i)$ for $i = 1, 2, \dots, n$. But, we have $w_1 = T(v_1) = 0$. Thus, $T(V)$ is spanned by only $n-1$ elements, $w_2, \dots, w_n$. Therefore, $\dim T(V) \leq n-1 < n = \dim V$. Thus, $T(V) \neq V$ and hence T is not onto, which is a contradiction. So, T must be regular. □

Remark 2.1.22. Above theorem indicates the difference between regular elements and singular elements of $A(V)$ in terms their range, in finite dimensional cases. In other words, by above theorem, $T \in A(V)$ is regular if and only if $\dim T(V) = \dim V$. Thus, we can compute and use $\dim(T(V))$ to check whether given $T \in A(V)$ is regular or singular.

Combining Theorems 2.1.19 and 2.1.21, when V is finite dimensional, we can say that T is regular if and only if T is both one-one and onto. In other words, if T fails to be either one-one or onto or both then T must be singular.

Definition 2.1.23. Let V be a finite dimensional vector space over F and $T \in A(V)$. Then the dimension of the range of T over F is called the *rank* of T. We denote the rank of T by

$r(T)$. That is,

$$r(T) = \dim(T(V)).$$

Exercise 2.1.24. Prove that $S \in A(V)$ is regular if and only if whenever $v_1, \dots, v_n \in V$ are linearly independent, then $S(v_1), S(v_2), \dots, S(v_n)$ are also linearly independent.

Solution. Seminar exercise. □

Lemma 2.1.25. *Let V be a finite dimensional vector space over F and $S, T \in A(V)$. Then*

1. *$r(TS) \leq r(T)$;*
2. *$r(ST) \leq r(T)$;*
3. *If S is regular then $r(ST) = r(TS) = r(T)$.*

Proof.

1. Since $S(V) \subset V$, $(TS)(V) = T(S(V)) \subset T(V)$. By Lemma 1.2.26, $\dim((TS)(V)) \leq \dim(T(V))$, i.e., $r(TS) \leq r(T)$.
2. Suppose that $r(T) = m$. Therefore, $T(V)$ has a basis of m elements $w_1, \dots, w_m$ and so

$$T(V) = L(\{w_1, \dots, w_m\}).$$

But then

$$\begin{aligned} S(T(V)) &= S\big(L(\{w_1, \dots, w_m\})\big) \\ &= L(\{Sw_1, \dots, Sw_m\}) \end{aligned}$$

Then, dimension of $(ST)(V) = S(T(V))$ is at most m. Hence

$$r(ST) = \dim((ST)(V)) \leq m = \dim(T(V)) = r(T).$$

3. Since S is regular, by Theorem 2.1.21, $S(V) = V$. Therefore, $(TS)(V) = T(S(V)) = T(V)$ and hence $r(TS) = r(T)$.
On the other hand, if $r(T) = m$ and $\{w_1, w_2, \dots, w_m\}$ is a basis of $T(V)$. Then,

$$T(V) = L(\{w_1, \dots, w_m\}).$$

Then as before,

$$\begin{aligned} S(T(V)) &= S\big(L(\{w_1, \dots, w_m\})\big) \\ &= L(\{Sw_1, \dots, Sw_m\}) \end{aligned}$$

Since S is regular, it maps linearly independent set to a linearly independent set (by above exercise). So, $\{Sw_1, \dots, Sw_m\}$ is linearly independent and hence forms a basis for $S(T(V))$ over F. Hence,

$$r(ST) = \dim(ST)(V) = \dim(S(T(V))) = m = \dim(T(V)) = r(T).$$

□

Corollary 2.1.26. *Let V be finite dimensional over F and $S,T \in A(V)$. Then*

1. $r(ST) \leq \min\{r(T), r(S)\}$.
2. *If S is regular then $r(T) = r(S^{-1}TS)$.*

Proof. First part easily follows from the above theorem. We shall show the second consequence. If S is regular, then by above theorem, we have

$$r(S^{-1}TS) = r((S^{-1}T)S) = r(S(S^{-1}T)) = r((S^{-1}S)T) = r(T).$$

□

Definition 2.1.27. Let V be a finite dimensional vector space over F. Two elements $S,T \in A(V)$ are said to be *similar* if there exists a regular element $C \in A(V)$ such that $S = C^{-1}TC$. If S and T are similar, we denote it by $S \sim T$.

Exercise 2.1.28. Show that '$\sim$' defined above is an equivalence relation, i.e., being similar is an equivalence relation on $A(V)$.

Solution. Seminar exercise. □

Exercise 2.1.29. If $S,T \in A(V)$ are similar then show that they have the same rank i.e., $r(S) = r(T)$. What about the converse?

2.2 Characteristic Roots

Definition 2.2.1. Let V be a vector space over F and $T \in A(V)$. A scalar $\lambda \in F$ is called a *characteristic root* (or *eigenvalue*) of T if there is a non-zero vector v in V such that $Tv = \lambda v$.

The vector $v \neq 0$ in V such that $Tv = \lambda v$ is called the *characteristic vector* (or *eigenvector*) of T corresponding to the eigenvalue λ.

Proposition 2.2.2. *Let V be a finite dimensional vector space over F. If $T \in A(V)$ and $\lambda \in F$, then λ is a characteristic root of T if and only if $T - \lambda I$ is singular.*

Proof. Suppose λ is a characteristic root of T then there exists a $v \neq 0$ in V such that $Tv = \lambda v$. Then $(T - \lambda I)(v) = 0$ which implies that T is not one-one. Hence, $T - \lambda I$ is singular.

Conversely, assume that $T - \lambda I$ is singular. Then $T - \lambda I$ is not one-one. So, there exists a $v \neq 0$ such that $(T - \lambda I)(v) = 0$. Then $Tv = \lambda v$ and hence λ is a characteristic of T. □

Lemma 2.2.3. *Let V be a finite dimensional vector space over F and $T \in A(V)$. If $\lambda \in F$ is a characteristic root of T, then for any polynomial $q(x) \in F[x]$, $q(\lambda)$ is a characteristic root of $q(T)$.*

Proof. Since λ is a characteristic root of T, there exists $v \in V$, $v \neq 0$ such that $Tv = \lambda v$. Then

$$T^2v = T(Tv) = T(\lambda v) = \lambda T(v) = \lambda(\lambda v) = \lambda^2 v.$$

Continuing this way, we get $T^k v = \lambda^k v$. Now, let $q(x) = \alpha_0 + \alpha_1 x + \cdots + \alpha_n x^n$ be any polynomial in $F[x]$. Then,

$$\begin{aligned} q(T)v &= (\alpha_0 I + \alpha_1 T + \cdots + \alpha_n T^n)v \\ &= \alpha_0 Iv + \alpha_1 Tv + \cdots + \alpha_n T^n v \\ &= \alpha_0 v + \alpha_1 \lambda v + \cdots + \alpha_n \lambda^n v \\ &= (\alpha_0 + \alpha_1 \lambda + \cdots + \alpha_n \lambda^n)v = q(\lambda)v. \end{aligned}$$

Thus, we have $q(T)v = q(\lambda)v$ which means that $q(\lambda)$ is a characteristic root of $q(T)$. □

Corollary 2.2.4. *Let V be a finite dimensional vector space over F and $T \in A(V)$. If $\lambda \in F$ is a characteristic root of T then λ is a root of the minimal polynomial of T.*

Proof. Let $p(x) \in F[x]$ be the minimal polynomial of T. Since λ is a characteristic root of T, there exists a $v \neq 0$ in V such that $Tv = \lambda v$. By above theorem, $p(T)v = p(\lambda)v$. Since $p(T) = 0$, we have $p(\lambda)(v) = 0 \Rightarrow p(\lambda) = 0$ (by properties of vector space and as $v \neq 0$). Hence, λ is a root of $p(x)$. □

Corollary 2.2.5. *Let V be an n-dimensional vector space over F and $T \in A(V)$. Then the number of characteristic roots of T is at most n^2.*

Proof. Since $\dim V = n$, as we have seen before there exits a $q(x) \in F[x]$ of degree at most n^2 such that $q(T) = 0$.

Let $p(x) \in F[x]$ be the minimal polynomial of T. Then

$$\deg(p(x)) \leq \deg(q(x)) \leq n^2.$$

Therefore, the number of roots of $p(x)$ is finite and at most n^2.

By above corollary, if λ is a characteristic root of T then λ is a root of $p(x)$. As the number of roots of $p(x)$ is at most n^2, the number of characteristic roots of T is also at most n^2. □

Theorem 2.2.6. *Let V be a finite dimensional vector space over F. Let $S, T \in A(V)$ and let S be regular. Then*

1. *T and $S^{-1}TS$ have the same minimal polynomial.*
2. *$\lambda \in F$ is a characteristic root of T if and only if it is a characteristic root of $S^{-1}TS$.*

Proof. 1. Since S is regular, $(S^{-1}TS)^2 = (S^{-1}TS)(S^{-1}TS) = S^{-1}T^2S$. Similarly, for any $k \in \mathbb{N}$, we have

$$(S^{-1}TS)^k = S^{-1}T^kS.$$

Now, let $q(x)=\alpha_0+\alpha_1x+\alpha_2x^2\cdots+\alpha_nx^n$ be any polynomial in $F[x]$. Then

$$\begin{aligned} q(S^{-1}TS) &= \alpha_0 I+\alpha_1(S^{-1}TS)+\alpha_2(S^{-1}TS)^2+\cdots+\alpha_n(S^{-1}TS)^n \\ &= \alpha_0(S^{-1}S)+\alpha_1(S^{-1}TS)+\alpha_2(S^{-1}T^2S)+\cdots+\alpha_n(S^{-1}T^nS) \\ &= S^{-1}(\alpha_0 I+\alpha_1 T+\alpha_2T^2+\cdots+\alpha_nT^n)S \\ &= S^{-1}q(T)S. \end{aligned}$$

Thus, $q(S^{-1}TS)=0$ if and only if $q(T)=0$. Therefore, T and $S^{-1}TS$ have the same minimal polyomial.

2. Let λ be a characteristic root of T. Then $Tv=\lambda v$ for some $v\neq 0$ in V. Since S is regular, by Theorem 2.1.19, S is one-one and since $v\neq 0$, we have $Sv\neq 0$. Let $u=S^{-1}v\neq 0$. Now,

$$\begin{aligned} (S^{-1}TS)(u) &= (S^{-1}TS)(S^{-1}v) \\ &= S^{-1}(Tv) \\ &= S^{-1}(\lambda v) \\ &= \lambda(S^{-1}v) \\ &= \lambda u. \end{aligned}$$

Thus, $(S^{-1}TS)(u)=\lambda u$. Therefore, λ is a characteristic root of $S^{-1}TS$.
Conversely, assume that λ is a characteristic root of $S^{-1}TS$. Then there exists a $v\neq 0$ in V such that $(S^{-1}TS)(v)=\lambda v$. Applying S to the left on both sides, we get

$$\begin{aligned} (TS)(v) &= S(\lambda v) \\ \Rightarrow T(Sv) &= \lambda(Sv). \end{aligned}$$

Since S is regular, it is one-one and since $v\neq 0$, $u=Sv\neq 0$. This implies, $Tu=\lambda u$ and hence λ is a characteristic root of T.

□

The following theorem gives the relation between the characteristic vectors of $T\in A(V)$ corresponding to distinct characteristic roots of T.

Theorem 2.2.7. *Let V be a finite dimensional vector space over F and $T\in A(V)$. If $\lambda_1,\lambda_2,\ldots,\lambda_k$ in F are distinct characteristic roots of T and $v_1,v_2,\ldots,v_k$ are characteristic vectors of T corresponding to $\lambda_1,\lambda_2,\ldots,\lambda_k$ respectively, then $v_1,v_2,\ldots,v_k$ are linearly independent.*

Proof. We give the proof by applying Principle of Mathematical Induction on k. Clearly, the result holds for $k=1$ as $v_1(\neq 0)$ is linearly independent.

Assume that the result holds for $k-1$. That is, if $v_1,v_2,\ldots,v_{k-1}$ are characteristic vectors of T corresponding to distinct characteristic roots $\lambda_1,\lambda_2,\ldots,\lambda_{k-1}$ respectively, then $v_1,v_2,\ldots,v_{k-1}$ are linearly independent.

Now, we prove the result for k. Let $v_1,v_2,\ldots,v_k$ be the characteristic vectors of T corresponding to distinct characteristic roots $\lambda_1,\lambda_2,\ldots,\lambda_k$ respectively. We have to show that $v_1,v_2,\ldots,v_k$ are linearly independent. Let $\alpha_1,\alpha_2,\ldots,\alpha_k\in F$ such that

$$\alpha_1v_1+\alpha_2v_2+\cdots+\alpha_kv_k=0. \tag{2.1}$$

Then we have to show that $\alpha_i = 0$ for all $i = 1,2,\dots,k$. Applying T, we get

$$\alpha_1 T v_1 + \alpha_2 T v_2 + \cdots + \alpha_k T v_k = 0.$$

Since $T v_i = \lambda_i v_i$, we have

$$\alpha_1 \lambda_1 v_1 + \alpha_2 \lambda_2 v_2 + \cdots + \alpha_k \lambda_k v_k = 0. \tag{2.2}$$

Multiplying equation (2.1) by λ_k and subtracting it from (2.2), we get

$$\alpha_1(\lambda_1 - \lambda_k)v_1 + \alpha_2(\lambda_2 - \lambda_k)v_2 + \cdots + \alpha_{k-1}(\lambda_{k-1} - \lambda_k)v_{k-1} = 0.$$

By induction hypothesis, since $v_1, v_2, \dots, v_{k-1}$ are linearly independent, we conclude that

$$\alpha_i(\lambda_i - \lambda_k) = 0 \qquad \text{for all } i = 1,2,\dots,k-1.$$

As, $\lambda_1, \lambda_2, \dots, \lambda_k$ are distinct characteristic roots, $\lambda_i - \lambda_k \neq 0$ and hence $\alpha_i = 0$ for all $i = 1,2,\dots,k-1$. Then equation (2.1) reduces to

$$\alpha_k v_k = 0.$$

Since, $v_k \neq 0$, we have $\alpha_k = 0$ and hence $v_1, v_2, \dots, v_k$ are linearly independent. This proves the result by induction. □

Corollary 2.2.8. *If $T \in A(V)$ and $\dim V = n$, then T can have at most n distinct characteristic roots.*

Proof. Any set of distinct characteristic roots of T gives a corresponding set of linearly independent characteristic vectors of T by above theorem. Since $\dim V = n$, any linearly independent set can have at most n elements. Hence, T can have at most n distinct characteristic roots. □

Corollary 2.2.9. *If $T \in A(V)$ and $\dim V = n$. If T has n distinct characteristic roots in F, then there is a basis of V over F consisting of characteristic roots of T.*

Proof. Let $\lambda_1, \lambda_2, \dots, \lambda_n$ in F be the n distinct characteristic roots of T. Then by above theorem, the set of corresponding characteristic vectors, $\{v_1, v_2, \dots, v_n\}$ is linearly independent. Since, $\dim V = n$, the set $\{v_1, v_2, \dots, v_n\}$ forms a basis of V over F. Thus, V has a basis consisting of characteristic vectors of T. □

2.3 Matrices

Let V be an n-dimensional vector space over F with basis $\{v_1, v_2, \dots, v_n\}$. If $T \in A(V)$ then T is completely determined by its value on the basis of V. Since $T : V \to V$, the elements

$Tv_1, Tv_2, \ldots, Tv_n$ must all be in V and hence can be uniquely expressed as a linear combination of $v_1, v_2, \ldots, v_n$ over F. Thus,

$$\begin{aligned} Tv_1 &= \alpha_{11}v_1 + \alpha_{21}v_2 + \cdots + \alpha_{n1}v_n \\ Tv_2 &= \alpha_{12}v_1 + \alpha_{22}v_2 + \cdots + \alpha_{n2}v_n \\ &\vdots \\ Tv_n &= \alpha_{1n}v_1 + \alpha_{2n}v_2 + \cdots + \alpha_{nn}v_n, \end{aligned}$$

where $\alpha_{ij} \in F$, $1 \leq i, j \leq n$.

This system of equations can be compactly written as follows:

$$Tv_j = \sum_{i=1}^{n} \alpha_{ij}v_i \qquad \text{for } j = 1, 2, \ldots, n.$$

Thus, the ordered set of n^2 elements $\alpha_{ij} \in F$ completely describes T.

Definition 2.3.1. Let V be an n-dmensional vector space over F with basis $v_1, v_2, \ldots, v_n$. If $T \in A(V)$ then the matrix of T in the basis $v_1, v_2, \ldots, v_n$ is denote by $m(T)$ and defined as

$$m(T) = \begin{pmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{n1} & \alpha_{n2} & \cdots & \alpha_{nn} \end{pmatrix},$$

where $Tv_j = \sum_{i=1}^{n} \alpha_{ij}v_i$. Instead of writing the whole square matrix every time we shall also denote it as (α_{ij}). Thus,

$$m(T) = (\alpha_{ij}), \text{ where } Tv_j = \sum_{i=1}^{n} \alpha_{ij}v_i.$$

Note: Note that if $Tv_j = \sum_{i=1}^{n} \alpha_{ij}v_i = \alpha_{1j}v_1 + \alpha_{2j}v_2 + \cdots + \alpha_{nj}v_n$, then we write the coefficients α_{ij} of v_i as a column vector in the matrix (α_{ij}), i.e., the j^{th} column of the matrix $m(T) = (\alpha_{ij})$ is written as $\begin{pmatrix} \alpha_{1j} \\ \alpha_{2j} \\ \vdots \\ \alpha_{nj} \end{pmatrix}$.

Exercise 2.3.2. Let F be a field and $M_n(F)$ be the collection of all $n \times n$ matrices over F. Show that $M_n(F)$ is an algebra with usual addition of matrices, scalar multiplication and product of matrices.

Question: Is it always true that given an element T in $A(V)$ we can always find a matrix (α_{ij}) in $M_n(F)$ corresponding to T and vice-versa?

This is answered by the following theorem:

Theorem 2.3.3. *If V is an n-dimensional vector space over F, then $A(V)$ and $M_n(F)$ are isomorphic as algebras over F.*

More precisely, if $\{v_1, v_2, \ldots, v_n\}$ is a basis of V over F, if $T \in A(V)$ and $m(T)$ is the *matrix of T in the basis $\{v_1, v_2, \ldots, v_n\}$ then the mapping $\phi : A(V) \to M_n(F)$ defined as $\phi(T) = m(T)$ is an algebra isomorphism of $A(V)$ onto $M_n(F)$.*

Proof. First we show that ϕ is an algebra homomorphism.
Let $S,T \in A(V)$ and $\alpha, \beta \in F$. Let the matrix of S and T be given by (α_{ij}) and (β_{ij}) respectively. That is, $\phi(S) = m(S) = (\alpha_{ij})$ and $\phi(T) = m(T) = (\beta_{ij})$, where

$$Sv_j = \sum_{i=1}^{n} \alpha_{ij} v_i \quad \text{and} \quad Tv_j = \sum_{i=1}^{n} \beta_{ij} v_i.$$

First we show that ϕ is a linear map, i.e., $\phi(\alpha S + \beta T) = \alpha\phi(S) + \beta\phi(T)$.
Now, to find $m(\alpha S + \beta T)$, consider

$$\begin{aligned}(\alpha S + \beta T)v_j &= \alpha S(v_j) + \beta T(v_j)\\ &= \alpha \sum_{i=1}^{n} \alpha_{ij} v_i + \beta \sum_{i=1}^{n} \beta_{ij} v_i\\ &= \sum_{i=1}^{n} (\alpha\alpha_{ij} + \beta\beta_{ij}) v_i\end{aligned}$$

Thus, $m(\alpha S + \beta T) = (\alpha\alpha_{ij} + \beta\beta_{ij})$ and hence

$$\begin{aligned}\phi(\alpha S + \beta T) &= m(\alpha S + \beta T)\\ &= (\alpha\alpha_{ij} + \beta\beta_{ij})\\ &= \alpha(\alpha_{ij}) + \beta(\beta_{ij})\\ &= \alpha m(S) + \beta m(T)\\ &= \alpha\phi(S) + \beta\phi(T).\end{aligned}$$

Thus, for any $S,T \in A(V)$ and $\alpha, \beta \in F$,

$$\phi(\alpha S + \beta T) = \alpha\phi(S) + \beta\phi(T).$$

Thus, ϕ is a linear map (a vector space homomorphism). To show that ϕ is an algebra homomorphism, it remains to show that $\phi(ST) = \phi(S)\phi(T)$, i.e., we have to show that $m(ST) = m(S)m(T)$.

