LINEAR ALGEBRA

(MTH1 C02)

I SEMESTER

2019 Admission)

M Sc MATHEMATICS

UNIVERSITY OF CALICUT
School of Distance Education
Calicut University P.O., Malappuram, Kerala, India - 673 635

190552

UNIVERSITY OF CALICUT

SCHOOL OF DISTANCE EDUCATION

Study Material

I SEMESTER

(2019 Admission)

M Sc MATHEMATICS

Core Course (MTH1 C02)

LINEAR ALGEBRA

Prepared by:

Dr. Kavitha. T
Assistant Professor,
Department of Mathematics,
Zamorin's Guruvayurappan College,
Kozhikode.

Scrutinized by:
Dr. Sini. P,
Assistant Professor,
Department of Mathematics,
University of Calicut.

Contents

Chapter 1

VECTOR SPACES OVER A FIELD

It is both meaningful and interesting to deal with linear combinations of objects in a set X. In the study of linear equations we consider linear combinations of the rows of a matrix.

1.1 Fields

We have studied algebraic properties of real numbers. The operation called addition associates two real numbers $x, y \in R$ to their sum $x + y \in R$. Also there is another operation called multiplication associates with each pair $x, y \in R$, an element $xy \in R$. Also we know how to add and multiply two complex numbers. Now let F denote either the set of real numbers or the set of complex numbers. Then addition and multiplication has the following properties.

1. Addition is commutative, $x + y = y + x$ for all $x, y \in F$.

2. Addition is associative, $x + (y + z) = (x + y) + z$ for all $x, y, z \in F$.

3. There is a unique element 0(zero)$\in F$ such that $x + 0 = x$, for every $x \in F$.

4. To each $x \in F$ there corresponds a unique element $-x$ in F such that $x + (-x) = 0$.

5. Multiplication is commutative, $xy = yx$ for all $x, y \in F$.

6. Multiplication is associative, $x(yz) = (xy)z$ for all $x, y, z \in F$.

7. There is a unique element 1(one)$\in F$ such that $x1 = x$, for every $x \in F$.

8. To each non-zero $x \in F$ there corresponds a unique element x^{-1} or $\frac{1}{x}$ in F such that $xx^{-1} = 1$.

9. Multiplication distributes over addition, that is, $x(y + z) = xy + xz$ for all $x, y, z \in F$.

Suppose we have a set F of objects $x, y, z, \ldots$ and two operations on the elements of F as follows. The first operation called addition associates with each pair of elements $x, y \in F$ an element $x + y \in F$, the second operation, called multiplication, associates with each pair $x, y \in F$ an element $xy \in F$ and these two operations satisfy all conditions above. The set F together these two operations is then called a **field**. With the usual operations of addition and multiplication, the set C of complex numbers is a field, as is the set R of Real numbers. A **subfield** of the field C is a set F of complex numbers which itself a field under the usual operations of addition and multiplication of complex numbers.

Example 1. 1. The set of **positive integers**: $1, 2, 3, \ldots$, is not a field. Since there exists no zero element.

2. The set of **integers**: $\ldots, -2, -1, 0, 1, 2, \ldots$ is not a subfield of C, because for integer n, $1/n$ is not an integer unless n is 1 or -1.

3. The set of rational numbers, that is numbers of the form p/q, where p and q are integers and $q \neq 0$, is a subfield of the complex numbers.

4. The set of all real numbers R is a subfield of C.

5. The set of all complex numbers of the form $x + y\sqrt{2}$, where x and y are rational is a subfield of C.

1.2 Vector Spaces

Definition 1.1. *(Vector space or Linear space) A vector space or linear space consists of the following:*

1. *a field F of scalars*
2. *a set V of objects called vectors*
3. *an operation, called vector addition which associates each pair of vectors $\alpha, \beta \in V$, a vector $\alpha + \beta \in V$, called the sum of α and β in such a way that*

   (a) *vector addition is commutative; $\alpha + \beta = \beta + \alpha$.*

   (b) *vector addition is associative; i.e. $\alpha + (\beta + \gamma) = (\alpha + \beta) + \gamma$ for $\alpha, \beta, \gamma \in V$.*

   (c) *There exists a unique vector $0 \in V$, called the zero vector, such that $\alpha + 0 = 0 + \alpha = \alpha$ for every $\alpha \in V$.*

   (d) *for each vector $\alpha \in V$, there is a unique vector α^{-1} (called the additive inverse of α) in V such that $\alpha + \alpha^{-1} = 0$.*

4. *an operation called scalar multiplication, which associates each scalar $c \in F$ and a vector $\alpha \in V$, a vector $c\alpha \in V$ in such a way that*

1. $1.\alpha = \alpha$ *for every* $\alpha \in V$
2. $(c_1 c_2)\alpha = c_1(c_2\alpha)$
3. $c(\alpha + \beta) = c\alpha + c\beta$
4. $(c_1 + c_2)\alpha = c_1\alpha + c_2\alpha$.

We say that V is a vector space over the field F.

Example 2. Let n be a positive integer. F^n is the set of all ordered n-tuples $(x_1, x_2, \ldots, x_n)$ where $x_i \in F$. Then show that F^n is a vector space over F.

Solution: We have $V = F^n = \{(x_1, x_2, \ldots, x_n),\ x_i \in F\}$.

Let $\alpha = (x_1, x_2, \ldots, x_n)$ and $\beta = (y_1, y_2, \ldots, y_n) \in F^n$ and $c \in F$.

Then addition in F^n is defined as $\alpha + \beta = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n)$ and the scalar multiplication in F^n is defined by $c.\alpha = (cx_1, cx_2, \ldots, cx_n)$. Now the properties of vector addition are:

1. $\alpha + \beta = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n) = (y_1 + x_1, y_2 + x_2, \ldots, y_n + x_n) = \beta + \alpha$ so that vector addition is commutative.

2. $\alpha + (\beta + \gamma) = (x_1, x_2, \ldots, x_n) + [(y_1, y_2, \ldots, y_n) + (z_1, z_2, \ldots, z_n)] = (x_1, x_2, \ldots, x_n) + [(y_1 + z_1, y_2 + z_2, \ldots, y_n + z_n)] = (x_1 + (y_1 + z_1), x_2 + (y_2 + z_2), \ldots, x_n + (y_n + z_n)) = ((x_1 + y_1) + z_1, (x_2 + y_2) + z_2, \ldots, (x_n + y_n) + z_n)$ (In each component we have used the associativity of addition in the field F) $= (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n) + (z_1, z_2, \ldots, z_n) = [(x_1, x_2, \ldots, x_n) + (y_1, y_2, \ldots, y_n)] + (z_1, z_2, \ldots, z_n) = (\alpha + \beta) + \gamma$; so that vector addition is associative.

3. Now $0 = (0, 0, \ldots, , 0) \in F_n$ and $\alpha + 0 = (x_1, x_2, \ldots, x_n) + (0, 0, \ldots, 0) = (x_1, x_2, \ldots, x_n) = \alpha$, so that 0 is the zero vector in F^n.

4. For $\alpha = (x_1, x_2, \ldots, x_n) \in F^n$, there is a unique vector $\alpha^{-1} = (-x_1, -x_2, \ldots, -x_n) \in F^n$, such that

$$\begin{aligned}
\alpha + (\alpha^{-1}) &= (x_1, x_2, \ldots, x_n) + (-x_1, -x_2, \ldots, -x_n) \\
&= (x_1 + (-x_1), x_2 + (-x_2), \ldots, x_n + (-x_n)) \\
&= (0, 0, \ldots, 0) = 0
\end{aligned}$$

For the scalar multiplication, $c\alpha = c(x_1, x_2, \ldots, x_n) = (cx_1, cx_2, \ldots, cx_n) \in F^n$, and the following properties hold:

1. $1.\alpha = 1.(x_1, x_2, \ldots, x_n) = (1.x_1, 1.x_2, \ldots, 1.x_n) = (x_1, x_2, \ldots, x_n) = \alpha.$

2. Let $\alpha = (x_1, \ldots, x_n) \in F^n)$

$$\begin{aligned}
c_1c_2\alpha &= c_1c_2(x_1, x_2, \ldots, x_n) \\
&= (c_1c_2x_1, c_1c_2x_2, \ldots, c_1c_2x_n) \\
&= c_1(c_2x_1, c_2x_2, \ldots, c_2x_n) \\
&= c_1(c_2\alpha).
\end{aligned}$$

3. Let $\alpha = (x_1, \ldots, x_n)$, $\beta = (y_1, \ldots, y_n) \in F^n)$

$$\begin{aligned}
c(\alpha + \beta) &= c[(x_1, x_2, \ldots, x_n) + (y_1, y_2, \ldots, y_n)] \\
&= c(x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n) \\
&= (c(x_1 + y_1), c(x_2 + y_2), \ldots, c(x_n + y_n)) \\
&= (cx_1 + cy_1, cx_2 + cy_2, \ldots, cx_n + cy_n) \\
&= (cx_1, cx_2, \ldots, cx_n) + (cy_1, cy_2, \ldots, cy_n) \\
&= c(x_1, x_2, \ldots, x_n) + c(y_1, y_2, \ldots, y_n) \\
&= c\alpha + c\beta.
\end{aligned}$$

4. Let $c_1, c_2 \in F$, and $\alpha = (x_1, \ldots, x_n) \in F^n$,

$$\begin{aligned}
(c_1 + c_2)\alpha &= (c_1 + c_2)(x_1, x_2, \ldots, x_n) \\
&= ((c_1 + c_2)x_1, (c_1 + c_2)x_2, \ldots, (c_1 + c_2)) \\
&= (c_1x_1 + c_2x_1, c_1x_2 + c_2x_2, \ldots, c_1x_n + c_2x_n) \\
&= c_1(x_1, x_2, \ldots, x_n) + c_2(x_1, x_2, \ldots, x_n) \\
&= c_1\alpha + c_2\alpha.
\end{aligned}$$

Thus F^n is a vector space over F.

Note that in Example 2 if we take $n = 1$ and $F = R$ we can see that R is a vector space over R; if we take $n = 2$ and $F = R$ we can see that R^2 is a vector space over R; taking $n = 3$, R^3 is a vector space over R and so on.

Remark 1. The vector addition and scalar multiplication defined has major role in determining whether a set is a vector space or not as in Examples 3 and 4.

Example 3. Show that $R^2 = R \times R$ is not a vector space over R when the vector addition is defined by $(x_1, x_2) + (y_1, y_2) = (x_1 + y_1, x_2 + y_2)$ and the scalar multiplication is defined by $c(x_1, x_2) = (cx_1, 0)$, where $x_1, x_2, y_1, y_2, c \in R$.

Solution: Note that condition (1) under the scalar multiplication is not satisfied as $1.\alpha = 1.(x_1, x_2) = (1.x_1, 0) = (x_1, 0) \neq \alpha$. Hence R^2 is not a vector space over R under the above defined operations.

Example 4. Show that $R^2 = R\times R$ is not a vector space over R when the vector addition is defined by $(x_1, x_2)+(y_1, y_2) = (x_1+y_1, x_2+y_2)$ and the scalar multiplication is defined by $c(x_1, x_2) = (cx_1, x_2)$, where $x_1, x_2, y_1, y_2, c \in R$.

Solution: We show that condition (4) under scalar multiplication in the definition of vector space is not satisfied by the multiplication defined in this example: For $c_1, c_2 \in F$ and $\alpha = (x_1, x_2) \in R^2$, by the given definition of multiplication, $(c_1 + c_2).\alpha = (c_1 + c_2)(x_1, x_2) = ((c_1 + c_2)x_1, x_2) = ((c_1x_1 + c_2x_1), x_2) \neq (c_1x_1, x_2) + (c_2x_1, x_2)$. Thus $(c_1 + c_2).\alpha \neq c_1\alpha + c_2\alpha$.

Example 5.

The space of $m \times n$ matrices, $F^{m\times n}$:

Let $F^{m\times n}$ be the set of all $m \times n$ matrices over the field F, where vector addition and scalar multiplication is defined as follows: For every $A, B \in$

$F^{m\times n}$ and for every $c \in R$

$$(A+B)_{ij} = A_{ij} + B_{ij}$$

and

$$(cA)_{ij} = c(A)_{ij}.$$

Example 6. The space of functions from any nonempty set to the field F. Let F be any field and X be any non empty set. Let V be the set of all functions from the set X into F. The sum of two vectors $f, g \in V$ is defined as $(f+g)(x) = f(x) + g(x)$. The product of the scalar $c \in F$ and a vector $f \in V$ is defined as $(cf)(x) = cf(x)$.
Verification: (a) Vector addition is commutative i.e. $(f+g)(x) = f(x) + g(x) = g(x) + f(x) = (g+f)(x)$. Since $f(x)$ and $g(x)$ are elements of F and by the commutativity of elements of field F.

(b) Vector addition is associative. Again using the associativity of elements of field, we can see that

$$\begin{aligned}((f+g)+h)(x) &= (f+g)(x)) + h(x)\\ &= [f(x)+g(x)] + h(x)\\ &= f(x) + [g(x)+h(x)]\\ &= f + (g+h)(x).\end{aligned}$$

(c) Identity element exists. $(f+0)(x) = f(x) + 0(x) = f(x)$ where 0 is the zero function. Hence the zero function, is the identity element.
(d) Existence of additive inverse: For every f in V , there exists a function $-f \in V$ which is given by $(-f)(x) = -f(x)$ such that $f + (-f) = 0$, the zero function because $(f + (-f))(x) = (f)(x) - f(x) = 0$.
(e) For every $f \in V$, $1.f = f$ as: $(1.f)(x) = 1.f(x) = f(x)$.

(f) We have $(c_1c_2f)(x) = c_1c_2f(x) = c_2c_1f(x)$, since $c_1,\ c_2,\ f(x) \in F$.

(g) We have $c(f+g)(x) = c(f(x)+g(x)) = cf(x)+cg(x)$.

(h) $(c_1+c_2)(f)(x) = (c_1+c_2)(f(x)) = (c_1f(x)+c_2f(x)) = (c_1f+c_2f)(x)$.

Example 7. Let F be a field and let V be the set of all functions f from F into F which have a rule of the form $f(x) = c_0 + c_1x + c_2x^2 + \ldots + c_nx^n$, where $c_0, c_1, \ldots, c_n$ are fixed scalars in F. A function of this type is called a polynomial function of F. Let addition and scalar multiplication be defined as $(f+g)(x) = f(x)+g(x)$ and $(cf)(x) = cf(x)$.
Solution: Let V be the set of all polynomials in the indeterminate x with coefficients in F . i.e.

$$V = \{f : F \to F | f(x) = a_0 + a_1x + a_2x^2 + \ldots + a_nx^n : a_i \in F\}$$

Let

$$\begin{aligned} f(x) &= a_0 + a_1x + a_2x^2 + \ldots + a_nx^n \\ g(x) &= b_0 + b_1x + b_2x^2 + \ldots + b_nx^n \\ h(x) &= p_0 + p_1x + p_2x^2 + \ldots + p_nx^n. \end{aligned}$$

be any three elements (i.e. polynomials) in V. Polynomial addition (vector addition) is defined by $(f+g)(x) = f(x)+g(x)$
$= a_0 + a_1x + a_2x^2 + \ldots + a_nx^n + b_0 + b_1x + b_2x^2 + \ldots + b_nx^n$
$= a_0+b_0+(a_1+b_1)x+(a_2+b_2)x^2+\ldots+(a_n+b_n)x^n$. and scalar multiplication is defined by $cf(x) = c(a_0 + a_1x + a_2x^2 + \ldots + a_nx^n) = ca_0 + ca_1x + ca_2x^2 + \ldots + ca_nx^n$. Clearly $f+g$ and cf also belong to V.
1. Commutativity of addition:

$$\begin{aligned} (f+g)(x) &= f(x)+g(x) \\ &= (a_0 + a_1x + a_2x^2 + \ldots + a_nx^n) + (b_0 + b_1x + b_2x^2 + \ldots + b_nx^n) \end{aligned}$$

$$
\begin{aligned}
&= (a_0 + b_0) + (a_1 + b_1)x + ... + (a_n + b_n)x^n. \\
&= (b_0 + a_0) + (b_1 + a_1)x + ... + (b_n + a_n)x^n. \\
&= g(x) + f(x) \\
&= (g + f)(x)
\end{aligned}
$$

2. 0, the zero polynomial, is the zero vector in V. 3. The inverse of f is $(-f) \in V$, where $(-f)(x) = -f(x) = -a_0 + -a_1x + -a_2x^2 + \ldots + -a_nx^n$ in V. The student can easily verify all other properties for vector space.

Example 8.

The field C of complex numbers may be regarded as a vector space over the field R of real numbers.

Is R a vector space over C?

Solution: No, since $c \in C$ and $x \in R$ does not imply $cx \in R$. In particular, $i \in C$ and $2 \in R$ but $2i$ does not belong to R.

Is R a vector space over Q?

Solution: Yes. Since for every $c \in Q$ and $x, y \in R$ we have $cx + y \in R$.

Is Q a vector space over R?

Solution: No, since $c \in R$ and $x \in Q$ does not imply $cx \in Q$. In particular $\sqrt{2} \in R$ and $1 \in Q$, but $\sqrt{2}.1 = \sqrt{2}$ does not belong to Q.

Exercises

1. Show that V, the set of all vectors in the real plane is a vector space over R.

2. Show that R^3 , the set of all vectors in the three dimensional real space is a vector space over R.

3. Prove that the set of integers is not a vector space over the set of rational numbers under usual addition and scalar multiplication.

4. Let $V = C[0,1]$ be the set of all real valued continuous functions defined over the closed unit interval $[0,1]$. Show that V is a real vector space under the vector addition and scalar multiplication defined as follows: For every $f, g \in V$ and for every $c \in R$. $(f+g)(x) = f(x) + g(x)$ for every $x \in [0,1]$
$(cf)(x) = cf(x)$ for every $x \in [0,1]$.

Linear combination of Vectors

Definition 1.2. *A vector $\beta \in V$ is said to be a linear combination of vectors $\alpha_1, \alpha_2, \ldots, \alpha_n$ in V if there exists scalars $c_1, c_2, \ldots, c_n$ in F such that*

$$\beta = c_1\alpha_1 + c_2\alpha_2 + \ldots + c_n\alpha_n = \sum_{i=1}^{n} c_i\alpha_i.$$

We know that $(1,0)$ and $(0,1)$ are unit vectors in R^2. And if $(x,y) \in R^2$, then $(x,y) = x(1,0) + y(0,1)$. That is any vector in R^2 is a linear combination of the two vectors $(1,0)$ and $(0,1)$. Similarly in the case of R^n, $(x_1, x_2, \ldots, x_n)$ can be written as a linear combination of the vectors

$$(1,0,0,\ldots,0), (0,1,0,\ldots,0), \ldots, (0,0,0,\ldots,1) \in R^n.$$

That is

$$(x_1, x_2, \ldots, x_n) = x_1(1,0,0,\ldots,0) + x_2(0,1,0,\ldots,0) + \ldots + x_n(0,0,0,\ldots,1).$$

1.3 Subspaces

In this section, we deal with some basic concepts in the study of vector spaces.

Definition 1.3. *Let V be a vector space over the field F. A subspace of V is a subset W of V which is itself a vector space over F with the operations of vector addition and scalar multiplication on V.*

Now we study a property of subspaces which helps us to check a given subset of a vector space V is a subspace or not.

Theorem 1.1. *A non-empty subset W of V is a subspace of V if and only if for each pair of vectors $\alpha, \beta \in W$ and $c \in F$, the vector $c\alpha + \beta$ is again in W.*

Proof. Assume that A non-empty subset W of V is a subspace of V. Let $\alpha, \beta \in W$. Now W is itself a vector space over F by the definition of a subspace. Thus $c\alpha \in W$ and $\beta \in W$ and thus their sum $c\alpha + \beta$ belongs to W.

Conversely assume that W is a non-empty subset of V such that $c\alpha + \beta \in W$ for all vectors $\alpha, \beta \in W$. To prove that W is a vector space over F. Since W is non-empty, there is a vector $\alpha \in W$ and hence $(-1)\alpha + \alpha = 0$ is in W. If α is any vector, and c is any scalar, then we have $c\alpha = c\alpha + 0$. Thus $c\alpha$ is in W. In particular, $(-1)\alpha + 0 = -\alpha$ is in W. Now if β is any other vector in W, then $\alpha + \beta = 1.\alpha + \beta \in W$. Thus sum of vectors, additive identity, additive inverse and scalar multiple is there in W and W is a subset of vector space V, Thus W is itself a vector space over F. That is W is a subspace of V. □

Example 9.

1. If V is any vector space, V is a subspace of V. Also the subset consisting of zero vector alone is a subspace of V called the zero subspace of V.

2. In F^n, the set W of n tuples $(x_1, x_2, \ldots, x_n)$ with $x_1 = 0$ is a subspace. For, $\alpha = (0, x_2, \ldots, x_n)$, $\beta = (0, y_2, \ldots, y_n)$ and $c \in F$, $c\alpha + \beta = (0, x_2 + y_2, \ldots x_n + y_n) \in W$.

3. The set W of n-tuples with $x_1 = 1 + x_2$ is not a subspace $(n \geq 2)$. is not a subspace of V. For, $\alpha = (1+x_2, x_2, \ldots, x_n)$ and $\beta = (1+y_2, y_2, \ldots, y_n)$ and $c = 1$, then $c\alpha + \beta = (2 + x_2 + y_2, x_2 + y_2, \ldots, x_n + y_n)$ which is not an element of W.

4. The space of polynomial functions over the field F is a subspace of the space of all functions from F into F.

5. An $n \times n$ square matrix A over a field F is symmetric if $A_{ij} = A_{ji}$ for each i and j. The symmetric matrices form a subspace of the space of all $n \times n$ matrices over F.

6. An $n \times n$ square matrix A over the field of complex numbers is Hermitian(or self adjoint) if $A_{jk} = \bar{A_{kj}}$ for each $j,\ k$. A 2×2 matrix is Hermitian if and only if it has the form

$$\begin{pmatrix} z & x+iy \\ x-iy & w \end{pmatrix}$$

where x, y, z and w are real numbers. The set of all Hermitian matrices is not a subspace of the space of all $n \times n$ matrices over C. For if A is Hermitian, its diagonal entries $A_{11}, A_{22}, \ldots$, are all real numbers, but the diagonal entries of iA are in general not real.
Note that the set of all $n \times n$ Hermitian matrices <u>is a subspace</u> of the space of all $n \times n$ matrices over $\underline{\mathrm{R}}$.

7. The solution space of a system of homogeneous linear equations. Let A be an $m \times n$ matrix over F. Then the set of all $n \times 1$ (column) matrices X over F such that $AX = 0$ is a subspace of the space of all $n \times 1$ matrices over F. To prove this we must show that $A(cX + Y) = 0$ when $AX = 0$, $AY = 0$ and c is an arbitrary scalar in F. This follows immediately from the following general fact.

Lemma 1.1. *If A is an $m \times n$ matrix over F and $B,\ C$ are $n \times p$ matrices over F then*

$$A(dB + C) = d(AB) + AC$$

for each scalar d in F.

Proof.

$$\begin{aligned}[A(dB+C)]_{ij} &= \sum_k A_{ik}(dB+C)_{kj} \\ &= \sum_k (dA_{ik}B_{kj} + A_{ik}C_{kj}) \\ &= d\sum_k A_{ik}B_{kj} + \sum_k A_{ik}C_{kj} \\ &= d(AB)_{ij} + (AC)_{ij} \\ &= [d(AB)+AC]_{ij}\end{aligned}$$

□

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

Theorem 1.2. *Let V be a vector space over the field F. The intersection of any collection of subspaces of V is a subspace of V.*

Proof Let $\{W_a/a \in \mathscr{A}\}$ be a collection of subspaces and, $W = \bigcap_{a\in\mathscr{A}} W_a$ be their intersection. Since each W_a is a subspace, zero vector exists in each W_a. So zero vector belongs to their intersection W. Thus W is non-empty. Consider $\alpha, \beta \in W$ and $c \in F$. Both $\alpha, \beta \in W_a$ and since each W_a is a subspace, the vector $c\alpha + \beta \in W_a$. Therefore $c\alpha + \beta$ belongs to their intersection $W = \bigcap_{a\in\mathscr{A}} W_a$. Hence W is a subspace of V.

Subspace Spanned by a Subset

Definition 1.4. *Let S be a set of vectors in a vector space V. The subspace spanned by S is defined as the intersection W of all subspaces of V which contains S. When S is a finite set of vectors, $S = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$. Then we denote W as the subspace spanned by the vectors $\alpha_1, \alpha_2, \ldots, \alpha_n$.*

Theorem 1.3. *The subspace spanned by a non-empty subset S of a vector space V is the set of all linear combinations of vectors in S.*

Proof. Let W be the subspace spanned by S and L be the set of all linear combinations of vectors in S. To prove that $W = L$. Let $\alpha_1, \alpha_2, \ldots, \alpha_n \in S$, Now $\alpha = c_1\alpha_1 + c_2\alpha_2 + \ldots + c_n\alpha_n$ is in W. Then each linear combinations of vectors in S belongs to W. Thus W contains the set L of all linear combinations of vectors in S. That is $L \subseteq W$.

Note that S is contained in L. Next we have to prove that $W \subseteq L$. In order to prove this, it is enough to prove that L is a subspace of V (Since W is the intersection of all subspaces of V containing S.) The set L contains S and is non empty. Let $\alpha, \beta \in L$, then each one is a linear combination of the vectors in S; say, $\alpha = c_1\alpha_1 + c_2\alpha_2 + \ldots + c_n\alpha_n$ and $\beta = d_1\beta_1 + d_2\beta_2 + \ldots + d_n\beta_n$ where α_i and β_j are vectors in S. Thus

$$\begin{aligned} c\alpha + \beta &= cc_1\alpha_1 + cc_2\alpha_2 + \ldots + cc_n\alpha_n + d_1\beta_1 + d_2\beta_2 + \ldots + d_n\beta_n \\ &= cc_1\alpha_1 + d_1\beta_1 + cc_2\alpha_2 + d_2\beta_2 + \ldots + cc_n\alpha_n + d_n\beta_n. \end{aligned}$$

This implies that $c\alpha + \beta$ is an element of L. Therefore L is a subspace of V which contains S. Also W is the smallest subspace of V containing S (since by definition it is the intersection of all subspaces containing S), which implies that $W \subseteq L$. Thus $L = W$. □

Definition 1.5. *If $S_1, S_2, \ldots, S_k$ are the subsets of a vector space V, the set of all sum $\alpha_1 + \alpha_2 + \ldots + \alpha_n$ where $\alpha_i \in S_i$ is called sum of the subsets of $S_1, S_2, \ldots, S_k$ and is denoted by $S_1, S_2, \ldots, S_k = \sum_{i=1}^{k} S_i$.*
If $W_1, W_2, \ldots, W_n$ are the subspaces of a vector space V, then the sum $W_1 + W_2 + \ldots + W_n$ is the set of all sum of vectors $\alpha_1, \alpha_2, \ldots, \alpha_n$ of $\alpha_i \in W_i$.

Example 10.

1. If $S = \{(1,0),(0,1)\}$ then $L = spanS = \{\alpha(1,0) + \beta(0,1) | \alpha, \beta \in R\} = \{(\alpha, \beta) | \alpha, \beta \in R\} = R^2$
2. $(3,7)$ belongs to the set of linear combination of the set $\{(1,2),(0,1)\}$ because $(3,7) = 3.(1,2) + 1.(0,1)$.

3. $(3, 7)$ is not a linear combination of the vectors $(1, 2)$, $(2, 4)$; for if $(3, 7) = c_1(1, 2) + c_2(2, 4)$. then

$$c_1 + 2c_2 = 3 \tag{1.1}$$

$$2c_1 + 4c_2 = 7 \tag{1.2}$$

The above system of equation is inconsistent so that there exist no c_1 and c_1 satisfying the equations (1) and (2). Hence we can conclude that $(3, 7)$ cannot be expressed as a linear combination of $(1, 2)$ and $(2, 4)$.

4. Let V be the space of all polynomial functions over the field F. Let S be a subset of V consisting of the polynomial functions $f_0, f_1, f_2, \ldots$ defined by $f_n(x) = x^n$, $n = 1, 2, 3, \ldots$ i.e.

$$f_0 = 1,\ f_1 = x,\ f_2 = x^2$$

The subspace spanned by S is the set of all linear combination of elements of S. i.e. elements of the form $a_0 + a_1x + a_2x^2 + \ldots$

Exercises

1. Show that $W = \{(x_1, x_2, ..., x_n)|x_i \in F, x_1 = 0\}$ is a subspace of F_n .
2. Which of the following sets U of vectors $u = (c_1, \ldots, c_n)$ in R_n are subspaces of R_n $(n \geq 3)$?
   (a) all u such that $c_1 \geq 0$.
   (b) all u such that $c_1 + 3c_2 = c_3$.
   (c) all u such that $c_2 = c_1^2$.
   (d) all u such that $c_1c_2 = 0$.
   (e) all u such that c_2 is rational.
3. Let W be the set of all vectors of the form $(c, 2c, -3c, c)$ in R_4. Show that W is a subspace of R_4 .
4. We know that C , the set of all complex numbers, is a vector space over R. Then show that $W = \{iy/y \in R\}$ is a subspace of C.

1.4 Bases and Dimension

In this section we assign a dimension to certain vector spaces. We will define dimension of a vector space by using the concept of a basis of a space.

Definition 1.6. *Let V be a vector space over R (or over the field F). A finite set vectors $\alpha_1, \alpha_2, \ldots, \alpha_n$ in V is said to be linearly dependent if there exist real numbers (or scalars) $c_1, c_2, \ldots, c_n$ not all zeros such that $c_1\alpha_1 + c_2\alpha_2 + \ldots + c_n\alpha_n = 0$, where 0 is the zero vector in the vector space V.*

Example 11.

The set $\{(1,0,1),(1,1,0),(-1,0,-1)\}$ is linearly dependent in R^3.
Now $c_1(1,0,1) + c_2(1,1,0) + c_3(-1,0,-1) = (0,0,0)$ implies $(c_1 + c_2 - c_3, c_2, c_1 - c_3) = (0,0,0)$ implies $c_1 + c_2 - c_3 = 0$, $c_2 = 0$, $c_1 - c_3 = 0$ implies $c_2 = 0, c_1 = c_3$, which implies that c_1 can take any arbitrary value, in particular take $c_1 = 1$, so that $c_1 = 1, c_2 = 0, c_3 = 1$ and so scalars are not all zeros and $1.(1,0,1) + 0.(1,0,1) + 1.(-1,0,-1) = (0,0,0) = 0$. Hence $\{(1,0,1),(1,1,0),(-1,0,-1)\}$ is linearly dependent.

The set $\{(1,2),(0,0)\}$ is linearly dependent in R^2 as there exist scalars 0 and 1 (not all zeros) such that $0.(1,2) + 1.(0,0) = (0,0)$.

Remark 2. The set $\{v_1, v_2, v_3\}$ of vectors is linearly dependent if and only if one of them is a linear combination of the other two vectors.

Solution: Suppose $\{v_1, v_2, v_3\}$ is linearly dependent. Then there exist scalars c_1, c_2, c_3 at least one of them, say c_1 not equal to 0 such that $c_1v_1 + c_2v_2 + c_3v_3 = 0$. or $v_1 = -\frac{c_2}{c_1}v_2 - \frac{c_3}{c_1}v_3$ which means that v_1 is a linear combination of the vectors v_2 and v_3.

Conversely suppose v_1 is a linear combination of the other two vectors v_2 and v_3. i.e. suppose that $v_1 = -c_2v_2 - c_3v_3$ for some scalars c_2 and c_3. Hence

v_1, v_2, v_3 is linearly dependent.

Definition 1.7. *(Linearly independent set) A set is said to be linearly independent if it is not linearly dependent. Hence a finite set* $\alpha_1, \alpha_2, \ldots, \alpha_n$ *of vectors in the vector space* V *is said to be linearly independent if* $c_1\alpha_1 + c_2\alpha_2 + \ldots + c_n\alpha_n = 0$ *implies that* $c_1 = c_2 = \ldots = c_n = 0$.

Note that the null set is always taken to be linearly independent.

Example 12.

$\{(1,0,1),(1,1,0),(1,1,-1)\}$ is linearly independent,
Solution: Now $c_1(1,0,1) + c_2(1,1,0) + c_3(1,1,-1) = (0,0,0)$ implies that $(c_1 + c_2 + c_3, c_2 + c_3, c_1 - c_3) = (0,0,0)$. $c_1 + c_2 + c_3 = 0$, $c_2 = -c_3$ and $c_1 = c_3$. Then $c_1 = c_2 = c_3 = 0$. Hence $\{(1,0,1),(1,1,0),(1,1,-1)\}$ is linearly independent.

Remark 3.

1. Any set which contains a linearly dependent set is linearly dependent.

2. Any subset of a linearly independent set is linearly independent.

3. $\{0\}$, the set consisting of zero vector alone, is linearly dependent as there exists such that $\alpha.0 = 0$, in particular $1.0 = 0$.

4. Any set that contains the zero vector 0 is linearly dependent.

Excercises

1. Determine whether the subset $\{(1,0,1),(1,1,1),(0,0,1)\}$ of R^3 is a linearly independent or not.

2. Which of the following subsets S of R^4 are linearly dependent?

   (a) $S = \{(1,0,0,0),(1,1,0,0),(1,1,1,1),(0,0,1,1)\}$.

   (b) $S = \{(1,-1,2,0),(1,1,2,0),(3,0,0,1),(2,1,-1,0)\}$.

3. Show that $\{1, x, x^2\}$ is a linearly independent subset of $P_2(x)$.
4. Prove that $\{1, 1+x, 2x+x^2\}$ is a linearly independent subset of $P_2(x)$.
5. Show that in R_3, the vectors $v_1 = (-1, 2, 1)$ and $v_2 = (3, 1, -2)$ are linearly independent.
6. Show that $\{(1, 2, 3), (2, 3, 1), (3, 1, 2)\}$ is a linearly independent subset of R^3.
7. Show that the vectors $v_1 = (1, 1, 2, 4), v_2 = (2, -1, -5, 2), v_3 = (1, -1, 4, 0)$ and $v_4 = (2, 1, 1, 6)$ are linearly independent in R^4.
8. In R^n show that $S = \{e_1, e_2, \ldots, e_n\}$, where $e_1 = (1, 0, 0, \ldots, 0)$, $e_2 = (0, 1, 0, \ldots, 0), \ldots, e_n = (0, 0, 0, \ldots, 0, 1)$ is linearly independent.

1.4.1 Basis

Definition 1.8. *Let V be a vector space. A basis for V is a linearly independent set of vectors in V which spans V.*

Example 13. 1. $\mathscr{B} = \{(1, 0), (0, 1)\}$ is a basis for R^2. Let $(x, y) \in R^2$. Then $(x, y) = x(1, 0) + y(0, 1)$. This means that any $(x, y) \in R^2$ can be written as a linear combination of these two vectors. This means that $\mathscr{B}$ spans R^2. On inspection one can verify that $\mathscr{B}$ is linearly independent. Thus $\mathscr{B}$ is a basis for R^2 and is called standard basis for R^2.

2. The set $\mathscr{B} = \{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$ is a basis for R^3.

Solution

To prove that $\mathscr{B}$ is linearly independant and it generates R^3. First we will show that $\mathscr{B}$ generates R^3. i.e. to show that every element in R^3 can be expressed as a linear combination of vectors in $\mathscr{B}$. To prove this let (x_1, x_2, x_3) be any element in R^3 and also write

$$(x_1, x_2, x_3) = c_1(1, 0, 0) + c_2(0, 1, 0) + c_3(0, 0, 1) \tag{1.3}$$

We have to verify whether it is possible to write the scalars c_1, c_2, c_3 in terms of x_1, x_2 and x_3. Now 1.3 is equivalent to write,

$$(x_1, x_2, x_3) = (c_1, c_2, c_3),$$

so that $x_1 = c_1$, $x_2 = c_3$ and $x_3 = c_3$. so $(x_1, x_2, x_3) = x_1(1, 0, 0) + x_2(0, 1, 0) + x_3(0, 0, 1)$, which implies that $\mathscr{B}$ generates R^3.

(ii) $\mathscr{B}$ is linearly independent, for $c_1(1, 0, 0) + c_2(0, 1, 0) + c_3(0, 0, 1) = (0, 0, 0)$

implies $(c_1, c_2, c_3) = (0, 0, 0)$ implies $c_1 = c_2 = c_3 = 0$. Hence $\mathscr{B}$ is a basis for R^3. This basis is called standard basis for R^3.

3. $\mathscr{B} = \{e_1, e_2, \ldots, e_n\}$ where e_i is the n tuple whose ith coordinate is 1 and all other coordinates are zero. That is

$$e_1 = (1, 0, 0, \ldots, 0), e_2 = (0, 1, 0, \ldots, 0), \ldots, e_n = (0, 0, 0, \ldots, 0, 1)$$

. Then $\mathscr{B}$ is a basis for R^n and is called standard basis for R^n.

4. Show that the set $\mathscr{B} = \{(1, 2, 1), (2, 1, 0), (1, -1, 2)\}$ is a basis for R^3.

Solution

(i). First let us prove that $\mathscr{B}$ generates R^3. i.e. to show that every element in R^3 can be expressed as a linear combination of elements of $\mathscr{B}$. To prove this let (x_1, x_2, x_3) be any element in R^3 and also write $(x_1, x_2, x_3) = c_1(1, 2, 1) + c_2(2, 1, 0) + c_3(1, -1, 2)$ We have to verify whether it is possible to write the scalars c_1, c_2, c_3 in terms of x_1, x_2 and x_3. Now $(x_1, x_2, x_3) = (c_1 + 2c_2 + c_3, 2c_1 + c_2 - c_3, c_1 + 2c_3)$, so that

$$x_1 = c_1 + 2c_2 + c_3,$$

$$x_2 = 2c_1 + c_2 - c_3,$$

$$x_3 = c_1 + 2c_3,$$

which on solving gives

$$c_1 = 1/9(-2x_1 + 4x_2 + 3x_3),$$

$$c_2 = 1/9(5x_1 - x_2 - 3x_3),$$

$$c_3 = 1/9(x_1 - 2x_2 + 3x_3),$$

which implies that $\mathscr{B}$ generates R^3.

(ii). $\mathscr{B}$ is linearly independent.

For $c_1(1,2,1) + c_2(2,1,0) + c_3(1,-1,2) = (0,0,0)$ implies

$$(c_1 + 2c_2 + c_3, 2c_1 + c_2 - c_3, c_1 + 2c_3) = (0,0,0).$$

This implies

$$c_1 + 2c_2 + c_3 = 0; 2c_1 + c_2 - c_3 = 0; c_1 + 2c_3 = 0$$

implies

$$c_1 = c_2 = c_3 = 0$$

on solving.

5. Let P be an invertible $n \times n$ matrix with entries in the field F. Then $P_1, P_2, \ldots, p_n$, the columns of P, form a basis for the space of column matrices, $F^{n\times 1}$.

Theorem 1.4. *Let V be a vector space, which is spanned by a finite set of vectors $\beta_1, \beta_2, \ldots, \beta_m$. Then any independent set of vectors in V is finite and contains no more than m elements.*

Proof. To prove this theorem we show that every subset S of V which contains more than m vectors is linearly dependent. Let S be such a set. In S

there are distinct vectors $\alpha_1, \alpha_2, \ldots, \alpha_n$, where $n > m$. Since $\beta_1, \beta_2, \ldots, \beta_m$ spans V, there exists scalars $A_{ij} \in F$ such that $\alpha_j = \sum_{i=1}^{m} A_{ij}\beta_i$. For any n scalars $x_1, x_2, \ldots, x_n$, we have

$$\begin{aligned} x_1\alpha_1 + x_2\alpha_2 + \ldots + x_n\alpha_n &= \sum_{j=1}^{n} x_j\alpha_j \\ &= \sum_{j=1}^{n} x_j \sum_{i=1}^{m} A_{ij}\beta_i \\ &= \sum_{j=1}^{n}\sum_{i=1}^{m} (A_{ij}x_j)\beta_i \\ &= \sum_{i=1}^{m} (\sum_{j=1}^{n} A_{ij}x_j)\beta_i. \end{aligned}$$

If $x_1\alpha_1 + x_2\alpha_2 + \ldots + x_n\alpha_n = 0 \Rightarrow \sum_{i=1}^{m}(\sum_{j=1}^{n} A_{ij}x_j)\beta_i = 0$, Since $\beta_i \neq 0$, we have $\sum_{j=1}^{n} A_{ij}x_j = 0$ for each $1 \leq i \leq m$. {We have the following theorem on the solutions of linear equations.

If A is an $m \times n$ matrix and $n > m$, then the homogeneous system of linear equations $AX = 0$ has a nontrivial solution.}

Since $n > m$ by this result, there exists scalars $x_1, x_2, \ldots, x_n$ not all zero such that $\sum_{j=1}^{n} A_{ij}x_j = 0$, $1 \leq i \leq m$. This shows that S is a linearly dependent set. Thus we proved that any independent set of vectors in V is finite and contains no more than m elements. □

Definition 1.9. *The number of elements in a basis of a vector space is called dimension of the vector space. Dimension of a vector space V is denoted by dim V. If dim $V = n$, then any subset of V which contains more than n vectors is linearly dependent.*

Example 14. 1. The dim R^3 is 3 .

2. The dim $P_2(x)$ is 3.

3. The dim $P_n(x)$ is $n+1$.

Corollary 1.1. *If V is a finite dimensional vector space, then any two bases of V have the same number of elements.*

Proof. Since V is finite dimensional, there exists a basis $\mathscr{B} = \{\beta_1, \beta_2, \ldots, \beta_m\}$ consisting of finite number of elements. Then every basis of V is finite and contains no more than m elements, by Theorem 1.4. Thus if $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ is any other basis, then $n \leq m$. Now take $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ as a basis. Then by the similar argument as above we get $m \leq n$. Thus $n = m$. □

Note that Corollary 1.1 says that in a finite dimensional vector space all its bases have the same number of elements.

Corollary 1.2. *Let V be a finite dimensional vector space and let $n = dimV$. Then*

1. *any subset of V which contains more than n vectors is linearly dependent.*
2. *no subset of V which contains less than n vectors can span V.*

Lemma 1.2. *Let S be a linearly independent subset of a vector space V. Suppose β is a vector in V which is not in the subspace spanned by S. Then the set obtained by adjoining β to S is linearly independent.*

Proof. Suppose $\alpha_1, \alpha_2, \ldots, \alpha_m$ are distinct elements of S. Suppose β is a vector in V which is not in the subspace spanned by S. Now to prove that $\{\alpha_1, \alpha_2, \ldots, \alpha_m, \beta\}$ is linearly independent. Let $c_1\alpha_1 + c_2\alpha_2 + \ldots + c_m\alpha_m + b\beta = 0$. If $b \neq 0$, then $\beta = \frac{-c_1}{b}\alpha_1 + \frac{-c_2}{b}\alpha_2 + \ldots + \frac{-c_m}{b}\alpha_m$. This means that β is a linear combination of $\alpha_1, \alpha_2, \ldots, \alpha_m$, which is a contradiction. Thus $b = 0$. So $c_1\alpha_1 + c_2\alpha_2 + \ldots + c_m\alpha_m = 0$, and since S is linearly independent set, each $c_i = 0$. This implies that $\{\alpha_1, \alpha_2, \ldots, \alpha_m, \beta\}$ is linearly independent. □

Theorem 1.5. *If W is a subspace of a finite dimensional vector space V, every linearly independent subset of W is finite and is part of a basis for W.*

Proof. Let S_0 be a linearly independent subset of W. If S is a linearly independent subset of W containing S_0, then S is also a linearly independent subset of V. Since V is finite dimensional, S contains no more than $dimV$ elements. We extend S_0, to a basis of W. If S_0 spans W, it is a basis for W. If not, we use the preceding Lemma to find a vector $\beta_1 \in W$ such that $S_1 = S_0 \cup \{\beta_1\}$ is independent. If S_1 spans W, it is a basis and S_0 is a part of the basis. If not, apply Lemma to obtain a vector β_2 in W such that $S_2 = S_1 \cup \{\beta_2\}$ is independent. Continuing the same argument we reach a set (in not more than dim V steps) $S_m = S_0 \cup \{\beta_1, \beta_2, \ldots, \beta_m\}$ which is a basis for W. □

Corollary 1.3. *If W is a proper subspace of a finite dimensional vector space V, then W is finite dimensional and $dimW < dimV$.*

Proof. Assume that W is a proper subspace of V, then there is an element $\alpha \neq 0$ in V which is not in W. Thus, by the previous Theorem, there is a basis of W which contains not more than $dimV$ elements. Hence W is finite dimensional and $dimW \leq dimV$. Adjoining α to any basis of W, by the Lemma, we obtain a linearly independent subset of V. Thus $dimW < dimV$. □

Corollary 1.4. *In a finite dimensional vector space, every non-empty linearly independent set of vectors is a part of a basis.*

Proof. Let V be a finite dimensional vector space of dimension n. Let $\{\alpha_1, \ldots, \alpha_k\}$ be a set of independent vectors in V. Let W be a subspace spanned by $\{\alpha_1, \ldots, \alpha_k\}$. Then by Theorem 1.5, we get every linearly independent subset of W is a part of a basis for W. In particular $\{\alpha_1, \ldots, \alpha_k\}$ is a linearly independent subset of W. Thus it is a part of a basis. $\square$

Corollary 1.5. *Let A be an $n \times n$ matrix over a field F, and suppose the row vectors of A form a linearly independent set of vectors in F^n. Then A is invertible.*

Proof. Let A be an $n \times n$ matrix over a field F, and suppose the row vectors of A form a linearly independent set of vectors in F^n. We have to prove that A is invertible. That is to prove that there exists a matrix B such that $BA = I$, where I is the identity matrix. Let $\alpha_1, \alpha_2, \ldots, \alpha_n$ be the row vectors of A, and suppose W is the subspace of F^n spanned by $\alpha_1, \alpha_2, \ldots, \alpha_n$. Since $\alpha_1, \alpha_2, \ldots, \alpha_n$ are linearly independent, the dimension of W is n. Then by corollary 1.3, we get $W = F^n$.
Now $e_i \in F^n$ implies that $e_i \in W$ and W is spanned by $\alpha_1, \alpha_2, \ldots, \alpha_n$. Hence there exists scalars B_{ij} in F such that $e_i = \sum\limits_{j=1}^{n} B_{ij}\alpha_j$, $1 \leq i \leq n$, where $e_1, e_2, \ldots, e_n$ is the standard basis of F^n. Thus for the matrix B with entries B_{ij} we have $BA = I$. $\square$

Theorem 1.6. *If W_1 and W_2 are finite dimensional subspace of a vector space V, then W_1+W_2 is finite dimensional and $dimW_1+dimW_2 = dim(W_1\cap W_2) + dim(W_1 + W_2)$.*

Proof. Since W_1 and W_2 are finite dimensional vector spaces, $W_1\cap W_2$ is also a finite dimensional vector space (by the Corollary above). Let $\{\alpha_1, \alpha_2, \ldots, \alpha_k\}$ be a basis of $W_1\cap W_2$. This basis is a part of the basis $\{\alpha_1, \alpha_2, \ldots, \alpha_k, \beta_1, \beta_2, \ldots, \beta_m\}$ for W_1 and a part of a basis $\{\alpha_1, \alpha_2, \ldots, \alpha_k, \gamma_1, \gamma_2, \ldots, \gamma_n\}$ for W_2. The subspace W_1+W_2 is spanned by the vectors $S = \{\alpha_1, \alpha_2, \ldots, \alpha_k, \beta_1, \beta_2, \ldots, \beta_m, \gamma_1, \gamma_2, \ldots, \gamma_n\}$. We prove that S is linearly independent. For suppose $\sum x_i\alpha_i + \sum y_j\beta_j + \sum z_r\gamma_r = 0$. Then $\sum x_i\alpha_i + \sum y_j\beta_j = -\sum z_r\gamma_r$. Means that $\sum z_r\gamma_r$ is a linear combination of α_i and β_j, which implies that $\sum z_r\gamma_r$ belongs to W_1. Also $\sum z_r\gamma_r$ is a linear combination of $\gamma_1, \gamma_2, \ldots, \gamma_n$. Therefore it belongs to W_2 also. This means that $\sum z_r\gamma_r$ belongs to $W_1 \cap W_2$. And therefore $\sum z_r\gamma_r = \sum c_i\alpha_i$ for certain scalars $c_1, c_2, \ldots, c_k$(Since $\{\alpha_1, \alpha_2, \ldots, \alpha_k\}$ is a basis of $W_1 \cap W_2$). Then $\sum z_r\gamma_r = \sum c_i\alpha_i \Rightarrow \sum z_r\gamma_r - \sum c_i\alpha_i = 0$. Since $\{\alpha_1, \alpha_2, \ldots, \alpha_k, \gamma_1, \gamma_2, \ldots, \gamma_n\}$ is linearly independent, we get each of the scalars $z_r = 0$. Thus $\sum x_i\alpha_i + \sum y_j\beta_j = 0$. Again since $\{\alpha_1, \alpha_2, \ldots, \alpha_k, \beta_1, \beta_2, \ldots, \beta_n\}$ is linearly independent, each $x_i = 0$ and each $y_j = 0$. Thus S is linearly independent. Hence S is a basis of $W_1 + W_2$. Finally

$$\begin{aligned} dimW_1 + dimW_2 &= (k+m) + (k+n) \\ &= k + (m+k+n) \\ &= dim(W_1 \cap W_2) + dim(W_1 + W_2). \end{aligned}$$

□

Example 15. Show that V is a vector space having m finite number of elements if and only if $m = p^n$ where p is prime and n is a non-negative integer.

Solution

Suppose there exists a vector space V having m number of elements. Also suppose that the underlying field is F. Let dimension of $V = n$.

Let $\{\alpha_1, \ldots, \alpha_n$ be one basis for V. Then the elements in V are exactly the linear combination of the n vectors $\alpha_1, \ldots, \alpha_n$. That is $a_1\alpha_1 + \cdots + a_n\alpha_n$, where $a_i, i = 1, 2, \ldots, n$ belongs to underlying field F. If F has infinite number of elements, then V must have infinite number of elements, which is not the case here. Therefore F has finite number of elements, or F is a finite field. We know that every finite field is isomorphic to Z_p , where p is a prime number. Thus F also must have p number of elements. Thus we can say that V contains p^n elements, where p is a prime (Reason: V contains p^n elements, because we have freedom to choose all the p elements of F as $a_1, a_2, \ldots, a_n$. Thus $\underbrace{p.p.\ldots.p}_{n\ times}$ elements will be there. Thus $m = p^n$.

Now suppose $m = p^n$. Then consider $V = Z_{p^n}$ and $F = Z_p$. Then V is a vector space over F having m number of elements.

Example 16. Does there exist a vector space having 8 number of elements?

Solution

Yes. As $8 = 2^3$ is an integral power of the prime 2.

Exercises

1. Let V be a vector space of all 2×2 matrices over the field F. Prove that V has dimension 4 by exhibiting a basis for V which has 4 elements.

2. Let V be a vector space over a subfield F of the complex numbers. Suppose α, β and γ are linearly independent vectors in V. Prove that $(\alpha + \beta), (\beta + \gamma)$ and $(\gamma + \alpha)$ are linearly independent.

3. Let V be a vector space over the field F. Suppose that there are a finite number of vectors $\alpha_1, \alpha_2, \ldots, \alpha_k$ in V which span V. Prove that V is finite dimensional.

4. Let V be the set of real numbers. Regard V as a vector space over the field of rational numbers, with the usual operations. Prove that this vector space is not finite dimensional.

1.5 Coordinates

One of the useful features of a basis $\mathscr{B}$ in an n-dimensional space V is that it enables one to introduce coordinates in V analogous to the natural coordinates x_i of a vector $\alpha = (x_1, x_2, \ldots, x_n)$ in the space F^n.

Definition 1.10. *If V is a finite dimensional vector space, an ordered basis for V is a finite sequence of vectors which is linearly independent and spans V.*

A basis $\mathscr{B}$ of a vector space is said to be an ordered basis, if the elements in the basis $\mathscr{B}$ are placed in some order.
For example, $B_1 = \{(1,0),(0,1)\}$ is an ordered basis for R^2, called the standard ordered basis for R^2. Then $B_2 = \{(0,1),(1,0)\}$ is another ordered basis for R^2. Note that even though elements (vectors) in B_1 and B_2 are the same, they are different in the order sense.

1.5.1 Representation of a vector in the matrix form relative to an ordered basis

Consider a vector $u = (u_1, u_2) \in R^2$. Let $B_1 = \{(1,0),(0,1)\}$ and $B_2 = \{(0,1),(1,0)\}$. In terms of the basis $B_1 = \{(1,0),(0,1)\}$, $u = (u_1, u_2)$ can be written as the linear combination, $u = (u_1, u_2) = u_1(1,0) + u_2(0,1)$, where u_1 and u_2 are scalars called co-ordinates or coefficients of u relative to B_1. We represent the vector $u = (u_1, u_2)$ by placing the coordinates in the column matrix $\begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$ which is called the co-ordinate or coefficient matrix of u relative to the ordered basis B_1 and is denoted by $[u]_{B_1}$. Hence

$$[u]_{B_1} = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \tag{1.4}$$

It can be seen that the co-ordinate matrix of $u = (u_1, u_2)$ in R^2 relative to the ordered basis $B_2 = \{(0,1),(1,0)\}$ is given by

$$[u]_{B_2} = \begin{pmatrix} u_2 \\ u_1 \end{pmatrix} \tag{1.5}$$

From 1.4 and 1.5, it is to be noted that co-ordinate matrix of a vector depends upon the choice of the ordered basis. In general, suppose V is a finite dimensional vector space over a filed F and $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ be an ordered basis for V. Let $\alpha \in V$, then there is a unique n-tuple $(x_1, x_2, \ldots, x_n)$ of scalars such that $\alpha = x_1\alpha_1 + x_2\alpha_2 + \ldots + x_n\alpha_n$ Then the co-ordinate matrix of α relative to the basis $\mathscr{B}$ is $[\alpha]_B = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$

Example 17. Consider R^3 with standard basis $(1,0,0)$, $(0,1,0)$ and $(0,0,1)$. Then $(3,2,-1) = 3(1,0,0) + 2(0,1,0) - 1(0,0,1)$. Then the coordinate matrix

of the vector $(3, 2, -1)$ relative to the standard ordered basis is given by
$$\begin{pmatrix} 3 \\ 2 \\ -1 \end{pmatrix}$$

Theorem 1.7. *Let V be an n-dimensional vector space over the field F, and let $\mathscr{B}$ and $\mathscr{B}'$ be two ordered basis of V. Then there is a unique, necessarily invertible, $n \times n$ matrix P with entries in F such that*

$$[\alpha]_{\mathscr{B}} = P[\alpha]_{\mathscr{B}'} \tag{1.6}$$

$$[\alpha]_{\mathscr{B}'} = P^{-1}[\alpha]_{\mathscr{B}} \tag{1.7}$$

for every vector $\alpha \in V$. The columns of P are given by $P_j = [\alpha'_j]_{\mathscr{B}}$, $j = 1, 2, \ldots, n$.

Proof. Suppose that $\mathscr{B} = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ and $\mathscr{B}' = \{\alpha'_1, \alpha'_2, \ldots, \alpha'_n\}$ are two ordered bases for V. Then there are unique scalars P_{ij} such that $\alpha'_j = P_{1,j}\alpha_1 + P_{2,j}\alpha_2 + \ldots + P_{n,j}\alpha_n$

Let $x'_1, x'_2, \ldots, x'_n$ be the coordinates of a given vector α in the ordered basis $\mathscr{B}'$. Then

$$\begin{aligned}
\alpha &= x'_1\alpha'_1 + x'_2\alpha'_2 + \ldots + x'_n\alpha'_n \\
&= \sum_{j=1}^{n} x'_j\alpha'_j \\
&= \sum_{j=1}^{n} x'_j \sum_{i=1}^{n} P_{ij}\alpha_i \\
&= \sum_{j=1}^{n}\sum_{i=1}^{n} (P_{ij}x'_j)\alpha_i \\
&= \sum_{i=1}^{n} (\sum_{j=1}^{n} P_{ij}x'_j)\alpha_i
\end{aligned}$$

Thus we obtain the relation

$$\alpha = \sum_{i=1}^{n} (\sum_{j=1}^{n} P_{ij} x'_j) \alpha_i$$

Since the coordinates $x_1, x_2, \ldots, x_n$ of α in the ordered basis $\mathscr{B}$ are uniquely determined, it follows that

$$x_i = \sum_{j=1}^{n} P_{ij} x'_j, \quad 1 \leq i \leq n.$$

Let P be the $n \times n$ matrix whose i, j entry is the scalar P_{ij} , and let X and X' be the coordinate matrices of the vector α in the ordered bases $\mathscr{B}$ and $\mathscr{B}'$. Then we get

$$X = PX'$$

Since $\mathscr{B}$ and $\mathscr{B}'$ are linearly independent sets, $X = 0$ if and only if $X' = 0$. This implies that P is invertible. Hence

$$X' = P^{-1}X.$$

Thus

$$[\alpha]_{\mathscr{B}} = P[\alpha]_{\mathscr{B}'}$$

$$[\alpha]_{\mathscr{B}'} = P^{-1}[\alpha]_{\mathscr{B}}.$$

□

Example 18. 1. Let F be a field and let $\alpha = (x_1, x_2, \ldots, x_n)$ be a vector in F^n. If $\mathscr{B}$ is the standard ordered basis of F^n, $\mathscr{B} = \{e_1, e_2, \ldots, e_n\}$, the co-ordinate matrix of the vector α in the basis $\mathscr{B}$ is given by $[\alpha]_{\mathscr{B}} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$, since

$$\alpha = (x_1, x_2, \ldots, x_n) = x_1 e_1 + x_2 e_2 + \ldots + x_n e_n.$$

2. Let R be the field of real numbers and let θ be a fixed real matrix. The matrix $\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$ is invertible with the inverse $P^{-1} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$. Thus for each θ the set $\mathscr{B}'$ consisting of the vectors $(\cos\theta, \sin\theta)$ and $(-\sin\theta, \cos\theta)$ is a basis for R^2. If $\alpha = (x_1, x_2)$, then

$$[\alpha]_{\mathscr{B}'} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

3. Let F be a subfield of the complex numbers. The matrix

$$\begin{pmatrix} -1 & 4 & 5 \\ 0 & 2 & -3 \\ 0 & 0 & 8 \end{pmatrix}$$

is invertible with inverse

$$P^{-1} = \begin{pmatrix} -1 & 2 & \frac{11}{8} \\ 0 & \frac{1}{2} & \frac{3}{16} \\ 0 & 0 & \frac{1}{8} \end{pmatrix}$$

Thus the column vectors of P $\mathscr{B}' = \{(-1, 0, 0), (4, 2, 0), (5, -3, 8)\}$ form a basis of F^3. The coordinates x_1', x_2', x_3' of the vector $\alpha = (x_1, x_2, x_3)$ in the basis $\mathscr{B}'$ are given by $\begin{pmatrix} x_1' \\ x_2' \\ x_3' \end{pmatrix} = P^{-1} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} -x_1 + 2x_2 + \frac{11}{8}x_3 \\ \frac{1}{2}x_2 + \frac{3}{16}x_3 \\ \frac{1}{8}x_3 \end{pmatrix}$.

4. If $\mathscr{B}$ is the standard basis for R^3 and $\mathscr{B}' = \{(1, 1, 0), (1, 0, 1), (0, 1, 1)\}$ be another ordered basis for R^3, then find P such that $[\alpha]_{\mathscr{B}} = P[\alpha]_{\mathscr{B}'}$.

   Solution

   We have

$$\mathscr{B} = \{(1, 0, 0), (0, 1, 0), (0, 0, 1)\} = \{\alpha_1, \alpha_2, \alpha_3\}$$

and

$$\mathscr{B}' = \{(1,1,0),(1,0,1),(0,1,1)\} = \{\alpha_1', \alpha_2', \alpha_3'\}$$

We can write

$$\alpha_1' = (1,1,0) = 1(1,0,0) + 1(0,1,0) + 0(0,0,1) = 1.\alpha_1 + 1.\alpha_2 + 0.\alpha_3$$

$$\alpha_2' = (1,0,1) = 1(1,0,0) + 0(0,1,0) + 1(0,0,1) = 1.\alpha_1 + 0.\alpha_2 + 1.\alpha_3$$

$$\alpha_3' = (0,1,1) = 0(1,0,0) + 1(0,1,0) + 1(0,0,1). = 0.\alpha_1 + 1.\alpha_2 + 1.\alpha_3$$

and this implies that

$$P_{11} = 1, P_{12} = 1, P_{13} = 0$$

$$P_{21} = 1, P_{22} = 0, P_{23} = 1$$

$$P_{11} = 0, P_{12} = 1, P_{13} = 1$$

Hence

$$P = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}$$

Here the rows of P are elements of $\mathscr{B}'$ itself because $\mathscr{B}$ is the standard basis.