We know that, ij^{th} entry of the product of two matrices (α_{ij}) and (β_{ij}) is given as

$$\gamma_{ij} = \sum_{k=1}^{n} \alpha_{ik}\beta_{kj}, \tag{2.3}$$

where $(\gamma_{ij}) = (\alpha_{ij})(\beta_{ij})$.

First we find the matrix of ST. For this, we consider $(ST)v_j$, $j = 1,2,\ldots,n$. Now,

$$\begin{aligned}(ST)v_j &= S(Tv_j)\\ &= S\left(\sum_{k=1}^{n} \beta_{kj} v_k\right)\\ &= \sum_{k=1}^{n} \beta_{kj} S(v_k) \qquad \text{(since } S \text{ is linear)}\end{aligned}$$

$$
\begin{aligned}
&= \sum_{k=1}^{n} \beta_{kj} \left(\sum_{i=1}^{n} \alpha_{ij} v_i \right) \\
&= \sum_{i=1}^{n} \left(\sum_{k=1}^{n} \alpha_{ik} \beta_{kj} \right) v_i \\
&= \sum_{i=1}^{n} \gamma_{ij} v_i \qquad\qquad \text{(by (2.3))}
\end{aligned}
$$

Therefore,

$$m(ST) = (\gamma_{ij}) = (\alpha_{ij})(\beta_{ij}) = m(S)m(T).$$

So, $\phi(ST) = \phi(S)\phi(T)$ and hence, ϕ is an algebra homomorphism.

Now we show that ϕ is one-one. Let $T \in A(V)$ such that $\phi(T) = 0$. Then $m(T) = 0$ and hence $Tv_j = 0$ for all $j = 1, 2, \dots, n$. Then for any $v \in V$, since $v_1, v_2, \dots, v_n$ is a basis of V, we have

$$
\begin{aligned}
v &= \mu_1 v_1 + \mu_2 v_2 + \cdots + \mu_n v_n \\
\Rightarrow T(v) &= \mu_1 T v_1 + \mu_2 T v_2 + \cdots + \mu_n T v_n = 0
\end{aligned}
$$

Thus, we have $Tv = 0$ for all $v \in V$ and hence $T = 0$ which proves that ϕ is one-one. Thus, ϕ is an isomorphism of $A(V)$ into $M_n(F)$.

Finally, we show that ϕ is onto. Let $(\alpha_{ij}) \in M_n(F)$ be any $n \times n$ matrix. We have to find a $T \in A(V)$ such that $\phi(T) = m(T) = (\alpha_{ij})$. Define T on the basis $\{v_1, v_2, \dots, v_n\}$ as

$$T v_j = \sum_{i=1}^{n} \alpha_{ij} v_i.$$

Extend T linearly on V, then we see that $T \in A(V)$ and $\phi(T) = (\alpha_{ij})$. Hence, ϕ is onto.

Another argument for showing that ϕ is onto:
We know that dimension of $A(V) = \text{Hom}(V,V)$ over F is n^2. Also, dimension of the vector space $M_n(F)$ over F is n^2. Thus, ϕ is isomorphism of $A(V)$ into $M_n(F)$ and $\dim A(V) = \dim M_n(F)$, and so ϕ is onto. Hence,

$$\boxed{A(V) \cong M_n(F).}$$

□

Let us consider an example to see, given a linear transformation, how to obtain a matrix.

Example 2.3.4. Let F be a field and $V = F_{n-1}[x]$ be the vector space of all the polynomials of degree at most $n-1$. Define $D : V \to V$ as follows:

$$D(\alpha_0 + \alpha_1 x + \alpha_2 x^2 + \cdots + \alpha_{n-1} x^{n-1}) = \alpha_1 + 2\alpha_2 x + \cdots + (n-1)\alpha_{n-2} x^{n-2}.$$

Then clearly $D \in A(V)$. D is nothing but the differentiation operator. We find the matrix of D with respect to the standard basis of $F_{n-1}[x]$, i.e., $v_1 = 1$, $v_2 = x$, $v_3 = x^2, \dots, v_n = x^{n-1}$. Now,

$$
\begin{aligned}
D(v_1) &= \quad D(1) = 0 = 0v_1 + 0v_2 + \cdots + 0v_n \\
D(v_2) &= \quad D(x) = 1 = 1v_1 + 0v_2 + \cdots + 0v_n \\
&\vdots \qquad\qquad \vdots \\
D(v_n) &= \quad D(x^{n-1}) = (n-1)x^{n-2} = 0v_1 + 0v_2 + \cdots + (n-1)v_{n-1} + 0v_n
\end{aligned}
$$

Therefore, by the definition of a matrix of a linear transformation, the matrix of D in the standard basis $\{1,x,x^2,\dots,x^{n-1}\}$ is given by

$$m_1(D)=\begin{pmatrix}0&1&0&0&\cdots&0\\0&0&2&0&\cdots&0\\0&0&0&3&\cdots&0\\\vdots&\vdots&\vdots&\vdots&\ddots&\vdots\\0&0&0&0&\cdots&(n-1)\\0&0&0&0&\cdots&0\end{pmatrix}.$$

We know that matrix of a linear transformation depends on the chosen (ordered) basis of V. In the next example, we examine the matrix of D defined above but with a different basis of V.

Example 2.3.5. Let $V=F_{n-1}[x]$ and $D\in A(V)$ defined as above. Let $u_1=1$, $u_2=1+x$, $u_3=1+x^2,\dots,u_n=1+x^{n-1}$. Check! that $u_1,u_2,\dots,u_n$ forms a basis of V over F. We find the matrix of D in this basis. Now,

$$\begin{aligned}D(u_1)=&\quad D(1) &&=0=0u_1+0u_2+\cdots+0u_n\\D(u_2)=&\quad D(1+x) &&=1=1u_1+0u_2+\cdots+0u_n\\D(u_3)=&\quad D(1+x^2) &&=2x=2(u_2-u_1)=-2u_1+2u_2+0u_3+\cdots+0u_n\\&\vdots\qquad\quad\vdots\\D(u_n)=&\quad D(1+x^{n-1}) &&=(n-1)x^{n-2}=(n-1)(u_n-u_1)\\& &&=-(n-1)u_1+0u_2+\cdots+0u_{n-2}+(n-1)u_{n-1}+0v_n\end{aligned}$$

Therefore, the matrix of D in the basis $\{1,1+x,1+x^2,\dots,1+x^{n-1}\}$ is given by

$$m_2(D)=\begin{pmatrix}0&1&-2&-3&\cdots&-(n-1)\\0&0&2&0&\cdots&0\\0&0&0&3&\cdots&0\\\vdots&\vdots&\vdots&\vdots&\ddots&\vdots\\0&0&0&0&\cdots&(n-1)\\0&0&0&0&\cdots&0\end{pmatrix}.$$

Definition 2.3.6. Two matrices A and B in $M_n(F)$ are said to be *similar* if there is an invertible matrix $C\in M_n(F)$ such that $A=C^{-1}BC$.

Exercise 2.3.7. If $T\in A(V)$ is invertible then prove that $\phi(T^{-1})=(\phi(T))^{-1}$.

Remark 2.3.8. We have seen from the above two examples that the same linear transformation D has different matrices $m_1(D)$ and $m_2(D)$ corresponding to two different bases of V. Is there any relation between $m_1(D)$ and $m_2(D)$? It is therefore an interesting question to ask, at this stage, that is there any relation between different matrices of the same linear transformation? This question is answered by the following given below.

Theorem 2.3.9. *Let V be an n-dimensional vector space over F and $T\in A(V)$. Suppose T has the matrix $m_1(T)$ with respect to the basis $B_1=\{v_1,v_2,\dots,v_n\}$ and the matrix $m_2(T)$ with respect to the basis $B_2=\{w_1,w_2,\dots,w_n\}$ of V over F, then $m_1(T)$ and $m_2(T)$ are* *similar i.e., there is an invertible element $C \in M_n(F)$ such that $m_2(T) = C^{-1}m_1(T)C$.*

Proof. Let $m_1(T) = (\alpha_{ij})$ and $m_2(T) = (\beta_{ij})$, where

$$Tv_j = \sum_{i=1}^{n} \alpha_{ij} v_i \quad \text{and} \quad Tw_j = \sum_{i=1}^{n} \beta_{ij}, \qquad j = 1, 2, \ldots, n.$$

Define $Sv_j = w_j$ for $j = 1, 2, \ldots, n$.
Extend S linearly on V, i.e., for $v = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n$ define

$$S(v) = \alpha_1 w_1 + \alpha_2 w_2 + \cdots + \alpha_n w_n.$$

Then $S \in A(V)$ and since S maps basis B_1 of V onto basis B_2 of V over F by Theorem 2.1.21, S is regular (because S is onto). Now,

$$\begin{aligned} TS(v_j) &= T(w_j) && (\text{since } Sv_j = w_j) \\ &= \sum_{i=1}^{n} \beta_{ij} w_i \\ &= \sum_{i=1}^{n} \beta_{ij} S v_i && (\text{since } w_i = Sv_i) \\ &= S\left(\sum_{i=1}^{n} \beta_{ij} v_i\right) && (\text{since } S \text{ is linear}). \end{aligned}$$

Since S is regular, multiplying both sides by S^{-1} (to the left), we get

$$(S^{-1}TS)(v_j) = \sum_{i=1}^{n} \beta_{ij} v_i.$$

This means the matrix of $(S^{-1}TS)$ with respect to the basis B_1 of V is (β_{ij}). Thus,

$$m_1(S^{-1}TS) = (\beta_{ij}) = m_2(T).$$

Since $\phi : A(V) \to M_n(F)$ given by $T \mapsto m(T)$ is an onto isomorphism, we have

$$\begin{aligned} & m_1(S^{-1}TS) = (\beta_{ij}) = m_2(T) \\ \Rightarrow\ & m_1(S^{-1}) m_1(T) m_1(S) = m_2(T) \\ \Rightarrow\ & m_1(S)^{-1} m_1(T) m_1(S) = m_2(T). \end{aligned}$$

Therefore, $m_1(T)$ and $m_2(T)$ are similar. Particularly,

$$C^{-1} m_1(T) C = m_2(T),$$

where the matrix C can be chosen to be $m_1(S)$. □

We try to understand the above theorem by an example. In Examples 2.3.4 and 2.3.5, we saw that D has two different matrices $m_1(D)$ and $m_2(D)$ with respect to different bases of $V = F_{n-1}[x]$. We shall now show that they are similar.

Example 2.3.10. As considered in Examples 2.3.4 and 2.3.5, let $V = F_3[x]$ and D be the differentiation operator defined by

$$D(\alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_3 x^3) = \alpha_1 + 2\alpha_2 x + 3\alpha_3 x^2.$$

Then, we have seen that, the matrix of D in the basis $v_1 = 1, v_2 = x, v_3 = x^2, v_4 = x^3$ is given by

$$m_1(D) = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

and the matrix of D with respect to basis $u_1 = 1, u_2 = 1+x, u_3 = 1+x^2, u_4 = 1+x^3$ is given by

$$m_2(D) = \begin{pmatrix} 0 & 1 & -2 & -3 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

We want to find a matrix C such that $C^{-1}m_1(D)C = m_2(D)$. For this we need to find an element $S \in A(V)$ such that $C = m_1(S)$. As defined in the above theorem, let the linear transformation S on $F_3[x]$ be defined as follows:

$$\begin{aligned} S(v_1) &= u_1 = 1 = v_1, \\ S(v_2) &= u_2 = 1 + x = v_1 + v_2, \\ S(v_3) &= u_3 = 1 + x^2 = v_1 + v_3, \\ S(v_4) &= u_4 = 1 + x^3 = v_1 + v_4. \end{aligned}$$

Then the matrix of S in the basis v_1, v_2, v_3, v_4 is

$$C = m_1(S) = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Then, by a little computation, one can check that

$$C^{-1} = m_1(S)^{-1} = \begin{pmatrix} 1 & -1 & -1 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

and

$$\begin{aligned} C^{-1}m_1(D)C &= \begin{pmatrix} 1 & -1 & -1 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 1 & -2 & -3 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix} = m_2(D). \qquad \text{(Verify!)} \end{aligned}$$

Proposition 2.3.11. *Let $B_1 = \{v_1, v_2, \ldots, v_n\}$ and $B_2 = \{w_1, w_1, \ldots, w_n\}$ be bases of a vector space V over F. Let $S, T \in A(V)$ and let $A = (\alpha_{ij}) \in M_n(F)$. Suppose that A is the matrix of T with respect to the basis B_1 and A is the matrix of S with respect to the basis B_2. Then S and T are similar.*

Proof. Since A is the matrix of T and S in basis B_1 and B_2 respectively, we have

$$Tv_j = \sum_{i=1}^{n} \alpha_{ij} v_i \quad \text{and} \quad Sw_j = \sum_{i=1}^{n} \alpha_{ij} w_i \qquad (1 \le j \le n).$$

Define $P : V \to V$ as follows: for $v \in V$, $v = \gamma_1 v_1 + \gamma_2 v_2 + \cdots + \gamma_n v_n$, put

$$P(v) = \gamma_1 w_1 + \gamma_2 w_2 + \cdots + \gamma_n w_n.$$

Thus, P maps basis B_1 onto B_2 and hence P is onto. Therefore, by Theorem 2.1.21, P is regular. Now, for $1 \le j \le n$,

$$\begin{aligned} SPv_j &= Sw_j && (\text{since } Pv_j = w_j) \\ &= \sum_{i=1}^{n} \alpha_{ij} w_i \\ &= \sum_{i=1}^{n} \alpha_{ij} P v_i && (\text{since } w_i = Pv_i) \\ &= P\left(\sum_{i=1}^{n} \alpha_{ij} v_i\right) && (\text{since } P \text{ is linear}) \\ &= PTv_j. \end{aligned}$$

Thus, $SP(v_j) = PT(v_j)$ for all $j = 1, 2, \ldots, n$. Since $B_1 = \{v_1, v_2, \ldots, v_n\}$ is a basis of V over F, we conclude that

$$SP = PT.$$

Since P is regular, multiplying by P^{-1} (to the left) on both sides in the above equality, we get

$$P^{-1}SP = T.$$

Hence, S and T are similar. □

We conclude this unit (chapter) by making the following remark:

Remark 2.3.12. By Theorem 2.3.9, we can say that if the same linear transformation on V has two different matrices in two different bases of V, then the two matrices must be similar. By Proposition 2.3.11, we say that if two different linear transformations on V has the same matrix in two different bases of V, then the two linear transformations must be similar.

CHAPTER 3

CANONICAL FORMS

We have seen the definition of similar linear transformations. It is an easy exercise (Exercise 2.1.28) to see that the relation of similarity is an equivalence relation and the equivalence class of an element of $A(V)$ is called its *similarity class*.

When can we say that two linear transformations are similar? We shall exhibit that the matrix of a linear transformation in some basis has particular nice form. These matrices are called *canonical forms*. To check whether two linear transformations are similar or not, we shall compare their canonical forms. If these forms are same then the linear transformations are similar. There are many possible canonical forms, for example, triangular form, Jordan form, etc. We shall study some of them in this unit.

3.1 Triangular Form

Definition 3.1.1. Let V be a vector space over F and $T \in A(V)$. A subspace W of V is called *invariant under* T if $T(W) \subset W$ (i.e. $Tw \in W$ for all $w \in W$).

Example 3.1.2. Let V be a vector space over F and let $I \in A(V)$ be the identity operator on V. Let W be a subspace of V. Then is invariant under I as $W = I(W) \subset W$.

Thus, every subspace W of V is invariant under the identity map I.

Example 3.1.3. Clearly, $W = \{0\}$ is invariant under any $T \in A(V)$ as $T\{0\} \subset \{0\}$. Also, $W = V$ is invariant under any $T \in A(V)$ as $T(V) \subset V$.

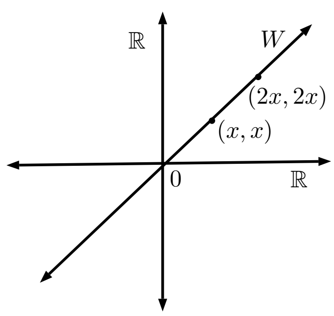
Thus, $\{0\}$ and V are always invariant under any given $T \in A(V)$.

Example 3.1.4.

Let $V = \mathbb{R}^2$ and $T : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by

$$T(x,y) = (x+y, x+y),$$

where $(x,y) \in \mathbb{R}^2$. Let $W = \{(x,x) : x \in \mathbb{R}\}$ be a subspace of $V = \mathbb{R}^2$. Then clearly, W is invariant under T defined above,



since for any $(x,x) \in W$, $T(x,x) = (2x,2x) \in W$. Thus, $T(W) \subset W$.

Theorem 3.1.5. *Let V be a finite dimensional vector space over F and let $T \in A(V)$. Let W be a subspace of V which is invariant under T. Then T induces a map $\overline{T} = V/W \to V/W$ defined by*

$$\overline{T}(v+W) = Tv + W, \qquad w \in W$$

such that $\overline{T}$ is well defined and $T \in A(V/W)$. Further, $\overline{T}$ satisfies very polynomial in $F[x]$ which is satisfied by T. In particular, if $p_1(x) \in F[x]$ is the minimal polynomial for $\overline{T}$ over F and $p(x) \in F[x]$ is the minimal polynomial for T, then $p_1(x)|p(x)$.

Proof. First we show that $\overline{T}$ is well-defined. Let $v_1 + W, v_2 + W \in V/W$ such that

$$\begin{aligned}
& v_1 + W = v_2 + W \\
\Rightarrow\ & v_1 - v_2 \in W \\
\Rightarrow\ & T(v_1 - v_2) = Tv_1 - Tv_2 \in W && (\because W \text{ is invariant under } T) \\
\Rightarrow\ & Tv_1 + W = Tv_2 + W.
\end{aligned}$$

This shows that $\overline{T}$ is well-defined.

Next, we show that $\overline{T}$ is a linear transformation i.e., $\overline{T} \in A(V/W)$.
Let $v_1 + W, v_2 + W \in V/W$ and $\alpha, \beta \in F$. Then

$$\begin{aligned}
\overline{T}(\alpha(v_1+W) + \beta(v_2+W)) &= \overline{T}((\alpha v_1 + \beta v_2) + W) \\
&= T(\alpha v_1 + \beta v_2) + W && (\text{by definition of } \overline{T}) \\
&= (\alpha T v_1 + \beta T v_2) + W && (\text{since } T \text{ is linear}) \\
&= \alpha(Tv_1 + W) + \beta(Tv_2 + W) \\
&= \alpha\overline{T}(v_1+W) + \beta\overline{T}(v_2+W) && (\text{by definition of } \overline{T}).
\end{aligned}$$

Thus, $\overline{T} : V/W \to V/W$ is a linear transformation.

Now, we show that $\overline{T}$ satisfies every polynomial satisfied by T. Before we can show this, we have to show the following:
If W is invariant under T then $T(W) \subset W$. Then, $T^2(W) = T(T(W)) \subset W$. Thus, W is invariant under T^2 also. Similarly, for any $k \in \mathbb{N}$, W is invariant under T^k. Now, for any $v + W \in V/W$,

$$\begin{aligned}
(\overline{T})^2 &= \overline{T}(\overline{T}(v+W)) \\
&= \overline{T}(Tv+W) \\
&= T(Tv) + W \\
&= T^2 v + W \\
&= \overline{T^2}(v+W).
\end{aligned}$$

Therefore, $(\overline{T})^2 = \overline{T^2}$ and if follows by induction that

$$(\overline{T})^k = \overline{T^k} \text{ for any } k \in \mathbb{N}. \tag{3.1}$$

Similarly, one can easily show the following:

- If W is invariant under S and T then W is invariant under $S+T$ and
$$\overline{S+T} = \overline{S} + \overline{T}. \tag{3.2}$$
- If W is invariant under T and $\alpha \in F$ then W is invariant under αT and
$$\overline{\alpha T} = \alpha \overline{T}. \tag{3.3}$$

We also note that, $\overline{I} \in A(V/W)$ is the identity map on $V/W$and the map $\overline{0}$ is the zero map on V/W.

Now, let $q(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_n x^n \in F[x]$ be a polynomial satisfied by T, i.e.,

$$\begin{aligned}
& q(T) = 0 \\
\Rightarrow\ & \alpha_0 I + \alpha_1 T + \alpha_2 T^2 + \alpha_n T^n = 0 \\
\Rightarrow\ & \overline{\alpha_0 I + \alpha_1 T + \alpha_2 T^2 + \cdots + \alpha_n T^n} = \overline{0} \\
\Rightarrow\ & \overline{\alpha_0 I} + \overline{\alpha_1 T} + \overline{\alpha_2 T^2} + \cdots + \overline{\alpha_n T^n} = \overline{0} && \text{(by equation (3.2))} \\
\Rightarrow\ & \alpha_0 \overline{I} + \alpha_1 \overline{T} + \alpha_2 \overline{T^2} + \cdots + \alpha_n \overline{T^n} = \overline{0} && \text{(by equation (3.3))} \\
\Rightarrow\ & \alpha_0 \overline{I} + \alpha_1 \overline{T} + \alpha_2 \overline{T}^2 + \cdots + \alpha_n \overline{T}^n = \overline{0} && \text{(by equation (3.3))} \\
\Rightarrow\ & q(\overline{T}) = \overline{0}.
\end{aligned}$$

Thus, any polynomial $q(x) \in F[x]$ which is satisfied by T is satisfied by $\overline{T}$.

Now, let $p_1(X) \in F[x]$ be minimal polynomial for $\overline{T}$ and $p(x) \in F[x]$ be minimal polynomial for T. Then, by above, $p(T) = 0 \Rightarrow p(\overline{T}) = 0$. Since, minimal polynomial for $\overline{T}$ divides every other polynomial satisfied by $\overline{T}$, we have $p_1(x) | p(x)$. □

Let V be finite dimensional over F and $T \in A(V)$. We have seen, by Corollary 2.2.4, that if $\lambda \in F$ is a characteristic root of T then λ is a root of minimal polynomial for T. The following lemma exactly states its converse.

Lemma 3.1.6. *Let V be a finite dimensional vector space over F and $T \in A(V)$. If $\lambda \in F$ is a root of the minimal polynomial for T, then λ is a characteristic root of T.*

Proof. Let $p(x) \in F[x]$ be the minimal polynomial for T and $\lambda \in F$ be a root $p(x)$, i.e. $p(\lambda) = 0$. Therefore,
$$p(x) = (x - \lambda)q(x),$$
where $0 \neq q(x) \in F[x]$ and $\deg q(x) < \deg p(x)$.
As $p(x)$ is the minimal polynomial for T and $\deg q(x) < \deg p(x)$, we have $q(T) \neq 0$. Then there exists some $w \in V$ such that
$$v = q(T)(w) \neq 0.$$
Now, $0 = p(T)(w) = (T - \lambda I)q(T)(w)$. This implies
$$(T - \lambda I)v = 0.$$
Hence, $Tv = \lambda v$ and hence λ is a characteristic root of T. □

Remark 3.1.7. Combining Corollary 2.2.4 and Lemma 3.1.6, we can say that $\lambda \in F$ is a root of minimal polynomial for $T \in A(V)$ if and only if λ is a characteristic root of T.

Definition 3.1.8. Let G be a field and $n \in \mathbb{N}$. A matrix $(\alpha_{ij}) \in M_n(F)$ is said to be *upper triangular* if all its entries below the main diagonal are 0, i.e.,

$$\alpha_{ij} = 0 \quad \text{for } i > j.$$

A matrix (α_{ij}) is said to be *lower triangular* if all its entries above the main diagonal are 0, i.e.,

$$\alpha_{ij} = 0 \quad \text{for } i < j.$$

Equivalently, for $T \in A(V)$, the matrix of T is said to be *upper triangular* if

$$\begin{aligned} Tv_1 &= \alpha_{11}v_1 \\ Tv_2 &= \alpha_{12}v_1 + \alpha_{22}v_2 \\ &\vdots \\ Tv_i &= \alpha_{1i}v_1 + \alpha_{2i}v_2 + \cdots + \alpha_{ii}v_i \\ &\vdots \\ Tv_n &= \alpha_{1n}v_1 + \alpha_{2n}v_2 + \cdots + \alpha_{nn}v_n, \end{aligned}$$

i.e., if Tv_i is a linear combination of only v_i and its preceding ones, $v_1, v_2, \ldots, v_{i-1}$ in the basis of V.

Theorem 3.1.9. *Let V be a finite dimensional vector space over F and $T \in A(V)$. If all the roots of the minimal polynomial for T are in F then there is a basis of V with respect to which the matrix of T is (upper) triangular.*

Since by Remark 3.1.7, we know that, $\lambda \in F$ is a characteristic root of T if and only if λ is a root of the minimal polynomial for T, the above theorem can be restated as follows:

Theorem 3.1.9. *Let V be a finite dimensional vector space over F and $T \in A(V)$. If all the characteristic roots of T are in F then there is a basis of V with respect to which the matrix of T is (upper) triangular.*

Proof. We shall prove the theorem by principle of mathematical induction on the dimension n of V.

• Let $\dim V = 1$.
Let $\lambda \in F$ be a characteristic root of T. Then there exists a $0 \neq v \in V$ such that $Tv = \lambda v$. Since, $v \neq 0$ and $\dim V = 1$, clearly $\{v\}$ is a basis of V. The matrix of T with respect to the basis $\{v\}$ is (λ) which is a triangular matrix.

• Assume that the result is true for $\dim V = n - 1$.

• Let $\dim V = n$.
Assume that all the characteristic roots of T are in F. Let $\lambda_1 \in F$ be a characteristic root of T. Then there is a $v_1 \neq 0$ in V such that $Tv_1 = \lambda_1 v_1$. Let

$$W = L(\{v_1\}) = \{\alpha v_1 : \alpha \in F\}.$$

Then W is a one-dimensional subspace of V. Also for any $w \in W$, $w = \alpha v_1$, we have

$$Tw = T(\alpha v_1) = \alpha(Tv_1) = \alpha(\lambda v_1) = \lambda(\alpha v_1) \in W.$$

Thus, W is invariant under T. Then by Theorem 3.1.5, T induces a map $\overline{T} : V/W \to V/W$ such that $\overline{T} \in A(V/W)$ and minimal polynomial for $\overline{T}$ divides the minimal polynomial for T. Since all the roots of minimal polynomial for $\overline{T}$ are the roots of the minimal polynomial for T, the map $\overline{T}$ also satisfies the (condition) hypothesis of the theorem.

Also, we know that

$$\dim(V/W) = \dim V - \dim W = n - 1.$$

Then by induction hypothesis there is a basis $\{v_2 + W, v_3 + W, \ldots, v_n + W\}$ of V/W such that the matrix of $\overline{T}$, say $(\alpha_{ij})_{2 \leq i,j \leq n}$, in this basis is upper triangular. Equivalently,

$$\begin{aligned}
\overline{T}(v_2 + W) &= \alpha_{22}(v_2 + W) \\
\overline{T}(v_3 + W) &= \alpha_{23}(v_2 + W) + \alpha_{33}(v_3 + W) \\
&\vdots \\
\overline{T}(v_n + W) &= \alpha_{2n}(v_2 + W) + \alpha_{3n}(v_3 + W) + \cdots + \alpha_{nn}(v_n + W).
\end{aligned}$$

Claim: $\{v_2, v_3, \ldots, v_n\}$ is linearly independent.
Let $\beta_2, \beta_3, \ldots, v_n \in F$ such that $\beta_2 v_2 + \beta_3 v_3 + \cdots + \beta_n v_n = 0$. Then

$$\begin{aligned}
&(\beta_2 v_2 + \beta_3 v_3 + \cdots + \beta_n v_n) + W = 0 + W = W \\
\Rightarrow\ &\beta_2(v_2 + W) + \beta_3(v_3 + W) + \cdots + \beta_n(v_n + W) = W
\end{aligned}$$

Since $v_2 + W, v_3 + W, \ldots, v_n + W$ are linearly independent (being basis of V/W), we have $\beta_2 = \beta_3 = \cdots = \beta_n = 0$. This proves our claim.

Also, since $v_2 + W, v_3 + W, \ldots, v_n + W$ are linearly independent, $v_i + W \neq W$ and hence $v_i \notin W$ for all $2 \leq i \leq n$. But $W = L(\{v_1]\})$ and so v_1 cannot be written as a linear combination of $v_2, \ldots, v_n$. Therefore, $\{v_1, v_2, \ldots, v_n\}$ must be linearly independent. Since, $\dim V = n$, the set $\{v_1, v_2, \ldots, v_n\}$ forms a basis of V.

Finally, we show that the matrix of T in the basis $\{v_1, v_2, \ldots, v_n\}$ is upper triangular.
We have, $Tv_1 = \lambda_1 v_1$. Take $\lambda_1 = \alpha_{11}$, then $Tv_1 = \alpha_{11} v_1$. Now, we have

$$\begin{aligned}
&\overline{T}(v_2 + W) = \alpha_{22} v_2 + W \\
\Rightarrow\ &Tv_2 - \alpha_{22} v_2 \in W \\
\Rightarrow\ &Tv_2 - \alpha_{22} v_2 = \alpha_{12} v_1 \qquad (\text{for some } \alpha_{12} \in F, \text{ since } W = L(\{v_1\}) \\
\Rightarrow\ &Tv_2 = \alpha_{12} v_1 + \alpha_{22} v_2.
\end{aligned}$$

Similarly, it follows that for any $1 \leq j \leq n$, there is $\alpha_{1j} \in F$ such that

$$Tv_j - \sum_{i=2}^{j} \alpha_{ij} v_j = \alpha_{1j} v_1 \in W.$$

Thus,

$$Tv_j = \sum_{i=1}^{j} \alpha_{ij} v_j.$$

Therefore, the matrix of T with respect to the basis $\{v_1, v_2, \ldots, v_n\}$ is upper triangular. □

Exercise 3.1.10. Let V be a finite dimensional vector space over F, $T \in A(V)$ and $m(T)$ be the matrix of T. Show that T and $m(T)$ have the same characteristic roots.

Alternatively Theorem 3.1.9 can also be stated in the following form:

Corollary 3.1.11. *If the matrix $A \in M_n(F)$ has all its characteristic roots in F, then A is similar to a triangular matrix, i.e., there is an invertible matrix $C \in M_n(F)$ such that $C^{-1}AC$ is a triangular matrix.*

Theorem 3.1.12. *Let V be an n-dimensional vector space over F. If $T \in A(V)$ has all its characteristic roots (or roots of the minimal polynomial) of T are in F, then T satisfies a polynomial of degree n over F.*

Proof. Since all the characteristic roots of T are in F, by (above) Theorem 3.1.9, we can find a basis $v_1, v_2, \dots, v_n$ of V such that the matrix of T in this basis is upper triangular, i.e.,

$$\begin{aligned} Tv_1 &= \alpha_{11}v_1 \\ Tv_2 &= \alpha_{12}v_1 + \alpha_{22}v_2 \\ &\vdots \\ Tv_n &= \alpha_{1n}v_1 + \alpha_{2n}v_2 + \cdots + \alpha_{nn}v_n. \end{aligned}$$

Now, from first equation, we have

$$(T - \alpha_{11}I)v_1 = 0 \Rightarrow (T - \alpha_{22}I)(T - \alpha_{11}I)v_1 = 0.$$

Also, by the second equation above, we have

$$\begin{aligned} (T - \alpha_{11}I)(T - \alpha_{22}I)v_2 &= (T - \alpha_{11}I)(Tv_2 - \alpha_{22}v_2) \\ &= (T - \alpha_{11}I)(\alpha_{12}v_1) \\ &= \alpha_{12}(T - \alpha_{11}I)v_1 \\ &= 0 \qquad \text{(by first equation)} \end{aligned}$$

Inductively, we get

$$(T - \alpha_{11}I)(T - \alpha_{22}I)\cdots(T - \alpha_{ii}I)v_i = 0$$

for all $i = 1, 2, \dots, n$. Therefore,

$$(T - \alpha_{11}I)(T - \alpha_{22}I)\cdots(T - \alpha_{nn}I) = 0.$$

Hence, T satisfies a polynomial $q(x) = (x - \alpha_{11})(x - \alpha_{22})\cdots(x - \alpha_{nn})$ of degree n over F. □

3.2 Canonical Forms: Nilpotent Tranformations

Definition 3.2.1. Let V be a vector space over F and $T \in A(V)$. Then T is called *nilpotent* if there is a positive integer n such that $T^n = 0$. The smallest such n is called the *index of* *nilpotence* of T, i.e., if n is the index of nilpotence then

$$T^n = 0 \text{ but } T^{n-1} \neq 0.$$

Exercise 3.2.2. If $T \in A(V)$ is nilpotent and if $\alpha \in F$ such that $\alpha \neq 0$ then prove that $\alpha I + T$ is regular and its inverse $(\alpha I + T)^{-1}$ is a polynomial in T.

Solution. Seminar exercise. □

Lemma 3.2.3. *Let V be a vector space over F and $T \in A(V)$ be nilpotent. Then $\alpha_0 + \alpha_1 T + \ldots + \alpha_m T^m$, where $\alpha_0, \alpha_1, \ldots, \alpha_m \in F$ is invertible, if $\alpha_0 \neq 0$.*

Proof. Note that, $\alpha_1 T + \ldots + \alpha_m T^m = (\alpha_1 I + \alpha_2 T + \ldots + \alpha_m T^{m-1})T$.
Since T is nilpotent, $T^n = 0$ for some $n \in \mathbb{N}$. Then

$$(\alpha_1 T + \ldots + \alpha_m T^m)^n = (\alpha_1 I + \alpha_2 T + \ldots + \alpha_m T^{m-1})^n T^n = 0.$$

Therefore, if T is nilpotent then $\alpha_1 T + \ldots + \alpha_m T^m$ is also nilpotent. Then by the above exercise, if $\alpha_0 \neq 0$ then $\alpha_0 I + \alpha_1 T + \ldots + \alpha_m T^m$ is invertible. □

Note: We observe that, by the above exercise, the inverse of $\alpha_0 + \alpha_1 T + \ldots + \alpha_m T^m$ is a polynomial in T.

Lemma 3.2.4. *Let V be a vector space over F and let $T \in A(V)$ be nilpotent with the index of nilpotence n_1. Let $v \in V$ be such that $T^{n_1-1}v \neq 0$. Let*

$$V_1 = L(\{v, Tv, \ldots, T^{n_1-1}v\}).$$

Then dimension of V_1 is n_1, V_1 is invariant under T and the matrix of $T\big|_{V_1}$ is

$$M_{n_1} = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix}_{n_1 \times n_1}.$$

Proof. Since V_1 is already the span of $v, Tv, \ldots, T^{n_1-1}v$, to show that $\dim V_1 = n_1$ we have to show that $v, Tv, \ldots, T^{n_1-1}v$ are linearly independent. Suppose if possible, $v, Tv, \ldots, T^{n_1-1}v$ are linearly dependent. Then for $\alpha_0, \alpha_1, \ldots, \alpha_{n_1-1} \in F$, we have

$$\alpha_0 v + \alpha_1 Tv + \cdots + \alpha_{n_1-1} T^{n_1-1} v = 0 \tag{3.4}$$

for some α_i's non-zero. Let s be the first integer such that $\alpha_s \neq 0$. Then (3.4) becomes

$$\alpha_s T^s v + \alpha_{s+1} T^{s+1} v + \cdots + \alpha_{n_1-1} T^{n_1-1} v = 0$$
$$\Rightarrow (\alpha_s I + \alpha_{s+1} T + \cdots + \alpha_{n_1-1} T^{n_1-1-s}) T^s v = 0. \tag{3.5}$$

Since T is nilpotent and $\alpha_s \neq 0$, by Lemma 3.2.3, $(\alpha_s I + \alpha_{s+1} T + \cdots + \alpha_{n_1-1} T^{n_1-1-s})$ is regular. Applying its inverse to the left on both sides of (3.5), we get $T^s v = 0$. Therefore,

$$T^{n_1-1} v = T^{n_1-1-s}(T^s v) = 0$$

which is a contradiction, since the index of nilpotence is given to be n_1. Hence, all α_i's must be 0 and so the set $\{v, Tv, \ldots, T^{n_1-1}v\}$ is linearly independent. Hence,

$$\dim V_1 = n_1.$$

Now, we show that $\underline{V_1 \text{ is invariant under } T}$.

$$\begin{aligned} T(V_1) &= T(L(\{v, Tv, \ldots, T^{n_1-1}v\})) \\ &= L(\{Tv, T^2 v, \ldots, T^{n_1-1}v\}) \qquad \text{(since } T^{n_1} v = 0) \\ &\subset L(\{v, Tv, \ldots, T^{n_1-1}v\}) = V_1. \end{aligned}$$

Therefore, V_1 is invariant under T and as a result, $T\big|_{V_1} : V_1 \to V_1$ is a homomorphism.

Now, we find the matrix of $T\big|_{V_1}$.

$$\begin{aligned} T\big|_{V_1}(v) &= Tv \; (= 0v + Tv + 0T^2 v + \cdots + 0T^{n_1-1}v) \\ T\big|_{V_1}(Tv) &= T(Tv) = T^2(v) \\ &\vdots \\ T\big|_{V_1}(T^{n_1-1}v) &= T(T^{n_1-1}v) = T^{n_1} v = 0. \end{aligned}$$

Therefore, the matrix of $T\big|_{V_1}$ is

$$m(T\big|_{V_1}) = M_{n_1} = \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}_{n_1 \times n_1}.$$

□

The lemma is used in proving the lemma succeeding it, i.e. to prove that there exists another subspace W invariant under T such that $V = V_1 \oplus W$. The proof of the following lemma is easy and is left as an exercise for the students.

Lemma 3.2.5. *Let V, T, n_1 and V_1 be as in previous lemma. If $u \in V_1$ be such that $T^{n_1-k}u = 0$ for some $0 < k \leq n_1$, then $u = T^k u_0$ for some $u_0 \in V_1$.*

Proof. Exercise. □

Let V, T, n_1 and V_1 be as in Lemma 3.2.4. Then we have the following result:

Lemma 3.2.6. *There exists a subspace W if V such that W is invariant under T and $V_1 \oplus W = V$.*

Proof. Let W be a subspace of V of largest possible dimension such that

- W is invariant under T;
- $V_1 \cap W = \{0\}$.

Since $V_1 \cap W = \{0\}$, to show that $V = V_1 \oplus W$, it remains to show that $V_1 + W = V$. Suppose if possible, we have $V_1 + W \subsetneq V$. Then there exists and element $z \in V$ such that $z \notin V_1 + W$.

Note that, n_1 being the index of nilpotence of T, we have $T^{n_1} = 0$. Therefore, $T^{n_1} z = 0 \in V_1 + W$. Then there exists an integer k, $1 \leq k \leq n_1$ such that

$$\begin{aligned} T^k z &\in V_1 + W \\ \text{but} \quad T^i z &\notin V_1 + W \qquad i < k. \end{aligned} \tag{3.6}$$

Thus, $T^k z = u + w$, where $u \in V_1$ and $w \in W$. But then,

$$0 = T^{n_1} z = T^{n_1 - k}(T^k z) = T^{n_1 - k}(u + w) = T^{n_1 - k} u + T^{n_1 - k} w.$$

Therefore, $T^{n_1 - k} u = -T^{n_1 - k} w$. Since V_1 and W are invariant under T, we have $T^{n_1 - k} u \in V_1$ and $T^{n_1 - k} w \in W$ and hence

$$T^{n_1 - k} u = -T^{n_1 - k} w \in V_1 \cap W = \{0\}.$$

This implies,

$$T^{n_1 - k} u = 0.$$

Then by previous lemma, $T^k u_0 = u$ for some $u_0 \in V_1$. Therefore,

$$T^k z = u + w = T^k u_0 + w \Rightarrow T^k (z - u_0) = w.$$

Take $z_1 = z - u_0$, then $T^k z_1 = w \in W$. Since W is invariant under T, we have

$$T^m z_1 \in W \quad \text{for all} \quad m \geq k.$$

Now, we show that for $i < k$, $T^i z_1 \notin W$. Suppose, if possible, $T^i z_1 \in W$ $(i < k)$. Then

$$\begin{aligned} T^i z &= T^i z - T^i u_0 + T^i u_0 \\ &= T^i z_1 + T^i u_0 \in V_1 + W \qquad (\text{since } T^i u_0 \in V_1 \text{ and } T^i z_1 \in W). \end{aligned}$$

This is not possible by our choice of k in (3.6). Therefore,

$$T^i z_1 \notin W \quad \text{for all } i < k. \tag{3.7}$$

Let W_1 be the subspace of V spanned by W and $z_1, T z_1, \ldots, T^{k-1} z_1$, i.e.,

$$W_1 = L(W \cup \{z_1, T z_1, \ldots, T^{k-1} z_1\}).$$

Since $T^i z_1 \notin W$ for $i < k$, W is a proper subspace of W_1, i.e., $W \subsetneq W_1$. Also, since $T^k z_1 \in W$ and W is invariant under T, we get that W_1 must be invariant under T.