Exercises

1. Show that the vectors $\alpha_1 = (1,1,0,0)$, $\alpha_2 = (0,0,1,1)$, $\alpha_3 = (1,0,0,4)$, $\alpha_4 = (0,0,0,2)$ form a basis for R^4. Find the coordinates of each of the standard basis vectors in the ordered basis $\{\alpha_1, \alpha_2, \alpha_3, \alpha_4\}$.

2. Find the coordinate matrix of the vector $(1,0,1)$ in the basis of C^3 consisting of the vectors $(2i,1,0), (2,-1,1), (0,1+i,1-i)$ in that order.

3. Let $\mathscr{B} = \{\alpha_1, \alpha_2, \alpha_3\}$ be an ordered basis for R^3 consisting of

$$\alpha_1 = (1,0,-1), \alpha_2 = (1,1,1), \alpha_3 = (1,0,0)$$

What are the coordinates of the vector (a,b,c) in the ordered basis $\mathscr{B}$?

Chapter 2

Linear Transformations

2.1 Linear Transformations

In this section we study linear functions from a vector space into another.

Definition 2.1. *Let V and W be vector spaces over the field F. A linear transformation from V into W is a function T from V into W such that*

$$T(c\alpha + \beta) = c(T(\alpha)) + T(\beta)$$

for all α and β in V and all scalars $c \in F$.

Example 19.

1. If V is any vector space, the identity transformation I defined by $I(\alpha) = \alpha$ is a linear transformation from V into V. The zero transformation 0, defined by $0.\alpha = 0$, is a linear transformation from V into V.

2. Let F be a field and let V be the space of polynomial functions f from F into F, given by

$$f(x) = c_0 + c_1x + \ldots + c_kx^k.$$

Let

$$Df(x) = c_1 + 2c_2x + \ldots + kc_kx^{k-1}.$$

Let $f, g \in V$ and $c \in F$. Take $f(x) = c_0 + c_1x + \ldots + c_kx^k$ and $g(x) = a_0 + a_1x + a_2x^2 + \ldots + a_kx^k$. Then

$$\begin{aligned}
D(cf+g)(x) &= D[(cc_0+a_0) + (cc_1+a_1)x + (cc_2+a_2)x^2 + \ldots + (cc_k+a_k)x^k] \\
&= (cc_1+a_1) + 2(cc_2+a_2)x + \ldots + k(cc_k+a_k)x^{k-1} \\
&= cc_1 + 2cc_2x + \ldots + kcc_kx^{k-1} + a_1 + 2a_2x + \ldots + ka_kx^{k-1} \\
&= c(c_1 + 2c_2x + \ldots + kc_kx^{k-1}) + a_1 + 2a_2x + \ldots + ka_kx^{k-1} \\
&= c.D(f(x)) + D(g(x)) = cDf + Dg.
\end{aligned}$$

That is D is a linear transformation from V into V- the differentiation transformation.

3. Let A be a fixed $m \times n$ matrix with entries in the field F. The function T defined by $T(X) = AX$ is a linear transformation from $F^{n\times 1}$ into $F^{m\times 1}$. The function U defined by $U(\alpha) = \alpha A$ is a linear transformation from F^m into F^n.

4. Let R be the field of real numbers and let V be the space of all functions from R into R which are continuous. Define T by $(Tf)(x) = \int_0^x f(t)dt$. Then for $f, g \in V$ and $c \in F$, we have $T(cf+g)(x) = \int_0^x (cf+g)(t)dt = \int_0^x cf(t)dt + \int_0^x g(t)dt = c. \int_0^x f(t)dt + \int_0^x g(t)dt = c.Tf(x) + Tg(x) = (c.Tf + Tg)(x)$. That is T is a linear transformation from V into V.

Remark 4. 1. If T is a linear transformation from V into W, then $T(0) = 0$.

Proof. Suppose T is a linear transformation. Then $T(0) = T(0+0) = T(0) + T(0) \Rightarrow T(0) - T(0) = T(0) + T(0) - T(0) \Rightarrow T(0) = 0$. $\square$

2. Linear transformation preserves linear combinations. That is if $\alpha_1, \ldots, \alpha_n$ are vectors in a vector space V and $c_1, \ldots, c_n$ are scalars, then

$$T(c_1\alpha_1 + \ldots + c_n\alpha_n) = c_1T(\alpha_1) + \ldots + c_nT(\alpha_n).$$

Theorem 2.1. *Let V be a finite-dimensional vector space over the field F and let $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ be an ordered basis for V. Let W be a vector space over the same field and let $\beta_1, \beta_2, \ldots, \beta_n$ be any vectors in W. Then there is precisely one linear transformation T from V into W such that $T(\alpha_j) = \beta_j$, $j = 1, 2, \ldots, n$.*

Proof. To prove there is some linear transformation T with $T(\alpha_j) = \beta_j$, we proceed as follows. Given $\alpha \in V$, there is a unique n-tuple $(x_1, \ldots, x_n)$ such that

$$\alpha = x_1\alpha_1 + \ldots + x_n\alpha_n,$$

since $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ is an ordered basis for V. For this vector α, we define

$$T(\alpha) = x_1\beta_1 + \ldots + x_n\beta_n.$$

Then T is a well defined rule(Since $x_i \in F$ and $\beta_i \in W \Rightarrow x_i\beta_i \in W$ for $i = 1, 2, \ldots, n \Rightarrow T(\alpha) = x_1\beta_1 + \ldots + x_n\beta_n \in W$.) Now $T(\alpha_j) = \beta_j$ for each j, since $\alpha_j = 0.\alpha_1 + 0.\alpha_2 + \ldots + 1.\alpha_j + \ldots + 0.\alpha_n$.

To prove that T is a linear transformation, let $\beta = y_1\alpha_1 + \ldots + y_n\alpha_n \in V$ and let c be any scalar. Then to prove that $T(c\alpha + \beta) = c.T(\alpha) + T(\beta)$. We can write $c\alpha + \beta = c(x_1\alpha_1 + \ldots + x_n\alpha_n) + y_1\alpha_1 + \ldots + y_n\alpha_n = (cx_1 + y_1)\alpha_1 + \ldots + (cx_n + y_n)\alpha_n$. Then

$$T(c\alpha + \beta) = (cx_1 + y_1)\beta_1 + \ldots + (cx_n + y_n)\beta_n \tag{2.1}$$

On the other hand $c(T\alpha) + T\beta = c(x_1\beta_1 + \ldots + x_n\beta_n) + y_1\beta_1 + \ldots + y_n\beta_n =$

$(cx_1 + y_1)\beta_1 + \ldots + (cx_n + y_n)\beta_n)$. That is

$$c(T\alpha) + T\beta = (cx_1 + y_1)\beta_1 + \ldots + (cx_n + y_n)\beta_n) \tag{2.2}$$

From equation 2.1 and 2.2, we get T is a linear transformation.

Next we have to prove that T is unique. Assume that U is a linear transformation from V into W with $U(\alpha_j) = \beta_j$, $j = 1, 2, \ldots, n$. Then for a vector $\alpha = x_1\alpha_1 + \ldots + x_n\alpha_n$, we have

$$\begin{aligned} U(\alpha) &= U(x_1\alpha_1 + \ldots + x_n\alpha_n) \\ &= x_1U(\alpha_1) + \ldots + x_nU(\alpha_n) \\ &= x_1\beta_1 + \ldots + x_n\beta_n. \end{aligned}$$

This means that U is exactly the same rule T which we defined above. This shows that the linear transformation T with $T(\alpha_j) = \beta_j$ is unique. □

Example 20. 1. The vectors

$$\alpha_1 = (1, 2)$$

$$\alpha_2 = (3, 4)$$

are linearly independent and therefore form a basis for R^2. According to Theorem 2.1, there is a unique linear transformation from R^2 into R^3 such that

$$T(\alpha_1) = (3, 2, 1)$$

$$T(\alpha_2) = (6, 5, 4).$$

If so we must able to find $T(1, 0)$. For that write $(1, 0)$ as a linear combination of α_1 and α_2. And then apply T. That is if $(1, 0) = a(1, 2) + b(3, 4)$, $a + 3b = 1$ and $2a + 4b = 0$. That is $a + 2b = 0 \Rightarrow a =$

$-2b$. Then $b = 1 \Rightarrow a = -2$. Thus $(1,0) = -2(1,2) + (3,4)$. Thus By the method in Theorem 2.1, we get

$$\begin{aligned} T(1,0) &= T(-2(1,2) + (3,4)) \\ &= -2T(1,2) + T(3,4) \\ &= -2(3,2,1) + (6,5,4) \\ &= (0,1,2). \end{aligned}$$

2. Let T be a linear transformation from the m-tuple space F^m into the space n-tuple space F^n. Theorem 2.1, tells us that T is uniquely determined by the sequence of vectors $\beta_1, \beta_2, \ldots, \beta_m$ where $\beta_i = Te_i$, $i = 1, 2, \ldots, m$. In short T is uniquely determined by the images of the standard basis vectors. That is $\alpha = (x_1, \ldots, x_m) = x_1 e_1 + \ldots + x_m e_m$. Then

$$\begin{aligned} T(\alpha) &= T(x_1 e_1 + \ldots + x_m e_m) \\ &= x_1 T(e_1) + x_2 T(e_2) + \ldots + x_m T(e_m) \\ &= x_1 \beta_1 + x_2 \beta_2 + \ldots + x_m \beta_m. \end{aligned}$$

If B is the $m \times n$ matrix which has row vectors $\beta_1, \beta_2, \ldots, \beta_m$. If $\beta_i = (B_{i1}, \ldots, B_{in})$, then $T(x_1, \ldots, x_n) = [x_1 \cdots x_m] = \begin{pmatrix} B_{11} & \cdots & B_{1n} \\ \vdots & & \vdots \\ B_{m1} & \cdots & B_{mn} \end{pmatrix}$

2.1.1 Rank Nullity Theorem

Definition 2.2. *Let V and W be vector spaces over the field F and let T be a linear transformation from V into W. The null space of T is the set of all vectors $\alpha \in V$ such that $T(\alpha) = 0$.*

If V is finite dimensional, then the rank of T is the dimension of the range of T and the nullity of T is the dimension of the null space of T.

Remark 5. Let V and W be two vector spaces over the field F and let T be a linear transformation from V into W. Then null space is a subspace of V. For, let $c \in F$, $\alpha, \beta \in N$, then $T(c\alpha+\beta) = c.T(\alpha)+T(\beta) = 0 \Rightarrow c\alpha+\beta \in N$.

The following is one of the most important results in linear algebra.

Theorem 2.2. *Let V and W be vector spaces over the field F and let T be a linear transformation from V into W. Suppose that V is finite dimensional. Then*
rank(T)+nullity(T)=dim V.

Proof. We know that N , the null space of T , is a subspace of V. Since V is finite dimensional, its subspace N is also finite dimensional and has a basis consisting of finite number of elements. Let $\{\alpha_1, \ldots, \alpha_k\}$ be a basis for the null space N of T. There are vectors $\alpha_{k+1}, \ldots, \alpha_n$ in V such that $\{\alpha_1, \ldots, \alpha_k, \alpha_{k+1}, \ldots, \alpha_n\}$ is a basis for V.
Claim: $\{T\alpha_{k+1}, \ldots, T\alpha_n\}$ is a basis for the range of T.
(i) $T\alpha_{k+1}, \ldots, T\alpha_n$ spans range of T.
Clearly the vectors $\{T\alpha_{k+1}, \ldots, T\alpha_n\}$ spans the range of T. Since $T(\alpha_j) = 0$ for $j \leq k$, we see that $T\alpha_{k+1}, \ldots, T\alpha_n$ span the range. To see that these vectors are independent, suppose we have scalars c_i such that $c_{k+1}T(\alpha_{k+1}) + \ldots + c_nT(\alpha_n) = 0$. This says that

$$T(c_{k+1}\alpha_{k+1} + \ldots c_n\alpha_n) = 0.$$

and accordingly the vector $\alpha = c_{k+1}\alpha_{k+1}+\ldots+c_n\alpha_n$ is in the null space of T. Since $\alpha_1, \ldots, \alpha_k$ forms a basis of N, there must be scalars $b_1, b_2, \ldots, b_k$ such that $\alpha = b_1\alpha_1+\ldots+b_k\alpha_k$. Thus $b_1\alpha_1+\ldots+b_k\alpha_k-(c_{k+1}\alpha_{k+1}+\ldots+c_n\alpha_n) = 0$

and since $\alpha_1, \ldots, \alpha_n$ are linearly independent we must have $b_1 = \cdots = b_k = c_{k+1} = \cdots = c_n = 0$. If r is the rank of T, the fact that $\{T(\alpha_{k+1}), \ldots, T(\alpha_n)\}$ forms a basis for the range of T tells us that $r = n - k$. Here k is the nullity of T and n is the dimension of V. Thus we get rank(T)+nullity(T)=dim V. $\square$

Example 21. Find the rank and nullity of the linear transformation $T : R^3 \to R^3$ defined by $T(x_1, x_2, x_3) = (x_1 + x_3, 2x_1 + x_2, x_1 + 2x_3)$.
Solution
We have from the definition of T, $T(1,0,0) = (1,2,1)$, $T(0,1,0) = (0,1,0)$ and $T(0,0,1) = (1,0,2)$. Thus

$$A = [T]_{\mathscr{B}} = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 2. \end{pmatrix}$$

By performing elementary transformations, we can see that Rank $A = 3$. Hence Rank of T is 3 and Nullity of T = dim $V-$ rank $T = 3 - 3 = 0$.

Example 22. Does there exist a linear operator T on a vector space V having dimension 7 such that the dimension of the range space and the null space are the same.
Solution
Suppose there exists such a closure operator. By rank-nullity theorem, we have rank T + nullity T = dim V. the dimension of the range space and the null space are the same, then we obtain,

$$\begin{aligned} RankT + NullityT &= dimV \\ \Rightarrow RankT + RankT &= 7 \\ \Rightarrow 2\ RankT &= 7 \end{aligned}$$

$$\Rightarrow RankT = 3.5.$$

which is not possible as the dimension of the range space (in general dimension of any finite dimensional vector space) must be an integer. So There does not exists a linear operator with rank 7.

Theorem 2.3. *If A is an $m \times n$ matrix with entries in the field F, then row rank (A)=column rank (A).*

Proof. Let T be a linear transformation from $F^{n\times 1}$ into $F^{m\times 1}$ defined by $T(X) = AX$. The null space of T is the solution space for the system $AX = 0$, that is the set of all column matrices X such that $AX = 0$. The range of T is the set of all $m \times 1$ column matrices Y such that $AX = Y$ has a solution for X. If $A_1, \ldots, A_n$ are the columns of A, then

$$AX = x_1A_1 + \ldots + x_nA_n.$$

so that range of T is the subspace of spanned by the columns of A. In other words, the range of T is the column space of A. Therefore $rank(T) = columnrank(A)$. Previous Theorem(Rank Nullity Theorem) tells us that if S the solution space of the system $AX = 0$, then

$$\dim S + \text{ column rank } (A) = n. \tag{2.3}$$

If r is the dimension of the row space of A, then the solution space S has a basis consisting of $n - r$ vectors:

$$\dim S = n - \text{ row rank } (A). \tag{2.4}$$

Thus from equations 2.3 and 2.4, we get

$$\text{row rank } (A) = \text{ column rank } (A).$$

□

2.1.2 Algebra of Linear Transformations

Theorem 2.4. *Let V and W be vector spaces over the field F. Let T and U be linear transformations from V into W. The function $(T+U)$ defined by $(T+U)(\alpha) = T(\alpha) + U(\alpha)$ is a linear transformation from V into W. If c is any elements of F, the function cT defined by $(cT)(\alpha) = c(T(\alpha))$ is a linear transformation from V into W. The set of all linear transformations from V into W, together with the addition and scalar multiplication defined above, is a vector space over the field F.*

Proof. Suppose T and U are linear transformations from V in to W and that we define as above. Then

$$\begin{aligned}
(T+U)(c\alpha+\beta) &= T(c\alpha+\beta) + U(c\alpha+\beta) \\
&= c(T\alpha) + T\beta + c(U\alpha) + U\beta \\
&= c(T\alpha + U\alpha) + T\beta + U\beta \\
&= c(T+U)(\alpha) + (T+U)(\beta).
\end{aligned}$$

which shows that $(T+U)$ is a linear transformation. Similarly

$$\begin{aligned}
(cT)(d\alpha+\beta) &= cT(d\alpha+\beta) \\
&= c[dT(\alpha) + T(\beta)] \\
&= cdT(\alpha) + c(T\beta) \\
&= d[cT\alpha] + cT\beta \\
&= d[(cT)(\alpha)] + (cT)\beta
\end{aligned}$$

which shows that cT is a linear transformation. Thus we have sum of two linear transformations and scalar multiple of a linear transformation are linear. Now one can prove conditions on the vector addition and scalar multiplication. $\square$

Denote the space of linear transformations from V into W by $L(V, W)$. Note that $L(V, W)$ is defined only when V and W are vector spaces over the same field.

Theorem 2.5. *Let V be an n-dimensional vector space over the field F, and let W be an m-dimensional vector space over F. Then the space $L(V, W)$ is finite dimensional and has dimension mn.*

Proof. Let $\mathscr{B} = \{\alpha_1, \ldots, \alpha_n\}$ and $\mathscr{B}' = \{\beta_1, \ldots, \beta_m\}$ be ordered bases for V and W respectively. For each pair of integers (p, q) with $1 \leq p \leq m$ and $1 \leq q \leq n$, we define a linear transformation $E^{p,q}$ from V into W by

$$E^{p,q}(\alpha_i) = \begin{cases} 0 & \text{if } i \neq q \\ \beta_p, & \text{if } i = q \end{cases} = \delta_{iq}\beta_p$$

By Theorem 2.1, we have a unique linear transformation from V into W satisfying these conditions. The claim is that the mn transformations $E^{p,q}$ form a basis for $L(V, W)$.

Let T be a linear transformation from V into W. For each j, $1 \leq j \leq n$, let $A_{ij}, \ldots, A_{mj}$ be the coordinates of the vector $T\alpha_j$ in the ordered basis $\mathscr{B}'$, i.e.,

$$T(\alpha_j) = \sum_{p=1}^{m} A_{pj}\beta_p. \tag{2.5}$$

We want to show that

$$T = \sum_{p=1}^{m}\sum_{q=1}^{n} A_{pq}E^{p,q} \tag{2.6}$$

Let U be the linear transformation in the right hand member of 2.6. Then for each j,

$$\begin{aligned}
U\alpha_j &= \sum_p \sum_q A_{pq} E^{p,q}(\alpha_j) \\
&= \sum_p \sum_q A_{pq} \delta_{jq} \beta_p \\
&= \sum_{p=1}^{m} A_{pj} \beta_p \\
&= T\alpha_j
\end{aligned}$$

and consequently $U = T$. Now by equation 2.6, we get $E^{p,q}$ span $L(V, W)$. Next we prove that $E^{p,q}$ are independent. If $\sum\limits_{p=1}^{m}\sum\limits_{q=1}^{n} A_{pq}E^{p,q} = 0$, then $U\alpha_j = 0$, for each j, so $\sum\limits_{p=1}^{m} A_{pj}\beta_p = 0$ and the independence of the β_p implies that $A_{pj} = 0$ for every p and j. □

Theorem 2.6. *Let V, W, and Z be vector spaces over the field F. Let T be a linear transformation from V into W and U a linear transformation from W into Z. Then the composed function UT defined by $UT(\alpha) = U(T(\alpha))$ is a linear transformation from V into Z.*

Proof.

$$\begin{aligned}
UT(c\alpha + \beta) &= U(T(c\alpha + \beta)) \\
&= U(cT\alpha + T\beta) \\
&= c(U(T\alpha)) + U(T\beta) \\
&= c(UT)(\alpha) + (UT)(\beta)
\end{aligned}$$

□

Note that a linear transformation from a vector space V to itself is called a linear operator on V.

Definition 2.3. *If V is a vector space over the field F, a linear operator on V is a linear transformation from V into V.*

Lemma 2.1. *Let V be a vector space over the field F, let U, T_1 and T_2 be linear operator on V, let c be an element of F.*

1. $IU = UI = U$

2. $U(T_1 + T_2) = UT_1 + UT_2$; $(T_1 + T_2)U = T_1U + T_2U$

3. $c(UT_1) = (cU)T_1 = U(cT_1)$.

Proof. 1. $IU(\alpha) = I(U(\alpha)) = U(\alpha)$. Similarly $UI(\alpha) = U(I(\alpha)) = U(\alpha)$. Thus $IU = UI = U$.

2.

$$\begin{aligned} U(T_1 + T_2)(\alpha) &= U((T_1 + T_2)(\alpha)) \\ &= U(T_1(\alpha) + T_2(\alpha)) \\ &= U(T_1(\alpha)) + U(T_2(\alpha)) \\ &= (UT_1)(\alpha) + (UT_2)(\alpha). \end{aligned}$$

$$\begin{aligned} [(T_1 + T_2)U](\alpha) &= (T_1 + T_2)(U(\alpha)) \\ &= T_1(U(\alpha)) + T_2(U(\alpha)) \\ &= (T_1U)(\alpha) + (T_2U)(\alpha). \end{aligned}$$

Thus $(T_1 + T_2)U = T_1U + T_2U$.

3. $c(UT_1)(\alpha) = c(U(T_1(\alpha))) = (cU)(T_1(\alpha)) = (cU)(T_1)(\alpha)$.

□

Remark 6. The vector space $L(V, V)$ together with the composition operation, is known as Linear algebra with identity.

Example 23.

If A is an $m \times n$ matrix with entries in F, we have the linear transformation T defined by $T(X) = AX$ from $F^{n\times 1}$ into $F^{m\times 1}$. If B is a $p \times m$ matrix, we have the linear transformation U from $F^{m\times 1}$ into $F^{p\times 1}$ defined by $U(Y) = BY$. The composition UT is easily described:

$$(UT)(X) = U(T(X)) = U(AX) = B(AX) = (BA)X. \tag{2.7}$$

Thus UT is left multiplication by the product matrix BA.

Let F be a field and V the vector space of all polynomial functions from F into F. Let D be the differentiation operator defined by Example 2. Let T be the linear operator multiplication by x:

$$(Tf)(x) = xf(x).$$

Then $DT \neq TD$. For, take $f(x) = a + bx + cx^2$, then

$$\begin{aligned} DT(f)(x) &= D(T(a + bx + cx^2)) \\ &= D(x(a + bx + cx^2)) \\ &= D(ax + bx^2 + cx^3) \\ &= a + 2bx + 3cx^2. \end{aligned}$$

But

$$\begin{aligned} (TD)f(x) &= T(D(a + bx + cx^2)) \\ &= T(b + 2cx) \\ &= x(b + 2cx) \\ &= bx + 2cx^2 \\ &\neq DT(f(x)). \end{aligned}$$

Exercises

1. Which of the following functions T from R^2 to R^2 are linear transformations?
   (a) $T(x_1, x_2) = (1 + x_1, x_2)$
   (b) $T(x_1, x_2) = (x_2, x_1)$
   (c) $T(x_1, x_2) = (x_1^2, x_2)$
   (d) $T(x_1, x_2) = (\sin x_1, x_2)$
   (e) $T(x_1, x_2) = (x_1 - x_2, 0)$
2. Find the range, rank, null space, and nullity for the zero transformation and the identity transformation of a finite dimensional space V.
3. Describe the range and null space for the differentiation transformation on a finite dimensional space V.
4. Is there a linear transformation T from R^3 to R^2 such that $T(1, -1, 1) = (1, 0)$ and $T(1, 1, 1) = (0, 1)$?

2.2 Invertible Linear Operators

Next we are going to discuss which linear operators T on the space V does there exist a linear operator T^{-1} such that $TT^{-1} = T^{-1}T = I$?

Definition 2.4. *A function $T : V \to W$ is called invertible if there exists a function U from W into V such that UT is the identity function on V and TU is the identity function on W. If T is invertible, the function U is unique and is denoted by T^{-1}.*
T is invertible if and only if

1. *T is one-one ($T\alpha = T\beta \Rightarrow \alpha = \beta$;)*
2. *T is onto (Range of $T = W$.)*

Theorem 2.7. *Let V and W be vector spaces over the field F and let T be a linear transformation from V into W. If T is invertible, then the inverse function T^{-1} is a linear transformation from W onto V.*

Proof. Assume that T is invertible. Then there exists an inverse function T^{-1} from W onto V such that $T^{-1}T$ is the identity function on V and TT^{-1} is the identity function on W. To prove that T^{-1} is linear.
Let $\beta_1, \beta_2 \in W$ and let c be a scalar. $\beta_1, \beta_2 \in W$ and T is onto implies that there exists $\alpha_1, \alpha_2 \in V$ such that $T(\alpha_1) = \beta_1$ and $T(\alpha_2) = \beta_2$. Since T is linear, $T(c\alpha_1 + \alpha_2) = cT(\alpha_1) + T(\alpha_2) = c\beta_1 + \beta_2$ which implies that $T^{-1}(c\beta_1 + \beta_2) = c\alpha_1 + \alpha_2$. Now $T(\alpha_1) = \beta_1$ and $T(\alpha_2) = \beta_2$ implies that $T^{-1}(\beta_1) = \alpha_1$ and $T^{-1}(\beta_2) = \alpha_2$ respectively. Thus we get

$$T^{-1}(c\beta_1 + \beta_2) = cT^{-1}(\beta_1) + T^{-1}(\beta_2).$$

Thus T^{-1} is linear. □

Example 24. Find the inverse of the linear operator T on 3 defined by $T(x_1, x_2, x_3) = (3x_1, x_2 - x_1, 2x_1 + x_2 - x_3)$.
Solution
$T(x_1, x_2, x_3) = (y_1, y_2, y_3)$. Then

$$3x_1 = y_1, \tag{2.8}$$

$$x_2 - x_1 = y_2, \tag{2.9}$$

$$2x_1 + x_2 - x_3 = y_3 \tag{2.10}$$

Now from equation(2.8), we get

$$x_1 = \frac{y_1}{3}.$$

And from equation(2.9),

$$x_2 = y_2 + x_1 = y_2 + (\frac{y_1}{3}).$$

Finally from equation (2.10),

$$x_3 = y_3 - 2x_1 - x_2 = y_3 - 2(\frac{y_1}{3}) - (y_2 + (\frac{y_1}{3})) = y_3 - y_2 - y_1.$$

$$T^{-1}(y_1, y_2, y_3) = (\frac{y_1}{3}, y_2 + (\frac{y_1}{3}), y_3 - y_2 - y_1).$$

Remark 7.

Suppose we have an invertible linear transformation T from V onto W and an invertible linear transformation U from W onto Z. Then UT is invertible and $(UT)^{-1} = T^{-1}U^{-1}$.

If T is linear, then $T(\alpha - \beta) = T\alpha - T\beta$. This implies that $T\alpha = T\beta$ if and only if $T(\alpha - \beta) = 0$.

Definition 2.5. *A linear transformation T is called non-singular if $T(\alpha) = 0$ implies that $\alpha = 0$. That is if null space of T is $\{0\}$.*

Note that T is one-one if and only if T is non-singular.

Theorem 2.8. *Let T be a linear transformation from V into W. Then T is non-singular if and only if T carries each linearly independent subset of V onto a linearly independent subset of W.*

Proof. First assume that T is non-singular. Let S be a linearly independent subset of V. If $\alpha_1, \ldots, \alpha_k$ are vectors in S, then

$$\begin{aligned} c_1T(\alpha_1) + \ldots + c_kT(\alpha_k) = 0 \ &\Rightarrow \ T(c_1\alpha_1 + \ldots + c_k\alpha_k) = 0. \\ &\Rightarrow \ c_1\alpha_1 + \ldots + c_k\alpha_k = 0, \end{aligned}$$

since T is non-singular. Then each $c_i = 0$, since $\alpha_1, \ldots, \alpha_k$ are linearly independent vectors. Thus $T\alpha_1, \ldots, T\alpha_k$ are linearly independent. This means that T carries linearly independent set into a linearly independent set.

Conversely assume that T carries linearly independent set into a linearly independent set. To prove that T is non-singular. Let α be a non zero vector in V. Then the set S consisting of one vector α is independent. The image of S is the set consisting of one vector $T\alpha$ is independent by our assumption.

Hence $T\alpha \neq 0$(since a set consisting of the zero vector alone is dependent). This shows that the null space of T is the zero subspace. That is T is non-singular. $\square$

Example 25. Let F be a field and let T be the linear operator on F^2 defined by $T(x_1, x_2) = (x_1 + x_2, x_1)$. Then $T(x_1, x_2) = 0 \Rightarrow (x_1 + x_2, x_1) = (0, 0) \Rightarrow x_1 = 0$ and $x_2 = 0$. That means T is non-singular. Thus T is one-one. Now we prove that T is onto. Let $z_1, z_2 \in F$ such that

$$x_1 + x_2 = z_1$$

and

$$x_1 = z_2$$

and solution is $x_1 = z_2$ and $x_2 = z_1 - z_2$. Both values of x_1 and x_2 lies in F. Means that T is onto. Thus T is invertible and its inverse is given by $T^{-1}(z_1, z_2) = (z_2, z_1 - z_2)$.