Due to our assumption that W is the invariant subspace of V of largest dimension such that $V_1 \cap W = \{0\}$ and $W \subsetneq W_1$, W_1 also being invariant, we must have

$$\{0\} \subsetneq V_1 \cap W_1.$$

Then there must be an element $w' \neq 0$ in $V_1 \cap W_1$ such that

$$w' = w_0 + \alpha_0 z_1 + \alpha_1 T z_1 + \cdots + \alpha_{k-1} T^{k-1} z_1$$

for some $w_0 \in W$ and $\alpha_i \in F$, $i = 0, 1, \ldots, k-1$. If all the α_i's are zero then we get $w' = w_0 \in V_1 \cap W = \{0\}$ ($\because w_0 \in W$) which is a contradiction as $w' \neq 0$. Therefore, at least one α_i must be non-zero. Let $1 \leq s \leq k-1$ be the smallest integer such that $\alpha_s \neq 0$. Then, we have

$$\begin{aligned} w' &= w_0 + \alpha_s T^s z_1 + \cdots + \alpha_{k-1} T^{k-1} z_1 \\ &= w_0 + (\alpha_s I + \alpha_{s+1} T + \cdots + \alpha_{k-1} T^{k-1-s}) T^s z_1 \in V_1 \quad (\because w' \in V_1 \cap W_1). \end{aligned} \tag{3.8}$$

Since $\alpha_s \neq 0$, by Exercise 3.2.2, the element $\alpha_s I + \alpha_{s+1} T + \cdots + \alpha_{k-1} T^{k-1-s}$ is invertible and its inverse, say R, is a polynomial in T. Since, R is a polynomial in T and W and V_1 are invariant under T, it is clear that W and V_1 are also invariant under R. Then applying R on both sides of (3.8), we get

$$Rw' = Rw_0 + T^s z_1 \in R(V_1) \subset V_1.$$

Therefore, $T^s z_1 \in V_1 + R(W) \subset V_1 + W$, where $s \leq k-1 < k$. This is contradiction to (3.7). Hence, $V = V_1 + W$ and since $V_1 \cap W = \{0\}$, we have $V = V_1 \oplus W$. □

Lemma 3.2.7. *Let V be a finite dimensional vector space over F and $T \in A(V)$. Let $V = V_1 \oplus V_2 \oplus \cdots \oplus V_k$, where each subspace V_i $(1 \leq i \leq k)$ is of dimension n_i and is invariant under T. Then there is a basis of V in which the matrix of T is of the form*

$$\begin{pmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_k \end{pmatrix}$$

where each A_i is an $n_i \times n_i$ matrix of $T_i = T\big|_{V_i} : V_i \to V_i$ i.e., matrix of the linear transformation induced by T on V_i $(1 \leq i \leq k)$.

Proof. For $i = 1, 2, \ldots, k$, choose a basis $\{v_1^i, v_2^i, \ldots, v_{n_i}^i\}$ of V_i. Since $V = V_1 \oplus V_2 \oplus \cdots \oplus V_k$, the basis of V is given by

$$\begin{aligned} B &= \bigcup_{i=1}^{k} \{v_1^i, v_2^i, \ldots, v_{n_i}^i\} \\ &= \{v_1^1, v_2^1, \ldots, v_{n_1}^1, v_1^2, v_2^2, \ldots, v_{n_2}^2, \ldots, v_1^k, v_2^k, \ldots, v_{n_k}^k\}. \end{aligned}$$

Since each V_i is invariant under T, we have $T(v_j^i) \in V_i$ for all $i = 1, 2, \ldots, k$ and for all $j = 1, 2, \ldots, n_i$. Thus, in the matrix of T in basis B, Tv_j^i is a linear combination of $v_1^i, v_2^i, \ldots, v_{n_i}^i$ only and other coefficients are zero.

Let A_i be the matrix of $T_i = T\big|_{V_i}$ in the basis $\{v_1^i, v_2^i, \ldots, v_{n_i}^i\}$. Then by the definition of the matrix of a linear transformation, the matrix of T in the basis B is of the form

$$\begin{pmatrix} A_1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & A_2 & 0 & \cdots & 0 & 0 \\ 0 & 0 & A_3 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & A_{k-1} & 0 \\ 0 & 0 & 0 & \cdots & 0 & A_k \end{pmatrix}.$$

□

Theorem 3.2.8. *Let V be a finite dimensional vector space over F and $T \in A(V)$ be nilpotent with index of nilpotence n_1. Then there is a basis of V in which the matrix of T is of the form*

$$\begin{pmatrix} M_{n_1} & 0 & \cdots & 0 \\ 0 & M_{n_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & M_{n_k} \end{pmatrix}$$

where $n_1 \geq n_2 \geq \cdots \geq n_k$ and $\dim V = n_1 + n_2 + \cdots + n_k$.

Proof. Since n_1 is the index of nilpotence of T, $T^{n_1-1} \neq 0$. Therefore, there exists $v \in V$ such that $T^{n_1-1}v \neq 0$. Let

$$V_1 = L(\{v, Tv, \ldots, T^{n_1-1}v\}).$$

Then by Lemma 3.2.6, there exists a subspace W_1 of V invariant under T, i.e., $T(W) \subset W_1$ such that

$$V = V_1 \oplus W_1.$$

Let $T_1 = T\big|_{V_1}$, then by Lemma 3.2.4, we have a basis of V_1 in which matrix of $T_1 = T\big|_{V_1}$ is M_{n_1}. Let T_2 be the linear transformation on W_1 induced by T, i.e., $T_2 = T\big|_{W_1}$. Since, W_1 is invariant under T, we can say that it is invariant under T_2. Also, for any $w \in W_1$

$$T_2^{n_1}w = T^{n_1}w = 0.$$

Therefore, the index of nilpotence of T_2, say n_2, (being smallest such integer) must less than or equal to n_1, i.e., $n_2 \leq n_1$. Also, we have $T_2(W_1) \subset W_1$.
Then there is a $w_1 \in W_1$ such that $T_2^{n_2-1}w_1 \neq 0$. Let

$$V_2 = L(\{w_1, T_2w_1, T_2^2w_1, \ldots, T_2^{n_2-1}w_1\}).$$

Then by Lemma 3.2.6, there is a subspace W_2 invariant under T_2 such that $W_1 = V_2 \oplus W_2$.
We take $T_2 = T\big|_{V_2}$. Then by Lemma 3.2.4, $\dim V_2 = n_2$ and the matrix of T_2 in the basis $\{w_1, T_2w_1, T_2^2w_1, \ldots, T_2^{n_2-1}w_1\}$ of V_2 is M_{n_2}.

Continuing this way, we get subspaces $V_1, V_2, \ldots, V_k$ of V such that $\dim V_i = n_i$ $(1 \leq i \leq k)$, $n_1 \geq n_2 \geq \cdots \geq n_k$,

$$T(V_i) \subset V_i$$

and the matrix of $T_i = T\big|_{V_i}$ is M_{n_i}. Then by (previous) Lemma 3.2.7 there is a basis of V in which the matrix of T is of the form

$$\begin{pmatrix} M_{n_1} & 0 & \cdots & 0 \\ 0 & M_{n_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & M_{n_k} \end{pmatrix}.$$

□

This form of the matrix above is called the *canonical representation* of the nilpotent linear transformation T.

Definition 3.2.9. The integers $n_1, n_2, \dots, n_k$ in the canonical representation of nilpotent $T \in A(V)$ are called *invariants* of T.

Definition 3.2.10. Let $T \in A(V)$ be nilpotent. A subspace M of V is called *cyclic with respect to T* if

1. $T^m(M) = \{0\}$ but $T^{m-1}(M) \neq \{0\}$,
2. there is an element $z \in M$ such that $\{z, Tz, \dots, T^{m-1}z\}$ forms a basis of M,

Example 3.2.11. The subspaces $V_1, V_2, \dots, V_k$ obtained in the proof of Theorem 3.2.8 are cyclic with respect to T.

Lemma 3.2.12. *Let M be a subspace of a vector space V and $T \in A(V)$ be nilpotent. If* $\dim(M) = m$ *and M is cyclic with respect to T, then* $\dim(T^k M) = m - k$ *for all $k \leq m$.*

Proof. Since M is cyclic, the set $\{z, Tz, \dots, T^{m-1}z\}$ forms a basis of M, i.e.,

$$M = L(\{z, Tz, \dots, T^{m-1}z\}).$$

Then,

$$T^k M = L(\{T^k z, T^{k+1}z, \dots, T^{m-1}z\}) \qquad (\text{since } T^m(M) = \{0\}).$$

Clearly, $\{T^k z, T^{k+1}z, \dots, T^{m-1}z\}$ is linearly independent as it is a subset of an linearly independent set $\{z, Tz, \dots, T^{m-1}z\}$. Therefore, $\{T^k z, T^{k+1}z, \dots, T^{m-1}z\}$ forms a basis of $T^k M$ and hence $\dim(T^k M) = m - k$. □

Theorem 3.2.13. *Let V be a finite dimensional vector space over F and $T \in A(V)$ be nilpotent. Then the invariants of T are unique.*

In other words,

if $V_1, V_2, \dots, V_k$ are cyclic with respect to T such that $V = V_1 \oplus V_2 \oplus \cdots \oplus V_k$ with $\dim V_i =$ *n_i $(1 \leq i \leq k)$, $n_1 \geq n_2 \geq \cdots \geq n_k$,*

and $U_1, U_2, \dots, U_l$ are cyclic with respect to T such that $V = U_1 \oplus U_2 \oplus \cdots \oplus U_l$ with $\dim U_i = m_i$ *$(1 \leq i \leq l)$, $m_1 \geq m_2 \geq \cdots \geq m_l$,* *then $k = l$ and $m_i = n_i$ for all i.*

Proof. Suppose $V_1, V_2, \ldots, V_k$ are cyclic with respect to T such that

$$V = V_1 \oplus V_2 \oplus \cdots \oplus V_k$$

with $\dim V_i = n_i$ $(1 \le i \le k)$, $n_1 \ge n_2 \ge \cdots \ge n_k$ and also $U_1, U_2, \ldots, U_l$ are cyclic with respect to T such that

$$V = U_1 \oplus U_2 \oplus \cdots \oplus U_l$$

with $\dim U_i = m_i$ $(1 \le i \le l)$, $m_1 \ge m_2 \ge \cdots \ge m_l$.

We have to show that $k = l$ and $m_j = n_j$ for all j. If this is not the case then let i be the first integer such that $m_i \ne n_i$. Without the loss of generality, we may assume that

$$m_i < n_i. \tag{3.9}$$

Since $V = V_1 \oplus V_2 \oplus \cdots \oplus V_k$,

$$T^{m_i}V = T^{m_i}V_1 \oplus T^{m_i}V_2 \oplus \cdots \oplus T^{m_i}V_k.$$

Therefore, by (above) Lemma 3.2.12 (and considering only upto V_i but not upto V_k), we have

$$\dim(T^{m_i}V) \ge (n_1 - m_i) + (n_2 - m_i) + \cdots + (n_i - m_i). \tag{3.10}$$

Also since $V = U_1 \oplus U_2 \oplus \cdots \oplus U_l$, we have

$$T^{m_i}V = T^{m_i}U_1 \oplus T^{m_i}U_2 \oplus \cdots \oplus T^{m_i}U_{i-1} \qquad (\because T^{m_i}U_j = \{0\},\ i \ge j).$$

Therefore, by previous lemma we have

$$\dim(T^{m_i}V) = (m_1 - m_i) + (m_2 - m_i) + \cdots + (m_{i-1} - m_i).$$

Since i was the first integer such that $m_i \ne n_i$, we have $m_1 = n_1$, $m_2 = n_2$, $ldots, m_{i-1} = n_{i-1}$. Then, the above equation becomes

$$\dim(T^{m_i}V) = (n_1 - m_i) + (n_2 - m_i) + \cdots + (n_{i-1} - m_i). \tag{3.11}$$

Substituting the value of $\dim(T^{m_i}V)$ from equation (3.11) in equation (3.10), we have

$$\begin{aligned}
&(n_1 - m_i) + (n_2 - m_i) + \cdots + (n_{i-1} - m_i) \ge (n_1 - m_i) + (n_2 - m_i) + \cdots + (n_i - m_i) \\
&\Rightarrow 0 \ge n_i - m_i \\
&\Rightarrow m_i \ge n_i.
\end{aligned}$$

This is contradiction to (3.9). Hence, $k = l$ and $m_i = n_i$ for all i. □

Theorem 3.2.14. *Two nilpotent linear transformations are similiar if and only if they have the same invariants.*

Proof. Let $S,T \in A(V)$ be two nilpotent linear transformations. Suppose S and T have different invariants, say $n_1,n_2,\dots,n_k$ and $m_1,m_2,\dots,m_l$ respectively. Then by previous theorem and remark, their respective matrices

$$m(S)=\begin{pmatrix} M_{n_1} & 0 & \cdots & 0\\ 0 & M_{n_2} & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & M_{n_k}\end{pmatrix} \text{ and } m(T)=\begin{pmatrix} M_{m_1} & 0 & \cdots & 0\\ 0 & M_{m_2} & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & M_{m_l}\end{pmatrix}$$

cannot be similar. Hence, S and T cannot be similar.

Conversely, suppose that S and T have the same invariants, say $n_1,n_2,\dots,n_k$. Then by Theorem 3.2.8, there are bases $B_1=\{v_1,v_2,\dots,v_n\}$ and $B_2=\{w_1,w_2,\dots,w_n\}$ of V such that the matrix of S in the basis B_1 and the matrix of T in the basis B_2 is

$$m_{B_1}(S)=\begin{pmatrix} M_{n_1} & 0 & \cdots & 0\\ 0 & M_{n_2} & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & M_{n_k}\end{pmatrix}=m_{B_2}(T).$$

We have seen, in Proposition 2.3.11, that if two linear transformations have same matrix in different bases, then they must be similar. Hence, S and T are similar. □

Let us compute a couple of examples of invariants.

Example 3.2.15. Find the invariants of the linear tranformation $T: F^3 \to F^3$ defined by $T(x,y,z)=(y,z,0)$, where $x,y,z\in F$.

Solution. It is clear to see that the index of nilpotence of T is 3 as $T^3(x,y,z)=(0,0,0)$ for all $(x,y,z)\in F^3$. Hence, the invariant of T is $\dim F^3=n_1=3$. Now,

$$\begin{aligned} T(1,0,0)&=(0,0,0)\\ T(0,1,0)&=(1,0,0)\\ T(0,0,1)&=(0,1,0)\end{aligned}$$

Thus, the matrix of T in standard basis of F^3 is given by

$$\begin{pmatrix}0&1&0\\0&0&1\\0&0&0\end{pmatrix}.$$

Note that $v=(0,0,1)$ and $T^{n_1-1}=T^2v=(1,0,0)\neq(0,0,0)$. Hence, v,Tv,T^2v forms a basis of $V_1=F^3$ and the matrix of T in the canonical form is

$$M_{n_1}=\begin{pmatrix}0&0&0\\1&0&0\\0&1&0\end{pmatrix}.$$

□

Example 3.2.16. Find the invariants of $T : \mathbb{R}^3 \to \mathbb{R}^3$ defined by

$$T(x_1, x_2, x_3) = (x_2, 0, 0).$$

Solution. The index of nilpotence of T is $n_1 = 2$ as clearly, $T^2 = 0$. Then the invariants (as in Theorem 3.2.8) n_1 and n_2 such that $n_1 \geq n_2$ and $\dim \mathbb{R}^3 = n_1 + n_2$ are $n_1 = 2,\ n_2 = 1$. Note that, as seen in Lemma 3.2.4 and Lemma 3.2.6, we can find $v = (0,1,0) \in \mathbb{R}^3$ such that $T^{n_1-1}v = Tv = (1,0,0) \neq (0,0,0)$. Then $\mathbb{R}^3$ can be written as

$$\mathbb{R}^3 = V_1 \oplus V_2,$$

where basis of $V_1 = \{v, Tv\} = \{(0,1,0),(1,0,0)\}$ and $V_2 = L(\{(0,0,1)\}$. Then the matrix of $T\big|_{V_1}$ is given by $M_{n_1} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$ while $M_{n_2} = \begin{pmatrix} 0 \end{pmatrix}$. Then the canonical representation of T is as follows:

$$\begin{pmatrix} M_{n_1} & 0 \\ 0 & M_{n_2} \end{pmatrix} = \left(\begin{array}{cc|c} 0 & 0 & 0 \\ 1 & 0 & 0 \\ \hline 0 & 0 & 0 \end{array}\right).$$

Note that this is not the matrix of T in the standard basis. The matrix of T in the standard basis of $\mathbb{R}^3$ is given by

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

□

3.3 Canonical Forms: Jordan (Decomposition) Form

Lemma 3.3.1. *Let V be a finite dimensional vector space over F and $T \in A(V)$. Suppose that $V = V_1 \oplus V_2$, where V_1 and V_2 are subspaces of V invariant under T. Let $T_1 = T\big|_{V_1}$ and $T_2 = T\big|_{V_2}$. If the minimal polynomial of T_1 over F is $p_1(x)$ while minimal polynomial of T_2 over F is $p_2(x)$ then the minimal polynomial of T over F is the least common multiple of $p_1(x)$ and $p_2(x)$.*

Proof. Let $p(x) \in F[x]$ be the minimal polynomial for T. Then $p(T) = 0$. Thus for $v_1 \in V_1$, by definition of T_1, we have $T_1 v_1 = T v_1$. Therefore,

$$p(T_1)v_1 = p(T)v_1 = 0.$$

Similarly, for any $v_2 \in V_2$, since $T_2 v_2 = T v_2$, we have

$$p(T_2)v_2 = p(T)v_2 = 0.$$

Thus, $p(T) = 0 \Rightarrow p(T_1) = 0$ and $p(T_2) = 0$.
Since $p_1(x)$ and $p_2(x)$ are minimal polynomials for T_1 and T_2 respectively,

$$p_1(x) | p(x) \text{ and } p_2(x) | p(x).$$

Let $q(x)$ be the L.C.M. of $p_1(x)$ and $p_2(x)$. Then,

$$q(x)|p(x). \tag{3.12}$$

On the other hand, we know that both $p_1(x)$ and $p_2(x)$ divides their least common multiple $q(x)$, i.e.,

$$p_1(x)|q(x) \text{ and } p_2(x)|q(x).$$

Therefore, $q(T_1) = 0$ and $q(T_2) = 0$.
Now, let $v \in V$. Since $V = V_1 \oplus V_2$, there exists $v_1 \in V_1$ and $v_2 \in V_2$ such that

$$v = v_1 + v_2.$$

Then,

$$\begin{aligned} q(T)v &= q(T)v_1 + q(T)v_2 \\ &= q(T_1)v_1 + q(T_2)v_2 = 0. \end{aligned}$$

Thus, $q(T) = 0$. Since $p(x)$ is the minimal polynomial for T, and T satisfies $q(x)$, we conclude that

$$p(x)|q(x). \tag{3.13}$$

By equations (3.12) and (3.13), we have $p(x) = q(x)$. Hence, the minimal polynomial for T is the least common multiple of $p_1(x)$ and $p_2(x)$. □

Corollary 3.3.2. *If $V = V_1 \oplus V_2 \oplus \cdots \oplus V_k$, where each V_i is invariant under T and if $p_i(x)$ is the minimal polynomial for $T_i = T\big|_{V_i}$ over F, then the minimal polynomial of T over F is the least common multiple of $p_1(x), p_2(x), \ldots, p_k(x)$.*

Notations: Suppose V is a finite dimensional vector space over F and $T \in A(V)$.

Suppose $p(x) \in F[x]$ is the minimal polynomial for T. Consider

$$p(x) = q_1(x)^{l_1} q_2(x)^{l_2} \cdots q_k(x)^{l_k},$$

where $q_i(x)$, $(i = 1, 2, \ldots, k)$ are distinct irreducible polynomials in $F[x]$ and $l_1, l_2, \ldots, l_k$ are positive integers. For $i = 1, 2, \ldots, k$, consider

$$V_i = \{v \in V : q_i(T)^{l_i} v = 0\} = \ker(q_i(T)^{l_i}).$$

Then V_i is a subspace of V which is invariant under T. This is because, if $v \in V_i$ then

$$q_i(T)^{l_i}(Tv) = T(q_i(T)^{l_i} v) = T(0) = 0.$$

Therefore, $Tv_i \in V_i$ and hence each V_i is invariant under T.