Theorem 2.9. *Let V and W be finite dimensional vector spaces over the field F such that dimV=dimW. If T is a linear transformation from V into W, the following are equivalent:*

1. *T is invertible.*
2. *T is non-singular*
3. *T is onto, that is Range of T is W.*

Proof. Let $n =$ dim $V =$ dim W. From Rank nullity Theorem, rank $T+$ nullity T=dim V=n. Now T is non-singular if and only if nullity $(T) = 0$, then rank $T = n$, that is if and only if Range of T is W. Thus T is non-singular if and only if T is onto. So if either condition (ii) or (iii) holds, the other is satisfied as well and T is invertible. $\square$

Theorem 2.10. *Let V and W be finite dimensional vector spaces over the field F such that dim$V$=dim$W$. If T is a linear transformation from V into W, the following are equivalent:*

1. *T is invertible.*
2. *T is non-singular*
3. *T is onto, that is Range of T is W.*
4. *If $\{\alpha_1, \ldots, \alpha_n\}$ is a basis for V, then $\{T\alpha_1, \ldots, T\alpha_n\}$ is a basis for W.*
5. *There is some basis $\{\alpha_1, \ldots, \alpha_n\}$ for V such that $\{T\alpha_1, \ldots, T\alpha_n\}$ is a basis for W.*

Proof. $(i) \Rightarrow (ii)$
If T is invertible, then T is non-singular.
$(ii) \Rightarrow (iii)$
Suppose T is non-singular. If $\{\alpha_1, \ldots, \alpha_n\}$ is a basis for V, they are linearly independent vectors. Since T is non-singular, it maps linearly independent set to linearly independent set. Now dimension of W is n, this set of vectors is a basis of W. Now let $\beta \in W$. There exists scalars $c_1, \ldots, c_n$ such that

$$\beta = c_1T\alpha_1 + \ldots + c_nT\alpha_n = T(c_1\alpha_1 + \ldots + c_n\alpha_n)$$

Here $c_1\alpha_1 + \ldots + c_n\alpha_n \in V$. Thus β belongs to the range of T. Means that T is onto.
$(iii) \Rightarrow (iv)$
Assume T is onto. If $\{\alpha_1, \ldots, \alpha_n\}$, is any basis for V, the vectors $\{T\alpha_1, \ldots, T\alpha_n\}$ span the range of T, which is all of W by assumption. Since the dimension of W is n, these n vectors must be linearly independent, that is must comprise a basis a basis for W. (iv) clearly implies (v). (v) implies (i). Suppose there

is some basis $\{\alpha_1, \ldots, \alpha_n\}$ for V such that $\{T\alpha_1, \ldots, T\alpha_n\}$ is a basis for W. To prove that T is one-one and onto. Since $\{T\alpha_1, \ldots, T\alpha_n\}$ span W and $\dim W = n$, it is clear that the range of T is all of W. Means that T is onto. If $\alpha = c_1\alpha_1 + \ldots + c_n\alpha_n$ is in the null space of T, then

$$T(c_1\alpha_1 + \ldots + c_n\alpha_n) = 0$$

or

$$c_1(T\alpha_1) + \ldots + c_nT\alpha_n = 0$$

and since $\{T\alpha_1, \ldots, T\alpha_n\}$ is a basis, it is linearly independent. Therefore each $c_i = 0$ and thus $\alpha = 0$. Means that T is non-singular or one-one. Thus T is invertible. □

Exercises

1. Let T be a unique linear operator on C^3 for which $T(e_1) = (1, 0, i), T(e_2) = (0, 1, 1)$, $T(e_3) = (i, 1, 0)$. Is T invertible?

2. Let T be a linear operator on R^3defined by $T(x_1, x_2, x_3) = (3x_1, x_1 - x_2, 2x_1 + x_2 + x_3)$. Is T invertible? If so, find a rule for T^{-1} like one which define T.

3. Find two linear operators T and U on R^2 such that $TU = 0$ but $UT \neq 0$.

2.3 Isomorphism

Definition 2.6. *If V and W are vector spaces over the field F, any one-one linear transformation T of V onto W is called an isomorphism of V onto W. If there exists an isomorphism of V onto W, we say that V is isomorphic to W.*

Theorem 2.11. *Every n-dimensional vector space over the field F is isomorphic to the space F^n.*

Proof. Let V be an n-dimensional space over the field F and let $\mathscr{B} = \{\alpha_1, \ldots, \alpha_n\}$ be an ordered basis for V. We define a function from V into F^n as follows: If $\alpha \in V$, let $T\alpha$ be the n-tuple $(x_1, \ldots, x_n)$ of coordinates of α relative to the ordered basis $\mathscr{B}$. Then we have verified that T is linear, one-one, and maps V onto F^n. □

Exercises

1. Let V be the set of complex numbers and let F be the field of real numbers. With the usual operations, V is a vector space over F. Describe explicitly an isomorphism of this space onto R^2.

2. Show that $F^{m\times n}$ is isomorphic to F^{mn}.

3. Let V and W be finite dimensional vector spaces over the field F. Prove that V and W are isomorphic if and only if $dim\ V = dim\ W$.

2.4 Representation of Transformations by matrices

Let V be an $n-$dimensional vector space over the field F and let W be an $m-$dimensional vector space over F. Let $\mathscr{B} = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ be an ordered basis for V and $\mathscr{B}' = \{\beta_1, \beta_2, \ldots, \beta_m\}$ be an ordered basis for W. If T is any linear transformation from V into W, then T is determined by its action on the vectors α_j. Each of the n vectors $T(\alpha_j)$ is uniquely expressible as a linear combination

$$T(\alpha_j) = \sum_{i=1}^{m} A_{ij}\beta_i. \tag{2.11}$$

of the β_i, the scalars $A_{1j}, \ldots, A_{mj}$ being the coordinates of the $T(\alpha_j)$ in the ordered basis $\mathscr{B}'$. Accordingly, the transformation T is determined by mn scalars A_{ij} via the formula 2.11. The $m \times n$ matrix A defined by $A(i,j) = A_{ij}$ is called the matrix of T relative to the pair of ordered bases $\mathscr{B}$ and $\mathscr{B}'$. Let's look at how the matrix A determines the linear transformation T.

Consider a vector $\alpha \in V$, then we can find scalars $x_1, \ldots, x_n$ such that $\alpha = x_1\alpha_1 + \ldots x_n\alpha_n$. Then

$$T(\alpha) \;=\; T(x_1\alpha_1 + \ldots x_n\alpha_n)$$

$$
\begin{aligned}
&= x_1T(\alpha_1) + \ldots x_nT(\alpha_n) \\
&= x_1(A_{11}\beta_1 + \ldots + A_{1m}\beta_m) + \ldots + x_n(A_{n1}\beta_1 + \ldots + A_{nm}\beta_m) \\
&= x_1\sum_{i=1}^{m} A_{1j}\beta_i + \ldots + x_n\sum_{i=1}^{m} A_{nj}\beta_i \\
&= \sum_{j=1}^{n} x_j \sum_{i=1}^{m} A_{ij}\beta_i \\
&= \sum_{i=1}^{m} \left(\sum_{j=1}^{n} A_{ij}x_j \right) \beta_i
\end{aligned}
$$

Here the scalar $\sum_{i=1}^{n} A_{ij}x_j$, is the entry in the i th row of the column matrix AX. Thus AX is the coordinate matrix of the vector $T(\alpha)$ in the ordered basis $\mathscr{B}'$, if X is the coordinate matrix of α in the ordered basis $\mathscr{B}$.

Also note that if A is any $m \times n$ matrix over the field F, then

$$
T(\sum_{i=1}^{n} x_j\alpha_j) = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} A_{ij}x_j \right) \beta_i \tag{2.12}
$$

defines a linear transformation T from V into W.

Theorem 2.12. *Let V be an n-dimensional vector space over the field F and W be an m-dimensional vector space over F. Let $\mathscr{B}$ be an ordered basis for V and $\mathscr{B}'$ an ordered basis for W. For each linear transformation T from V into W, there is an $m \times n$ matrix A with entries in F such that*

$$
[T\alpha]_{\mathscr{B}'} = A[\alpha]_{\mathscr{B}}
$$

for every vector $\alpha \in V$. Furthermore, the map $T \to A$ is a one-one correspondence between the set of all linear transformations from V into W and the set of all $m \times n$ matrices over the field F.

Note that the matrix A in the above theorem is known as matrix of T relative to the ordered bases $\mathscr{B}$ and $\mathscr{B}'$. Columns $A_1, \ldots, A_n$ of the matrix A is given by

$$
A_j = [T\alpha_j]_{\mathscr{B}'} \qquad j = 1, \ldots, n.
$$

Note 2.2. If U is another linear transformation from V into W and $B = [B_1, \ldots, B_n]$ is the matrix of U relative to the ordered bases $\mathscr{B}$ and $\mathscr{B}'$.

$$\begin{aligned} cA_j + B_j &= c[T\alpha_j]_{\mathscr{B}'} + [U\alpha_j]_{\mathscr{B}'} \\ &= [T\alpha_j + U\alpha_j]_{\mathscr{B}'} \\ &= [(cT + U)_{\alpha_j}]_{\mathscr{B}'} \end{aligned}$$

then $cA + B$ is the matrix of $cT + U$ relative to $\mathscr{B}$, $\mathscr{B}'$.

Theorem 2.13. *Let V be an n-dimensional vector space over the field F and let W be an m-dimensional vector space over F. For each pair of ordered bases $\mathscr{B}$, $\mathscr{B}'$ for V and W respectively, the function which assigns to a linear transformation T its matrix relative to $\mathscr{B}$, $\mathscr{B}'$ is an isomorphism between the space $L(V, W)$ and the space of all $m \times n$ matrices over the field F .*

Proof. Consider the function f which assigns to a linear transformation T its matrix relative to $\mathscr{B}$, $\mathscr{B}'$. Then f is a function from the space $L(V, W)$ to the space of all $m \times n$ matrices over the field F. We have from note 2.2, f is linear. Also by Theorem 2.12, f is a one-one correspondence. And corresponding to an $m \times n$ matrix A, there is a linear transformation from V in to W defined by Equation 2.12. Hence f is onto. Thus f is an isomorphism. □

Note 2.3. Consider representation by matrices of linear transformations from the space into itself, that is linear operators on a space V. Here we use the same ordered basis in each case, that is to take $\mathscr{B} = \mathscr{B}'$. We can call the representing matrix as the matrix of T relative to the ordered basis $\mathscr{B}$.

Example 26. 1. Let F be a field and let T be the operator on F^2 defined by $T(x_1, x_2) = (x_2, x_1)$. We are going to find the matrix representing T relative to the standard ordered basis $\mathscr{B} = \{e_1, e_2\}$, where $e_1 = (1, 0)$

and $e_2 = (0, 1)$. Here we have both bases are same. So find images of e_1, e_2 under T and represent those images in terms of e_1 and e_2.

$$T(e_1) = T(1, 0) = (0, 1) = 0.e_1 + 1.e_2.$$

$$T(e_2) = T(0, 1) = (1, 0) = 1.e_1 + 0.e_2.$$

So the columns of the matrix of T are $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and the matrix of T in the ordered basis $\mathscr{B}$ is given by,

$$[T]_{\mathscr{B}} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

2. Let V be the space of $n \times 1$ column matrices over the field F. Let W be the space of $m \times 1$ matrices over F, and let A be a fixed $m \times n$ matrix over F. Let T be the linear transformation of V into W defined by $T(X) = AX$.

3. Let V be the space of all polynomial functions from R into R of the form $f(x) = c_0 + c_1x + c_2x^2 + c_3x^3$. That is the space of all polynomial functions of degree three or less. The differentiation operator D that maps V into V, since D is degree decreasing. Let $\mathscr{B} = \{f_1, f_2, f_3, f_4\}$, where $f_1(x) = 1$ for all x.
$f_2(x) = x$ for all x.
$f_3(x) = x^2$ for all x.
$f_4(x) = x^3$ for all x.
Then

$$Df_1(x) = 0 = 0f_1 + 0f_2 + 0f_3 + 0f_4$$

$$Df_2(x) = 1 = 1f_1 + 0f_2 + 0f_3 + 0f_4$$

$$Df_3(x) = 2x = 0f_1 + 2f_2 + 0f_3 + 0f_4$$

$$Df_4(x) = 3x^2 = 0f_1 + 0f_2 + 3f_3 + 0f_4.$$

So the columns of matrix are given by $\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 3 \\ 0 \end{bmatrix}$, and thus matrix of D is given by $\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$.

Theorem 2.14. *Let V, W and Z be finite dimensional vector spaces over the field F, let T be a linear transformation from V into W and U be a linear transformation from W into Z. If $\mathscr{B}$, $\mathscr{B}'$ and $\mathscr{B}''$ are ordered bases for the spaces V, W and Z respectively, if A is the matrix of T relative to the pair $\mathscr{B}$, $\mathscr{B}'$ and B is the matrix of U relative to the pair $\mathscr{B}'$ $\mathscr{B}''$, then the matrix of the composition UT relative to the pair $\mathscr{B}$, $\mathscr{B}''$ is the product matrix $C = BA$.*

Definition 2.7. *Let A and B be $n \times n$ matrices over the field F. We say that B is similar to A if there is an invertible $n \times n$ matrix P over F such that $B = P^{-1}AP$.*

In the next theorem we are saying that if V is an n-dimensional vector space over F and $\mathscr{B}$, $\mathscr{B}'$ are two ordered bases for V, then for each linear operator T on V the matrix $[T]_{\mathscr{B}'}$ is similar to $[T]_{\mathscr{B}}$.

Theorem 2.15. *Let V be a finite dimensional vector space over the field F, and let $\mathscr{B} = \{\alpha_1, \ldots, \alpha_n\}$ and $\mathscr{B}' = \{\alpha'_1, \ldots, \alpha'_n\}$ be ordered base for V.*

Suppose T is a linear operator on V. If $P = [P_1, \ldots, P_n]$ is the $n \times n$ matrix with columns $P_j = [\alpha'_j]_{\mathscr{B}}$, then $[T]_{\mathscr{B}'} = P^{-1}[T]_{\mathscr{B}}P$.
If U is the invertible operator on V defined by $U(\alpha_j) = \alpha'_j$, $j = 1, 2, \ldots, n$, then $[T]_{\mathscr{B}'} = [U]_\beta^{-1}[T]_{\mathscr{B}}[U]_\beta$.

Example 27. Let T be the linear operator on R^2 defined by $T(x_1, x_2) = (x_1, 0)$. Then what is the matrix of T with respect to the standard basis $\{(1,0), (0,1)\}$?
We have $T(x_1, x_2) = (x_1, 0)$ implies that
$T(1,0) = (1,0) = 1.(1,0) + 0.(0,1)$ and
$T(0,1) = (0,0) = 0.(1,0) + 0.(0,1)$.
Thus we get matrix of T relative to the standard basis as $[T]_{\mathscr{B}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.
Suppose $\mathscr{B}'$ is the ordered basis for R^2 consisting of the vectors $e'_1 = (1,1)$, $e'_2 = (2,1)$.
Then we have $(1,1) = 1.(1,0) + 1.(0,1) = e_1 + e_2$
and $(2,1) = 2.(1,0) + 1.(0,1) = 2e_1 + e_2$.
Now $P = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix}$
Then we can easily write $P^{-1} = \begin{bmatrix} -1 & 2 \\ 1 & -1 \end{bmatrix}$.
By above Theorem we can find marix of T relative the basis consisting of $e'_1 = (1,1)$, $e'_2 = (2,1)$.

$$\begin{aligned} [T]_{\mathscr{B}'} &= P^{-1}[T]_{\mathscr{B}}P \\ &= \begin{bmatrix} -1 & 2 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} -1 & 2 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix} \end{aligned}$$

$$= \begin{bmatrix} -1 & -2 \\ 1 & 2 \end{bmatrix}$$

2.5 Linear functionals

Definition 2.8. *If V is a vector space over the field F, a linear transformation f from V into the scalar field F is called a linear functional on V.*

Example 28. 1. Consider vector space R^2 over R.

Then the map defined by $T(x_1, x_2) = x_1 + x_2$ is a linear functional.

2. Consider vector space F^n over F.

Then the map defined by $f(x_1, x_2, \ldots, x_n) = c_1x_1 + c_2x_2 + \ldots + c_nx_n$ is a linear functional on F^n.

Now consider the standard basis of F^n. We can find the corresponding matrix of f relative to this basis.

We have $f(1, 0, \ldots, 0) = c_1$

$f(0, 1, \ldots, 0) = c_2$

.

.

.

$f(0, 0, \ldots, 1) = c_n$.

That is $f(e_i) = c_i, i = 1, 2, \ldots, n$.

Thus $\begin{bmatrix} c_1 & c_2 & \ldots & c_n \end{bmatrix}$ is the matrix of f relative to the standard ordered basis for F^n and the basis $\{1\}$ for F.

We have

$$(x_1, x_2, \ldots, x_n) = x_1e_1 + x_2e_2 + \ldots + x_ne_n$$

$$f(x_1, x_2, \ldots, x_n) = f(x_1e_1 + x_2e_2 + \ldots + x_ne_n)$$

$$= x_1 f(e_1) + x_2 f(e_2) + \ldots + x_n f(e_n)$$
$$= x_1 c_1 + x_2 c_2 + \ldots + x_n c_n$$

Thus every linear functional on F^n is of the form $f(x_1, x_2, \ldots, x_n) = c_1 x_1 + c_2 x_2 + \ldots + c_n x_n$.

3. Let n be a positive integer and F a field. If A is an $n \times n$ matrix with entries in F, the trace of A is the sum of the diagonal entries. Thus Trace of $A = tr\, A = A_{11} + A_{22} + \ldots + A_{nn}$. For example

$$tr \begin{bmatrix} 1 & 6 & 3 \\ 3 & 2 & 1 \\ 5 & 1 & 2 \end{bmatrix} = 1 + 2 + 2 = 5$$

We can easily verify that tr is a linear functional.

$$\begin{aligned} tr\,(cA + B) &= cA_{11} + B_{11} + cA_{22} + B_{22} + \ldots + cA_{nn} + B_{nn} \\ &= cA_{11} + cA_{22} + \ldots + cA_{nn} + B_{11} + B_{22} + \ldots + B_{nn} \\ &= c(A_{11} + A_{22} + \ldots + A_{nn}) + B_{11} + B_{22} + \ldots + B_{nn} \\ &= c.tr\, A + tr\, B \end{aligned}$$

4. Let V be the space of all polynomial functions from the field F into itself. Let t be an element of F. If we define $L_t(p) = p(t)$, then L_t is a linear functional on V. For example, let p denotes the polynomial $p(x) = 1 + 2x + x^3$ and $t = 2$, then $L_t(p) = p(t) = 1 + 2.2 + 2^3 = 13$.

5. Let $[a, b]$ be a closed interval on the real line and let $C[a, b]$ be the space of continuous real valued functions on [a,b]. We have $\int_a^b g(t)dt$ is a real number. Then

$$L(g) = \int_a^b g(t)dt$$

is a linear functional on $C[a, b]$.

$$L(cg+h) = \int_a^b (cg+h)(t)dt = \int_a^b (cg(t))dt + \int_a^b (h(t))dt = cL(g)+L(h)$$

Thus L is a linear functional on $C[a, b]$.

Dual Space

Definition 2.9. *If V is a vector space, the collection of all linear functionals on V forms a vector space and is the space $L(V, F)$. It is denoted by V^* and call it the dual space of V. Thus $L(V, F) = V^*$*

Dual Space:$L(V, F) = V^*$

If V is finite dimensional, dim V^* =dim V.

<u>Dual basis</u>

Let $\mathscr{B} = \{\alpha_1, \ldots, \alpha_n\}$ be a basis for V. Then there is a unique linear functional f_i on V such that $f_i(\alpha_j) = \delta_{ij}$, where $\delta_{ij} = 1$, if $i = j$ and 0 if $i \neq j$. Thus from the basis $\mathscr{B}$, we get a set of n distinct linear functionals $\{f_1, \ldots, f_n\}$ on V. We can prove that

$$\mathscr{B}' = \{f_1, \ldots, f_n\}$$

is a basis for V^* and is called dual basis of $\mathscr{B}$.
For this suppose

$$c_1 f_1 + \ldots + c_n f_n = 0,$$

then

$$(c_1 f_1 + \ldots + c_n f_n)(\alpha_j) = 0(\alpha_j) = 0, \text{ for } j = 1, 2, \ldots, n$$

we have

$$f_j(\alpha_j) = 1 \text{ and } f_i(\alpha_j) = 0 \text{ for } i \neq j,$$

therefore from equation 2.5, we get $c_j = 0, j = 1, 2, \ldots, n$. Thus $\{f_1, \ldots, f_n\}$ is linearly independent. Since we know that V^* has dimension n, $\{f_1, \ldots, f_n\}$ is a basis for V^*. This basis is called the dual basis of $\mathscr{B}$.

Dual Basis: $\mathscr{B}' = \{f_1, \ldots, f_n\}$ where $f_i(\alpha_j) = \delta_{ij}$, $\delta_{ij} = 1$, if $i = j$ and 0 if $i \neq j$.

Theorem 2.16. *Let V be a finite dimensional vector space over the field F and let $\mathscr{B} = \{\alpha_1, \ldots, \alpha_n\}$ be a basis for V. Then there is a unique dual basis $\mathscr{B}^* = \{f_1, \ldots, f_n\}$ in V^* such that $f_i(\alpha_j) = \delta_{ij}$. For each linear functional f on V we have*

$$f = \sum_{i=1}^{n} f(\alpha_i) f_i$$

and

$$\alpha = \sum_{i=1}^{n} f_i(\alpha)\alpha_i.$$

Proof. We already proved that $\mathscr{B}'$ is basis for V^*. Now if

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

Then

$$\begin{aligned} f(\alpha_j) &= \sum_{i=1}^{n} c_i f_i(\alpha_j) \\ &= \sum_{i=1}^{n} c_i \delta_{ij} \\ &= c_j \end{aligned}$$

That is $c_j = f(\alpha_j)$. Hence

$$f = \sum_{i=1}^{n} f(\alpha_i) f_i \tag{2.13}$$

Next we have to prove that this representation is unique. For that take a vector in V

$$\alpha = x_1\alpha_1 + \ldots + x_n\alpha_n = \sum_{i=1}^{n} x_i\alpha_i.$$

Then

$$\begin{aligned} f_j(\alpha) &= f_j(x_1\alpha_1 + \ldots + x_n\alpha_n) \\ &= x_1 f_j(\alpha_1) + \ldots + x_n f_j(\alpha_n) \end{aligned}$$

$$= \quad x_j$$

That is

$$x_j = f_j(\alpha)$$

Thus the unique representation of α is given by

$$\alpha = \sum_{i=1}^{n} f_i(\alpha)\alpha_i. \tag{2.14}$$

□

Example 29. 1. Let V be the vector space of all polynomial functions from R into R which have degree less than or equal to 2. Let t_1, t_2, and t_3 be any three distinct real numbers and let $L_i(p) = p(t_i)$. Then L_1, L_2 and L_3 are functionals on V. Suppose $c_1L_1 + c_2L_2 + c_3L_3 = L$. Consider particular functions $1, x, x^2$,

$L_1(1) = 1$

$L_2(1) = 1$

$L_3(1) = 1$

$L_1(x) = p(t_1) = t_1$

$L_2(x) = p(t_2) = t_2$

$L_3(1) = p(t_3) = t_3$

$L_1(x^2) = p(t_1) = t_1^2$

$L_2(x^2) = p(t_2) = t_2^2$

$L_3(x^2) = p(t_3) = t_3^2$

$$L(1) = c_1L_1(1) + c_2L_2(1) + c_3L_3(1) = c_1.1 + c_2.1 + c_3.1$$

$$L(x) = (c_1L_1 + c_2L_2 + c_3L_3(x)) = c_1L_1(x) + c_2L_2(x) + c_3L_3(x) = c_1t_1 + c_2t_2 + c_3t_3$$

$$L(x^2) = (c_1L_1 + c_2L_2 + c_3L_3(x^2)) = c_1L_1(x^2) + c_2L_2(x^2) + c_3L_3(x^2) = c_1t_1^2 + c_2t_2^2 + c_3t_3^2$$

If $L = 0$, ie if $L(p) = 0$ for each $p \in V$, then applying L to the particular functions $1, x, x^2$ we get,

$$c_1 + c_2 + c_3 = 0$$

$$c_1t_1 + c_2t_2 + c_3t_3 = 0$$

$$c_1t_1^2 + c_2t_2^2 + c_3t_3^2 = 0$$

The corresponding coefficient matrix $\begin{bmatrix} 1 & 1 & 1 \\ t_1 & t_2 & t_3 \\ t_1^2 & t_2^2 & t_3^2 \end{bmatrix}$ is invertible when t_1, t_2 and t_3 are distinct. This implies that $c_1 = c_2 = c_3 = 0$. This implies that three functionals are linearly independent. Now the L_i are independent, and since V has dimension 3, these 3 functionals $\{L_1, L_2, L_3\}$ form a basis for V^*.

Dual basis of this basis will be of the form $\{p_1, p_2, p_3\}$ [Dual basis is a subset of V=set of all polynomials]. $\{p_1, p_2, p_3\}$ will satisfy
$L_i(p_j) = \delta_{ij}$ by the previous theorem
$\Rightarrow p_j(t_i) = \delta_{ij}$, by the definition of L_i, $\quad p_1(x) = \frac{(x-t_2)(x-t_3)}{(t_1-t_2)(t_1-t_3)}$
$p_2(x) = \frac{(x-t_1)(x-t_3)}{(t_2-t_1)(t_2-t_3)}$
$p_3(x) = \frac{(x-t_1)(x-t_2)}{(t_3-t_1)(t_3-t_2)}$

Remark 8. If f is a non-zero linear functional, then the rank of f is 1 because the range of f is a non zero subspace of the scalar field and must be the scalar field. If the underlying space V is finite dimensional, the rank-nullity theorem tells us that the null space has dimension $\boxed{\dim N_f = \dim V - 1}$

Note 2.4. In a vector space of dimension n, a subspace of dimension $n - 1$ is called a hyperspace. The null space of a functional is always a hyperspace.

Definition 2.10. *If V is a vector space over the field F and S is a subset of V, the annihilator of S is the set S^0 of linear functionals f on V such that $f(\alpha) = 0$ for every $\alpha \in S$.*

Note 2.5. S^0 is a subspace of V^*. If $S = \{0\}$, the zero subspace, then $S^0 = V^*$. If $S = V^*$, then $S^0 = \{0\}$, the zero subspace.

Theorem 2.17. *Let V be a finite dimensional vector space over the field F, and let W be a subspace of V. Then $dimW + dimW^0 = dimV$.*

Proof. Let k be the dimension of W and $\{\alpha_1, \ldots, \alpha_k\}$ be a basis of V. Let $\{f_1, \ldots, f_n\}$ be a basis for V^* which is dual to this basis for V.
Claim: $\{f_{k+1}, \ldots, f_n\}$ is a basis for the annihilator W^0.
$f_i(\alpha_j) = \delta_{ij}$, by the property of dual basis.
$\delta_{k+1,k-1} = 0$ and $\delta_{ij} = 0$ if $i \geq k+1$ and $j \leq k$.
Suppose $\alpha = c_1\alpha_1 + c_2\alpha_2 + \ldots + c_{n-k}\alpha_n$.
Then for $i \geq k+1$, $f_i(\alpha) = c_1 f_i(\alpha_1) + c_2 f_i(\alpha_2) + \ldots + c_n f_i(\alpha_n) = 0$.
This means that $f_{k+1}, f_{k+2}, \ldots, f_n$ belongs to W^0. The functionals $f_{k+1}, f_{k+2}, \ldots, f_n$ are linearly independent since $\{f_1, \ldots, f_n\}$ is a basis of V^*.
Now we have to show that $f_{k+1}, f_{k+2}, \ldots, f_n$ span W^0. Let $f \in W^0$. We have to express f as a linear combination of $f_{k+1}, f_{k+2}, \ldots, f_n$.
Since W^0 is a subset of V^*, $f \in W^0$ implies that $f \in V^*$. Then we know that f is a linear combination of $f_1, f_2, \ldots, f_n$ (Since it is basis for V^*). That is

$$f = \sum_{i=1}^{n} f(\alpha_i) f_i.$$

Now $f \in W^0$ implies that $f(\alpha_i) = 0$ for $i \leq k$ and thus

$$f = \sum_{i=k+1}^{n} f(\alpha_i) f_i.$$

This means that f can be written as a linear combination of $f_{k+1}, f_{k+2}, \ldots, f_n$.
Thus we get $\{f_{k+1}, f_{k+2}, \ldots, f_n\}$ is a basis for W^0.

Thus if $dimV = n$ and $dimW = k$, then dim $W^0 = n - k$. In other words, $dimW + dimW^0 = dimV$. $\square$

Corollary 2.1. *If W is a k-dimensional subspace of an $n-k$ dimensional vector space V, then W is the intersection of $n-k$ hyper spaces in V.*

Proof. Let W be a k-dimensional subspace of an n dimensional vector space V, then from the proof of the Theorem 2.17, it follows that W is exactly the set of vectors such that $f_i(\alpha) = 0$, for $i = k+1, \ldots, n$. That is

$$W = \{\alpha / f_i(\alpha) = 0, \text{ for all } i = k+1, \ldots, n\}.$$

That is

$$W = \bigcap_{i=k+1}^{n} N_{f_i}.$$

where N_{f_i} is the null space of f_i.

$$\begin{aligned}
\alpha \in W \quad &\Rightarrow \quad \alpha = x_1\alpha_1 + \ldots + x_k\alpha_k \ \textit{where}\ \{\alpha_1, \ldots, \alpha_k\}\ \textit{is a basis for W.} \\
&\Rightarrow \quad \alpha = x_1 f_i(\alpha_1) + \ldots + x_k f_i(\alpha_k) \\
&\Rightarrow \quad = 0\ \textit{for all}\ i \geq k+1\ \textit{since}\ f_{k+1}, \ldots, f_n \in W^0. \\
&\Rightarrow \quad \alpha \in N_{f_i}\ \textit{for all}\ i \geq k+1 \\
&\Rightarrow \quad \alpha \in \bigcap_{i=k+1}^{n} N_{f_i}.
\end{aligned}$$

$$W \subseteq \bigcap_{i=k+1}^{n} N_{f_i} \tag{2.15}$$

Conversely

$$\begin{aligned}
\alpha \in \bigcap_{i=k+1}^{n} N_{f_i} \quad &\Rightarrow \quad \alpha \in N_{f_i}\ \textit{for all}\ i \geq k+1 \\
&\Rightarrow \quad f_i(\alpha) = 0\ \textit{for all}\ i = k+1, \ldots, n
\end{aligned}$$

We have

$$\alpha = \sum_{i=1}^{n} f_i(\alpha)\alpha_i \text{ by above theorem}$$

Then $\alpha = \sum\limits_{i=1}^{k} f_i(\alpha)\alpha_i$ since $f_i(\alpha) = 0 \; for \; all \; i = k+1, \ldots, n$. Which implies $\alpha \in W$. Thus

$$\bigcap_{i=k+1}^{n} N_{f_i} \subseteq W \tag{2.16}$$

Thus by the Equation 2.15 and 2.16, we get $W = \bigcap\limits_{i=k+1}^{n} N_{f_i}$. Since each f_i is a non-zero linear functional on V, the null space N_{f_i} is a hyperspace in V. Thus W is the intersection of $n-k$ hyper spaces in V. □

Corollary 2.2. *If W_1 and W_2 are subspaces of a finite- dimensional vector space, then $W_1 = W_2$ if and only if $W_1^0 = W_2^0$.*

Proof. If $W_1 = W_2$, then obviously $W_1^0 = W_2^0$. If $W_1 \neq W_2$, then one of the two subspaces contains a vector which is not in the other. Suppose $\alpha \in W_2$ but not in W_1. Then there is a linear functional f such that $f(\beta) = 0$ for all $\beta \in W$, but $f(\alpha) \neq 0$. Then $f \in W_1^0$, but not in W_2^0. Therefore $W_1^0 \neq W_2^0$. Thus $W_1^0 = W_2^0$ implies that $W_1 = W_2$. □

Note 2.6. A system of homogeneous linear equations from the point of view of linear functionals. Suppose we want to find solutions of the following system of linear equations,

$$\begin{aligned} A_{11}x_1 + \ldots + A_{1n}x_n &= 0 \\ &\vdots \\ A_{m1}x_1 + \ldots + A_{mn}x_n &= 0 \end{aligned}$$

if we let f_i, $i = 1, 2, \ldots, m$, be the linear functional on F^n defined by

$$f_i(x_1, \ldots, x_n) = A_{i1}x_1 + \ldots + A_{in}x_n$$

then we are seeking the subspace of F^n of all α such that $f_i(\alpha) = 0$, $i = 1, \ldots, m$. In other words we are seeking the subspace annihilated by

$f_1, \ldots, f_m$. Row reduction of the coefficient matrix provides us with a systematic method of finding this subspace. The n-tuple $(A_{i1}, \ldots, A_{in})$ gives the coordinates of the linear functional f_i relative to the basis which is dual to the standard basis for F^n. The row space of the coefficient matrix may thus be regarded as the space of linear functionals spanned by $f_1, \ldots, f_m$. The solution space is the subspace annihilated by this space of functionals.