Theorem 3.3.3. *With the above notations, $V_i \neq \{0\}$, for each $i = 1, 2, \ldots, k$, $V = V_1 \oplus V_2 \oplus \cdots \oplus V_k$, and the minimal polynomial of $T_i = T\big|_{V_i}$ is $q_i(x)^{l_i}$.*

Proof. First we show that $V_i \neq \{0\}$ for all $i = 1, 2, \ldots, k$. For $k = 1$ there is nothing to prove. So, we may assume $k > 1$. Consider

$$\begin{aligned} h_1(x) &= q_2(x)^{l_2} q_3(x)^{l_3} \cdots q_k(x)^{l_k}, \\ h_2(x) &= q_1(x)^{l_1} q_3(x)^{l_3} \cdots q_k(x)^{l_k}, \\ &\vdots \\ h_i(x) &= q_1(x)^{l_1} \cdots q_{i-1}(x)^{l_{i-1}} q_{i+1}(x)^{l_{i+1}} \cdots q_k(x)^{l_k}. \end{aligned}$$

That is

$$h_i(x) = \frac{p(x)}{q_i(x)^{l_i}}, \qquad i = 1, 2, \ldots, k.$$

Clearly, for $i = 1, 2, \ldots, k$, $\deg h_i(x) < \deg p(x)$. Since $p(x)$ is the minimal polynomial for T, we have $h_i(T) \neq 0$ and so $h_i(T)(V) \neq \{0\}$. Now, for $v \in V$,

$$q_i(T)^{l_i} h_i(T)v = p(T)v = 0.$$

Therefore, $h_i(T)v \in \ker(q_i(T)^{l_i}) = V_i$ and hence we have

$$\{0\} \neq h_i(T)(V) \subset V_i. \tag{3.14}$$

$$\therefore \boxed{V_i \neq \{0\}.}$$

Now, we show that $V = V_1 \oplus V_2 \oplus \cdots \oplus V_k$. For this we have to show that for $1 \leq i \leq k$:

- Every $v \in V$ can be written as $v = v_1 + v_2 + \cdots + v_k$, where $v_i \in V_i$.
- If $u_1 + u_2 + \cdots + u_k = 0$ with $u_i \in V_i$ then each $u_i = 0$.

Since GCD of $h_1(x), h_2(x), \ldots, h_k(x)$ is 1, there exists $a_1(x), a_2(x), \ldots, a_k(x) \in F[x]$ such that

$$a_1(x)h_1(x) + a_2(x)h_2(x) + \cdots + a_k(x)h_k(x) = 1.$$

Therefore,

$$a_1(T)h_1(T) + a_2(T)h_2(T) + \cdots + a_k(T)h_k(T) = I.$$

Then for any $v \in V$,

$$v = a_1(T)h_1(T)v + a_2(T)h_2(T)v + \cdots + a_k(T)h_k(T)v.$$

Now, by (3.14), $h_i(T) \subset V_i$ for all $i = 1, 2, \ldots, k$ and V_i is invariant under T (by above description in Notations), we have

$$a_i(T)h_i(T)v \in V_i.$$

Therefore,

$$v = v_1 + v_2 + \cdots + v_k,$$

where $v_i = a_i(T)h_i(T)v \in V_i,\ i = 1, 2, \ldots, k$. Thus, we can write V as

$$V = V_1 + V_2 + \cdots + V_k.$$

To show that V is the direct sum, i.e., $V = V_1 \oplus V_2 \oplus \cdots \oplus V_k$ it is enough to show that if

$$u_1 + u_2 + \cdots + u_k = 0 \tag{3.15}$$

with $u_i \in V_i$ then each $u_i = 0$.

Suppose if possible, some u_i is not 0, say $u_1 \neq 0$. Then by (3.15), we get

$$h_1(T)u_1 + h_1(T)u_2 + \cdots + h_1(T)u_k = h_1(T)0 = 0. \tag{3.16}$$

Now, for $2 \leq i \leq k$, $u_i \in V_i = \ker(q_i(T)^{l_i}) \Rightarrow q_i(T)^{l_i} u_i = 0$ and so

$$h_(T)u_i = q_2(T)^{l_2} q_3(T)^{l_3} \cdots q_k(T)^{l_k} u_i = 0.$$

So equation (3.16) reduces to $h_1(T)u_1 = 0$. Also since $u_1 \in V_1 = \ker(q_1(T)^{l_1})$, we have $q_1(T)^{l_1} u_1 = 0$. Now, as GCD of $h_1(x)$ and $q_1(x)^{l_1}$ is 1, there exists $\lambda(x), \mu(x) \in F[x]$ such that

$$\begin{aligned}
&\lambda(x)h_1(x) + \mu(x)q_1(x)^{l_1} = 1.\\
\therefore\ &\lambda(T)h_1(T) + \mu(T)q_1(T)^{l_1} = I.\\
\therefore\ &\lambda(T)h_1(T)u_1 + \mu(T)q_1(T)^{l_1}u_1 = u_1.\\
\therefore\ &u_1 = 0.
\end{aligned}$$

This is a contradiction to our assumption that $u_1 \neq 0$. Therefore, all u_i's must be 0 and hence

$$\boxed{V = V_1 \oplus V_2 \oplus \cdots \oplus V_k.}$$

Now, consider $T_i = T\big|_{V_i}$. Then we show that the minimal polynomial for T_i is $q_i(x)^{l_i}$.

By definition of $V_i (= \ker(q_i(T)^{l_i})$, since $q_i(T)^{l_i}(V_i) = \{0\}$ we have $q_i(T_i)^{l_i} = 0$. Therefore, the minimal polynomial of T_i must divide $q_i(x)^{l_i}$ and hence it will be of the form $q_i(x)^{f_i}$, where $f_i \leq l_i$. Then by Corollary 3.3.2, the minimal polynomial of T over F is the LCM of $q_1(x)^{f_1}$, $q_2(x)^{f_2}, \ldots, q_k(x)^{f_k}$ and since $q_i(x)$ are distinct and irreducible, the minimal polynomial must be their product, i.e.,

$$q_1(x)^{f_1}, q_2(x)^{f_2}, \ldots, q_k(x)^{f_k}.$$

But (by notations given above), the minimal polynomial for T is

$$p(x) = q_1(x)^{l_1} q_2(x)^{l_2} \cdots q_k(x)^{l_k}$$

and so, we have

$$q_1(x)^{l_1} q_2(x)^{l_2} \cdots q_k(x)^{l_k} = q_1(x)^{f_1}, q_2(x)^{f_2}, \ldots, q_k(x)^{f_k}.$$

Therefore, $f_i = l_i$ for all $i = 1, 2, \ldots, k$. Hence, $\boxed{q_i(x)^{l_i} \text{ is the minimal polynomial for } T_i.}$ □

Remark 3.3.4. We know that, by Remark 3.1.7, $\lambda \in F$ is a root of the minimal polynomial for $T \in A(V)$ if and only if λ is a characteristic root of T. If all the characteristic roots of T lie in F then the minimal polynomial $p(x) \in F[x]$ for T has the following nice form:

$$p(x) = (x - \lambda_1)^{l_1}(x - \lambda_2)^{l_2} \cdots (x - \lambda_k)^{l_k},$$

where $\lambda_1, \lambda_2, \ldots, \lambda_k$ are distinct characteristic roots of T. The irreducible factor $q_i(x)$ is now $q_i(x) = x - \lambda_i$. Note that, λ_i is the only characteristic root of T on V_i.

Corollary 3.3.5. *Let V be a finite dimensional vector space over F and $T \in A(V)$ be such that all the characteristic roots of T lie in F (i.e., all the roots of the minimal polynomial for T are in F). If $\lambda_1, \lambda_2, \ldots, \lambda_k \in F$ are the* ***distinct*** *characteristic roots of T, then the minimal polynomial $p(x)$ for T is*

$$p(x) = (x-\lambda_1)^{l_1}(x-\lambda_2)^{l_2}\cdots(x-\lambda_k)^{l_k}$$

and

$$V = V_1 \oplus V_2 \oplus \cdots \oplus V_k,$$

where $V_i = \ker(T-\lambda_i I)^{l_i} = \{v \in V : (T-\lambda_i I)^{l_i} v = 0\}$, $i = 1, 2, \ldots, k$.

We are now in the position to state the condition when the matrix of $T \in A(V)$ will be in special recognizable form. Before we exhibit this condition, consider the following definition:

Definition 3.3.6. The matrix

$$\begin{pmatrix} \lambda & 0 & 0 & \cdots & 0 & 0 \\ 1 & \lambda & 0 & \cdots & 0 & 0 \\ 0 & 1 & \lambda & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda & 0 \\ 0 & 0 & 0 & \cdots & 1 & \lambda \end{pmatrix} \tag{3.17}$$

with λ's on the diagonal, 1's on the subdiagonal (entries just under the diagonal), and 0's elsewhere, is called the basic *Jordan block belonging to* λ.

Theorem 3.3.7 (Jordan Decomposition)**.** *Let V be a finite dimensional vector space over F and $T \in A(V)$ be such that all its distinct characteristic roots are in F (i.e., all the roots of minimal polynomial for T are in F). Then there is a basis of V in which the matrix of T is of the form*

$$J = \begin{pmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_k \end{pmatrix}$$

where for $i = 1, 2, \ldots, k$ each

$$J_i = \begin{pmatrix} B_{i_1} & & & \\ & B_{i_2} & & \\ & & \ddots & \\ & & & B_{i_{r_i}} \end{pmatrix}$$

and where $B_{i_1}, B_{i_2}, \ldots, B_{i_{r_i}}$ are basic Jordan blocks belonging to λ_i (as given in (3.17)).

Proof. Suppose $\lambda_1, \lambda_2, \ldots, \lambda_k \in F$ are distinct characteristic roots of T (or roots of the minimal polynomial). Then by previous corollary, the minimal polynomial for T is of the form

$$\begin{aligned} p(x) &= (x-\lambda_1)^{l_1}(x-\lambda_2)^{l_2}\cdots(x-\lambda_k)^{l_k}, \\ V &= V_1 \oplus V_2 \oplus \cdots \oplus V_k, \end{aligned}$$

where $V_i = \ker(T-\lambda_i I)^{l_i}$, $i = 1,2,\ldots,k$. Also, each V_i is invariant under T and the minimal polynomial for $T_i = T\big|_{V_i}$ is $(x-\lambda_i)^{l_i}$. Therefore, for $i = 1,2,\ldots,k$,

$$(T_i - \lambda_i I)^{l_i} = 0.$$

Thus, $T_i - \lambda_i I$ is nilpotent and hence by Theorem 3.2.8 there is a basis of V_i in which the matrix of $T_i - \lambda_i I$ is of the form

$$\begin{pmatrix} M_{i_1} & 0 & \cdots & 0 \\ 0 & M_{i_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & M_{i_{r_i}} \end{pmatrix}.$$

Since, T_i can be written as $T_i = (T_i - \lambda_i I) + \lambda_i I$, the matrix of T_i can be written as

$$\begin{aligned} m(T_i) &= m(T_i - \lambda_i I) + m(\lambda_i I) \\ &= \begin{pmatrix} M_{i_1} & & & \mathbf{0} \\ & M_{i_2} & & \\ & & \ddots & \\ \mathbf{0} & & & M_{i_{r_i}} \end{pmatrix} + \begin{pmatrix} \lambda_i & & & \mathbf{0} \\ & \lambda_i & & \\ & & \ddots & \\ \mathbf{0} & & & \lambda_i \end{pmatrix} \\ &= \begin{pmatrix} B_{i_1} & & & \mathbf{0} \\ & B_{i_2} & & \\ & & \ddots & \\ \mathbf{0} & & & B_{i_{r_i}} \end{pmatrix} = J_i. \end{aligned}$$

Since $V = V_1 \oplus V_2 \oplus \cdots \oplus V_k$ and each V_i is invariant under T, by Lemma 3.2.7 there is a basis of V in which the matrix of T is of the form

$$J = \begin{pmatrix} J_1 & 0 & \cdots & 0 \\ 0 & J_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & J_k \end{pmatrix}.$$

□

The above matrix J is called the *Jordan canonical form* of T. We conclude this section by leaving the following result as an exercise:

Exercise 3.3.8. Two linear transformation in $A_F(V)$ having all their characteristic roots in F are similar if and only if they have the same Jordan form (except for the order of their characteristic roots).

CHAPTER 4

TRACE, TRANSPOSE, DETERMINANTS AND CLASSIFICATION OF QUADRATICS

4.1 Trace and Transpose

Definition 4.1.1. Let F be a field and $A \in M_n(F)$. If $A = (\alpha_{ij})$, then the trace of A is

$$\operatorname{tr}(A) = \sum_{i=1}^{n} \alpha_{ii},$$

i.e., the trace of a matrix A is the sum of all the diagonal entries of A.

Note that, trace of a matrix is an element of F. Thus, trace (of a matrix is a) function

$$\operatorname{tr} : M_n(F) \to F \text{ is given by } A \mapsto \operatorname{tr}(A).$$

The following lemma lists some of the fundamental properties of the trace function. The first two properties proves that trace function 'tr' is a linear function.

Lemma 4.1.2. *Let F be a field. Then for $A, B \in M_n(F)$ and $\lambda \in F$,*

1. $\operatorname{tr}(\lambda A) = \lambda \operatorname{tr}(A)$.
2. $\operatorname{tr}(A + B) = \operatorname{tr}(A) + \operatorname{tr}(B)$.
3. $\operatorname{tr}(AB) = \operatorname{tr}(BA)$.

Proof. Suppose $A = (\alpha_{ij})$ and $B = (\beta_{ij})$. Then

1. $\lambda A = (\lambda \alpha_{ij})$. Therefore,

$$\operatorname{tr}(\lambda A) = \sum_{i=1}^{n} \lambda \alpha_{ii} = \lambda \sum_{i=1}^{n} \alpha_{ii} = \lambda \operatorname{tr}(A).$$

2. $A+B=(\alpha_{ij}+\beta ij)$. Therefore,

$$\begin{aligned}\operatorname{tr}(A+B) &= \sum_{i=1}^{n}(\alpha_{ii}+\beta_{ii})\\ &= \sum_{i=1}^{n}\alpha_{ii}+\sum_{i=1}^{n}\beta_{ii}\\ &= \operatorname{tr}(A)+\operatorname{tr}(B).\end{aligned}$$

Thus, tr is a linear map on $M_n(F)$.

3. Suppose $AB=(\gamma_{ij})$, where $\gamma_{ij}=\sum_{k=1}^{n}\alpha_{ik}\beta_{kj}$ and $BA=(\mu_{ij})$, where $\mu_{ij}=\sum_{k=1}^{n}\beta_{ik}\alpha_{kj}$. Then,

$$\begin{aligned}\operatorname{tr}(AB) &= \sum_{i=1}^{n}\gamma_{ii}\\ &= \sum_{i=1}^{n}\left(\sum_{k=1}^{n}\alpha_{ik}\beta_{ki}\right)\\ &= \sum_{k=1}^{n}\left(\sum_{i=1}^{n}\beta_{ki}\alpha_{ik}\right) && \text{(interchanging the order of summation)}\\ &= \sum_{k=1}^{n}\mu_{kk}=\operatorname{tr}(BA).\end{aligned}$$

□

Corollary 4.1.3. *Let F be a field and $A\in M_n(F)$. Let $C\in M_n(F)$ be an invertible matrix. Then* $\operatorname{tr}(C^{-1}AC)=\operatorname{tr}(A)$.

In other words, "two similar matrices have the same trace".

Proof. If A and B are similar matrices then there exist an invertible matrix C such that $B=C^{-1}AC$. Then,

$$\begin{aligned}\operatorname{tr}(B) &= \operatorname{tr}(C^{-1}AC)\\ &= \operatorname{tr}((C^{-1}A)C)\\ &= \operatorname{tr}(C(C^{-1}A)) && \text{(by property 3 above)}\\ &= \operatorname{tr}(A).\end{aligned}$$

□

Using the definition of trace of a matrix, we now define trace of a linear transformation.

Definition 4.1.4. Let $T\in A(V)$. Then $\operatorname{tr}(T)$, the *trace of* T, is the trace of $m_1(T)$, where $m_1(T)$ is the matrix of T in some basis of V.

Remark 4.1.5. Since by previous corollary, trace of similar matrices is same, the definition of trace of a linear transformation T makes sense. For if $m_2(T)$ is the matrix of T in some other basis of V then $m_1(T)$ and $m_2(T)$ are similar matrices and their trace being same, $\text{tr}(T)$ remains the same.

We recall the definition of a splitting field.

Definition 4.1.6. Suppose F is a field and $p(x) \in F[x]$. The smallest field K such that F is a subfield of K and K contains all the roots of $p(x)$ is called the splitting field of $p(x)$.
Example: The splitting field of $x^2+1 \in \mathbb{R}[x]$ or in $\mathbb{Q}[x]$ is $\mathbb{C}$. The splitting field of $x^2-2 \in \mathbb{Q}[x]$ is $\mathbb{Q}(\sqrt{2})$.

Lemma 4.1.7. *If $T \in A(V)$, then $\text{tr}(T)$ is the sum of all the characteristic roots of T counted according to their multiplicity.*

Proof. Let A be the matrix of T and $p(x) \in F[x]$ be the minimal polynomial for T (i.e. for A). Let K be the splitting field of $p(x)$. Then all the roots of $p(x)$ are in K, i.e., all the characteristic roots of T are in K. Then by Theorem 3.3.7, the matrix A is similar to Jordan form matrix $J \in M_n(K)$ of T. Then,

$$\text{tr}(T) = \text{tr}(A) = \text{tr}(J).$$

Since in matrix J the characteristic roots of T appear on the diagonal, the $\text{tr}(J)$ is the sum of all the diagonal entries of the matrix J counted according to their multiplicities. □

As an application of the above lemma, we make the following remark:

Remark 4.1.8. *The trace of a nilpotent linear transformation (or matrix) be* 0.

By the above lemma, trace of a linear transformation (or its matrix) is the sum of its characteristic roots. We know that, 0 is the only characteristic root of a nilpotent linear transformation and hence its trace is 0.

Question 4.1.9. What about the converse of the above remark, i.e. if the trace of a linear transformation (or its matrix) is 0, can we say that it is nilpotent?

Solution. The converse is not true in general. If trace of a matrix is 0, it may not necessarily be nilpotent. Consider the following example: □

Example 4.1.10. Let $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ then $\text{tr}(A) = 0$. However, $A^2 = I$, $A^3 = A$. So A is not nilpotent.

Thus, we now know that if $\text{tr}(T) = 0$ then T may not be nilpotent. Now, the question remains that under what additional conditions we can conclude that T is nilpotent. This is answered by the following lemma:

Lemma 4.1.11. *Let F be a field of characteristic* 0*, V be a vector space over F and $T \in A(V)$. Then T is nilpotent if and only if $\text{tr}(T^i) = 0$ for all $i \geq 1$.*

Proof. If T is nilpotent, then T^i is also nilpotent and hence $\mathrm{tr}(T^i) = 0$ for all $i \geq 1$.

Conversely, suppose $\mathrm{tr}(T^i) = 0$ for all $i \geq 1$. Let $p(x) = \alpha_0 + \alpha_1 x + \cdots + \alpha_m x^m \in F[x]$ be the minimal polynomial for T. Then,

$$0 = p(T) = \alpha_0 I + \alpha_1 T + \cdots + \alpha_m T^m.$$

Taking trace on both the sides, we get

$$0 = \alpha_0 \,\mathrm{tr}(I) + \alpha \,\mathrm{tr}(T) + \cdots + \alpha_m \,\mathrm{tr}(T^m).$$

Since $\mathrm{tr}(T^i) = 0$ for all $i \geq 1$, from above equation, we have

$$\alpha_0 n = 0 \Rightarrow \alpha_0 = 0 \quad (\because \text{ characteristic of } F \text{ is } 0).$$

Since the constant term in the minimal polynomial for T is 0, by Theorem 2.1.15, T is singular and hence by Theorem 2.1.19, 0 is a characteristic root of T.

Let K be the splitting field of $p(x)$. We can consider T as a matrix in $M_n(F)$ and hence in $M_n(K)$. Since K is the extension of the field F such that all the roots of $p(x)$ (i.e. characteristic roots of T) are in K, by Theorem 3.1.9, we can bring T to a triangular form in $M_n(K)$. Since 0 is a characteristic root of T, the matrix T is of the form

$$\left(\begin{array}{c|ccc} 0 & \beta_2 & \cdots & \beta_n \\ \hline 0 & \alpha_2 & & \\ \vdots & & \ddots & * \\ 0 & 0 & & \alpha_n \end{array}\right) = \left(\begin{array}{c|c} 0 & * \\ \hline 0 & T_2 \end{array}\right),$$

where

$$T_2 = \begin{pmatrix} \alpha_2 & & * \\ & \ddots & \\ 0 & & \alpha_n \end{pmatrix}$$

is an $(n-1) \times (n-1)$ (upper triangular) matrix and * denotes the entries in K. Now,

$$T^k = \left(\begin{array}{c|c} 0 & * \\ \hline 0 & T_2^k \end{array}\right)$$

and hence $0 = \mathrm{tr}(T^k) = \mathrm{tr}(T_2^k)$. Thus, T_2 is an $(n-1) \times (n-1)$ matrix with $\mathrm{tr}(T_2^k) = 0$ for all $k \geq 1$. Repeating the same argument for T_2 and continuing this way, we get $\alpha_2 = \cdots = \alpha_n = 0$. Thus, T can be converted to a triangular matrix with all the entries of main diagonal equal to 0. Then (by a seminar exercise) T is nilpotent. □

Exercise 4.1.12. Prove that there do not exists $A, B \in M_n(F)$ such that $AB - BA = I$, where F is a field of characteristic 0.

Solution. Suppose there exists $A, B \in M_n(F)$ such that $AB - BA = I$. Then taking trace on both the sides, we get

$$\begin{aligned} n = \mathrm{tr}(I) &= \mathrm{tr}(AB - BA) \\ &= \mathrm{tr}(AB) - \mathrm{tr}(BA) && (\because \mathrm{tr}(A+B) = \mathrm{tr}(A) + \mathrm{tr}(B)) \\ &= \mathrm{tr}(AB) - \mathrm{tr}(AB) && (\because \mathrm{tr}(AB) = \mathrm{tr}(BA)) \\ &= 0 \end{aligned}$$

which is not possible since characteristic of the field F is 0. Hence, the result. □

As an immediate application of Lemma 4.1.11, we have the following result usually known as *Jacobson lemma*.

Lemma 4.1.13 (Jacobson lemma)**.** *Let F be a field of characteristic 0 and V be a vector space over F. If $S,T \in A(V)$ such that $ST-TS$ commutes with S, then $ST-TS$ is nilpotent.*

Proof. For any $k \geq 1$, we compute $(ST-TS)^k$. Now,

$$\begin{aligned}(ST-TS)^k &= (ST-TS)^{k-1}(ST-TS)\\ &= (ST-TS)^{k-1}ST-(ST-TS)^{k-1}TS\\ &= S(ST-TS)^{k-1}T-(ST-TS)^{k-1}TS \qquad (\because ST-TS \text{ commutes with } S)\\ &= SB-BS,\end{aligned}$$

where $B=(ST-TS)^{k-1}T$. Hence,

$$\begin{aligned}\operatorname{tr}((ST-TS)^k) &= \operatorname{tr}(SB-BS)\\ &= tr(SB)-\operatorname{tr}(BS) \qquad (\because \operatorname{tr}(A+B)=\operatorname{tr}(A)+\operatorname{tr}(B))\\ &= \operatorname{tr}(SB)-\operatorname{tr}(SB) \qquad (\because \operatorname{tr}(AB)=\operatorname{tr}(BA))\\ &= 0\end{aligned}$$

for all $k \geq 1$. Then by previous lemma, $ST-TS$ is nilpotent. □

Definition 4.1.14. If $A=(\alpha_{ij}) \in M_n(F)$ then the *transpose* of A, denoted by A', is the matrix $A'=(\gamma_{ij})$, where $\gamma_{ij}=\alpha_{ji}$ for all $1 \leq i,j \leq n$.
Thus, the transpose of A a matrix is obtained by interchanging rows and columns of A.

Some of the basic properties of the transpose are given in the following lemma:

Lemma 4.1.15. *For all $A,B \in M_n(F)$ and $\lambda \in F$, prove that*

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

Proof. Homework (given as a seminar exercise). □

Exercises 4.1.16. Prove the following:

1. Let $p(x) \in F[x]$. Then $p(A)=0$ if and only if $p(A')=0$. Hence, the minimal polynomial for A and A' are same.
2. A is invertible if and only if A' is invertible and
$$(A')^{-1} = (A^{-1})'.$$
3. λ is a characteristic root of A if and only if λ is a characteristic root of A'.

Definition 4.1.17. A matrix $A \in M_n(F)$ is said to be a *symmetric matrix* if $A' = A$.

Thus, if $A = (\alpha_{ij})$ is symmetric then

$$\alpha_{ij} = \alpha_{ji} \quad (\text{for all } i, j).$$

Definition 4.1.18. A matrix $A \in M_n(F)$ is said to be a *skew-symmetric matrix* if $A' = -A$. Thus, if $A = (\alpha_{ij})$ is skew-symmetric then

$$\alpha_{ij} = -\alpha_{ji} \quad (\text{for all } i, j, i \neq j).$$

Lemma 4.1.19. *Let F be a field with characteristic different from 2. Then every matrix $M_n(F)$ can be uniquely written as a sum of a symmetric and a skew-symmetric matrix.*

Proof. Let $A \in M_n(F)$ and let $A = B + C$, where $B \in M_n(F)$ is any symmetric matrix and $C \in M_n(F)$ be any skew-symmetric matrix, i.e., $B' = B$ and $C' = C$. Now,

$$A' = (B+C)' = B' + C' = B - C.$$

Therefore, $A + A' = 2B$ and $A - A' = 2C$. Hence,

$$B = \frac{(A+A')}{2} \text{ and } C = \frac{(A-A')}{2}.$$

Thus, A can be written as

$$A = B + C = \frac{(A+A')}{2} + \frac{(A-A')}{2}.$$

It is easy to verify that $B = \frac{(A+A')}{2}$ is a symmetric matrix and $C = \frac{(A-A')}{2}$ is a skew-symmetric matrix. □

4.2 Determinants

Definition 4.2.1. Let F be a field and $A \in M_n(F)$. Then the *determinant of* A, written as $\det(A)$, is the element of F defined as

$$\det(A) = \sum_{\sigma \in S_n} (-1)^{\sigma} \alpha_{1\sigma(1)} \alpha_{2\sigma(2)} \cdots \alpha_{n\sigma(n)},$$

where

$$(-1)^{\sigma} = \begin{cases} 1 & \text{if } \sigma \text{ is an even permuation;} \\ -1 & \text{if } \sigma \text{ is an odd permutation.} \end{cases}$$

Remark 4.2.2. The determinant function is a function on $M_n(F)$ and takes values in F,

$$\det : M_n(F) \to F \text{ defined by } A \mapsto \det(A).$$

Lemma 4.2.3. *The determinant of a (lower) triangular matrix is the product of its entries on the main diagonal, i.e.,*

if $A = (\alpha_{ij}) \in M_n(F)$ is a lower triangular matrix, then $\det(A) = \alpha_{11}\alpha_{22}\cdots\alpha_{nn}$.

Proof. Note that, since $A = (\alpha_{ij})$ is a lower triangular matrix,

$$\alpha_{ij} = 0 \text{ if } j > i. \tag{4.1}$$

Now by definition of determinant,

$$\det(A) = \sum_{\sigma \in S_n} (-1)^\sigma \alpha_{1\sigma(1)}\alpha_{2\sigma(2)}\cdots\alpha_{n\sigma(n)}. \tag{4.2}$$

If $\sigma(1) \neq 1$. Then obviously $\sigma(1) > 1$ and hence by equation (4.1), the element $\alpha_{1\sigma(1)} = 0$. Thus, in the expansion of $\det(A)$ given in (4.2) above, the non-zero contribution comes from only those terms where $\sigma(1) = 1$.

Now, σ is a permutation (which is a one-one function) and $\sigma(1) = 1$ and so $\sigma(2) \neq 1$. If $\sigma(2) > 2$, then again because of the condition (4.1), $\alpha_{2\sigma(2)} = 0$. Thus, to get a non-zero contribution in the expansion (4.2) of $\det(A)$, $\sigma(2) = 2$.

Continuing this way, we get $\sigma(i) = i$ for all $i = 1, 2, \ldots, n$. Then σ is the identity permutation which is an even permutation. Hence, by (4.2) we have

$$\det(A) = \alpha_{11}\alpha_{22}\cdots\alpha_{nn}.$$

□

Lemma 4.2.4. *Determinants of a matrix and its transpose are same, i.e., if A is in $M_n(F)$ and A' is its transpose then*

$$\det(A) = \det(A').$$

Proof. Let $A = (\alpha_{ij}) \in M_n(F)$. Then $A' = (\beta_{ij})$, where $\beta_{ij} = \alpha_{ji}$ for all $i, j = 1, 2, \ldots, n$ and

$$\begin{aligned}
\det(A') &= \sum_{\sigma \in S_n} (-1)^\sigma \beta_{1\sigma(1)}\beta_{2\sigma(2)}\cdots\beta_{n\sigma(n)} \\
&= \sum_{\sigma \in S_n} (-1)^\sigma \alpha_{\sigma(1)1}\alpha_{\sigma(2)2}\cdots\alpha_{\sigma(n)n} \\
&= \sum_{\sigma \in S_n} (-1)^\sigma \alpha_{1\sigma^{-1}(1)}\alpha_{2\sigma^{-1}(2)}\cdots\alpha_{n\sigma^{-1}(n)} \\
&= \sum_{\sigma^{-1} \in S_n} (-1)^{\sigma^{-1}} \alpha_{1\sigma^{-1}(1)}\alpha_{2\sigma^{-1}(2)}\cdots\alpha_{n\sigma^{-1}(n)} \\
&= \sum_{\sigma \in S_n} (-1)^\sigma \alpha_{1\sigma(1)}\alpha_{2\sigma(2)}\cdots\alpha_{n\sigma(n)} \\
&= \det(A).
\end{aligned}$$

Note that the above sum is over all permutations $\sigma \in S_n$ and if σ is a permutation then so is σ^{-1}. □

Corollary 4.2.5. *The determinant of an upper triangular matrix is the product of its entries on the main diagonal, i.e.,*

if $A = (\alpha_{ij}) \in M_n(F)$ *is a lower triangular matrix, then* $\det(A) = \alpha_{11}\alpha_{22}\cdots\alpha_{nn}$.

Proof. Let $A = (\alpha_{ij}) \in M_n(F)$ be an upper triangular matrix. Then A' will be a lower triangular matrix. Then, by by lemma 4.2.3

$$\det(A') = \alpha_{11}\alpha_{22}\cdots\alpha_{nn}.$$

But by previous lemma, $\det(A) = \det(A')$ and hence,

$$\det(A) = \alpha_{11}\alpha_{22}\cdots\alpha_{nn}.$$

□

Notation: Given a matrix $A = (\alpha_{ij}) \in M_n(F)$, consider the vector $v_1 = (\alpha_{11}, \alpha_{12}, \ldots, \alpha_{1n})$ to be the first row of the matrix A. Similarly, let $v_2 = (\alpha_{21}, \alpha_{22}, \ldots, \alpha_{2n})$ be the second row vector and so on.Then we may also denote

$$\det(A) = d(v_1, v_2, \ldots, v_n).$$

With this notation, we now state the next lemma which says that *'if all the elements of in one row of a matrix* $A \in M_n(F)$ *is multiplied by a fixed element* $\gamma \in F$ *then* $\det(A)$ *is itself multiplied by* λ*'.*

Lemma 4.2.6. *If* $A \in M_n(F)$ *and* $\lambda \in F$ *then*

$$d(v_1, \ldots, v_{i-1}, \lambda v_i, v_{i+1}, \ldots, v_n) = \lambda\, d(v_1, \ldots, v_{i-1}, v_i, v_{i+1}, \ldots, v_n).$$

Proof.

$$\begin{aligned} & d(v_1, \ldots, v_{i-1}, \lambda v_i, v_{i+1}, \ldots, v_n) \\ = & \sum_{\sigma \in S_n} (-1)^{\sigma} \alpha_{1\sigma(1)} \cdots \alpha_{i-1,\sigma(i-1)} (\lambda \alpha_{i\sigma(i)}) \alpha_{i+1,\sigma(i+1)} \cdots \alpha_{n\sigma(n)} \\ = & \lambda \sum_{\sigma \in S_n} (-1)^{\sigma} \alpha_{1\sigma(1)} \cdots \alpha_{i\sigma(i)} \cdots \alpha_{n\sigma(n)} \\ = & \lambda\, d(v_1, \ldots, v_{i-1}, v_i, v_{i+1}, \ldots, v_n). \end{aligned}$$

□

Lemma 4.2.7. *If* $A \in M_n(F)$ *and* $u_i = (\beta_{i1}, \beta_{i2}, \ldots, \beta_{in})$ *then*

$$\begin{aligned} & d(v_1, \ldots, v_{i-1}, v_i + u_i, v_{i+1}, \ldots, v_n) \\ = & d(v_1, \ldots, v_{i-1}, v_i, v_{i+1}, \ldots, v_n) + d(v_1, \ldots, v_{i-1}, u_i, v_{i+1}, \ldots, v_n). \end{aligned}$$

Proof.

$$
\begin{aligned}
& d(v_1,\dots,v_{i-1},v_i+u_i,v_{i+1},\dots,v_n) \\
= & \sum_{\sigma\in S_n}(-1)^\sigma \alpha_{1\sigma(1)}\cdots\alpha_{i-1,\sigma(i-1)}(\alpha_{i\sigma(i)}+\beta_{i\sigma(i)})\alpha_{i+1,\sigma(i+1)}\cdots\alpha_{n\sigma(n)} \\
= & \sum_{\sigma\in S_n}(-1)^\sigma \alpha_{1\sigma(1)}\cdots\alpha_{i-1,\sigma(i-1)}\alpha_{i\sigma(i)}\alpha_{i+1,\sigma(i+1)}\cdots\alpha_{n\sigma(n)} \\
& \quad + \sum_{\sigma\in S_n}(-1)^\sigma \alpha_{1\sigma(1)}\cdots\alpha_{i-1,\sigma(i-1)}\beta_{i\sigma(i)}\alpha_{i+1,\sigma(i+1)}\cdots\alpha_{n\sigma(n)} \\
= & d(v_1,\dots,v_{i-1},v_i,v_{i+1},\dots,v_n)+d(v_1,\dots,v_{i-1},u_i,v_{i+1},\dots,v_n).
\end{aligned}
$$

□

Remark 4.2.8. The above lemma **does not** say that $\det(A+B)=\det(A)+\det(B)$; as this is *false*. For example,

$$A=\begin{pmatrix}1&0\\0&0\end{pmatrix} \text{ and } B=\begin{pmatrix}0&0\\0&1\end{pmatrix},$$

then $\det(A)=\det(B)=0$ but $A+B=I$ and hence $\det(A+B)=1$.

The lemma says the following:
Suppose A and B in $M_n(F)$ have all rows same except ith row. Also, all the rows of the new matrix are same as that of A and B except the ith row. If the ith row of the new matrix is the sum of the ith row of A and B, then the determinant of the new matrix is $\det(A)+\det(B)$. Consider the following example:

Example 4.2.9. If

$$A=\begin{pmatrix}1&2\\3&4\end{pmatrix} \text{ and } B=\begin{pmatrix}1&1\\3&4\end{pmatrix},$$

then $\det(A)=-2,\ \det(B)=1$, and $\det\begin{pmatrix}2&3\\3&4\end{pmatrix}=-1=\det(A)+\det(B)$.
Note that the second row of all the three matrices are same while the first row of the new matrix is the sum of the entries in first row of A and B.

Lemma 4.2.10. *If two rows of $A=(\alpha_{ij})\in M_n(F)$ are equal (i.e., $v_r=v_s$ for $r\neq s$), then* $\det(A)=0$.

Proof. Since $r\neq s$, for any permutation $\sigma\in S_n$, we have $\sigma(r)\neq\sigma(s)$ (as a permutation is one-one). Also, since rth row and sth row of A are same,

$$\alpha_{rj}=\alpha_{sj} \ \text{ for all } j=1,2,\dots,n. \tag{4.3}$$

Given $\sigma\in S_n$, consider the transposition $\tau=(\sigma(r),\sigma(s))$, i.e.,

$$
\begin{aligned}
\tau(\sigma(r)) &= \sigma(s),\\
\tau(\sigma(s)) &= \sigma(r),\\
\tau(\sigma(i)) &= \sigma(i) \text{ for all } i\neq r, i\neq s.
\end{aligned}
$$

Then by (4.3) and by above relations, we have

$$\alpha_{r\sigma(r)} = \alpha_{s\sigma(r)} = \alpha_{s\tau\sigma(s)}$$

and

$$\alpha_{s\sigma(s)} = \alpha_{r\sigma(s)} = \alpha_{r\tau\sigma(r)}.$$

Therefore,

$$\alpha_{1\sigma(1)}\alpha_{2\sigma(2)}\cdots\alpha_{r\sigma(r)}\cdots\alpha_{s\sigma(s)}\cdots\alpha_{n\sigma(n)} = \alpha_{1\tau\sigma(1)}\alpha_{2\tau\sigma(2)}\cdots\alpha_{s\tau\sigma(s)}\cdots\alpha_{r\tau\sigma(r)}\cdots\alpha_{n\tau\sigma(n)}.$$

Since τ is a transposition, it is an odd permutation and hence

$$(-1)^{\tau\sigma} = -(-1)^{\sigma}.$$

Therefore, from above we have

$$\begin{aligned} &(-1)^{\sigma}\alpha_{1\sigma(1)}\alpha_{2\sigma(2)}\cdots\alpha_{r\sigma(r)}\cdots\alpha_{s\sigma(s)}\cdots\alpha_{n\sigma(n)} \\ = &-(-1)^{\tau\sigma}\alpha_{1\tau\sigma(1)}\alpha_{2\tau\sigma(2)}\cdots\alpha_{s\tau\sigma(s)}\cdots\alpha_{r\tau\sigma(r)}\cdots\alpha_{n\tau\sigma(n)}. \end{aligned} \tag{4.4}$$

In the expansion

$$\det(A) = \sum_{\sigma\in S_n} (-1)^{\sigma}\alpha_{1\sigma(1)}\alpha_{2\sigma(2)}\cdots\alpha_{n\sigma(n)},$$

we pair the term $(-1)^{\sigma}\alpha_{1\sigma(1)}\cdots\alpha_{n\sigma(n)}$ with the term $(-1)^{\tau\sigma}\alpha_{1\tau\sigma(1)}\cdots\alpha_{n\sigma(n)}$. Then by (4.4) the paired terms cancel each other out in the sum and hence $\det(A) = 0$. □

Lemma 4.2.11. *Interchanging two rows of a matrix changes the sign of its determinant.*

Proof. Since two (rth and sth) rows are equal, by (above) Lemma 4.2.10,

$$d(v_1,\ldots,v_{r-1},v_r+v_s,v_{r+1},\ldots,v_{s-1},v_r+v_s,v_{s+1},\ldots,v_n) = 0.$$

Also, by Lemma 4.2.7, we have

$$\begin{aligned} 0 &= d(v_1,\ldots,v_{r-1},v_r+v_s,v_{r+1},\ldots,v_{s-1},v_r+v_s,v_{s+1},\ldots,v_n) \\ &= d(v_1,\ldots,v_{r-1},v_r,v_{r+1},\ldots,v_{s-1},v_r,v_{s+1},\ldots,v_n) \qquad \text{(two rows are same)} \\ &\quad + d(v_1,\ldots,v_{r-1},v_r,v_{r+1},\ldots,v_{s-1},v_s,v_{s+1},\ldots,v_n) \\ &\quad + d(v_1,\ldots,v_{r-1},v_s,v_{r+1},\ldots,v_{s-1},v_r,v_{s+1},\ldots,v_n) \\ &\quad + d(v_1,\ldots,v_{r-1},v_s,v_{r+1},\ldots,v_{s-1},v_s,v_{s+1},\ldots,v_n) \qquad \text{(two rows are same)} \\ &= d(v_1,\ldots,v_{r-1},v_r,v_{r+1},\ldots,v_{s-1},v_s,v_{s+1},\ldots,v_n) \\ &\quad + d(v_1,\ldots,v_{r-1},v_s,v_{r+1},\ldots,v_{s-1},v_r,v_{s+1},\ldots,v_n). \end{aligned}$$

Therefore,

$$\begin{aligned} &d(v_1,\ldots,v_{r-1},v_r,v_{r+1},\ldots,v_{s-1},v_s,v_{s+1},\ldots,v_n) \\ = &-d(v_1,\ldots,v_{r-1},v_s,v_{r+1},\ldots,v_{s-1},v_r,v_{s+1},\ldots,v_n). \end{aligned}$$

□

Corollary 4.2.12. *If the matrix B is obtained from A by permuting rows of A then*

$$\det(B) = \pm\det(A),$$

where the sign is

$$\begin{array}{rl} +1 & \text{if the permutation is even;} \\ -1 & \text{if the permutation is even.} \end{array}$$

[i.e., for $\sigma \in S_n$, $d(v_1,\ldots,v_n) = (-1)^{\sigma} d(v_{\sigma(1)},\ldots,v_{\sigma(n)})$].