Now look at the system of equations from the 'dual' point of view. suppose that we are given m vectors in F^n. $\alpha_i = (A_{i1}, \ldots, A_{in})$ and we wish to find the annihilator of the subspace spanned by these vectors. Since a typical linear functional on F^n has the form $f(x_1, \ldots, x_n) = c_1x_1 + \ldots + c_nx_n$ the condition that f be in this annihilator is that $\sum\limits_{j=1}^{n} A_{ij}c_j = 0$, $i = 1, 2, \ldots, m$ that is, that $(c_1, \ldots, c_n)$ be solution of the system $AX = 0$. From this point of view, row-reduction gives us a systematic method of finding the annihilator of the subspace spanned by a given finite set of vectors in F^n.

Example 30. 1. Let $\alpha_1 = (1, 0, -1, 2)$, $\alpha_2 = (2, 3, 1, 1)$ and let W be the subspace of R^4 spanned by α_1 and α_2. Which linear functionals are in the annihilator of W? Find a basis for W.

Solution

Let $A = \begin{bmatrix} 1 & 0 & -1 & 2 \\ 2 & 3 & 1 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -1 & 2 \\ 0 & 3 & 3 & -3 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -1 & 2 \\ 0 & 1 & 1 & -1 \end{bmatrix}$

we have W is the raw space of A. Thus the vectors

$$\beta_1 = (1, 0, -1, 2)$$

and

$$\beta_2 = (0, 1, 1, -1)$$

form a basis for W. Thus

$$dim W = 2,$$

So

$$dimW^0 = 4 - 2 = 2.$$

A linear functional on R^4 is of the form

$$f(x_1, x_2, x_3, x_4) = c_1x_1 + c_2x_2 + c_3x_3 + c_4x_4 \tag{2.17}$$

$f \in W^0 \Leftrightarrow f(\alpha) = 0$ for every $\alpha \in W$. Then $f(\beta_1) = 0$ and $f(\beta_2) = 0$. Now

$$f(\beta_1) = 0 \Rightarrow f(1, 0, -1, 2) = 0$$

and

$$f(\beta_2) = 0 \Rightarrow f(0, 1, 1, -1) = 0.$$

Thus $c_1 - c_3 + 2c_4 = 0$ and $c_2 + c_3 - c_4 = 0$. Now put $c_3 = a$ and $c_4 = b$. Then we get

$$c_1 = a - b$$

and

$$c_2 = b - a.$$

Now substitute c_1, c_2, c_3 and c_4 in Equation 2.17, then we get W^0 is the set of all linear functionals f of the form $(a-b)x_1+(b-a)x_2+ax_3+bx_4$, where $a, b \in R$.

2. Let W_1 and W_2 be subspaces of a finite-dimensional vector space V. Prove that

$$(W_1 + W_2)^0 = W_1^0 + W_2^0$$

Proof.

$$W_1 \subseteq W_1 + W_2 \Rightarrow (W_1 + W_2)^0 \subseteq W_1^0$$

$$W_2 \subseteq W_1 + W_2 \Rightarrow (W_1 + W_2)^0 \subseteq W_2^0$$

. Thus

$$(W_1 + W_2)^0 \subseteq W_1^0 \cap W_2^0. \tag{2.18}$$

To prove the reverse inclusion, let $f \in W_1^0 \cap W_2^0$. Then $f \in W_1^0$ and $f \in W_2^0$. Now

$$f \in W_1^0 \Rightarrow f(\alpha) = 0 \text{ for every } \alpha \in W_1$$

and

$$f \in W_2^0 \Rightarrow f(\beta) = 0 \text{ for every } \beta \in W_2.$$

Then $f(\alpha)+f(\beta) = 0$ for every $\alpha \in W_1$ and $\beta \in W_2$. Then $f(\alpha+\beta) = 0$ for every $\alpha \in W_1$ and $\beta \in W_2$. This means that $f \in W_1^0 + W_2^0$. Thus

$$W_1^0 \cap W_2^0 \subseteq W_1^0 + W_2^0 \tag{2.19}$$

From equations 2.18 and 2.19, we get

$$(W_1 + W_2)^0 = W_1^0 + W_2^0.$$

$\square$

3. Find the subspace annihilated by three linear functionals f_1, f_2 and f_3 on R^4 defined by
   $f_1(x_1, x_2, x_3, x_4) = x_1 + 2x_2 + 2x_3 + x_4$
   $f_2(x_1, x_2, x_3, x_4) = 2x_2 + x_4$
   $f_1(x_1, x_2, x_3, x_4) = -2x_1 - 4x_3 + 3x_4.$
   This system of equations can be written as

$$\begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 2 & 1 \\ 0 & 2 & 0 & 1 \\ -2 & 0 & -4 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$$

By the note already give above, to find the subspace which they annihilate can be found by finding the row reduced echelon form of the matrix

$$\begin{bmatrix} 1 & 2 & 2 & 1 \\ 0 & 2 & 0 & 1 \\ -2 & 0 & -4 & 3 \end{bmatrix}$$

It can be deduced into

$$\begin{bmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Thus the linear functionals

$g_1(x_1, x_2, x_3, x_4) = x_1 + 2x_3$

$g_2(x_1, x_2, x_3, x_4) = x_2$

$g_3(x_1, x_2, x_3, x_4) = x_4$ span the same subspace of $(R^4)^*$ and annihilate the same subspace of R^4 as do f_1, f_2, f_3. The subspace annihilated consists of the vectors with $x_1 + 2x_3 = 0 \Rightarrow x_1 = -2x_3$, $x_2 = 0$ and $x_4 = 0$. Thus $\{(-2x_3, 0, x_3, 0)/x_3 \in R\}$ is the required subspace.

Exercises

1. In R^3, let $\alpha_1 = (1, 0, 1)$, $\alpha_2 = (0, 1, -2)$, $\alpha_3 = (-1, -1, 0)$
   (a) If f is a linear functional on R^3 such that $f(\alpha_1) = 1$, $f(\alpha_2) = -1$, $f(\alpha_3) = 3$ and if $\alpha = (a, b, c)$, then find $f(\alpha)$.
   (b) Describe explicitly a linear functional f on R^3 such that $f(\alpha_1) = f(\alpha_2) = 0$ but $f(\alpha_3) \neq 0$.
   (c) Let f be any linear functional such that $f(\alpha_1) = f(\alpha_2) = 0$ and $f(\alpha_3) \neq 0$. If $\alpha = (2, 3, -1)$, show that $f(\alpha) \neq 0$.
2. Let $\mathscr{B} = \{\alpha_1, \alpha_2, \alpha_3\}$ be the basis for C^3 defined by $\alpha_1 = (1, 0, -1)$, $\alpha_2 = (1, 1, 1)$, $\alpha_3 = (2, 2, 0)$. Find a dual basis of $\mathscr{B}$.
3. Let $\alpha_1 = (1, 0, -1, 2)$ and $\alpha_2 = (2, 3, 1, 1)$ and let W be the subspace of R^4 spanned by α_1 and α_2. Which linear functionals f such that $f(x_1, x_2, x_3, x_4) = c_1x_1 + c_2x_2 + c_3x_3 + c_4x_4$ are in the annihilator of W.

4. Let W be the subspace of R^5 which is spanned by the vectors
   $\alpha_1 = e_1 + 2e_2 + e_3$,
   $\alpha_2 = e_2 + 3e_3 + 3e_4 + e_5$,
   $\alpha_3 = e_1 + 4e_2 + 6e_3 + 4e_4 + e_5$.
   Find a basis for W^0.

5. Let W_1 and W_2 be subspaces of a finite dimensional vector space V. The prove that
$$(W_1 \cap W_2)^0 = W_1^0 + W_2^0.$$

2.6 Double Dual

If α is a vector in V, then α induces a linear functional L_α on V^* defined by $L_\alpha(f) = f(\alpha)$ for $\alpha \in V^*$.

Note 2.7. We can prove that L_α is a linear transformation defined from V^* to F. Thus $L_\alpha \in (V^*)^* = V^{**}$.

$$\begin{aligned} L_\alpha(cf + g) &= (cf + g)(\alpha) \\ &= cf(\alpha) + g(\alpha) \text{ since f linear} \\ &= c.L_\alpha f + L_\alpha g \end{aligned}$$

Theorem 2.18. *Let V be a finite dimensional vector space over the field F. For each vector $\alpha \in V$, define $L_\alpha(f) = f(\alpha)$, $f \in V^*$. The mapping $\alpha \to L_\alpha$ is then an isomorphism of V onto V^{**}.*

Proof. We showed that for each α the function L_α is linear in Note 2.7. Suppose $\alpha, \beta \in V$ and $c \in F$, and let $\gamma = c\alpha + \beta$. Then for each $f \in V^*$,

$$\begin{aligned} L_\alpha(f) &= f(\gamma) \\ &= f(c\alpha + \beta) \\ &= c.f(\alpha) + f(\beta) \\ &= c.L_\alpha(f) + L_\beta(f) \end{aligned}$$

Thus $L_\gamma = c.L_\alpha + L_\beta$. Thus we get the map $\alpha \to L_\alpha$ is a linear transformation from V in to V^{**}. This map is one to one since

$$\begin{aligned} L_\alpha = 0 &\Leftrightarrow L_\alpha(f) = 0 \text{ for every } f \in V^* \\ &\Leftrightarrow f(\alpha) = 0 \text{ for every } f \in V^* \\ &\Leftrightarrow \alpha = 0 \end{aligned}$$

Thus we get $\alpha \to L_\alpha$ is a non-singular linear transformation from V into V^{**}, and since

$$dimV^{**} = dimV^* = dimV$$

we get this transformation is invertible, and therefore is an isomorphism of V onto V^{**}. □

Corollary 2.3. *Let V be a finite dimensional vector space over the field F. If L is a linear functional on the dual space V^* of V, then there is a unique vector $\alpha \in V$ such that $L(f) = f(\alpha)$ for every $f \in V^*$.*

Proof. Let L be a linear functional on the dual space V^* of V. Then $L \in V^{**}$. By the above theorem, $\alpha \to L_\alpha$ is an isomorphism of V onto V^{**}. There is a unique vector $\alpha \in V$ such that $L_\alpha = L$. Then for each $f \in V^*$,

$$\begin{aligned} L(f) &= L_\alpha(f) \\ &= f(\alpha) \end{aligned}$$

Hence the result. □

Corollary 2.4. *Let V be a finite dimensional vector space over the field F. Each basis for V^* is the dual basis of some basis for V.*

Proof. Let $B^* = \{f_1, f_2, \ldots, f_n\}$ be any basis for V^*. By theorem 2.16, there is a basis $\{L_1, \ldots, L_n\}$ for V^{**} such that $L_i(f_j) = \delta_{ij}$. By the above corollary, for each i, there is a vector $\alpha_i \in V$ such that $L_i = L\alpha_i \Rightarrow L_i(f) = f(\alpha_i)$, for all $f \in V^*$, that is such that $L_i = L_{\alpha_i}$. Thus $\{\alpha_1, \ldots, \alpha_n\}$ is a basis for V and that $\mathscr{B}^*$ is the dual basis of this basis. □

Remark 9. 1. In theorem 2.18, we proved that $\alpha \to L_\alpha$ is an isomorphism of V onto V^{**}. Therefore by identifying with α we can assume that $V^{**} = V$. Thus the dual space of V^* is V and the dual space of V is V^*. That is each is the dual space of the other.

2. If E is a subset of V^*, then the annihilator E^0 is technically a subspace of V^{**}. By identifying V^{**} with V, we can assume that E^0 is a subspace of V. That is

$$\begin{aligned} E^0 &= \{L \in V^{**}/L(f) = 0, \text{ for all } f \in E\} \\ &= \{L_\alpha \in V^{**}/L_\alpha(f) = 0, \text{ for all } f \in E\} \\ &= \{\alpha \in V/f(\alpha) = 0, \text{ for all } f \in E\}\, (by\, identifying\, \alpha\, with\, L_\alpha) \end{aligned}$$

Theorem 2.19. *If S is any subset of a finite dimensional vector space V, then $(S^0)^0$ is the subspace spanned by S.*

Proof. Let W be the subspace spanned by S. Then

$$\begin{aligned} f \in S^0 &\Rightarrow f(\alpha) = 0 \text{ for all } \alpha \in S \\ &\Rightarrow f(\alpha) = 0 \text{ for all } \alpha \in W (\textit{Since } W \textit{ is spanned by } S) \\ &\Rightarrow f \in W^0. \end{aligned}$$

$$\begin{aligned} f \in W^0 &\Rightarrow f(\alpha) = 0 \text{ for all } \alpha \in W \\ &\Rightarrow f(\alpha) = 0 \text{ for all } \alpha \in S (\text{Since } S \subseteq W) \\ &\Rightarrow f \in S^0. \end{aligned}$$

Thus $S^0 = W^0$. We have by Theorem 2.17, dim W+ dim W^0 =dim V. Then

$$dimW^0 + dim(W^0)^0 = dimV^*.$$

Now

$$dimV = dimV^*$$

Thus

$$dimW^0 + dim(W^0)^0 = dimW + dimW^0 \Rightarrow dimW = dim(W^0)^0. \quad (2.20)$$

Also W is a subspace of W^{00}, since

$$\begin{aligned} \alpha \in W &\Rightarrow f(\alpha) = 0 \; for \; all \; f \in W^0 \\ &\Rightarrow \alpha \in W^{00}. \end{aligned}$$

Since W is a subspace of W^{00}, we see that $W = W^{00}$. □

Definition 2.11. *A hyperspace in a vector space is a maximal proper subspace of V. That is N is a hyperspace in V if it satisfies two conditions:*

1. *N is a proper subspace of V,*

2. *If W is a proper subspace of V containing N, then either $W = N$ or $W = V$.*

Theorem 2.20. *If f is a non-zero linear functional on the vector space V, then the null space of f is a hyperspace in V. Conversely every hyperspace in V is the nullspace of a non-zero linear functional on V.*

Proof. Let f be a non-zero linear functional on V and N_f its null space. Let α be a vector in V which is not in N_f. That is α is such that $f(\alpha) \neq 0$. We have to prove that subspace spanned by α and N_f is V.

Any vector in the subspace has the form $r + c.\alpha$, $r \in N_f$ and $c \in F$.
Let β be any vector in V, define

$$c = \frac{f(\beta)}{f(\alpha)},$$

which makes sense since $f(\alpha) \neq 0$
Let $r = \beta - c\alpha$, then

$$\begin{aligned} f(r) &= f(\beta - c\alpha) \\ &= f(\beta) - c.f(\alpha) \\ &= f(\beta) - \frac{f(\beta)}{f(\alpha)}.f(\alpha) \\ &= 0 \end{aligned}$$

$f(r) = 0 \Rightarrow r \in N_f$. Thus any vector β can be written as $r + c.\alpha$, $r \in N_f$ and $c \in F$. This shows that the subspace spanned by N_f and α is V, for every $\alpha \in V$, which is not in N_f. Hence N_f is a maximal proper subspace of V that is a hyperspace in V.

Conversely suppose that N is a hyper space in V. Fix a vector α in V, which is not in N. Then the subspace spanned by N and α in V, since N is a maximal proper subspace of V. So every vector $\beta \in V$ has the form $\beta = r + c.\alpha$, $r \in V$, $c \in F$. This vector r and the scalar c are uniquely determined by β. Since if $\beta = r' + c'.\alpha$, $r' \in V$, $c' \in F$, then

$$\begin{aligned} 0 &= (r - r') + (c - c')\alpha \\ &\Rightarrow (r - r') = (c' - c)\alpha \end{aligned}$$

If $(c' - c) \neq 0$, then $\alpha = \frac{(r-r')}{(c'-c)}$ would belong to N, which is against our assumption. Hence $c' - c = 0$. Thus $c' = c$ and $r' = r$. Thus for every $\beta \in V$, there is a unique scalar c in F such that $\beta - c\alpha \in N$.

Define $g(\beta)$ as this scalar c that is $g(\beta) = c$, if $\beta - c.\alpha \in N$. Thus g is a linear functional on V and the nullspace of g is N. Also

$$\begin{aligned} \beta \in N_g &\Leftrightarrow g(\beta) = 0 \\ &\Leftrightarrow \beta - 0.\alpha \in N \\ &\Leftrightarrow \beta \in N. \end{aligned}$$

Thus the nullspace of g is N. Since null space of g not equal to V, we get $g \neq 0$. Hence N is the null space of a non-zero linear functional g on V. □

Lemma 2.8. *If f and g are linear functions on a vector space V, then g is a scalar multiple of f if and only if the null space of g contains the null space of f, that is if and only if $f(\alpha) = 0$ implies $g(\alpha) = 0$.*

Proof. If then g is a scalar multiple of f , then $f(\alpha) = 0$ implies that $g(\alpha) = 0$.

Conversely assume that $f(\alpha) = 0$ implies that $g(\alpha) = 0$. Consider the case when $f = 0$,, then $g = 0$ as well and g is trivially a scalar multiple of f. Suppose $f \neq 0$, then the null space N_f is a hyperspace in V. Choose some vector $\alpha \in V$ with $f(\alpha \neq 0)$ and let

$$c = \frac{f(\alpha)}{g(\alpha)}.$$

Define $h = g - cf$. Let $\beta \in N_f$, then $f(\beta) = 0$. Thus by assumption, $g(\beta) = 0$. Thus the linear functional $h = g - cf$ is 0 on N_f, since both f and g are 0 there and $h(\beta) = g(\beta) - cf(\beta)$. Thus h is 0 on the subspace spanned by N_f and α, and that is V. We conclude that $h = 0$, that is $g = cf$. Hence the result. □

Theorem 2.21. *Let $g, f_1, \ldots, f_r$ be linear functional on a vector space V with respective null spaces $N, N_1, \ldots, N_r$. Then g is a linear combination of $f_1, \ldots, f_r$ if and only if N contains the intersection $N_1 \cap \ldots \cap N_r$.*

Proof. Assume that g is a linear combination of $f_1, \ldots, f_r$. That is

$$g = c_1 f_1 + \ldots + c_r f_r. \tag{2.21}$$

Let N denotes the null space of g. Let $N_1, \ldots, N_r$ denotes the null space of $f_1, \ldots, f_r$. Then if $f_i(\alpha) = 0$ for each $i = 1, 2 \ldots, r$, then clearly $g(\alpha) = 0$. from equation 2.21.

$$\begin{aligned}
\alpha \in N_1 \cap \ldots \cap N_r \quad &\Rightarrow \quad \alpha \in N_i \text{ for } i = 1, 2, \ldots, r \\
&\Rightarrow \quad f_i(\alpha) = 0 \text{ for } i = 1, 2, \ldots, r \\
&\Rightarrow \quad g(\alpha) = 0 \\
&\Rightarrow \quad \alpha \in N
\end{aligned}$$

$$\text{Thus } N_1 \cap \ldots \cap N_r \subseteq N$$

Therefore N contains $N_1 \cap \ldots \cap N_r$.

We shall prove the converse by Mathematical induction on the number r. Consider the case when $r = 1$. By the preceding lemma, the result is true when $r = 1$.

Suppose that the result is true when $r = k - 1$. Let $g, f_1, \ldots, f_k$ be linear functionals with respective null spaces $N, N_1, \ldots, N_k$ such that $N_1 \cap \ldots \cap N_k$ is contained in N. Let $g', f_1', \ldots, f_{k-1}'$ be the restrictions of $g, f_1, \ldots, f_{k-1}$ respectively to the subspace N. Then $g', f_1', \ldots, f_{k-1}'$ are linear functionals on the vector space N_k. Also if $f_i'(\alpha) = 0$ for $i = 1, 2, \ldots, k - 1$, for some $\alpha \in N_k$, then $\alpha \in N_1 \cap \ldots \cap N_{k-1} \cap N_k$.

$$\begin{aligned}
\alpha \in N_k \text{ and } f_i'(\alpha) = 0 \text{ for } i = 1, 2, \ldots, k-1 \quad &\Rightarrow \quad f_1(\alpha) = 0, \ldots, f_{k-1}(\alpha) = 0 \\
&\Rightarrow \quad \alpha \in N_1, \ldots, \alpha \in N_{k-1} \\
&\Rightarrow \quad \alpha \in N_1 \cap \ldots \cap N_{k-1} \text{ and } \alpha \in N_k \\
&\Rightarrow \quad \alpha \in N_1 \cap \ldots \cap N_k
\end{aligned}$$

$$\begin{aligned}
&\Rightarrow \alpha \in N \text{ since } N \text{ contains } \alpha \in N_1 \cap \ldots \cap N_k \\
&\Rightarrow g(\alpha) = 0 \\
&\Rightarrow g'(\alpha) = 0
\end{aligned}$$

This implies that the null space of g' contains the intersection of the nullspaces of $f_1', f_2', \ldots, f_{k-1}'$. By the induction hypothesis for the case $r = k - 1$, we have g' is a linear combination of $f_1', f_2', \ldots, f_{k-1}'$. Therefore there exists scalars $c_1, c_2, \ldots, c_{k-1}$ such that

$$g' = \sum_{i=1}^{k-1} c_i f_i'.$$

Now define $h = g - \sum_{i=1}^{k-1} c_i f_i$. Then h is a linear functional on V. Now

$$\begin{aligned}
f_k(\alpha) = 0 \;&\Rightarrow\; \alpha \in N_k \\
&\Rightarrow\; g'(\alpha) = \sum_{i=1}^{k-1} c_i f_i'(\alpha) \\
&\Rightarrow\; g(\alpha) = \sum_{i=1}^{k-1} c_i f_i(\alpha) \\
&\Rightarrow\; g(\alpha) - \sum_{i=1}^{k-1} c_i f_i(\alpha) = 0 \\
&\Rightarrow\; h(\alpha) = 0
\end{aligned}$$

Thus the null space of h contains the null space of f_k. Hence by previous lemma, h is a scalar multiple of $f(k)$. So let $h = c_k f_k$. Thus $g - \sum_{i=1}^{k-1} c_i f_i = c_k f_k \Rightarrow g = \sum_{i=1}^{k-1} c_i f_i + c_k f_k = \sum_{i=1}^{k} c_i f_i$. Thus g is a linear combination of $f_1, \ldots, f_k$. Hence the result. $\square$

Example 31. Using above theorem, prove the following: If W is a subspace of a finite dimensional vector space V and if $\{g_1, g_2, \ldots, g_r\}$ is any basis for W^0, then

$$W = \bigcap_{i=1}^{r} N_{g_i}$$

Solution: Let $\alpha \in W$. Since $g_1, g_2, \ldots, g_r$ are in W^0, $g_1(\alpha) = 0$, $g_2(\alpha) = 0, \ldots, g_r(\alpha) = 0$. This means that $\alpha \in N_{g_1}$, $\alpha \in N_{g_2}, \ldots$, $\alpha \in N_{g_r}$. Hence $\alpha \in \bigcap\limits_{i=1}^{r} N_{g_i}$. Thus

$$W \subseteq \bigcap_{i=1}^{r} N_{g_i}. \tag{2.22}$$

Conversely let $\alpha \in \bigcap\limits_{i=1}^{r} N_{g_i}$. Then $\alpha \in N_{g_1}$, $\alpha \in N_{g_2}, \ldots$, $\alpha \in N_{g_r}$. Means that $g_1(\alpha) = 0$, $g_2(\alpha) = 0, \ldots, g_r(\alpha) = 0$,. since $\{g_1, g_2, \ldots, g_r\}$ is any basis for W^0, every $g \in W^0$ can be written as a linear combination of $g_1, g_2, \ldots, g_r$. Since each $g_i(\alpha) = 0$,, we get

$$g(\alpha) = 0 \text{ for } g \in W^0.$$

Thus

$$\alpha \in W^{0^0} = W.$$

Hence $\alpha \in W$. Hence we get

$$\bigcap_{i=1}^{r} N_{g_i} \subseteq W. \tag{2.23}$$

From equations 2.22 and 2.23 we can conclude that

$$W = \bigcap_{i=1}^{r} N_{g_i}.$$

Exersices

1. Let n be a positive integer and F be a field. Let W be the set of all vectors $(x_1, \ldots, x_n) \in F^n$ such that $x_1 + \ldots + x_n = 0$. Then show that W^0 consists of all linear functionals f of the form

$$f(x_1, \ldots, x_n) = c\sum_{i=1}^{n} x_j.$$

2.7 The Transpose of a Linear Transformation

Suppose V and W be two vector spaces over the same field F. Let T be a linear transformation V into W. Then T induces a linear transformation from W^* into V^*. Suppose g is a linear functional on W, and let

$$f(\alpha) = g(T(\alpha)), \alpha \in V \tag{2.24}$$

Then Equation 2.24 defines a function f from V into F, namely the composition of T, a function from V into W, with g, a function from W into F. Since both T and g are linear and composition is linear, f is a linear functional on V. Thus T provides us with a rule T^t which associates with each linear functional g on W a linear functional $f = T^t g$ on V, defined by Equation 2.24. Note that T^t is a linear transformation from W^* into V^*, for if g_1 and g_2 are in W^* and c is a scalar

$$\begin{aligned}
[T^t(cg_1 + g_2)](\alpha) &= (cg_1 + g_2)(T\alpha) \\
&= cg_1(T\alpha) + g_2(T(\alpha)) \\
&= c(T^t g_1)(\alpha) + (T^t g_2)(\alpha)
\end{aligned}$$

Thus $T^t(cg_1 + g_2) = c.T^t(g_1) + T^t(g_2)$.

Theorem 2.22. *Let V and W be vector spaces over the field F. For each linear transformation T from V into W, there is a unique linear transformation T^t from W^* into V^* such that $(T^t g)(\alpha) = g(T\alpha)$ for every $g \in W^*$ and $\alpha \in V$.*

If g_1 and g_2 are in W^* and c is a scalar

$$\begin{aligned}
[T^t(cg_1 + g_2)](\alpha) &= (cg_1 + g_2)(T\alpha) \\
&= cg_1(T\alpha) + g_2(T(\alpha)) \\
&= c(T^t g_1)(\alpha) + (T^t g_2)(\alpha)
\end{aligned}$$

Thus $T^t(cg_1 + g_2) = c.T^t(g_1) + T^t(g_2)$. We call T^t the Transpose of T or adjoint of T.

Theorem 2.23. *Let V and W be vector spaces over the field F. Let T be a linear transformation from V into W. The null space of T^t is the annihilator of the range of T. If V and W are finite dimensional, then* *(a) rank* (T^t)*=rank* (T)

(b) the range of T^t *is the annihilator of the null space of* T*.*

Proof. If $g \in W^*$, the by definition,

$$T^t g(\alpha) = g(T\alpha).$$

$$\begin{aligned} g \in \text{ null space of } T^t &\Leftrightarrow T^t(g) = 0 \\ &\Leftrightarrow T^t g(\alpha) = 0 \quad \forall \alpha \in V \\ &\Leftrightarrow g(T\alpha) = 0 \quad \forall \alpha \in V \\ &\Leftrightarrow g \text{ is the annihilator of the range of } T \end{aligned}$$

Thus the null space of T^t is the annihilator of the range of T. Suppose $\dim V = n$ and $\dim W = m$.

(a) Let r be the rank of T, that is the range of T is r. By Theorem 2.17, dimension of annihilator of the range of T is $m - r$. We observed in the beginning of the proof that null space of T^t is the annihilator of the range of T. Hence nullity of $T^t = m - r$. By rank nullity theorem, rank T^t+ Nullity $T^t = dimW = m$.
Hence rank $T^t = m - (m - r) = r$.

(b) Let N be the null space of T. If f is a linear functional in the range of T^t, then $f = T^t g$ for some $g \in W^*$. Hence for all $\alpha \in N$, then

$$\begin{aligned} f(\alpha) &= (T^t g)(\alpha) \\ &= g(T\alpha) \\ &= g(0) \\ &= 0 \end{aligned}$$

Hence $f \in N^0$. Thus the range of T^t is a subspace of N^0. But

$$\begin{aligned} dimN^0 &= dimV - dimN \\ &= dimV - NullityT \\ &= RankT \\ &= Rank(T^t) \\ &= \text{ dimension of the range of } T^t \end{aligned}$$

Hence dimension of the annihilator of the null space of T is same as dimension of the range of T^t. Thus we get the range of T^t is the annihilator of the null space of T.