Corollary 4.2.13. *For $\lambda \in F$,*

$$\begin{aligned} & d(v_1,\ldots,v_{r-1},v_r+\lambda v_s,v_{r+1},\ldots,v_{s-1},v_s,v_{s+1},\ldots,v_n) \\ = \quad & d(v_1,\ldots,v_{r-1},v_r,v_{r+1},\ldots,v_{s-1},v_s,v_{s+1},\ldots,v_n) \end{aligned}$$

Proof. By previous lemmas,

$$\begin{aligned} & d(v_1,\ldots,v_{r-1},v_r+\lambda v_s,v_{r+1},\ldots,v_{s-1},v_s,v_{s+1},\ldots,v_n) \\ &= d(v_1,\ldots,v_{r-1},v_r,v_{r+1},\ldots,v_{s-1},v_s,v_{s+1},\ldots,v_n) \\ &\qquad + d(v_1,\ldots,v_{r-1},\lambda v_s,v_{r+1},\ldots,v_{s-1},v_s,v_{s+1},\ldots,v_n) \\ &= d(v_1,\ldots,v_{r-1},v_r,v_{r+1},\ldots,v_{s-1},v_s,v_{s+1},\ldots,v_n) \\ &\qquad + \lambda\, d(v_1,\ldots,v_{r-1},v_s,v_{r+1},\ldots,v_{s-1},v_s,v_{s+1},\ldots,v_n) \quad \text{(two rows are same)} \\ &= d(v_1,\ldots,v_{r-1},v_r,v_{r+1},\ldots,v_{s-1},v_s,v_{s+1},\ldots,v_n). \end{aligned}$$

□

Remark 4.2.14. Let $A \in M_n(F)$ and let A' or A^{T} denote its transpose. By Lemma 4.2.4, we know that,

$$\det(A) = \det(A^{\mathrm{T}}).$$

Then all the properties of determinants for rows discussed above also hold for columns.

Theorem 4.2.15. *For $A, B \in M_n(F)$,*

$$\boxed{\det(AB) = \det(A)\det(B)}.$$

Proof. Let $A = (\alpha_{ij})$ and $B = (\beta_{ij})$. For $i = 1,2,\ldots,n$, let

$$\begin{aligned} v_i &= (\alpha_{i1},\alpha_{i2},\ldots,\alpha_{in}) \ -\text{be the } i\text{th row of } A \text{ and} \\ u_i &= (\beta_{i1},\beta_{i2},\ldots,\beta_{in}) \ -\text{be the } i\text{th row of } B. \end{aligned}$$

For $i = 1,2,\ldots,n$, let

$$w_i = \alpha_{i1}u_1 + \alpha_{i2}u_2 + \cdots + \alpha_{in}u_n.$$

Then

$$w_i = \alpha_{i1}(\beta_{11},\beta_{12},\ldots,\beta_{1n}) + \alpha_{i2}(\beta_{21},\beta_{22},\ldots,\beta_{2n}) + \cdots + \alpha_{in}(\beta_{n1},\beta_{n2},\ldots,\beta_{nn})$$

$$= (\alpha_{i1}\beta_{11} + \alpha_{i2}\beta_{21} + \cdots + \alpha_{in}\beta_{n1},$$
$$\alpha_{i1}\beta_{12} + \alpha_{i2}\beta_{22} + \cdots + \alpha_{in}\beta_{n2}, \ldots, \alpha_{i1}\beta_{1n} + \alpha_{i2}\beta_{2n} + \cdots + \alpha_{in}\beta_{nn}).$$

Thus, w_i is the ith row of the matrix AB.

$$\begin{aligned}
\therefore \det(AB) &= d(w_1, w_2, \ldots, w_n) \\
&= d(\alpha_{11}u_1 + \alpha_{12}u_2 + \cdots + \alpha_{1n}u_n, \\
&\qquad \alpha_{21}u_1 + \alpha_{22}u_2 + \cdots + \alpha_{2n}u_n, \ldots, \\
&\qquad \alpha_{n1}u_1 + \alpha_{n2}u_2 + \cdots + \alpha_{nn}u_n) \\
&= \sum_{i_1, i_2, \ldots, i_n = 1}^{n} \alpha_{1i_1}\alpha_{2i_2}\cdots\alpha_{ni_n} d(u_{i_1}, u_{i_2}, \ldots, u_{i_n}),
\end{aligned} \tag{4.5}$$

where $i_1, i_2, \ldots, i_n$ runs independently from 1 to n in the above multiple sum.

If $i_r = i_s$, then $u_r = u_s$ and hence $d(u_1, \ldots, u_r, \ldots, u_s, \ldots, u_n) = 0$. Thus, in the above sum only those terms will give non-zero contribution for which $i_1, i_2, \ldots, i_n$ are distinct. Take

$$\sigma = \begin{pmatrix} 1 & 2 & \cdots & n \\ i_1 & i_2 & \cdots & i_n \end{pmatrix} \in S_n \qquad (\because i_1, i_2, \ldots, i_n \text{ are distinct}).$$

Then,

$$\begin{aligned}
d(u_{i_1}, u_{i_2}, \ldots, u_{i_n}) &= d(u_{\sigma(1)}, u_{\sigma(2)}, \ldots, u_{\sigma(n)}) \\
&= (-1)^{\sigma} d(u_1, u_2, \ldots, u_n).
\end{aligned}$$

Therefore, from equation (4.5), we get

$$\begin{aligned}
\det(AB) &= \sum_{\sigma \in S_n} (-1)^{\sigma} \alpha_{1\sigma(1)}\alpha_{2\sigma(2)}\cdots\alpha_{n\sigma(n)} d(u_1, u_2, \ldots, u_n) \\
&= d(u_1, u_2, \ldots, u_n) \sum_{\sigma \in S_n} (-1)^{\sigma} \alpha_{1\sigma(1)}\alpha_{2\sigma(2)}\cdots\alpha_{n\sigma(n)} \\
&= \det(B)\det(A) \\
&= \det(A)\det(B).
\end{aligned}$$

□

Corollary 4.2.16. *If $A \in M_n(F)$ is regular then* $\det(A) \neq 0$ *and* $\det(A^{-1}) = \frac{1}{\det(A)}$.

Proof. If A is regular then there exists $A^{-1} \in M_n(F)$ such that $AA^{-1} = I$. Then

$$\det(AA^{-1}) = \det(I) = 1.$$

Therefore, by above theorem

$$\det(A)\det(A^{-1}) = 1.$$

Hence, $\det(A) \neq 0$ and $\det(A^{-1}) = \frac{1}{\det(A)}$. □

Corollary 4.2.17. *Determinants of similar matrices are same, i.e., if If $A, B \in M_n(F)$ are similar, then* $\det(A) = \det(B)$.

Proof. Since A and B are similar, there is a regular $C \in M_n(F)$ such that $A = C^{-1}BC$. Then,

$$\begin{aligned} \det(A) &= \det(C^{-1}BC) \\ &= \det(C^{-1})\det(B)\det(C) && \text{(by above theorem)} \\ &= \frac{1}{\det(C)}\det(B)\det(C) && \text{(by above corollary)} \\ &= \det(B). \end{aligned}$$

□

The above corollary allows us to define the determinant of a linear transformation.

Definition 4.2.18. Let V be a finite dimensional vector space and $T \in A(V)$. The determinant of T, denoted by $\det(T)$, is the determinant of the matrix of T, i.e., $\det(m(T))$.

Remark 4.2.19. If $m_1(T)$ and $m_2(T)$ are matrices of T in two different basis of V, then $m_2(T) = C^{-1}m_1(T)C$ and hence, by Corollary 4.2.17, $\det(m_1(T)) = \det(m_2(T))$. Thus, the definition of $\det(T)$ is independent of the choice of basis of V.

4.2.1 Cramer's Rule

Theorem 4.2.20 (Cramer's Rule)**.** *Consider the system of n linear equations:*

$$\begin{aligned} \alpha_{11}x_1 + \alpha_{12}x_2 + \cdots + \alpha_{1n}x_n &= \beta_1, \\ \alpha_{21}x_1 + \alpha_{22}x_2 + \cdots + \alpha_{2n}x_n &= \beta_2, \\ \vdots \qquad\qquad &\quad \vdots \\ \alpha_{n1}x_1 + \alpha_{n2}x_2 + \cdots + \alpha_{nn}x_n &= \beta_n, \end{aligned}$$

where $\beta_1, \beta_2, \ldots, \beta_n \in F$. Let $A = (\alpha_{ij}) \in M_n(F)$ be the matrix of the system *and let $\Delta = \det(A)$ be the* determinant of the system. *If $\Delta \neq 0$, then the above system has a unique solution*

$$x_i = \frac{\Delta_i}{\Delta},$$

where Δ_i is the determinant obtained by replacing ith column in Δ by $\begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{pmatrix}$.

Proof. Let $x_1, x_2, \ldots, x_n$ be a solution of the above system. Then for $1 \le i \le n$,

$$x_i\Delta = x_i \begin{vmatrix} \alpha_{11} & \alpha_{12} & \ldots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \ldots & \alpha_{2n} \\ \vdots & \vdots & & \vdots \\ \alpha_{n1} & \alpha_{n2} & \ldots & \alpha_{nn} \end{vmatrix} = \begin{vmatrix} \alpha_{11} & \ldots & \alpha_{1i}x_i & \ldots & \alpha_{1n} \\ \alpha_{21} & \ldots & \alpha_{2i}x_i & \ldots & \alpha_{2n} \\ \vdots & & \vdots & & \vdots \\ \alpha_{n1} & \ldots & \alpha_{ni}x_i & \ldots & \alpha_{nn} \end{vmatrix}$$

We know that (by Lemma 4.2.6 and Lemma 4.2.7), we can add any multiple of a column (or a row) to another without changing the value of the determinant. Therefore, we can write

$$\begin{aligned} x_i\Delta &= \begin{vmatrix} \alpha_{11} & \ldots & \alpha_{1,i-1} & (\alpha_{11}x_1 + \alpha_{12}x_2 + \cdots + \alpha_{1n}x_n) & \alpha_{1,i+1} & \ldots & \alpha_{1n} \\ \alpha_{21} & \ldots & \alpha_{1,i-1} & (\alpha_{21}x_1 + \alpha_{22}x_2 + \cdots + \alpha_{2n}x_n) & \alpha_{1,i+1} & \ldots & \alpha_{2n} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ \alpha_{n1} & \ldots & \alpha_{1,i-1} & (\alpha_{n1}x_1 + \alpha_{n2}x_2 + \cdots + \alpha_{nn}x_n) & \alpha_{1,i+1} & \ldots & \alpha_{nn} \end{vmatrix} \\ &= \begin{vmatrix} \alpha_{11} & \ldots & \alpha_{1,i-1} & \beta_1 & \alpha_{1,i+1} & \ldots & \alpha_{1n} \\ \alpha_{21} & \ldots & \alpha_{2,i-1} & \beta_2 & \alpha_{2,i+1} & \ldots & \alpha_{2n} \\ \vdots & & \vdots & & \vdots & & \\ \alpha_{n1} & \ldots & \alpha_{n,i-1} & \beta_n & \alpha_{n,i+1} & \ldots & \alpha_{nn} \end{vmatrix} \\ &= \Delta_i. \end{aligned}$$

Since $\Delta \neq 0$, we therefore have,

$$x_i = \frac{\Delta_i}{\Delta}, \quad i = 1, 2, \ldots, n.$$

□

Corollary 4.2.21. *If* $\det(A) \neq 0$ *then A is regular.*

Proof. Let $A = (\alpha_{ij}) \in M_n(F)$. Define $T : F^{(n)} \to F^{(n)}$ by

$$\begin{aligned} T(x_1, x_2, \ldots, x_n) = (&\alpha_{11}x_1 + \alpha_{12}x_2 + \cdots + \alpha_{1n}x_n, \\ &\alpha_{21}x_1 + \alpha_{22}x_2 + \cdots + \alpha_{2n}x_n, \ldots, \alpha_{n1}x_1 + \alpha_{n2}x_2 + \cdots + \alpha_{nn}x_n). \end{aligned} \tag{4.6}$$

Then T is a homomorphism and clearly the matrix of T in the standard basis of $F^{(n)}$ is A, i.e., $m(T) = A$.

Let $(\beta_1, \beta_2, \ldots, \beta_n) \in F^{(n)}$ be arbitrary. Now, consider the system of linear equations

$$\begin{aligned} \alpha_{11}x_1 + \alpha_{12}x_2 + \cdots + \alpha_{1n}x_n &= \beta_1, \\ \vdots \qquad\qquad & \quad \vdots \end{aligned}$$

$$\alpha_{n1}x_1 + \alpha_{n2}x_2 + \cdots + \alpha_{nn}x_n = \beta_n$$

Since $\det(A) \neq 0$, by Cramer's rule, the above system has a unique solution. That is, there are $x_1, x_2, \ldots, x_n$ satisfying above system of linear equations. But from the definition of T in (4.6), this means

$$T(x_1, x_2, \ldots, x_n) = (\beta_1, \beta_2, \ldots, \beta_n).$$

Therefore, T is onto and so T is regular. Hence, the matrix of T, $m(T) = A$ is regular (invertible). □

Combining Corolllary 4.2.16 and Corollary 4.2.21, we can state the following result:

Theorem 4.2.22. *$A \in M_n(F)$ is invertible if and only if* $\det(A) \neq 0$.

Proof. Proof of Corolllary 4.2.16 and Corollary 4.2.21. □

Proposition 4.2.23. *Let $A \in M_n(F)$. Then the determinant of A is the product of the characteristic roots of A counted according to their multiplicities.*

Proof. Same as proof of Lemma 4.1.7. □

4.3 Quadratic forms

We are grateful to Prof. P. A. Dabhi and Prof. A. B. Patel for providing us notes for this section.

Definition 4.3.1. Let V be a vector space over $\mathbb{R}$. A map $f : V \times V \to \mathbb{R}$ is called a *bilinear map* if for every $u, v, w \in V$ and $\alpha \in \mathbb{R}$ the following hold.

1. $f(u, v) = f(v, u)$.
2. $f(\alpha u, v) = \alpha f(u, v)$.
3. $f(u + v, w) = f(u, w) + f(v, w)$.

Remark 4.3.2. Observe the linearity in the second variable too:

$$\begin{aligned} f(u, \alpha v + \beta w) &= f(\alpha v + \beta w, u) && \text{(by property 1 above)} \\ &= \alpha f(v, u) + \beta f(w, u) && \text{(by property 2 and 3 above)} \\ &= \alpha f(u, v) + \beta f(u, v) && \text{(by property 1 above).} \end{aligned}$$

Theorem 4.3.3. *Let $f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ be a map. Then f is bilinear if and only if there exist*

$\alpha_{ij} \in \mathbb{R}$, $1 \le i, j \le n$ with $\alpha_{ij} = \alpha_{ji}$ such that

$$f(x,y) = \sum_{i,j=1}^{n} \alpha_{ij} x_i y_j.$$

Proof. Suppose that there exist $\alpha_{ij} \in \mathbb{R}$, $1 \le i, j \le n$ with $\alpha_{ij} = \alpha_{ji}$ such that

$$f(x,y) = \sum_{i,j=1}^{n} \alpha_{ij} x_i y_j.$$

Then clearly, f is bilinear (Verify!).

Conversely, assume that f is a bilinear map. Let $\{e_1, e_2, \ldots, e_n\}$ be the standard basis of $\mathbb{R}^n$. Let $\alpha_{ij} = f(e_i, e_j)$. Since f is symmetric, we have

$$\alpha_{ij} = f(e_i, e_j) = f(e_j, e_i) = \alpha_{ji}.$$

Now, let $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_n)$ be in $\mathbb{R}^n$. Then

$$\begin{aligned} f(x,y) &= f\left(\sum_{i=1}^{n} x_i e_i, \sum_{j=1}^{n} y_j e_j\right) \\ &= \sum_{i=1}^{n} x_i f\left(e_i, \sum_{j=1}^{n} y_j e_j\right) \\ &= \sum_{i=1}^{n} x_i \left(\sum_{j=1}^{n} y_j f(e_i, e_j)\right) \\ &= \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_{ij} x_i y_j = \sum_{i,j=1}^{n} \alpha_{ij} x_i y_j. \end{aligned}$$

□

Definition 4.3.4. Let V be a vector space over $\mathbb{R}$ and $f : V \times V \to \mathbb{R}$ be a bilinear map. Then a map $g : V \to \mathbb{R}$ defined by $g(v) = f(v, v)$ for every $v \in V$ is called a *quadratic form.*

Corollary 4.3.5. *Let $g : \mathbb{R}^n \to \mathbb{R}$ be a map. Then g is a quadratic form if and only if there are scalars $\alpha_{ij} \in \mathbb{R}$, $1 \le i, j \le n$ with $\alpha_{ij} = \alpha_{ji}$ such that*

$$g(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_{ij} x_i x_j.$$

As a consequence of Corollary 4.3.5 we have the following result:

Corollary 4.3.6. *A map $g : \mathbb{R}^n \to \mathbb{R}$ is a quadratic form if and only if there is a unique $n \times n$ real symmetric matrix $A = (\alpha_{ij})$ such that $g(x) = x^T A x$ for every $x \in \mathbb{R}^n$.* (We consider elements of $\mathbb{R}^n$ as column matrices).

Example 4.3.7. Find the symmetric matrices associated with the following quadratic forms:

1. $9x_1^2 - x_2^2 + 4x_3^2 + 6x_1x_2 - 8x_1x_3 + 2x_2x_3$.
2. $xy + xz + yz$.
3. $x_1^2 + x_2^2 - x_3^2 - x_4^2 + 2x_1x_2 - 10x_1x_4 + 4x_3x_4$.
4. $-y^2 - 2z^2 + 4xy + 8xz - 14yz$.

Solution. The matrices associated with the above quadratic forms are:

1. $\begin{pmatrix} 9 & 3 & -4 \\ 3 & -1 & 1 \\ -4 & 1 & 4 \end{pmatrix}$.
2. $\begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}$.
3. $\begin{pmatrix} 1 & 1 & 0 & -5 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 2 \\ -5 & 0 & 2 & -1 \end{pmatrix}$.
4. $\begin{pmatrix} 0 & 2 & 4 \\ 2 & -1 & -7 \\ 4 & -7 & -2 \end{pmatrix}$.

□

Definition 4.3.8. A *quadratic equation* in $\mathbb{R}^n$ is an equation in n variables $x_1, x_2, \ldots, x_n$ of the form

$$f(x_1, x_2, \ldots, x_n) = \sum_{i,j=1}^{n} \alpha_{ij} x_i x_j + \sum_{i=1}^{n} b_i x_i + c = 0,$$

where (α_{ij}) is a real symmetric matrix and $b_i, c \in \mathbb{R}$.

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a quadratic equation. Let $A = (\alpha_{ij})$ and $B = (b_1, b_2, \ldots, b_n) \in \mathbb{R}^n$ and $c \in \mathbb{R}$. Then the above quadratic equation can be written in the form

$$f(x) = x^{\mathrm{T}} A x + B x + c = 0.$$

Definition 4.3.9. A quadratic equation $f(x) = 0$ in $\mathbb{R}^n$ is said to be *consistent* if it has a solution, i.e., there exists $x_0 \in \mathbb{R}^n$ such that $f(x_0) = 0$. If a quadratic equation is not consistent (i.e., it does not have a solution), then it is called *inconsistent*.

Definition 4.3.10. The solution set of a consistent quadratic equation $f(x) = x^{\mathrm{T}}Ax + Bx + c = 0$ over $\mathbb{R}^n$ is called a *level surface*.

In particular, if $n = 2$, then the level surfaces are called *quadratic curves*. When $n = 3$, the level surfaces are called *quadratic surfaces*.

Now we state some definitions and results which are useful to us in identifying the level surfaces in $\mathbb{R}^2$ and $\mathbb{R}^3$.

Definition 4.3.11. Let F be a field. A matrix $A \in M_n(F)$ is said to be *diagonalizable* if there exists a matrix $C \in M_n(F)$ such that $C^{-1}AC$ is a diagonal matrix.

Remark 4.3.12. When a matrix $A \in M_n(F)$ is diagonalizable, the entries on the main diagonal of the matrix $C^{-1}AC$ are precisely the characteristic roots of A.

Theorem 4.3.13. *Let F be a field and $n \in \mathbb{N}$. A matrix $A \in M_n(F)$ is diagonalizable if and only if A has n linearly independent characteristic vectors.*

Definition 4.3.14. A matrix $A \in M_n(F)$ is said to be *orthogonal* if $AA^{\mathrm{T}} = I$.

Theorem 4.3.15. *Let $A \in M_n(\mathbb{R})$ be a symmetric matrix. Then all the characteristic roots of A are real and A is diagonalizable. In fact, there exists an orthogonal matrix P such that $P^{-1}AP$ is a diagonal matrix.*

Definition 4.3.16. Let F be a field and let $A \in M_n(F)$ $(n \in \mathbb{N})$. Then the equation $\det(A - \lambda I) = 0$ is called the *characteristic equation* (or *secular equation*) of A.