□

Theorem 2.24. *Let V and W be vector spaces over a field F. Let $\mathscr{B}$ be an ordered basis for V with dual basis $\mathscr{B}^*$ and let $\mathscr{B}'$ be an ordered basis for V with dual basis $\mathscr{B}'^*$. Let T be a linear transformation from V into W. Let A be the matrix of T relative to $\mathscr{B}$, $\mathscr{B}'$ and let B be the matrix of T^t relative to $\mathscr{B}'^*$, $\mathscr{B}^*$. Then $B_{ij} = A_{ij}$.*

Proof. Let

$$\mathscr{B} = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$$

$$\mathscr{B}' = \{\beta_1, \beta_2, \ldots, \beta_m\},$$

$$\mathscr{B}^* = \{f_1, f_2, \ldots, f_n\}$$

$$\mathscr{B}'^* = \{g_1, g_2, \ldots, g_m\}.$$

By the definition,

$$T\alpha_j = \sum_{i=1}^{m} A_{ij}\beta_i, \; j = 1, 2, \ldots, n. \tag{2.25}$$

$$T^t g_j = \sum_{i=1}^{m} B_{ij} f_i,\ j = 1, 2, \ldots, m. \tag{2.26}$$

We have

$$\begin{aligned} T^t g_j(\alpha_i) &= g_j T(\alpha_i) \\ &= g_j(\sum_{k=1}^{m} A_{ki}\beta_k) \\ &= \sum_{k=1}^{m} A_{ki} g_j(\beta_k) \\ &= \sum_{k=1}^{m} A_{ki}\delta_{jk} \\ &= A_{ji} \end{aligned}$$

Thus

$$T^t g_j(\alpha_i) = A_{ji} \tag{2.27}$$

From Equation2.13, we have for any linear functional f on V,

$$f = \sum_{i=1}^{n} fi(\alpha) f_i.$$

apply this to the linear functional $f = T^t g_j$ and use the fact that $T^t g_j = \sum_{i=1}^{n} = A_{ji}$, we have

$$T^t g_j = \sum_{i=1}^{n} A_{ji} f_i \tag{2.28}$$

From Equations 2.26 and 2.28, we get $B_{ij} = A_{ji}$. □

If T is a linear transformation from V into W whose matrix in some pair of bases is a matrix A, then the transpose transformation T^t is represented in the dual pair of bases by the transpose matrix A^t.

Definition 2.12. *If A is an $m \times n$ matrix over the field F, the transpose A is the $n \times m$ matrix A^t defined by $A^t_{ij} = A_{ji}$.*

Theorem 2.25. *Let A be any $m \times n$ matrix over the field F. Then the row rank of A is equal to the column rank of A.*

Proof. Let $\mathscr{B} = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ be the standard ordered basis for F^n and $\mathscr{B}' = \{\beta_1, \beta_2, \ldots, \beta_m\}$, be the standard ordered basis for F^m. Let T be the linear transformation from F^n to F^m such that the matrix of T relative to the pair $\mathscr{B}, \mathscr{B}'$ is A, That is $T(x_1, x_2, \ldots, x_n) = (y_1, y_2, \ldots, y_m)$ where

$$\sum_{j=1}^{n} A_{ij} x_j.$$

The column vectors of the matrix A spans the range of T. Hence column space of A is same as the range of T. Therefore rank$T$ = column rank (A).

Relative to the dual bases $\mathscr{B}'^*$ and $\mathscr{B}^*$, matrix of the transformation T^t is the matrix A^T. Hence rankT^t = column rank (A^T). But the the rows of A are columns of A^T Thus

$$\begin{aligned}
\text{Row rank of A} &= \text{Rank } (T^t) \\
&= \text{Rank } (T) \\
&= \text{column rank of } A
\end{aligned}$$

Thus Row rank of A = column rank of A. □

Definition 2.13. *If A is an $m \times n$ matrix over the field F and T is the linear transformation from F^n into F^m, then rank(T) = row rank (A) = column rank (A) and we call this number simply the rank of A.*

Example 32. Find the rank and nullity of the linear transformation $T : R^3 \to R^3$ defined by

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

Consider the standard ordered basis $B = \{e_1, e_2, e_3\}$. Then matrix of T with respect to this basis is

$$T_B = \begin{bmatrix} 1 & -1 & 2 \\ 2 & 1 & 0 \\ -1 & -2 & 2 \end{bmatrix}$$

This matrix can be reduced to

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

Thus rank $A = 2$ and therefore rank$T = 2$. Nullity $T = 1$.

Exercises

1. Let F be a field and let f be the linear functional on F^2 defined by $f(x_1, x_2) = ax_1 + bx_2$. For each of the linear operators T, let $g = T^t f$, and find $g(x_1, x_2)$.

   (a) $T(x_1, x_2) = (0, x_2)$;

   (b) $T(x_1, x_2) = (2x_2, x_1)$;

   (c) $T(x_1, x_2) = (x_1 + x_2, x_1 - x_2)$.

2. Let V be the vector space of all polynomial functions over the field of real numbers and let f be the linear functional on V defined by

$$f(p) = \int_a^b p(x)dx.$$

   If D is the differentiation operator on V, what is $D^t f$?

3. Let V be the space of all $n \times n$ matrices over a field F and let B be a fixed $n \times n$ matrix. If T is the linear operator on V defined by $T(A) = AB - BA$, and if f is the trace function, what is $T^t f$?

Chapter 3

Elementary Canonical Forms

3.1 Characteristic values

Definition 3.1. *Let V be a vector space over the field F and let T be a linear operator on V. A characteristic value of T is a scalar c in F such that there is a non-zero vector $\alpha \in V$ with $T(\alpha) = c.\alpha$. If c is a characteristic value of T then any $\alpha \in V$ such that $T(\alpha) = c.\alpha$ is called a characteristic vector of T associated with the characteristic value c.*

Note 3.1. The collection of all α such that $T(\alpha) = c.\alpha$ is called the characteristic space associated with c. Characteristic values are often called characteristic roots, latent roots eigenvalues, proper values, or spectral values. In this book we shall use only the name characteristic values.

If T is any linear operator and c is any scalar, the set of all characteristic vectors of T is a subspace of V. It is the null space of the linear transformation $(T-cI)$. We call c a characteristic value of T if this subspace is different from the zero subspace, That is, if $(T-cI)$ fails to be one-one. If the underlying space V is finite-dimensional, $(T-cI)$ fails to be one-one precisely when its determinant is different from 0.

Theorem 3.1. *Let T be a linear operator on a finite-dimensional space V*

and let c be a scalar. The following are equivalent.

1. *c is a characteristic value of T.*

2. *The operator $(T - cI)$ is singular (not invertible).*

3. *det $(T - cI) = 0$.*

Proof. (1)$\Rightarrow$ (2)

Assume that c is a characteristic value of T.

Then

$$\begin{aligned}
c \text{ is a characteristic value of } T \quad &\Leftrightarrow \quad \text{there exists } \alpha \neq 0, \text{ in } V \text{ such that } T(\alpha) = c.\alpha. \\
&\Leftrightarrow \quad \text{there exists } \alpha \neq 0, \text{ in } V \text{ such that } T(\alpha) - c.\alpha = 0 \\
&\Leftrightarrow \quad T - cI \text{ is not one-one} \\
&\Leftrightarrow \quad T - cI \text{ is not invertible, since V is finite dimensional}
\end{aligned}$$

(2) $\Leftrightarrow$ (3)

Assume that the operator $(T - cI)$ is not invertible. That is if and only if the matrix of $T - cI$ with respect to any basis is not invertible. And that is if and only if determinant of $T - cI$ is not equal to zero. Hence the result. $\square$

The determinant criteria is very important and we use this to trace out characteristic values of T. We have det$T - cI$ is a polynomial of degree n in the variable c, we can get characteristic values as the roots of this polynomial.

If $\mathscr{B}$ is any ordered basis for V, and $A = [T]_B$, then $(T - cI)$ is invertible if and only if the matrix $(A - cI)$ is invertible. Using this idea we define characteristic value of a matrix as follows:

Definition 3.2. *If A is an $n \times n$ matrix over the field F, a characteristic value of A in F is a scalar c in F such that the matrix $(A - cI)$ is singular (not invertible).*

Since c is a characteristic value of A if and only if $\det(A–cI) = 0$, or equivalently if and only if $\det(cI - A) = 0$, we form the matrix $(xI - A)$ with polynomial entries, and consider the polynomial $f = det(xI - A)$. Clearly the characteristic values of A in F are just the scalars $c \in F$ such that $f(c) = 0$. For this reason f is called the characteristic polynomial of A. It is important to note that f is a monic polynomial which has degree exactly n.

Lemma 3.2. *Similar matrices have the same characteristic polynomial.*

Proof. Suppose A and B are $n \times n$ matrices over the field F and $B = P^{-1}AP$. Then

$$\begin{aligned} xI - B &= xI - (P^{-1}AP) \\ &= P^{-1}PxI - P^{-1}AP \\ &= P^{-1}xIP - P^{-1}AP \\ &= P^{-1}[xI - A]P \end{aligned}$$

Thus

$$\begin{aligned} det(xI - B) &= det(P^{-1}[xI - A]P) \\ &= det(P^{-1})det[xI - A]det(P) \\ &= det(P^{-1})det(P)det[xI - A] \\ &= det(P^{-1}P)det[xI - A] \\ &= det[xI - A]. \end{aligned}$$

□

Remark 10. Let V be a finite dimensional vector space over the field F with dimension n. Let T be a linear operator on V. Let $\mathscr{B}$ and $\mathscr{B}'$ be two ordered bases for V. Then matrices associated with these bases are similar matrices as we have noted earlier. By above lemma similar matrices have same characteristic polynomial. So we define the **characteristic polynomial** of the linear operator T as the characteristic polynomial of any $n \times n$ matrix which represents T in some ordered basis for V.

Just as for matrices, the roots of the characteristic polynomial of T are the characteristic values of T. Since the characteristic polynomial is of degree n over the field F, the linear operator cannot have more than n distinct characteristic values. In some cases T cannot have any characteristic value at all.

Example 33. Let $T : R^2 \to R^2$ be defined as $T(x_1, x_2) = (-x_2, x_1)$. Then the matrix related to standard ordered basis is $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ The characteristic polynomial for T is given by

$$det(xI - A) = det \begin{bmatrix} x & 1 \\ -1 & x \end{bmatrix} = x^2 + 1$$

This polynomial has no roots in R. So T has no characteristic values in R.

If U is the linear operator on C^2 which is represented by A in the standard ordered basis, then U has two characteristic values i and $-i$.

The above example says that while we are discussing characteristic values, we have to specify the scalar field involved. The above matrix has no characteristic value in R, while it has two characteristic values in C.

Example 34. Let A be the real 3×3 matrix $\begin{bmatrix} 3 & 1 & -1 \\ 2 & 2 & -1 \\ 2 & 2 & 0 \end{bmatrix}$ The characteristic polynomial of A is given by $\det(xI - A) = 0$.

$$det(xI - A) = \begin{vmatrix} x-3 & -1 & 1 \\ -2 & x-2 & 1 \\ -2 & -2 & x \end{vmatrix}$$

$$
\begin{aligned}
&= (x-3)[(x-2)x+2]+1(-2x+2)+1[4+2(x-2)\\
&= x^3-2x^2+2x-3x^2+6x-6-2x+2+4+2x-4\\
&= x^3-5x^2+8x-4\\
&= (x-1)(x^2-4x+4)\\
&= (x-1)(x-2)^2
\end{aligned}
$$

Therefore characteristic values are given by $x=1$ and $x=2$.

Now we have to determine characteristic vector corresponding to these eigenvalues. Let T be a linear operator, which is represented in the standard ordered basis by the matrix A. Then $T: R^3 \to R^3$ is given by:

$$
\begin{aligned}
T(x_1,x_2,x_3) &= \begin{bmatrix} 3 & 1 & -1 \\ 2 & 2 & -1 \\ 2 & 2 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\\
&= (3x_1+x_2-x_3, 2x_1+2x_2-x_3, 2x_1+2x_2)
\end{aligned}
$$

<u>The characteristic vectors associated with the characteristic value 1:</u>

We have to find $(x_1,x_2,x_3) \in R^3$ such that $T(x_1,x_2,x_3) = 1.(x_1,x_2,x_3)$ We have

$$
\begin{aligned}
A-I &= \begin{bmatrix} 3 & 1 & -1 \\ 2 & 2 & -1 \\ 2 & 2 & 0 \end{bmatrix} - \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\\
&= \begin{bmatrix} 2 & 1 & -1 \\ 2 & 1 & -1 \\ 2 & 2 & -1 \end{bmatrix}
\end{aligned}
$$

Since first two rows of this matrix is same, its rank≤ 2.

$$
\begin{aligned}
\begin{bmatrix} 2 & 1 & -1 \\ 2 & 1 & -1 \\ 2 & 2 & -1 \end{bmatrix} &\sim \begin{bmatrix} 1 & 1/2 & -1/2 \\ 1 & 1/2 & -1/2 \\ 1 & 1 & -1/2 \end{bmatrix}\\
&\sim \begin{bmatrix} 1 & 1/2 & -1/2 \\ 0 & 0 & 0 \\ 0 & -1/2 & 0 \end{bmatrix}
\end{aligned}
$$

$$\sim \begin{bmatrix} 1 & 1/2 & -1/2 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

This implies that Rank of $T-I$ is 2. Since dimension of R^3 is 3, we have nullity of $T-I$ is 1. Thus the null space of $T-I$ is one dimensional. In other words the space of characteristic vectors of T associated with characteristic value 1 is one dimensional.

We have by inspection,

$$\begin{aligned} (T-I)(1,0,2) &= \begin{bmatrix} 1 & 1/2 & -1/2 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} \\ &= \begin{bmatrix} 1-1 \\ 0 \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \end{aligned}$$

Thus $(1,0,2)$ is in the null space of $T-I$. Thus $(1,0,2)$ is a characteristic vector. Since characteristic space is one dimensional, the vector $(1,0,2)$ spans the same.

<u>The characteristic vectors associated with the characteristic value 2:</u>

We have to find $(x_1,x_2,x_3) \in R^3$ such that $T(x_1,x_2,x_3) = 2.(x_1,x_2,x_3)$

We have

$$\begin{aligned} A-2I &= \begin{bmatrix} 3 & 1 & -1 \\ 2 & 2 & -1 \\ 2 & 2 & 0 \end{bmatrix} - \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 1 & -1 \\ 2 & 0 & -1 \\ 2 & 2 & -2 \end{bmatrix} \end{aligned}$$

$$\begin{aligned} \begin{bmatrix} 1 & 1 & -1 \\ 2 & 0 & -1 \\ 2 & 2 & -2 \end{bmatrix} &\sim \begin{bmatrix} 1 & 1 & -1 \\ 2 & 0 & -1 \\ 1 & 1 & -1 \end{bmatrix} \\ &\sim \begin{bmatrix} 1 & 1 & -1 \\ 1 & 1 & -1 \\ 2 & 0 & -1 \end{bmatrix} \end{aligned}$$

$$\sim \begin{bmatrix} 1 & 1 & -1 \\ 0 & 0 & 0 \\ 0 & -2 & 1 \end{bmatrix}$$

Hence rank of $A - 2I$ is 2 and nullity is 1. Hence null space of $A - 2I$ is one dimensional.

$$\begin{aligned}
(T - 2I)(1,1,2) &= \begin{bmatrix} 1 & 1 & -1 \\ 0 & 0 & 0 \\ 0 & -2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} \\
&= \begin{bmatrix} 1+1-2 \\ 0 \\ -2+2 \end{bmatrix} \\
&= \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}
\end{aligned}$$

Thus $(1,1,2)$ is in null space of $T - 2I$. Evidently $T\alpha = 2\alpha$ if and only if α is a scalar multiple of $(1,1,2)$.

Remark 11. 1. Characteristic polynomial of the identity operator will be given by

$$\begin{aligned}
\det\,(xI - I) &= \begin{vmatrix} x & 0 & \dots & 0 \\ 0 & x & \dots & 0 \\ \vdots & & & \\ 0 & 0 & \dots & x \end{vmatrix} - \begin{vmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & & & \\ 0 & 0 & \dots & 1 \end{vmatrix} \\
&= \begin{vmatrix} (x-1) & 0 & \dots & 0 \\ 0 & (x-1) & \dots & 0 \\ \vdots & & & \\ 0 & 0 & \dots & (x-1) \end{vmatrix} \\
&= (x-1)(x-1)\dots(x-1) \\
&= (x-1)^n
\end{aligned}$$

2. Characteristic polynomial of the zero operator will be given by

$$\det\,(xI - 0I) \;=\; \begin{vmatrix} x & 0 & \dots & 0 \\ 0 & x & \dots & 0 \\ \vdots & & & \\ 0 & 0 & \dots & x \end{vmatrix} - 0 = (x)(x)\dots(x) = x^n$$

Definition 3.3. *Let T be a linear operator on the finite dimensional space V. We say that T is* ***diagonalizable*** *if there exists a basis for V each vector of which is a characteristic vector of T.*

The following gives idea about the name diagonalizable:
If there exists a basis $\mathscr{B} = \{\alpha_1, \alpha_2, \dots, \alpha_n\}$ for V such that each α_i is a characteristic vector of T, then the matrix of T in the ordered basis β is a diagonal matrix. If $T(\alpha_i) = c_i\alpha_i$, then

$$[T]_{\mathscr{B}} = \begin{bmatrix} c_1 & 0 & \dots & 0 \\ 0 & c_2 & \dots & 0 \\ \vdots & & & \\ 0 & 0 & \dots & c_n \end{bmatrix}$$

Note that the scalars $c_1, c_2, \dots, c_n$ need not be distinct.
Thus the linear operator T is diagonalizable, if there is a basis $\mathscr{B}$ for V consisting of characteristic vectors of T and then the matrix of T in the ordered basis $\mathscr{B}$ is a diagonal matrix for which the entries in the main diagonal are characteristic values of T.

We can also say that T is diagonalizable if the characteristic vectors of T span V. This is because, we can select a basis out of any spanning set of vectors.

Remark 12.

1. In Example 33, we have a linear operator on R^2 which is not diagonalizable. That operator has no characteristic values and hence no characteristic vectors. Therefore there exists no basis for R^2 consisting of characteristic vectors of T.

2. In Example 34, we have a linear operator on R^3 which is not diagonalizable. That operator has two characteristic values 1 and 2, and

the space of characteristic vectors corresponding to these characteristic values are one dimensional. And R^3 is 3 dimensional. Hence a set containing characteristic vectors alone cannot span R^3. Therefore there exists no basis for R^3 consisting of characteristic vectors alone. Thus T is not diagonalizable.

Lemma 3.3. *Suppose that $T\alpha = c\alpha$. If f is any polynomial, then $f(T)\alpha = f(c)\alpha$.*

Proof. Let $f(x) = c_0 + c_1x + c_2x^2 + \ldots + c_nx^n$. Then $f(T) = c_0I + c_1T + c_2T^2 + \ldots + c_nT^n$.
We have $T\alpha = c\alpha$.
Then $T^2\alpha = T(T\alpha) = T(c\alpha) = c.T(\alpha) = c^2\alpha$.
Similarly we can prove that $T^3\alpha = c^3\alpha, \ldots, T^n\alpha = c^n\alpha$
Thus

$$\begin{aligned} f(T)(\alpha) &= c_0I(\alpha) + c_1T(\alpha) + c_2T^2(\alpha) + \ldots + c_nT^n(\alpha) \\ &= c_0\alpha + c_1c.\alpha + c_2c^2\alpha + \ldots + c_nc^n(\alpha) \\ &= (c_0 + c_1c + c_2c^2 + \ldots + c_nc^n)(\alpha) \\ &= f(c)\alpha. \end{aligned}$$

$\square$

Lemma 3.4. *Let T be a linear operator on the finite dimensional space V. Let $c_1, \ldots, c_k$ be the distinct characteristic values of T and let W_i be the space of characteristic vectors associated with the characteristic value c_i. If $W = W_1 + \ldots + W_k$, then*

$$dimW = dimW_1 + \ldots + dimW_k.$$

In fact, if $\mathscr{B}_i$ is an ordered basis for W_i, then $\mathscr{B} = (\mathscr{B}_1, \ldots, \mathscr{B}_k)$ is an ordered basis for W.

Proof. We know that if $W = W_1 + \ldots + W_k$, then W is the subspace spanned by $W_1 \cup \ldots \cup W_k$. In other words W is the subspace of V spanned by all characteristic vectors of T. Now suppose that for each i, we have a vector β_i in W_i and assume that

$$\beta_1 + \beta_2 + \ldots + \beta_k = 0$$

We will show that each $\beta_i = 0$.

Each β_i is a characteristic vector, therefore $T\beta_i = c_i\beta_i$ for each $i = 1, 2, \ldots, k$.

Let f be any polynomial. Then previous lemma ensures that $f(T)\beta_i = f(c_i)\beta_i$. Now

$$\begin{aligned} 0 &= f(T)0 \\ &= f(T)(\beta_1 + \beta_2 + \ldots + \beta_k) \\ &= f(T)\beta_1 + f(T)\beta_2 + \ldots + f(T)\beta_k \\ &= f(c_1)\beta_1 + f(c_2)\beta_2 + \ldots + f(c_k)\beta_k \end{aligned}$$

Thus for any polynomial f, we have

$$f(c_1)\beta_1 + f(c_2)\beta_2 + \ldots + f(c_k)\beta_k = 0.$$

Choose polynomials $f_1, \ldots, f_k$ such that $f_i(c_j) = \delta_{ij}$

For each $i = 1, 2, \ldots, k$, we have

$$\begin{aligned} 0 &= f_i(T).0 \\ &= \sum_{j=1}^{k} \delta_{ij}\beta_j \\ &= \beta_j \end{aligned}$$

Now let $\mathscr{B}_i$ be an ordered basis for W_i and let $\mathscr{B}$ be the sequence $\mathscr{B} = \{\mathscr{B}_1, \ldots, \mathscr{B}_k\}$. Since $\mathscr{B}_i$ spans W_i for each i, we see that the $\mathscr{B}$ spans W. Now we show that $\mathscr{B}$ is linearly independent. Any linear relation between the vectors of $\mathscr{B}$ is of the form $\beta_1 + \ldots + \beta_k = 0.$, where $\beta_i \in W_i$ is a linear combination of elements of $\mathscr{B}_i$. Then from the above argument each $\beta_i = 0$. Since each $\mathscr{B}_i$ is linearly independent(Being a basis) , we see that we have only the trivial linear relation between the vectors in $\mathscr{B}$. Hence $\mathscr{B}$ is linearly independent. Hence it is a basis for W. □

Theorem 3.2. *Let T be a linear operator on the finite dimensional space V. Let $c_1, \ldots, c_k$ be the distinct characteristic values of T and let W_i be the null space of $T - c_iI$. the following are equivalent.*

1. *T is diagonalizable.*
2. *The characteristic polynomial for T is*

$$f = (x - c_1)^{d_1} \ldots (x - c_k)^{d_k}$$

and $dimW_i = d_i, i = 1, \ldots, k$.

3. *$dimW_1 + \ldots + dimW_k = dimV$.*

Proof. $(i) \Rightarrow (ii)$
Suppose T is diagonalizable. Then there is basis for V consisting of characteristic vectors alone. Then the matrix of T in the ordered basis $\mathscr{B}$ is a diagonal matrix whose diagonal entries are the characteristic values $c_1, c_2, \ldots, c_k$ where each c_i repeated a certain number of times. If c_i is repeated d times in the

main diagonal of the matrix $[T]_{\mathscr{B}}$, then it has the following form

$$\begin{bmatrix} c_1 I_1 & 0 & \dots & 0 \\ 0 & c_2 I_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & c_k I_k \end{bmatrix},$$

where I_1 is the $d_1 \times d_1$ identity matrix, I_2 is the $d_2 \times d_2$ identity matrix and so on.

Fom this matrix, we can see that the characteristic polynomial of T is

$$f = \det\,(xI - [T]_{\mathscr{B}}) = (x - c_1)^{d_1}(x - c_2)^{d_2} \dots (x - c_k)^{d_k}.$$

Now we prove that $\dim W_i = d_i, i = 1, 2, \dots, k$. Since W_i is the null space of $T - c_i I$, the dimension of W_i is the nullity of $T - c_i I$. But the matrix of $T - c_i I$ in ordered basis is a diagonal matrix $[T - c_i I]_{\mathscr{B}}$. The nullity of a diagonal matrix is equal to the number of zero entries which it has on its main diagonal, and the matrix $[T - c_i I]_{\mathscr{B}}$ has d_i zeros on its main diagonal. Hence $\dim W_i = d_i, i = 1, 2, \dots, k$.

$(2) \Rightarrow (3)$

Assume that the characteristic polynomial for T is

$$f = (x - c_1)^{d_1} \dots (x - c_k)^{d_k}$$

and $\dim W_i = d_i, i = 1, \dots, k$.

Since T is a linear operator on the finite dimensional vector space V, we know that the characteristic polynomial f of T is of degree n, where $n = \dim V$. But the degree of the polynomial f is $d_1 + \dots + d_k$. Hence $d_1 + \dots + d_k = n = dimV$. That is $\dim W_1 + \dots + \dim\ W_n = dimV$.

$(3) \Rightarrow (1)$

Suppose that dim$W_1 + \ldots +$ dim $W_n = dimV$. Let $W = W_1 + \ldots + W_k$. Then W is the subspace of V spanned by $W_1 \cup \ldots \cup W_k$. In other word, W is the subspace of V spanned by all the characteristic vectors of T. By the previous lemma, dim$W_1 + \ldots +$ dim $W_n = dimW$. Thus we can conclude that $dimW = dimV$. Hence $W = V$. Thus the subspace spanned by all the characteristic vectors of T is V. Thus T is diagonalizable. $\square$

Example 35. Let T be a linear operator on R^3 which is represented in the standard ordered basis by the matrix

$$\begin{bmatrix} 5 & -6 & -6 \\ -1 & 4 & 2 \\ 3 & -6 & -4 \end{bmatrix}$$

The characteristic polynomial of T is det$xI - A = \begin{vmatrix} x-5 & 6 & 6 \\ 1 & x-4 & -2 \\ -3 & 6 & x+4 \end{vmatrix} =$ $(x-1)(x-2)^2$.

Thus the characteristic values of T are 1 and 2. Let W_1 and W_2 be the characteristic spaces corresponding to the characteristic value 1 and 2 respectively.

W_1 is the null space of $T - I$. We have

$$A - I = \begin{bmatrix} 4 & -6 & -6 \\ -1 & 3 & 2 \\ 3 & -6 & -5 \end{bmatrix}$$

and

$$A - 2I = \begin{bmatrix} 3 & -6 & -6 \\ -1 & 2 & 2 \\ 3 & -6 & -6 \end{bmatrix}$$

Here the rank of $T - I$ is 2 and rank of $T - 2I$ is 1. Hence the nullity of $T - I$ is 1 and nullity of $T - 2I$ is 2. Hence the dim W_1=1 and dim W_2=2. Thus dim W_1+dim $W_2 = 1 + 2 = 3$ =dim R^3. Hence by the above theorem, T is diagonalizable.

Now we will find the basis of R^3 consisting of characteristic vectors alone.

We have

$$(A-I)(x_1, x_2, x_3) = \begin{bmatrix} 4 & -6 & -6 \\ -1 & 3 & 2 \\ 3 & -6 & -5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = (4x_1-6x_2-6x_3, -x_1+3x_2+2x_3, 3x_1-6x_2-5x_3)$$

The vector $\alpha_1 = (3, -1, 3)$ is in the null space of $T - I$. Since W_1 is one dimensional, $\{\alpha_1\}$ is a basis of W_1. Since

$$(T-2I)(x_1, x_2, x_3) = \begin{bmatrix} 4 & -6 & -6 \\ -1 & 3 & 2 \\ 3 & -6 & -5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = (3x_1-6x_2-6x_3, -x_1+2x_2+2x_3, 3x_1-6x_2-6x_3)$$

The vectors $\alpha_2 = (2, 1, 0)$, $\alpha_3 = (2, 0, 1)$ are in the null space of $T - 2I$. So both are in W_2. They are linearly independent also. Therefore $\{\alpha_2, \alpha_3\}$ is a basis for W_2. Thus $B^{'} = \{(3, -1, 3), (2, 1, 0), (2, 0, 1)\}$ is a basis for R^3. Since $T\alpha_1 = \alpha_1$, $T\alpha_2 = 2\alpha_2$ and $T\alpha_3 = 2\alpha_3$,

$$D_1 = [T]_{B'} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

which is a diagonal matrix. Since A is the matrix of T in the standard ordered basis $\{e_1, e_2, e_3\}$, the matrix D is similar to A. That is there is an invertible matrix P such that $P^{-1}AP = D$. Columns of P are the coordinates of $\alpha_1, \alpha_2, \alpha_3$.

Exercises

1. Let T be a linear operator on R^3 which is represented in the standard ordered basis by the matrix

$$\begin{bmatrix} -9 & 4 & 4 \\ -8 & 3 & 4 \\ -16 & 8 & 7 \end{bmatrix}$$

Prove that T is diagonalizable by exhibiting a basis for R^3, each vector of which is a characteristic vector of T.

2. Let T be a linear operator on a finite dimensional vector space V. Show that non-zero characteristic vectors of T are linearly independent.

3. Let T be a linear operator on the n-dimensional vector space V and suppose that T has n distinct characteristic values. Prove that T is diagonalizable.

4. Let A an $m \times n$ triangular matrix over the field F. Prove that the characteristic values of A are the diagonal entries of A.

5. Let T be a linear operator on R^4 which is represented in the standard ordered basis by the matrix

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ a & 0 & 0 & 0 \\ 0 & b & 0 & 0 \\ 0 & 0 & c & 0 \end{bmatrix}.$$

What are the values of a, b, c such that T is diagonalizable?

6. Let

$$A = \begin{bmatrix} 6 & -3 & -2 \\ 4 & -1 & -2 \\ 10 & -5 & -3 \end{bmatrix}.$$

Is A similar over the field R to a diagonal matrix? Is A similar over the field C to a diagonal matrix?

3.2 Annihilating Polynomials

Definition 3.4. *Let V be a vector space over the field F and T be a linear operation on V. Let p be a polynomial over the field F. That is $p \in F[x]$. Then we say that p annihilates T if $p(T) = 0$.*

Remark 13. If V is a finite dimensional vector space over F and if T is a linear operator in V, then there exists a non-zero polynomial f in $F[x]$ which annihilates T.

Proof. Let V be finite dimensional.
Since T is a linear operator on V, we have $T \in L(V, V)$. We know that $L(V, V)$ is a vector space of dimension n^2. Therefore the $n^2 + 1$ elements $I, T, T^2, \ldots, T^{n^2}$ in $L(V, V)$ are linearly dependant. Hence there are scalars not all of which are zero such that $c_0 I + c_1 T + c_2 T^2 + \ldots + c_{n^2} T^{n^2} = 0$. Let f be the polynomial $f(x) = c_0 + c_1 x + \ldots + c_{n^2} x^{n^2}$ Then f is a non zero polynomial in $F[x]$ and $f(T) = 0$. □

Let us recall the definition,

Definition 3.5. *If R is commutative ring. a subset U of R is a called an ideal (ie ideal) in R if*

1. *U is a subgroup of R under addition.*

2. *For all $u \in U$, $r \in R$ implies that ur and ru belongs to U.*

If F is a field, the set $F[x]$ is a commutative ring. (Actually $F[x]$ is an integral domain).