Theorem 4.3.17. *Let F be a field and let $A \in M_n(F)$. Then*

1. (Caley - Hamilton Theorem)*: The matrix A satisfies its characteristic equation.*
2. *The roots of the characteristic equation of A are the characteristic roots of A.*

Theorem 4.3.18 (Principal Axes Theorem)**.** *Let $x^{\mathrm{T}}Ax$ be the quadratic form in n variables. Then there is a change of coordinates of x into $y = P^{\mathrm{T}}x$ such that*

$$x^{\mathrm{T}}Ax = y^{\mathrm{T}}Dy = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2,$$

where P is an orthogonal matrix such that $P^{\mathrm{T}}AP = D$ is a diagonal matrix with diagonal entries $\lambda_1, \lambda_2, \ldots, \lambda_n$.
[The axes are $v_i = P^{\mathrm{T}}e_i$ $(1 \le i \le n)$.]

Definition 4.3.19. Let $A \in M_n(\mathbb{R})$ be symmetric. Then A (or the quadratic form $x^{\mathrm{T}}Ax$) is called

1. *positive definite* if $x^{\mathrm{T}}Ax > 0$ for every $0 \neq x \in \mathbb{R}^n$.
2. *positive semidefinite* if $x^{\mathrm{T}}Ax \ge 0$ for every $x \in \mathbb{R}^n$.
3. *negative definite* if $x^{\mathrm{T}}Ax < 0$ for every $0 \neq x \in \mathbb{R}^n$.
4. *negative semidefinite* if $x^{\mathrm{T}}Ax \le 0$ for every $x \in \mathbb{R}^n$.
5. *indefinite* is $x^{\mathrm{T}}Ax$ takes both negative and positive values.

Theorem 4.3.20. *Let $A \in M_n(\mathbb{R})$ be a symmetric matrix. Then*

1. *A is positive definite if and only if all the characteristic roots of A are positive.*
2. *A is positive semidefinite if and only if all the characteristic roots of A are non-negative.*
3. *A is negative definite if and only if all the characteristic roots of A are negative.*
4. *A is negative semidefinite if and only if all the characteristic roots of A are non-*

positive.

5. *A is indefinite if and only if A has positive as well as negative characteristic roots.*

Definition 4.3.21. The *inertia* of a real symmetric matrix $A \in M_n(\mathbb{R})$ is a triple of integers denoted by $In(A) = (p,q,k)$, where p,q and k are the number of positive, negative and zero characteristic roots of A respectively.

The inertia $In(A)$, for symmetric $A \in M_n(\mathbb{R})$, determines the geometric type of the quadratic surface $x^{\mathrm{T}}Ax = c$ in the following sense. Since $In(-A) = (q,p,k)$ if $In(A) = (p,q,k)$, it suffices to consider the cases $c \geq 0$ and $p > 0$. Excluding those inconsistent cases, we have the following characterization of the solution sets for $n = 2$ and $n = 3$.

Table 4.1: Level surfaces for $n = 2$

$In(A) = (p,q,k)$	$c > 0$	$c = 0$
$(2,0,0)$	Ellipse	A point
$(1,1,0)$	Hyperbola	Pair of lines
$(1,0,1)$	Two parallel lines	A line

Table 4.2: Level surfaces for $n = 3$

$In(A) = (p,q,k)$	$c > 0$	$c = 0$
$(3,0,0)$	Ellipsoid	A point
$(2,1,0)$	Hyperboloid with one sheet	A cone
$(2,0,1)$	Elliptical cylinder	A line
$(1,1,1)$	Hyperbolic cylinder	Pair of intersecting planes
$(1,2,0)$	Hyperboloid with two sheets	A cone
$(1,0,2)$	Parabolic cylinder	A plane

4.3.1 Some standard conics and quadratic surfaces

We recall below some standard quadratic curves and quadratic surfaces along with their equations.

- **Circle**:

$$x^2 + y^2 = 1.$$

$$\{(x,y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}.$$

- **Cylinder**:

$$\{(x,y,z) \in \mathbb{R}^3 : x^2 + y^2 = 1\}.$$

• **Hyperbolic cylinder**:

$$\{(x,y,z) \in \mathbb{R}^3 : x^2 - y^2 = 1\}.$$

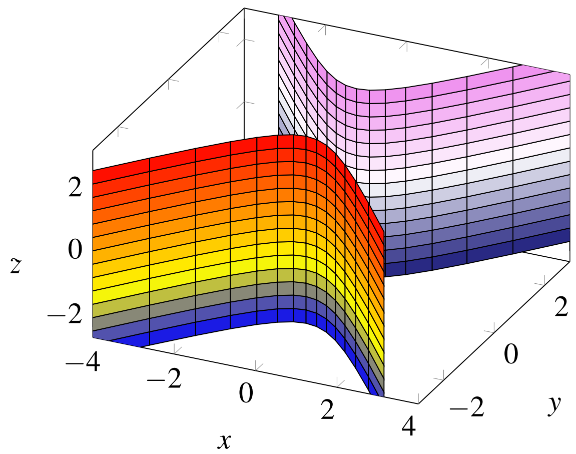


Figure 4.1: Hyperbolic cylinder

• **Elliptical cylinder**:

$$\left\{(x,y,z) \in \mathbb{R}^3 : \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1\right\}.$$

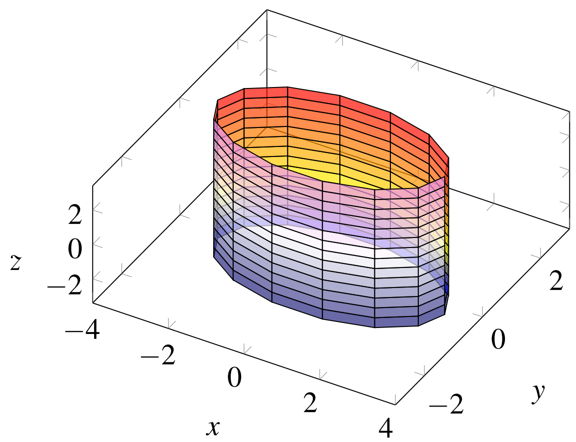


Figure 4.2: Elliptic cylinder

• **Hyperboloid with 1 sheet**:

$$x^2 + y^2 - z^2 = 1.$$

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1.$$

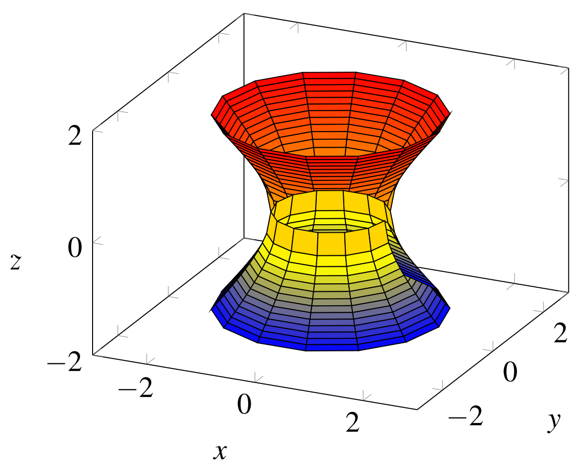


Figure 4.3: Hyperboloid with 1 sheet

• **Hyperboloid with 2 sheets**:

$$x^2 - y^2 - z^2 = 1.$$

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1.$$

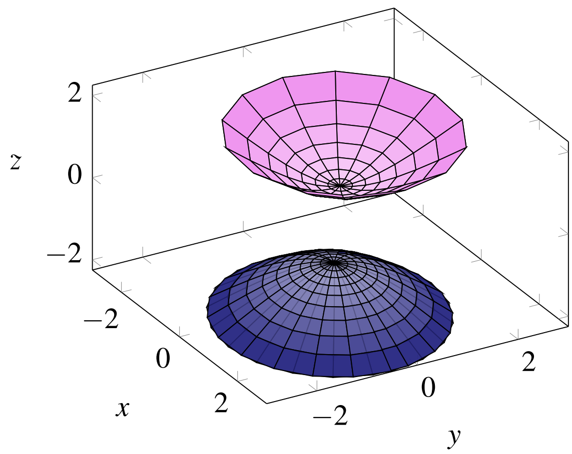


Figure 4.4: Hyperboloid with 2 sheets

• **Ellipsoid**:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1.$$

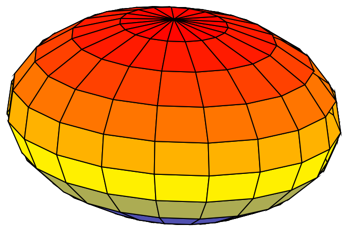


Figure 4.5: Ellipsoid

• **Cone**:

$$x^2 + y^2 = z^2.$$

$$ax^2 + by^2 = z^2 \quad (a \text{ and } b \text{ are positive}).$$

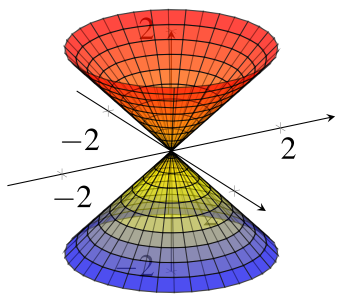


Figure 4.6: Cone

• **Paraboloid**:

$$z = x^2 + y^2.$$

$$z = ax^2 + by^2 \quad (a, b > 0).$$

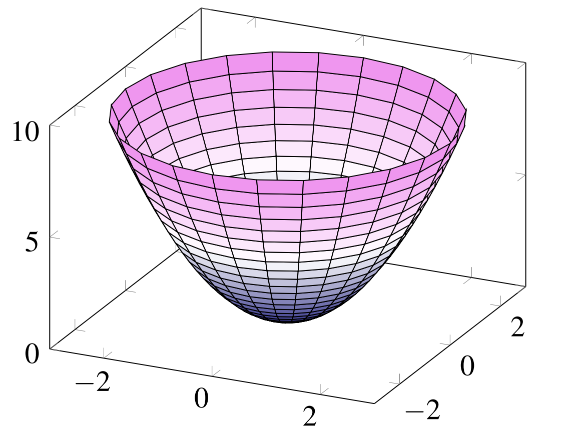


Figure 4.7: Paraboloid

• **Hyperbolic Paraboloid**:

$$z = x^2 - y^2.$$

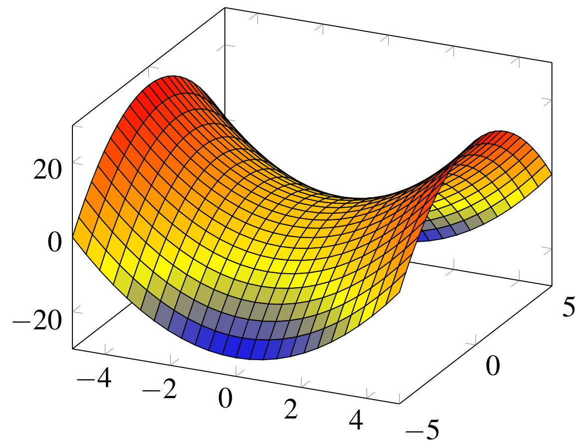


Figure 4.8: Hyperbolic Paraboloid

• **Parabolic cylinder**:

$$\{(x, y, z) \in \mathbb{R}^3 : y = x^2\}.$$

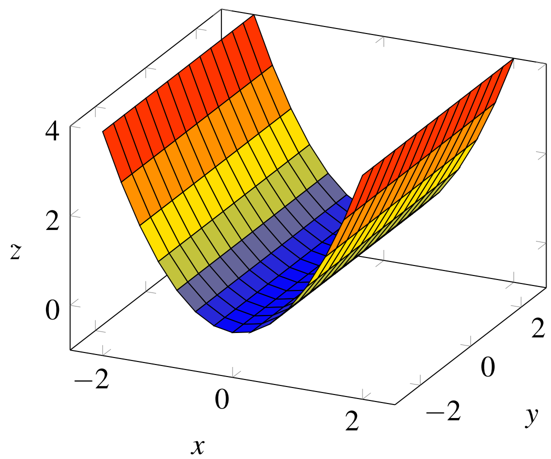


Figure 4.9: Parabolic cylinder

• **Hyperbola**:

$$x^2 - y^2 = 1.$$
$$\{(x,y) \in \mathbb{R}^2 : x^2 - y^2 = 1\}.$$
$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1.$$

• **Ellipse**:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1.$$

• **Parabola**:

$$y^2 = 4ax.$$

• **Pair of intersecting planes**:

$$x^2 - y^2 = 0 \quad \text{in } \mathbb{R}^3.$$

• **Pair of intersecting lines**:

$$ax^2 - by^2 = 0 \quad (a,b > 0) \text{ in } \mathbb{R}^2.$$

The intersecting lines are

$$\sqrt{a}x - \sqrt{b}y = 0 \text{ and } \sqrt{a}x + \sqrt{b}y = 0.$$

• **Pair of parallel lines**:

$$x^2 = c \quad (c > 0).$$

The parallel lines are

$$x = \pm\sqrt{c}.$$

Example 4.3.22. Identify the surface given by $11x^2 + 6xy + 19y^2 = 80$. Also convert it to the standard form by finding the orthogonal matrix P.

Solution. The symmetric matrix associated with above quadratic form is

$$A = \begin{pmatrix} 11 & 3 \\ 3 & 19 \end{pmatrix}.$$

We find the characteristic roots of A. We know that characteristic roots of A are the solutions of $\det(A - \lambda I) = 0$. Now,

$$\begin{aligned}
& \det(A - \lambda I) = 0 \\
\Rightarrow\ & \det\left(\begin{pmatrix} 11 & 3 \\ 3 & 19 \end{pmatrix} - \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}\right) = 0 \\
\Rightarrow\ & \begin{vmatrix} 11-\lambda & 3 \\ 3 & 19-\lambda \end{vmatrix} = 0 \\
\Rightarrow\ & (11-\lambda)(19-\lambda) - 9 = 0 \\
\Rightarrow\ & \lambda^2 - 30\lambda + 200 = 0 \\
\Rightarrow\ & (\lambda - 10)(\lambda - 20) = 0
\end{aligned}$$

Therefore 10 and 20 are the characteristic roots of A. Let $\begin{pmatrix} x \\ y \end{pmatrix}$ be a characteristic vector of A corresponding to the characteristic root 10. Then

$$\begin{pmatrix} 11 & 3 \\ 3 & 19 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = 10\begin{pmatrix} x \\ y \end{pmatrix}.$$

$$\therefore 11x + 3y = 10x \text{ and } 3x + 19y = 10y.$$

i.e., $x + 3y = 0$. Taking $y = 1$, we have $x = -3$. Therefore, $\frac{1}{\sqrt{10}}\begin{pmatrix} -3 \\ 1 \end{pmatrix} = \begin{pmatrix} -\frac{3}{\sqrt{10}} \\ \frac{1}{\sqrt{10}} \end{pmatrix}$ is an characteristic vector of A corresponding to the characteristic root 10.

Similarly for $\lambda = 20$,

$$\begin{pmatrix} 11 & 3 \\ 3 & 19 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = 20\begin{pmatrix} x \\ y \end{pmatrix}.$$

$$\therefore 11x + 3y = 20x \text{ and } 3x + 19y = 20y.$$

i.e., $3x - y = 0$. Taking $x = 1$, we have $y = 3$. Hence, $\begin{pmatrix} \frac{1}{\sqrt{10}} \\ \frac{3}{\sqrt{10}} \end{pmatrix}$ is an characteristic vector of A corresponding to the characteristic root 20. Therefore, the orthogonal matrix P (i.e., $PP^{\mathrm{T}} = I$) is

$$P = \begin{pmatrix} -\frac{3}{\sqrt{10}} & \frac{1}{\sqrt{10}} \\ \frac{1}{\sqrt{10}} & \frac{3}{\sqrt{10}} \end{pmatrix}.$$

Now, let

$$\begin{pmatrix} x \\ y \end{pmatrix} = P\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} -\frac{3}{\sqrt{10}} & \frac{1}{\sqrt{10}} \\ \frac{1}{\sqrt{10}} & \frac{3}{\sqrt{10}} \end{pmatrix}\begin{pmatrix} x' \\ y' \end{pmatrix}, i.e.,$$

$$x = -\frac{3}{\sqrt{10}}x' + \frac{1}{\sqrt{10}}y' \text{ and } y = \frac{1}{\sqrt{10}}x' - \frac{3}{\sqrt{10}}y'.$$

Substituting these values in the given equation $11x^2 + 6xy + 19y^2 = 80$, we get $10x'^2 + 20y'^2 = 80$ or $x'^2 + 2y'^2 = 8$. Therefore, the standard form of the given quadratic equation is $\boxed{x'^2 + 2y'^2 = 8}$ and it is an *ellipse*.

• **Ellipse**: $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$.
Here, in this example, we have

$$\frac{x'^2}{(2\sqrt{2})^2} + \frac{y'^2}{2^2} = 1.$$

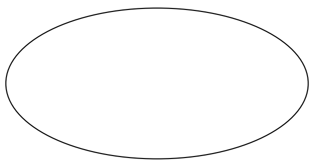


Figure 4.10: An ellipse

□

Example 4.3.23. Reduce the quadratic form $x_1^2 - x_3^2 - 4x_1x_2 + 4x_2x_3$ into standard form by finding an orthogonal matrix P. Hence determine the surface given by $x_1^2 - x_3^2 - 4x_1x_2 + 4x_2x_3 = 1$.

Solution. The matrix associated with the above quadratic form is $A = \begin{pmatrix} 1 & -2 & 0 \\ -2 & 0 & 2 \\ 0 & 2 & -1 \end{pmatrix}$.
Check that the roots of $\det(A - \lambda I) = 0$ are 3, -3 and 0 and characteristic vectors corresponding to these characteristic roots are $\begin{pmatrix} \frac{-2}{3} \\ \frac{2}{3} \\ \frac{1}{3} \end{pmatrix}$, $\begin{pmatrix} \frac{-1}{3} \\ \frac{-2}{3} \\ \frac{2}{3} \end{pmatrix}$ and $\begin{pmatrix} \frac{2}{3} \\ \frac{1}{3} \\ \frac{2}{3} \end{pmatrix}$ respectively.
Hence, the required orthogonal matrix is

$$P = \begin{pmatrix} \frac{-2}{3} & \frac{-1}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{-2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} & \frac{2}{3} \end{pmatrix}.$$

Let $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = P \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$, i.e., $x_1 = \frac{-2}{3}y_1 - \frac{1}{3}y_2 + \frac{2}{3}y_3$, $x_2 = \frac{2}{3}y_1 - \frac{2}{3}y_2 + \frac{1}{3}y_3$ and $x_3 = \frac{1}{3}y_1 + \frac{2}{3}y_2 + \frac{2}{3}y_3$. Substituting these values in the quadratic form we get $3y_1^2 - 3y_2^2$. Therefore, the standard form for the quadratic equation is $\boxed{3y_1^2 - 3y_2^2 = 1}$ and it is a *hyperbolic cylinder*. □

Example 4.3.24. Describe the conic C whose equation is $5x^2 - 4xy + 8y^2 + 4\sqrt{5}x - 16\sqrt{5}y + 4 = 0$.

Solution. The matrix associated with the above quadratic equation is $A = \begin{pmatrix} 5 & -2 \\ -2 & 8 \end{pmatrix}$. Check that 4 and 9 are the characteristic roots of A and the corresponding characteristic vectors are $\begin{pmatrix} \frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{pmatrix}$ and $\begin{pmatrix} \frac{-1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{pmatrix}$. Hence, the orthogonal matrix

$$P = \begin{pmatrix} \frac{2}{\sqrt{5}} & \frac{-1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{pmatrix}.$$

Let $\begin{pmatrix} x' \\ y' \end{pmatrix} = P \begin{pmatrix} x \\ y \end{pmatrix}$, i.e., $x' = \frac{2}{\sqrt{5}}x - \frac{1}{\sqrt{5}}y$ and $y' = \frac{1}{\sqrt{5}}x + \frac{2}{\sqrt{5}}y$. Substituting it in the given quadratic form we get $4x'^2 + 9y'^2 + 4\sqrt{5}(\frac{2}{\sqrt{5}}x' - \frac{1}{\sqrt{5}}y') - 16\sqrt{5}(\frac{1}{\sqrt{5}}x' + \frac{2}{\sqrt{5}}y') + 4 = 0$, i.e., $4x'^2 + 9y'^2 - 8x' - 36y' + 4 = 0$. Therefore, $\boxed{4(x'-1)^2 + 9(y'-2)^2 = 36}$, which is again an *ellipse*. □