Lemma 3.5. *Let V be a finite dimensional vector space over the field F and T is a linear operators on V, the set of all polynomials in $F[x]$ which annihilates T is a non zero ideal in $F[x]$.*

Proof. Let K be the set of all polynomials in $F[x]$ which annihilates T. Thus $K = \{f \in F[x]/f(T) = 0\}$. Hence K is a set group of $F[x]$ under the operation of addition Also

$$\begin{aligned} f, g \in K &\Rightarrow f(T) = 0, g(T) = 0 \\ &\Rightarrow (f - g)(T) = 0 \\ &\Rightarrow (f - g) \in K. \end{aligned}$$

Hence K is an ideal in $F[x]$. By remark 13, K is a non-zero ideal in $F[x]$. □

Remark 14. If M is a non-zero ideal in $F[x]$, then M is a principal ideal. There exists a unique monic polynomial p is M such that p generates M. This means that there exist a unique polynomial P in M such that

1. p is a monic polynomial.
2. p generates M. That is $M =< p >$(that is $f = pg$ for some $g \in F[x]$).
3. p is of minimal degree in M. That is there exists no polynomial in M having degree less than the degree of p.

Remark 15. Let V be an n dimensional vector space over the field F and let T be a linear operator on V. The set of all polynomials in $F[x]$ which annihilates T is a non-zero ideal in $F[x]$. By Remark 14, this ideal has a unique generator p. That is there exists a unique monic polynomial p of minimal degree in this ideal which generates this ideal. This polynomial p has the following property.

1. If f is any polynomial in $F[x]$, then $f(t) = 0$ if and only if $f = pg$ for some polynomial $g \in F[x]$.

We shall call this polynomial p the minimal polynomial for the linear operator T.

Definition 3.6. *Let T be a linear operator on a finite dimensional vector space V over the field F. The minimal polynomial for T is the unique monic generator of the ideal of polynomials over F, which annihilates T.*

If p is the minimal polynomial for T, Then p has the following properties.

1. p is a monic polynomial over F.
2. $p(T) = 0$
3. No polynomial having degree less than the degree of p can annihilate T.
4. The polynomial f annihilate T if and only if $f = pg$, for some polynomial $g \in F[x]$.

Definition 3.7. *If A is an $m \times n$ matrix over the field F, the minimal polynomial for A is that unique monic generator of the ideal of polynomials over F which annihilates A.*

If T is the linear operator which is represented in some ordered basis by the matrix A, there the minimal polynomial for T is same as the minimal polynomial of A.

Theorem 3.3. *Let T be a linear operator on an n dimensional vector space V (or let A be an $m \times n$ matrix). The characteristic and minimal polynomial for T (or for A) have the same roots except for multiplicities.*

Proof. Let p be the minimal polynomial for T. Then $P(T) = 0$. Let c be a scalar in F. We have to prove that c is a root of the polynomial p if and only if c is a root of the characteristic polynomial of T. That is $p(c) = 0$ if and only if c is a characteristic value of T. Suppose that $p(c) = 0$. Then $p(x) = (x - c)q(x)$ where q is some polynomial. since degree of $q(x)$ is less than the degree of p, by the definition of p, we have $q(T) \neq 0$
choose a vector β such that $q(T)\beta \neq 0$. Let $\alpha = q(T)\beta$. Then

$$0 = q(T)\beta$$

$$= (T - cI)q(T)\beta$$
$$= (T - cI)\alpha.$$

Hence c is a characteristic value of T. conversely suppose that c is a characteristic value of T. Then there is a non zero vector α in V such that $T\alpha = c\alpha$. By previous lemma, $P(T)\alpha = p(c)\alpha$. Since $P(T) = 0$ and $\alpha \neq 0$, we have $p(c) = 0$. Hence the result. $\square$

Note 3.6. Let T be a diagonalizable linear operator on a finite dimensional vector space V. Let $c_1, c_2, \ldots, c_k$ be the distinct characteristic values of T. Then the minimal polynomial for T is $p(x) = (x - c_1)(x - c_2)\ldots(x - c_k)$.

Proof. Since $c_1, c_2, \ldots, c_k$ are the characteristic values of T, They are the roots of the characteristic polynomial for T. By the above theorem, the minimal polynomial and the characteristic polynomial have the same roots. Hence $c_1, c_2, \ldots, c_k$ are the roots of the minimal polynomial for T. The monic polynomial for T. The monic polynomial of minimal degree with roots $c_1, c_2, \ldots, c_k$.

If α is a characteristic vector, then one of the operators $(T - c_1I), \ldots, (T - c_kI)$ sends α into 0. Therefore $(T - c_1I)\ldots.(T - c_kI)\alpha = 0$ for every characteristic vector α. There is a basis for the underlying space which consists of characteristic vectors of T, hence $p(T) = (T - c_1I)\ldots.(T - c_kI) = 0$ Thus p is a polynomial of minimal degree with roots $c_1, c_2, \ldots c_k$ which annihilates T. Hence p is the minimal polynomial for T. $\square$

Example 36. 1. Consider the linear operator T on R^3, which is represented in the standard ordered basis by the matrix

$$\begin{bmatrix} 5 & -6 & -6 \\ -1 & 4 & 2 \\ 3 & -6 & -4 \end{bmatrix}$$

The linear operator T is diagonalizable. The characteristic polynomial for T is $f = (x-1)(x-2)^2$. Hence 1 and 2 are the distinct characteristic values of T. since T is diagonalizable, the minimal polynomial for T is $(x-1)(x-2)$. we can easily verify that $(A-I)(A-2I) = 0$. Note that here the minimal polynomial divides the characteristic polynomial.

2. Consider the linear operator T on R^3 which is represented in the standard ordered basis by the Matrix

$$A = \begin{bmatrix} 3 & 1 & -1 \\ 2 & 2 & -1 \\ 2 & 2 & 0 \end{bmatrix}.$$

The characteristic polynomial for T is $f = (x-1)(x-2)^2$ which is same as that of the above operator. Hence 1 and 2 are the roots of the characteristic polynomial. The linear operator T is not diagonalizable. So we cannot say that the minimal polynomial for T is $(x-1)(x-2)$. Since 1 and 2 are the roots of the characteristic polynomial, they are the roots of the minimal polynomial for T. So the minimal polynomial for T is of the form $(x-1)^r(x-2)^k$, where $r \geq 1, k \geq 1$.

First we check the polynomial $(x-1)(x-2)$.

We have $(A-I)(A-2I) \neq 0$. Hence $(T-I)(T-2I) \neq 0$. There fore the polynomial $(x-1)(x-2)$ does not annihilate T. Hence $(x-1)(x-2)$ is not the minimal polynomial for T. Therefore the minimal polynomial for T has degree at least 3. So the next choices for the minimal polynomial are $(x-1)^2(x-2)$, $(x-1)(x-2)^2$. Here we can early verify that $(T-I)(T-2I)^2 = 0$. Hence $(x-1)(x-2)^2$ is the minimal polynomial for T.

Note that here, the minimal polynomial for T is same as the characteristic polynomial for T.

3. Consider the linear operator T on R^2 which is represented in the standard ordered basis by the matrix

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

The characteristic polynomial for T is $x^2 + 1$. This polynomial has no roots in the field R. So we cannot find the minimal polynomial for T by concentrating on the matrix A. consider the matrix A as a matrix over the field of complex numbers. Roots of the characteristic polynomial are i and $-i$ over C. The roots of this minimal polynomial are i and $-i$, and hence the roots of the minimal polynomial for A are also i and $-i$. The lowest degree polynomial with roots i and $-i$ is $x^2 + 1$. Also we can easily verify that $A^2 + I = 0$. Thus is the minimal polynomial for A. Hence the minimal polynomial for T is also $x^2 + 1$. Note that here also the characteristic and minimal polynomials are the same.

Theorem 3.4. *(Cayley Hamilton Theorem) Let T be a linear operator on a finite dimensional vector space V. If f is the characteristic polynomial for T, then $f(T) = 0$ in other words, the minimal polynomial for T divides the characteristic polynomial for T.*

Proof. Let K be the set of all polynomials over the field F in the variable T. Then K is a commutative ring with unity.

Choose an ordered basis $\{\alpha_1, \ldots, \alpha_n\}$ for V and let A be the matrix of T in this ordered basis. Then

$$T(\alpha_i) = \sum_{j=1}^{n} A_{ij}\alpha_j, \qquad 1 \leq i \leq n.$$

These operations can be written in the equivalent form

$$\sum_{j=1}^{n} (\delta_{ij}T - A_{ij}I)\alpha_j = 0, \quad 1 \leq i \leq n.$$

For $1 \leq i \leq n, \quad 1 \leq j \leq n$, let

$$B_{ij} = \delta_{ij}T - A_{ij}I.$$

Thus we can write

$$\sum_{j=1}^{n} B_{ij}\alpha_j = 0, \quad \alpha_j = 0, \quad 1 \leq i \leq n.$$

Let B be the $m \times n$ matrix with entries B_{ij}. Since f is the characteristic polynomials for T, we have f is the characteristic polynomial for A.
Therefore

$$\begin{aligned} f(x) &= det(xI - A) && (3.1) \\ &= det(\delta_{ij}x - A_{ji}) && (3.2) \end{aligned}$$

Therefore

$$\begin{aligned} f(T) &= det(\delta_{ij}T - A_{ji}I) && (3.3) \\ &= det[B_{ij}] && (3.4) \\ &= detB. && (3.5) \end{aligned}$$

We have to prove that $f(T) = 0$ To prove $f(T)$ is the zero linear operator on V, it is enough to prove that $f(T)\alpha_k = 0$ for all $k = 1, 2, \ldots, n$. So we shall prove that $[det(B)]\alpha_k = 0$ for all $k = 1, 2, \ldots, n$.
The vectors $\alpha_1, \ldots, \alpha_n$ Satisfies

$$\sum_{j=1}^{n} B_{ij}\alpha_j = 0, \quad , \quad 1 \leq i \leq n.$$

Let $B^{\sim} = AdjB$.

Since each $B^{\sim}_{ki}$ is a linear operator on V, we have $B^{\sim}_{ki}(\sum\limits_{j=1}^{n} B_{ij}\alpha_j) = 0$, for each pair k, i with $1 \leq k \leq n$, $1 \leq i \leq n$. That is for each pair k, i, with $1 \leq k, i \leq n$,

$$\sum_{j=1}^{n} B^{\sim}_{ki} B_{ij} \alpha_j = 0,$$

Summing on i we get,

$$\sum_{i=1}^{n}\sum_{j=1}^{n} B^{\sim}_{ki} B_{ij} \alpha_j = 0,$$

$$\sum_{j=1}^{n}\left(\sum_{i=1}^{n} B^{\sim}_{ki} B_{ij}\right) \alpha_j = 0, \quad 1 \leq k \leq n. \tag{3.6}$$

Since $B^{\sim} = AdjB$, we have

$$B^{\sim} B = (detB)I.$$

Hence $\sum_{i=1}^{n} B^{\sim}_{ki} B_{ij} = \delta_{kj} det(B)$. Hence from equation 3.6, we have $\sum_{i=1}^{n} \delta_{kj}(detB)\alpha_j = 0, \quad 1 \leq k \leq n$. That is $\det B\alpha_k = 0, \quad 1 \leq k \leq n$. Hence the proof. □

Remark 16. By Theorems 3.3 and 3.4, the characteristic and minimal polynomials for an operator T have the same roots and the minimal polynomial divides the characteristic polynomial. Hence if f is the characteristic polynomial and p is the minimal polynomial and $c_1, \ldots, c_k$ are distinct characteristic values of T and if

$$f = (x - c_1)^{d_1} \ldots (x - c_k)^{d_k}$$

then

$$P = (x - c_1)^{r_1} \ldots (x - c_k)^{r_k}, \quad 1 \leq r_j \leq d_j.$$

Example 37. Let A be the 4×4(rational) matrix

$$\begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}$$

The powers of A can be calculated as

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

and

$$A^3 = \begin{bmatrix} 0 & 4 & 0 & 4 \\ 4 & 0 & 4 & 0 \\ 0 & 4 & 0 & 4 \\ 4 & 0 & 4 & 0 \end{bmatrix} = 4 \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}$$

Thus $A^3 = 4A$. And if $p = x^3 - 4x = x(x+2)(x-2)$, then $p(A) = 0$. The minimal polynomial for A must divide the polynomials p. That minimal polynomial for A is not of degree 1. Any monic polynomial of degree 1, is of the form $(x+k)$. If this polynomial is the minimal polynomial of A, there we must have $A + KI = 0$. That is $A = kI$ so that A is a scalar multiple of the identity matrix which is not true. Hence the minimal polynomial for A is not of degree 1. So the candidates for the minimal polynomial are $x(x–2)$, $x(x+2)$, x^2-4. The three quadratic polynomials can be eliminated because it is obvious at a glance that $A^2 \neq -2A, \quad A^2 \neq 2A, A^2 \neq 4I$. Hence p is the minimal polynomial for A. The roots of the minimal polynomial for A are $0, 2$ and -2 and roots of the characteristic polynomials are also $0, 2and-2$. Hence the factors of the characteristic polynomials for A are x, $x-2$, $x+2$. Since the characteristic polynomials for A is a 4th degree polynomial one of these factors must be repeated. We have $0, 2,$ and -2 are the characteristic values of A. We have $rank(A) = 2$. Nullity of $A = 2$ Nullity of $(T) = 2$. Nullity $(T - 0I) = 2$ Thus the characteristic space associated with the characteristic values 0 is 2 dimensional. Therefore if

W_1, W_2, and W_3 are the space of A associated with the characteristic values $0, 2$ and -2 respectively , then dim $W_1 = 2$, and hence dim W_1 + dim $W_2 + dim\ W_3 = 4$ Hence A is diagonalizable. and $d_1 =$ dim $W_1 = 2$, $d_2 =$ dim $W_2 = 1$, $d_3 =$dim $W_3 = 1$. Hence the characteristic polynomial for A is $f = (x-0)^{d_1}(x-2)^{d_2}(x+2)^{d_3} = x^2(x-2)(x+2) = x^2(x^2-4)$.

Since A is diagonalizable, A is similar to the diagonal matrix

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & -2 \end{bmatrix}.$$

Exercises

1. Let V be a finite dimensional vector space. What is the minimal polynomial for the identity operator on V? What is the minimal polynomial for the zero operator?

2. Let a, b, and c be elements of a field F, and let A be the following 3×3 matrix over F:
$$\begin{bmatrix} 0 & 0 & c \\ 1 & 0 & b \\ 0 & 1 & a \end{bmatrix}.$$
Prove that the characteristic polynomial for A is $x^3 - ax^2 - bx - c$ and that this is also the minimal polynomial for A.

3. Find a 3×3 matrix for which the minimal polynomial is x^2.

4. Let A be the 4×4 real matrix
$$\begin{bmatrix} 1 & 1 & 0 & 0 \\ -1 & -1 & 0 & 0 \\ -2 & -2 & 2 & 1 \\ 1 & 1 & -1 & 0 \end{bmatrix}.$$
Show that the characteristic polynomial for A is $x^2(x-1)^2$ and that it is also the minimal polynomial.

3.3 Invariant Subspaces

In this section, we are analysing a linear operator. These ideas will help us to characterize diagonalizable operators in terms of their minimal polynomials.

Definition 3.8. *Let V be a vector space and T a linear operator on V. If W is a subspace of V, we say that W is invariant under T if for each vector $\alpha \in W$, the vector $T(\alpha)$ is in W. That is, if $T(W)$ is contained in W.*

Example 38.

1. If T is any linear operator on V, then V is invariant under T, as is the zero subspace. The range of T and the null space of T are also invariant under T.

2. Let F be a field and let D be the differentiation operator on the space $F[x]$ of polynomials over F. Let n be a positive integer and let W be the subspace of polynomials of degree not greater that n. Then W is invariant under D. This is just another way of saying that D is 'degree decreasing'.

3. Here is a very useful generalization of Example 1. Let T be a linear operator on V. Let U be any linear operator on V which commutes with T. That is, $TU = UT$. Let W be the range of U and let N be the null space of U. Both W and N are invariant under T. If α is in the range of U, say $\alpha = U\beta$, then $T(\alpha) = T(U\beta) = U(T(\beta))$. Then $T(\alpha)$ is in the range of U. If $\alpha \in N$, then $U(T(\alpha)) = T(U(\alpha)) = T(0) = 0$, then $T(\alpha)$ is in N.

4. Let T be the linear operator on R^2 which is represented in the standard ordered basis by the matrix

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

The characteristic polynomial is given by x^2+1 and it has no roots in R. Hence T has no (real) characteristic values. The only subspace of R^2 which are invariant under T are R^2 and the zero subspace. Any other invariant subspace would necessarily have dimension 1. Suppose W is any other subspace of R^2 which are invariant under T. Then W is one dimensional. Then there is a non-zero vector $\alpha \in W$ such that every vector in W is a scalar multiple of α. If W is invariant under T, we get $T(\alpha) = c\alpha$ for some scalar c. This means that α is a characteristic vector, but A has no real characteristic values.

Remark 17. 1. When the subspace W is invariant under a operator T, then T induces a linear operator T_W on the space W. The linear operator T_W is defined by $T_W(\alpha) = T(\alpha)$, for $\alpha \in W$, but T_W is quite a different object from T since its domain is W not V.

2. When V is finite-dimensional, the variance of W under T has a simple matrix interpretation. Let W be a subspace of V such that W is invariant under T. Let $\mathscr{B}' = \{\alpha_1, \ldots, \alpha_r\}$ be a basis for W and let $\mathscr{B} = \{\alpha_1, \ldots, \alpha_r, \alpha_{r+1}, \ldots, \alpha_n\}$ be a basis for V. Let $A = [T]_{\mathscr{B}}$. Then $T\alpha_j = \sum\limits_{i=1}^{n} A_{ij}\alpha_i,\ j = 1, 2, \ldots, n$. Since W is invariant under T, the vector $T\alpha_j$ belongs to W for $j = 1, 2, \ldots, r$. Hence $T\alpha_1, \ldots, T\alpha_r$ are linear combination of $\alpha_1, \ldots, \alpha_r$. Thus $T\alpha_j = \sum\limits_{i=1}^{r} A_{ij}\alpha_i,\ j = 1, 2, \ldots, r$. Thus $A_{ij} = 0$ for $i > r, \quad j \leq r$. Hence the matrix A takes the block form

$$A = \begin{bmatrix} B & C \\ 0 & D \end{bmatrix}$$

where B is an $r \times r$ matrix, C is an $r \times (n-r)$ matrix, and D is an $(n-r) \times (n-r)$ matrix. The matrix B is precisely the matrix of the induced operator T_W in the ordered basis $\mathscr{B}'$.

Example 39. 1. Find all invariant subspaces of the linear transformations $T : R^2 \to R^2$ defined by $T(x, y) = (x - y, 2x + 2y)$. The matrix of the linear operator T in the standard ordered basis is given by

$$\begin{bmatrix} 1 & -1 \\ 2 & 2 \end{bmatrix}.$$

The characteristic polynomial for T is given by $x^2 - 3x + 4$. Its roots are

$$x = \frac{3 + i\sqrt{7}}{2}, \frac{3 - i\sqrt{7}}{2}.$$

So the linear operator T has no real roots. That is characteristic polynomial has no characteristic roots. Hence as in the above example, the only subspace of R^2 which are invariant under T are the zero subspace and R^2.

Lemma 3.7. *Let W be an invariant subspace for T. The characteristic polynomial for the restriction operator T_W divides the characteristic polynomial for T. The minimal polynomial for T_W divides the minimal polynomial for T.*

Proof. If $\mathscr{B}' = \{\alpha_1, \ldots, \alpha_r\}$ is a basis for W and $\mathscr{B} = \{\alpha_1, \ldots, \alpha_r, \alpha_{r+1}, \ldots, \alpha_n\}$ is a basis for V and if $A = [T]_{\mathscr{B}}$ and $B = [T]_{\mathscr{B}'}$, then we have A is in the block form

$$A = \begin{bmatrix} B & C \\ 0 & D \end{bmatrix},$$

where B is the $r \times r$ matrix, C is $r \times (n - r)$ matrix, C is $r \times (n - r)$ matrix and D is $(n - r) \times (n - r)$ matrix. The by the property of determination of matrices in the block form we have, $det(xI - A) = det(xI - B)det(xI - D)$. since $det(xI - A)$ is the characteristic polynomial for T and $det(xI - B)$

is the characteristic polynomial for T_W, we can see that the characteristic polynomial for T_W divides the characteristic polynomial for T.

For $k = 1, 2, \ldots$ the kth power of A is also in the block form

$$A = \begin{bmatrix} B^k & C_k \\ 0 & D^k \end{bmatrix},$$

where C_k is some $r \times (n-r)$ matrix. Hence any polynomial P we have $f(A)$ is in the block form

$$f(A) = \begin{bmatrix} f(B) & C' \\ 0 & f(D) \end{bmatrix}$$

Hence $f(A) = 0$ implies that $f(B) = 0$. In other words if f annihilates A, f annihilates B. Since minimal polynomial for T annihilates A, minimal polynomial for T annihilates B. Hence the minimal polynomial for T_W divides the minimal polynomial for T. □

Definition 3.9. *Let W be an invariant subspace for T and let α be a vector in V. The $T-$**conductor of** α **into** W is the set $S_T(\alpha; W)$, which consists of all polynomials g (over the scalar field) such that $g(T)\alpha$ is in W.*

We usually drop the subscript T and write $S(\alpha; W)$ since the operator T is fixed throughout our discussions. It is the set of all polynomials g over the scalar field F such that the linear operator $g(T)$ leads α in to W. When $W = \{0\}$, the conductor $S(\alpha; W)$ is called the T-annihilator of α.

Lemma 3.8. *If W is an invariant subspace for T, then W is invariant under every polynomial in T. Thus, for each α in V, the conductor $S(\alpha; W)$ is an ideal in $F[x]$.*

Proof. If β is in W, then $T(\beta)$ is in W. Since W is invariant under T, we have $T(T(\beta))$ is in W. That is $T^2(\beta) \in W$. By induction, $T^k(\beta) \in W$. for every k. Taking linear combination, we can see that W is invariant under every polynomial in T. Now let f, g be any two polynomials in $F[x]$. Then $f, g \in S(\alpha; W)$ and $c \in F$,

$$\begin{aligned} f(T)\alpha \in W, \quad g(T)\alpha \in W &\Rightarrow cf(T)\alpha + g(T)\alpha \in W && (3.7) \\ &\Rightarrow [cf(T) + g(T)]\alpha \in W && (3.8) \\ &\Rightarrow [(cf + g)(T)]\alpha \in W && (3.9) \\ &\Rightarrow cf + g \in S(\alpha; W). && (3.10) \end{aligned}$$

Hence $S(\alpha; W)$ is a subspace of $F[x]$. Now $f \in F[x]$ and $g \in S(\alpha; W)$ implies that $g(T)\alpha \in W$. Hence $f(T)[g(T)\alpha] \in W$. Since W is invariant under every polynomial in T, $[f(T)g(T)]\alpha \in W$. Then $[fg(T)]\alpha \in W$, which implies that $fg \in S(\alpha; W)$. Thus $S(\alpha; W)$ is an ideal in $F[x]$. □

Remark 18. The unique monic generator of the ideal $S(\alpha; W)$ is also called the T-conductor of α into W (the T-annihilator in case $W = \{0\}$). The T-conductor of α into W is the monic polynomial g of least degree such the $g(T)\alpha$ is in W. A polynomial f is in $S(\alpha; W)$ if and only if g divides f. Note that the conductor $S(\alpha; W)$ always contains the minimal polynomial for T; hence, every T-conductor divides the minimal polynomial for T.

As the first illustration of how to use the conductor $S(\alpha; W)$, we shall characterize triangulable operators. The linear operator T is called triangulable if there is an ordered basis in which T is represented by a triangular matrix.

Lemma 3.9. *Let V be a finite-dimensional vector space over the field F. Let T be a linear operator on V such that the minimal polynomial for T is a product of linear factors*

$$p = (x - c_1)^{r_1}(x - c_k)^{r_k}, \quad c_i \in F$$

Let W be a proper ($W \neq V$) subspace of V which is invariant under T. There exists a vector $\alpha \in V$ such that

1. *α is not in W;*

2. *$(T - cI)\alpha$ is in W, for some characteristic value c of the operator T.*

Proof. What (a) and (b) say is that the $T-$conductor of α into W is a linear polynomial. Let β be any vector in V which is not in W. Let g be the $T-$conductor of β into W. Then g divides p, the minimal polynomial for T. Since β is not in W, the polynomial g is not constant. Therefore,

$$g = (x - c_1)^{e_1}(x - c_k)^{e_k}, \quad c_i \in F$$

where at least one of the integers e_i is positive. Choose j so that $e_i > 0$. Then $x - c_j$ divides g

$$g = (x - c_k)h$$

. By the definition of g, the vector $\alpha = h(T)\beta$ cannot be in W. But

$$\begin{aligned}(T - c_jI)\alpha &= (T - cI)h(T)\beta \\ &= g(T)\beta\end{aligned}$$

is in W. □

Theorem 3.5. *Let V be a finite-dimensional vector space over the field F and let T be a linear operator on V. Then T is triangulable if and only if the minimal polynomial for T is a product of linear polynomials over F.*

Proof. Suppose that the minimal polynomial factors

$$p = (x - c_1)^{r_1}(x - c_k)^{r_k}, \quad c_i \in F$$

By repeated application of the lemma 3.9, we shall arrive at an ordered basis $\mathscr{B} = \{\alpha_1, \ldots, \alpha_n\}$ which the matrix representing T is upper-triangular:

$$[T]_{\mathscr{B}} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \ldots & a_{1n} \\ 0 & a_{22} & a_{23} & \ldots & a_{2n} \\ 0 & 0 & a_{33} & \ldots & a_{3n} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \ldots & a_{nn} \end{bmatrix} \tag{3.11}$$

This means that

$$T_{\alpha_j} = a_{1j}\alpha_1 + \ldots + a_{jj}\alpha_j \qquad 1 \leq j \leq n. \tag{3.12}$$

That is, $T\alpha_j$ is in the subspace spanned by $\alpha_1, \ldots, \alpha_j$. To find $\alpha_1, \ldots, \alpha_n$, we start by applying the Lemma 3.9 to the subspace $W = \{0\}$, to obtain the vector α_1 Then apply the lemma 3.9 to W_1, the space spanned by α_1 and we obtain W_2. Continue in that way. After $\alpha_1, \ldots, \alpha_n$ have been found, it is the triangular-type relations 3.12 for $j = 1, \ldots, i$ which ensure that the subspace spanned by $\alpha_1, \ldots, \alpha_n$ invariant under T.

If T is triangulable, it is evident that the characteristic polynomial for T has the form

$$f = (x - c_1)^{d_1} \ldots (x - c_k)^{d_k}, \quad c_i \in F$$

Just look at the triangular matrix given by 3.11. The diagonal entries $a_{11}, \ldots, a_{nn}$ are the characteristic values, with c_i repeated d_i times. But, if f can be so factored, so can the minimal polynomial p, because it divides f. □

Corollary 3.1. *Let F be an algebraically closed field, for example: the complex number field. Every $n \times n$ matrix over F is similar over F to a triangular matrix.*

Proof. Let A be any $n \times n$ matrix over F. The minimal polynomial for A is a polynomial over F. Since F is algebraically closed field, the minimal polynomial can be factored into product of linear factors over F. Hence A is triangulable or A is similar over F to a triangular matrix. $\square$

Example 40.

1. Let T be a linear operator on R^2 which is represented in the standard ordered basis by

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

   Is T triangulable?

   Solution: The characteristic polynomial for T is $f = x^2+1$ and minimal polynomial for T is $p = x^2+1$. Then the polynomial can not be written as a product of linear polynomials over R. Hence T is not triangulable. (Note that if we consider T as a linear operator on C^2, then T is Triangulable.)

2. Show that the matrix

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

   is not similar over R to a triangular matrix.

   Solution: The characteristic polynomial for A is $f = (x-1)(x-2)+2 = x^2-3x+4$. Therefore f has no root in R. Hence minimal polynomial has no root in R. Therefore the minimal polynomial for A is $p = x^2-3x+4$. So the minimal polynomial can not be factored into a product of linear polynomials in R. Hence A is not similar to a triangular matrix.

3. Let T be a linear operator on R^3 which is represented in the standard

basis by the matrix

$$\begin{bmatrix} 3 & 1 & -1 \\ 2 & 2 & -1 \\ 2 & 2 & 0 \end{bmatrix}.$$

Show that T is not diagonalizable but T is triangulable.

Solution: The characteristic polynomial is given by $(x-1)(x-2)^2$. The minimal polynomial is $(x-1)(x-2)^2$. Let W_1 and W_2 be the characteristic spaces associated with the characteristic values 1 and 2 respectively. Here $\dim W_1 + \dim W_2 = 2 \neq 3 (= dim V)$. Hence T is not diagonalizable.

minimal polynomial for T is

$$(x-1)(x-2)^2 = (x-1)(x-2)(x-2),$$

a product of linear factors. Hence the operator associated with the matrix A is triangulable.

Theorem 3.6. *Let V be a finite-dimensional vector space over the field F and let T be a linear operator on V. Then T is diagonalizable if and only if the minimal polynomial for T has the form*

$$p = (x-c_1)(x-c_k),$$

where c_i are distinct elements of F.

Proof. We have noted earlier that, if T is diagonalizable, its minimal polynomial is a product of distinct linear factors. Conversely assume that the minimal polynomial for T has the form

$$p = (x-c_1)(x-c_k),$$

where c_i are distinct elements of F. Let W be the subspace spanned by all of the characteristic vectors of T. Then $W = W_1 + W_2 + \ldots + W_k$ Where W_i is the characteristic space associated with the characteristic value c_i. We have to prove that $W = V$.

Suppose $W \neq V$. Then W is a proper subspace of V which is invariant under T. Then by applying Lemma 3.9, choose a vector α not in W and a characteristic value c_j of T such that the vector $\beta = (T - c_j I)\alpha$ is in W. Since $\beta \in W$, and since $W = W_1 + W_2 + \ldots + W_k$, we have $\beta = \beta_1 + \beta_2 + \ldots + \beta_k$, where $\beta_i \in W_i$ for $i = 1, 2, \ldots, k$. Then $T\beta_i = c_i\beta_i$ for $i = 1, 2, \ldots, k$. Then for any polynomial h we have $h(T)\beta_i = h(c_i)\beta_i$. Therefore

$$\begin{aligned} h(T)\beta &= h(T)\beta_1 + h(T)\beta_2 + \ldots + h(T)\beta_k \\ &= h(c_1)\beta_1 + h(c_2)\beta_2 + \ldots + h(c_k)\beta_k \end{aligned}$$

Thus

$$h(T)\beta \in W. \tag{3.13}$$

We have $p = (x - c_j)q$ where q is a polynomial. Also $q - q(c_j) = (x - c_j)h$ where h is a polynomial. Now

$$\begin{aligned} q(T)\alpha - q(c_j)\alpha &= h(T)(T - c_j I)\alpha \\ &= h(T)\beta \end{aligned}$$

This implies that

$$q(T)\alpha - q(c_j)\alpha = h(T)\beta \in W. \tag{3.14}$$

Since p is the minimal polynomial for T we have $p(T) = 0$. Therefore $0 = P(T)\alpha = (T - c_j I)q(T)\alpha$. This implies that $q(T)\alpha$ is a characteristic vector of T associated with the characteristic value c_j. Hence from equation 2.14, $q(T)\alpha \in W$. Since $\alpha \notin W$, it follows that $q(c_j) = 0$. Since $p = (x - c_j)q$, this contradicts the fact that p has distinct roots. Hence $W = V$. Thus T is diagonalizable. $\square$

The above theorem 3.6 is useful in a computational way. Suppose we have a linear operator T, represented by the matrix A in some ordered basis, and we wish to know if T is diagonalizable. We compute the characteristic polynomial f. If we can factor f:

$$f = (x - c_1)^{d_1} \dots (x - c_k)^{d_k}$$

we have two different methods for determining whether or not T is diagonalizable. One method is to see whether (for each i) we can find d_i independent characteristic vectors associated with the characteristic value c_i. The other method is to check whether or not

$$(T - c_1 I) \dots (T - c_k I)$$

is the zero operator.

Theorem 3.5 provides a different proof of the Cayley-Hamilton theorem. That theorem is easy for a triangular matrix. Hence, via Theorem 3.5, we obtain the result for any matrix over an algebraically closed field. Any field is a subfield of an algebraically closed field. If one knows that result, one obtains a proof of the Cayley-Hamilton theorem for matrices over any field.

Exercises

1. Let T be the linear operator on R^2, the matrix of which in the standard ordered basis is
$$\begin{bmatrix} 1 & -1 \\ 2 & 2 \end{bmatrix}$$
   (a) Prove that the only subspaces of R^2 invariant under T are R^2 and the zero subspace.

   (b) If U is the linear operator on C^2 , the matrix of which in the standard ordered basis is A, show that U has 1-dimensional invariant subspaces.

2. Let W be an invariant subspace for T. Prove that the minimal polynomial for the restriction operator T_W divides the minimal polynomial for T, without referring to matrices.

3. Let c be a characteristic value of T and let W be the space of characteristic vectors associated with the characteristic value c. What is the restriction operator T_W ?

4. Let T be a diagonalizable linear operator on the n-dimensional vector space V, and let W be a subspace which is invariant under T. Prove that the restriction operator T_W is diagonalizable. T is diagonalizable if and only if T is annihilated by some polynomial over C which has distinct roots.

5. Let T be linear operator on a finite-dimensional vector space over the field of complex numbers. Prove that T is diagonalizable if and only if T is annihilated by some polynomial over C which has distinct roots.

6. Let T be a linear operator on V. If every subspace of V is invariant under T, then T is a scalar multiple of the identity operator.

3.4 Direct- Sum Decompositions

In this section we are trying to decompose the underlying space V into a sum of invariant subspaces for T such that the restriction operators on those subspaces are simple.

Definition 3.10. *Let $W_1, \ldots, W_k$ be subspaces of the vector space V. We say that $W_1, \ldots, W_k$ are independent if $\alpha_1 + \ldots + \alpha_k = 0$, $\alpha_i \in W_i$ implies that each α_i is* 0*.*

For $k = 2$, the meaning of independence is 0 intersection, i.e., W_1 and W_2 are independent if and only if $W_1 \cap W_2 = 0$. If $k > 2$, the independence of $W_1, \ldots, W_k$ says much more than $W_1 \cap \ldots \cap W_k = 0$. It says that each W_j intersects the sum of the other subspaces W_i only in the zero vector. The significance of independence is this. Let $W = W_1 + \ldots + W_k$ be the subspace spanned by $W_1, \ldots, W_k$. Each vector α in W can be expressed as a sum

$$\alpha = \alpha_1 + \ldots + \alpha_k, \qquad \alpha_i \in W_i$$

. If $W_1, \ldots, W_k$ are independent, then that expression for α is unique; for if

$$\alpha = \beta_1 + \ldots + \beta_k, \qquad \beta_i \in W_i$$

, then

$$0 = (\alpha_1 - \beta_1) + \ldots + (\alpha_k - \beta_k),$$

Hence $\alpha_i - \beta_i = 0, \quad i = 1, 2, \ldots, k$. Thus when $W_1, \ldots, W_k$ are independent we can operate with vectors in W as k-tuples $(\alpha_1, \ldots, \alpha_k)$, $\alpha_i \in W_i$ in the same way as we operate with vectors in R^k as k-tuples of numbers.

Lemma 3.10. *Let V be a finite-dimensional vector space. Let $W_1, \ldots, W_k$ be subspaces of V and let $W = W_1 + \ldots + W_k$. The following are equivalent.*

1. *$W_1, \ldots, W_k$ are independent.*

2. *For each j such that $2 \leq j \leq k$, we have $W_j \cap (W_1 + \ldots + W_{j-1}) = \{0\}$.*

3. *If $\mathscr{B}$ is an ordered basis for W_i, $1 \leq i \leq k$, then the sequence*

$$\mathscr{B} = (\mathscr{B}, \ldots, \mathscr{B})$$

an ordered basis for W.

Proof. (1) $\Leftrightarrow$ (2).
Assume (1). Suppose that $W_1, \ldots, W_k$ are independent. For each j, $2 \leq j \leq k$, we have $\alpha \in W_j \cap (W_1 + \ldots + W_{j-1})$.

$$\begin{aligned}
\alpha \in W_j \cap (W_1 + \ldots + W_{j-1}) &\Rightarrow \alpha \in W_j \text{ and } \alpha \in (W_1 + \ldots + W_{j-1}) \\
&\Rightarrow \alpha_1 + \ldots + \alpha_{j-1}, \qquad \alpha_i \in W_i \\
&\Rightarrow \alpha_1 + \ldots + \alpha_{j-1} - \alpha + 0 + \ldots + 0 = 0 \\
&\Rightarrow \alpha_1 = 0,\ \alpha_2 = 0, \ldots, \alpha_{j-1} = 0,\ \alpha = 0
\end{aligned}$$

Since $W_1, W_2, \ldots, W_k$ are independent, $\alpha = 0$. Hence $W_j \cap (W_1 + \ldots + W_{j-1}) = \{0\}$.
Conversely suppose that for each j, $2 \leq j \leq k$, we have $W_j \cap (W_1 + \ldots + W_{j-1}) = \{0\}$.
Suppose that $\alpha_1 + \ldots + \alpha_k = 0$, where $\alpha_i \in W_i$. Let j be the largest index such that $\alpha \neq 0$. Therefore $\alpha_1 + \ldots + \alpha_k = 0$.

$$\begin{aligned}
\alpha_1 + \ldots + \alpha_j = 0 &\Rightarrow \alpha_j = -\alpha_1 - \alpha_2 - \ldots - \alpha_{j-1} \\
&\in W_j \cap (W_1 + \ldots + W_{j-1}) = \{0\} \\
&\Rightarrow \alpha_j = 0
\end{aligned}$$

$$\Rightarrow \quad \alpha_i = 0 \text{ for all } i.$$

Hence $W_1, W_2, \ldots, W_k$ are independent. Hence (1) and (2) are equivalent. To prove (1) $\rightarrow$ (3). Suppose that $W_1, W_2, \ldots, W_k$ are independent. Let $\mathscr{B}_i$ be an ordered basis for W_i, $1 \leq i \leq k$. Let

$$\mathscr{B} = \{\mathscr{B}_1, \mathscr{B}_2, \ldots, \mathscr{B}_k\}.$$

Any linear relation between vectors of $\mathscr{B}$ is of the form $\beta_1 + \beta_2 + \ldots + \beta_k = 0$ where β_i is a linear combination of vectors in $\mathscr{B}_i$ for each i. Then $\beta_i \in W_i$ for each i.

Since $W_1, W_2, \ldots, W_k$ are independent, we have $\beta_i = 0$ for each i.

Since $\mathscr{B}_i$ is linearly independent it follows that there is only a trivial linear relation between the vectors of $\mathscr{B}$. Hence $\mathscr{B}$ is linearly independent. Since each $\mathscr{B}_i$ spans W_i for each W_i, $\mathscr{B}$ spans $W_1 + W_2 + \ldots + W_k$. Thus $\mathscr{B}$ is a basis for W.

Now to prove Assume that if $\mathscr{B}_i$ is an ordered basis of W_i, $1 \leq j \leq k$, then $\mathscr{B} = \{\mathscr{B}_1, \mathscr{B}_2, \ldots, \mathscr{B}_k\}$ is an ordered basis of $W_1 + W_2 + \ldots + W_k$. Let $\alpha_1 + \alpha_2 + \ldots + \alpha_k = 0$ where $\alpha_i \in W_i$. To prove that each $\alpha_i = 0$. Since $\mathscr{B}_i$ is an ordered basis of W_i, $1 \leq j \leq k$, each α_i is a linear combination of vectors in $\mathscr{B}_i$. Hence $\alpha_1 + \alpha_2 + \ldots + \alpha_k$ is a linear combination of vectors in $\mathscr{B}$. Since $\mathscr{B}$ is linearly independent, it follows that each coefficient of this linear combination is *zero*. Hence $\alpha_i = 0$ for each i. Hence the result.

$\square$

If any of the conditions of the last Lemma hold, we say that the sum $W_1 + W_2 + \ldots + W_k$ is direct or that W is the direct sum of $W_1, W_2, \ldots, W_k$ and we write

$$W = W_1 \oplus \ldots \oplus W_k.$$

Example 41. 1. Let V be a finite dimensional vector space over the field

F and let $\{\alpha_1, \ldots, \alpha_n\}$ be any basis for V. If W_i is the one-dimensional subspace spanned by α_i, then $V = W_1 \oplus \ldots \oplus W_n$

2. Let n be a positive integer and F be a subfield of the complex numbers and let V be the space of all $n \times n$ matrices over F. Let W_1 be the subspace of all symmetric matrices(Matrices with $A^t = A$). Let W_2 be the subspace of all skew- symmetric matrices(Matrices with $A^t = -A$). Then $V = W_1 \oplus W_2$.

$$\begin{aligned} A &= A_1 + A_2 \\ A_1 &= \frac{1}{2}(A + A^t) \\ A_2 &= \frac{1}{2}(A - A^t). \end{aligned}$$

3. Let T be any linear operator on a finite dimensional space V. Let $c_1, \ldots, c_k$ be the distinct characteristic values of T and let W_i be the characteristic spaces associated with the characteristic values c_i. Then $W_1, \ldots, W_k$ are independent. If T is diagonalizable, then $V = W_1 \oplus \ldots \oplus W_k$.

Definition 3.11. *If V is a vector space, a projection of V is linear operator E on V such that $E^2 = E$.*

For example, Consider the operator on R^2 given by $E(x_1, x_2) = (x_1, 0)$. Then $E^2(x_1, x_2) = E(E(x_1, x_2)) = E(x_1, 0) = (x_1, 0) = E(x_1, x_2)$. E is a projection of R^2.

Properties of E Suppose that E is a projection. Let R be the range of E and N be the nullspace of E.

1. vector $\beta \in R$ if and only if $E\beta = \beta$.

Proof. For let $\beta \in R$. Then $\beta = E\alpha$ for some $\alpha \in V$. Then

$$E\beta = E(E\alpha) = E^2\alpha = E\alpha = \beta.$$

Thus $E\beta = \beta$.

Suppose $E\beta = \beta$., then β is in the range of E. $\square$

2. $V = R \oplus N$

Proof. Let $\alpha \in V$. Let E denotes the projection, then $E^2 = E$. Then $E\alpha \in R$. Now

$$E(\alpha - E\alpha) = E(\alpha) - E^2\alpha = E(\alpha) - E(\alpha) = 0.$$

This implies that $\alpha - E\alpha \in N$.

$$\alpha = E\alpha - E\alpha + \alpha$$

where $E\alpha \in R$ and $\alpha - E\alpha \in N$. Thus $V = R + N$. Now assume $R \cap N \neq \phi$, let $\alpha \in R \cap N$. Then $\alpha \in R$ and $\alpha \in N$. Thus $\alpha \in R$ if and only if $\alpha = E\alpha$. We have $\alpha \in N$ if and only if $E\alpha = 0$. This implies that $\alpha = 0$. Thus $R \cap N = \phi$. Thus we can conclude that $V = R \oplus N$. $\square$

3. The unique expression for α as a sum of vectors in R and N is given by
$$\alpha = E\alpha - E\alpha + \alpha$$
where $E\alpha \in R$ and $\alpha - E\alpha \in N$.

From (1), (2) and (3) it is easy to see the following:

If R and N are subspaces of V such that $V = R \oplus N$, there is one and only one projection operator E which has range R and nullspace N. That operator E is called the projection on R along N.

Theorem 3.7. *If $V = W_1 \oplus \ldots \oplus W_k$, then there exists k linear operators $E_1, \ldots, E_k$ on V such that*

(i). *each E_i is a projection.*

(ii). $E_iE_j = 0$ *if* $i \neq j$

(iii). $I = E_1 + \ldots + E_k$

(iv). *the range of* E_i *is* W_i.

Conversely if $E_1, \ldots, E_k$ *are* k *linear operators which satisfy the conditions* (i), (ii) *and* (iii) *and if we let* W_i *be the range of* E_i, *then* $V = W_1 \oplus \ldots \oplus W_k$.

Proof. Assume that $V = W_1 \oplus \ldots \oplus W_k$. For each j, we have to define an operator E_j on V. Let $\alpha = \alpha_1 + \ldots + \alpha_n \in V$ with $\alpha_i \in W_i$. Define $E_j(\alpha) = \alpha_j$. Then E_j is a function from V to V and is well defined. Since $\alpha_j \in W_j$, $E_j(\alpha) \in W_j$. That is Range of E_j is W_j. Then $E_j(c\alpha + \beta) = c\alpha_j + \beta_j = cE(\alpha) + E(\beta)$. Thus E is linear. Then

$$E_j^2(\alpha) = E_j(E_j(\alpha)) = E_j(\alpha_j) = \alpha_j = E_j(\alpha).$$

The null space of E_j is the set of all α such that $E_j(\alpha) = 0$. That is the set of all α such that $\alpha_j = 0$. That is α is actually a sum of vectors from the spaces W_i with $i \neq j$. That is the null space of E_j is the subspace

$$(W_1 + \ldots + W_{j-1} + W_{j+1} + \ldots + W_k).$$

We have $E_j(\alpha) = \alpha_j$. Then $\alpha = E_1(\alpha) + E_2(\alpha) + \ldots + E_k(\alpha)$ for each $\alpha \in V$. That means

$$I = E_1 + E_2 + \ldots + E_k.$$

If $i \neq j$, then range of E_j is the subspace W_j which is contained in the null space of E_i. Then $E_i(E_j(\alpha)) = E_i(\alpha_j) = 0$. That means $E_iE_j = 0$ if $i \neq j$.

Conversely assume that $E_1, \ldots, E_k$ are k linear operators which satisfy the conditions (i), (ii) and (iii) and let W_i be the range of E_i. Let $\alpha \in V$. Then by condition (iii),

$$\alpha = E_1(\alpha) + \ldots + E_k(\alpha), \tag{3.15}$$

and W_i be the range of E_i implies that $\alpha \in W_1 + \ldots + W_k$. This expression for α is unique because if

$$\alpha = \alpha_1 + \ldots + \alpha_k \tag{3.16}$$

with $\alpha_i \in W_i$, say $\alpha_i = E_i\beta_i$, then using condition (i) and (ii) we have

$$\begin{aligned} E_j\alpha &= \sum_{i=1}^{k} E_j\alpha_j \\ &= \sum_{i=1}^{k} E_jE_i\beta_i \\ &= E_j^2\beta_j \\ &= E_j\beta_j \\ &= \alpha_j. \end{aligned}$$

That is the expression for α in 3.15 and 3.16 are the same. Thus V is the direct sum of the W_i. □

Exercises

1. Let V be a finite-dimensional vector space and let W_1 be any subspace of V. Prove that there is a subspace W_2 of V such that $V = W_1 \oplus W_2$.
2. Let V be a finite dimensional vector space and let $W_1, \ldots, W_k$ be subspaces of V such that $V = W_1 + \ldots + W_k$ and $dimV = dimW_1 + \ldots + dimW_k$. Prove that $V = W_1 \oplus \ldots \oplus W_k$.
3. Find a projection E which projects R^2 onto the subspace spanned by $(1, -1)$ along the subspace spanned by $(1, 2)$.
4. If E is a projection on R along N then $I - E$ is the projection on N along R.
5. Let $E_1, \ldots, E_k$ be operators on the space V such that $E_1 + \ldots + E_k = I$.
   (i) Prove that if $E_iE_j = 0$ for $i \neq j$, then $E_i^2 = E_i$ for each i.
   (ii) When $k = 2$, prove the converse of (i). That is $E_1 + E_2 = I$ and $E_1^2 = E_1$, $E_2^2 = E_2$, then $E_1E_2 = 0$.

3.5 Invariant Direct Sums

In this section we study the direct sum decomposition $V = W_1 \oplus \ldots \oplus W_k$, where each of the subspaces W_i is invariant under some given linear operator T. Given a decomposition of V T induces a linear operator T_i on each W_i by restriction. Thus T acts on V as follows. If $\alpha \in V$, we have unique vectors $\alpha_1, \alpha_2, \ldots, \alpha_k$ with $\alpha_i \in W_i$ such that

$$\alpha = \alpha_1 + \alpha_2 + \ldots + \alpha_k,$$

then

$$T\alpha = T_1\alpha_1 + T_2\alpha_2 + \ldots + T\alpha_k.$$

Then we say that T is the direct sum of the operators $T_1, T_2, \ldots T_k$. Here T_i are not linear operators on the space V but on the various subspaces W_i.

Theorem 3.8. *Let T be a linear operator on the space V, and let $W_1, \ldots, W_k$ and $E_1, \ldots, E_k$ be as in Theorem 3.7. Then a necessary and sufficient condition that each subspace W_i be invariant under T is that T commute with each of the projections E_i. That is*

$$TE_i = E_iT, \qquad i = 1, 2, \ldots, k.$$

Proof. Suppose that each subspace W_i be invariant under T. Let $\alpha \in V$. Then

$$\alpha = E_1\alpha + \ldots + E_k\alpha$$

$$T\alpha = TE_1\alpha + \ldots + TE_k\alpha.$$

Since $E_i\alpha$ is in W_i, which is invariant under T, we have $TE_i\alpha = E_i\beta_i$ for some vector β_i. Then

$$\begin{aligned} E_jTE_i\alpha &= E_jE_i\beta_i \\ &= \begin{cases} 0, & if \quad i \neq j \\ E_j\beta_j & if \quad i = j. \end{cases} \end{aligned}$$

Thus

$$E_jT\alpha = E_jTE_1\alpha + \ldots + E_jTE_k\alpha$$

$$
\begin{aligned}
&= E_j\beta_j \\
&= TE_j\alpha
\end{aligned}
$$

This holds for each $\alpha \in V$, so $E_jT = TE_j$.

Conversely suppose that T commutes with each E_i. Let α be in W_j. Then $E_j\alpha = \alpha$, and

$$T\alpha = T(E_j\alpha) = E_j(T\alpha).$$

Then $T\alpha$ is in the range of E_j. But range of E_j is W_j. Thus $T\alpha$ is in W_j. Hence W_j is invariant under T. □

Theorem 3.9. *Let T be a linear operator on a finite dimensional space V. If T is diagonalizable and if $c_1, \ldots, c_k$ are the distinct characteristic values of T, then there exist linear operators $E_1, \ldots, E_k$ on V such that*

(i) $T = c_1E_1 + \ldots + c_kE_k$;

(ii) $I = E_1 + \ldots + E_k$

(iii) $E_iE_j = 0,\ i \neq j$

(iv) $E_i^2 = E_i$

(v) *the range of E_i is the characteristic space for T associated with c_i.*

Conversely if there exist k distinct scalars $c_1, c_2, \ldots, c_k$ and k non-zero linear operators $E_1, \ldots, E_k$ which satisfy conditions (i),(ii) and (iii), then T is diagonalizable, $c_1, c_2, \ldots, c_k$ are distinct characteristic values of T and conditions (iv) and (v) are satisfied also.

Proof. Suppose that T is diagonalizable, with distinct characteristic values $c_1, \ldots, c_k$. Let W_i be the space of characteristic vectors associated with the characteristic value c_i. As we have seen, $V = W_1 \oplus \ldots + W_k$. Let $E_1, \ldots, E_k$

be the projections associated with this decomposition, as in Theorem 9. Then (ii),(iii),(iv) and (v) are satisfied. To verify (i), proceed as follows. For each Now suppose that we are given a linear operator T along with distinct scalars i c and non-zero operators E_i which satisfy (i),(ii) and (iii). For each $\alpha \in V$,

$$\alpha = E_1\alpha + \ldots + E_k\alpha$$

and so

$$T\alpha = TE_1\alpha + \ldots + TE_k\alpha = c_1E_1\alpha + \ldots + c_kE_k\alpha.$$

In other words

$$T = c_1E_1 + \ldots + c_kE_k$$

Now suppose that we are given a linear operator T along with distinct scalars c_i and non-zero operators E_i which satisfy (i), (ii) and (iii). Since $E_iE_j = 0$ when $i \neq j$, we multiply both sides of $I = E_1 + \ldots + E_k$ by E_i and obtain immediately $E_i^2 = E_i$. Multiplying $T = 0_1E_1 + \ldots + c_kE_k$ by E_i, we then have

$$TE_i = c_1E_iE_1 + \ldots + c_kE_iE_k = c_iE_i.$$

That is $TE_i = c_iE_i \Rightarrow (T - c_iI)E_i = 0$. Now let $\beta \in$ the range of E_i. Then $\beta = E_i(\alpha)$. Then

$$\begin{aligned} c_iE_i(\alpha) = TE_i(\alpha) \quad &\Rightarrow \quad c_i\beta = T\beta \\ &\Rightarrow \quad (T - c_i)\beta = 0 \\ &\Rightarrow \quad \beta \in \text{ null space of } T - c_i. \end{aligned}$$

This means that any vector in the range of E_i is in the null space of $T - c_iI$. Since we have assumed $E_i \neq 0$, this proves that there is a non-zero vector in the null space of $T - c_iI$. That is c_i is a characteristic value of T. Furthermore, the c_i are all of the characteristic values of T, for if c is any scalar, then

$$T - cI = c_1E_1 + \ldots + c_kE_k - cI = (c_1 - c)E_1 + \ldots + (c_k - c)E_k.$$

Thus if $(T - cI)\alpha = 0$, we must have $(c_i - c)E_i\alpha = 0$. If α is not the zero vector, then $E_i\alpha \neq 0$ for some i, so that for this i we have $c_i - c = 0$.

Every non-zero vector in the range of E_i is a characteristic vector of T, and $I = E_1 + \ldots + E_k$ shows that these characteristic values span V. Thus T is diagonalizable.

Now we have to prove that the null space of $T - c_iI$ is the range of E_i. If

$$\begin{aligned}
\alpha \in N(T - c_iI) &\Rightarrow (T - c_iI)\alpha = 0 \\
&\Rightarrow T\alpha = c_i\alpha \\
&\Rightarrow (c_1E_1 + \ldots + c_kE_k)\alpha = c_i\alpha \\
&\Rightarrow (c_1 - c_i)E_1\alpha + \ldots + (c_k - c_i)E_k\alpha = 0 \\
&\Rightarrow (c_j - c_i)E_j\alpha = 0 \text{ for each } j \\
&\Rightarrow E_j\alpha = 0, j \neq i.
\end{aligned}$$

Since $\alpha = E_1\alpha + \ldots + E_k\alpha$, and $E_j\alpha = 0$ for $j \neq i$, we have $\alpha = E_i\alpha$ which proves that α is in the range of E_i. □

Exercises

1. Let E be a projection of V and let T be a linear operator on V. Prove that the range of E is invariant under T if and only if $ETE = TE$. Prove that both the range and null space of E are invariant under T if and only if $ET = TE$.

2. Let T be the linear operator on R^2, the matrix of which in the standard ordered basis is
$$\begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}.$$
Let W_1 be the subspace of R^2 spanned by the vector $e_1 = (1, 0)$. Then
a) Prove that W_1 is invariant under T.
b) Prove that there is no subspace W_2 which is invariant under T and which is complementary to W_1: $R^2 = W_1 \oplus W_2$.

3. Let T be linear operator on a finite-dimensional vector space V. Let R be the range of T and let N be the null space of T. Prove that R

and N are independent if and only if $V = R \oplus N$. 3. Let T be a linear operator on V. Suppose $V = W_1 \oplus \ldots W_k$, where each W_i is invariant under T. Let T_i be the induced (restriction) operator on W_i.
1. Prove that det $(T) = \det(T_1) + \ldots + \det(T_k)$.
2. Prove that the characteristic polynomial for f is the product of the characteristic polynomials for $f_1, \ldots, f_k$.
3. Prove that the minimal polynomial for T is the least common multiple of the minimal polynomials for $T_1, T_2, \ldots, T_k$.

4. Let T be a linear operator on V which commutes with every projection operator on V. What can you say about T?.

Chapter 4

Inner Product Spaces

Here we study vector spaces in which it makes sense to speak of the length of a vector and the angle between two vectors. We will do this by introducing a scalar valued function on pairs of vectors known as inner product.

4.1 Inner Products

An inner product on a vector space is a function with properties similar to the dot product in R^3. And in terms of such an inner product one can also define length and angle. In this section we discuss the definition and examples of inner products and establish a few basic properties of inner products. Then we discuss length and orthogonality.

Definition 4.1. *Let F be the field of real numbers or the field of complex numbers, and V be a vector space over F. An **inner product** on V is a function which assigns to each ordered pair of vectors $\alpha, \beta \in V$ a scalar $(\alpha|\beta) \in F$ in such a way that for all α, β, γ in V and all scalars c*

1. $(\alpha + \beta|\gamma) = (\alpha|\gamma) + (\beta|\gamma)$;

2. $(c\alpha|\beta) = c(\alpha|\beta)$;

3. $(\beta|\alpha) = \overline{(\alpha|\beta)}$, *the bar denotes conjugation;*

4. $(\alpha|\alpha) > 0$ *if* $\alpha \neq 0$.

From the above definition, we can see that the above conditions (1), (2) and (3) imply that

$$(\alpha|(c\beta + \gamma)) = \bar{c}(\alpha|\beta) + (\alpha|\gamma) \tag{4.1}$$

Proof.

$$\begin{aligned}
(\alpha|(c\beta + \gamma)) &= \overline{((c\beta + \gamma)|\alpha)} \\
&= \overline{(c\beta|\alpha) + (\gamma|\alpha)} \\
&= \bar{c}\overline{(\beta|\alpha)} + \overline{(\gamma|\alpha)} \\
&= \bar{c}(\alpha|\beta) + (\alpha|\gamma)
\end{aligned}$$

□

When F is the field R of real numbers, the complex conjugates appearing in (3) and equation 4.1 are superfluous, however in the complex case they are necessary for consistency of the conditions for,

Without complex conjugates $(i\alpha|i\alpha) = i^2(\alpha|\alpha) = -(\alpha|\alpha)$,

by condition (4), $(\alpha|\alpha) > 0$. Thus we get $(i\alpha|i\alpha) = -(\alpha|\alpha) < 0$. This is a contradiction to condition (4) of the definition.

Note that Throughout this chapter, F is either the field of real numbers or the field of complex numbers.

Example 42. 1. Scalar or dot product of vectors in R^3. The scalar product of vectors $\alpha = (x_1, x_2, x_3)$ and $\beta = (y_1, y_2, y_3)$ in R^3 is the real number

$$(\alpha|\beta) = x_1y_1 + x_2y_2 + x_3y_3.$$

2. On F^n there is an inner product which we call the standard inner product. It is defined on $\alpha = (x_1, x_2, \ldots, x_n)$ and $\beta = (y_1, y_2, \ldots, y_n)$ by

$$(\alpha|\beta) = x_1\overline{y_1} + x_2\overline{y_2} + \ldots + x_n\overline{y_n} = \sum_j x_j\overline{y_j}$$

In the real case, the standard inner product is often called the dot product or scalar product and is denoted by $\alpha \cdot \beta$.

3. For $\alpha = (x_1, x_2)$ and $\beta = (y_1, y_2)$ in R^2, let

$$(\alpha|\beta) = x_1y_1 - x_2y_1 - x_1y_2 + 4x_2y_2.$$

$$\begin{aligned}
(\alpha + \beta|\gamma) &= (x_1 + y_1)z_1 - (x_2 + y_2)z_1 - (x_1 + y_1)z_2 + 4(x_2 + y_2)z_2 \\
&= x_1z_1 + y_1z_1 - x_2z_1 - y_2z_1 - x_1z_2 - y_1z_2 + 4x_2z_2 + 4y_2z_2 \\
&= (x_1z_1 - x_2z_1 - x_1z_2 + 4x_2z_2) + (y_1z_1 - y_2z_1 - y_1z_2 + 4y_2z_2) \\
&= (\alpha|\gamma) + (\beta|\gamma).
\end{aligned}$$

Thus condition (1) is satisfied.

$$\begin{aligned}
(c\alpha + \beta) &= cx_1y_1 - cx_2y_1 - cx_1y_2 + 4cx_2y_2 \\
&= c[x_1y_1 - x_2y_1 - x_1y_2 + 4x_2y_2] \\
&= c(\alpha + \beta)
\end{aligned}$$

Thus condition (2) is satisfied.

Condition (3) is satisfied since x_1, x_2, y_1 and y_2 are real numbers. $(\alpha|\alpha) = x_1^2 - x_2x_1 - x_1x_2 + 4x_2^2 = (x_1 - x_2)^2 + 3x_2^2$. This follows that $(\alpha|\alpha) > 0$ if $\alpha \neq 0$. Thus condition (4) is satisfied. Hence $(\alpha|\beta)$ is an inner product on R^2.

4. Let V be $F^{n\times n}$, the space of all $n \times n$ matrices over F. Then V is isomorphic to F^{n^2} in a natural way. Then

$$(A|B) = \sum_{j,k} A_{jk}\bar{B_{jk}}$$

defines an inner product on V. If B^* denotes the conjugate transpose matrix B^* of B defined as $B^*_{kj}\bar{B_{jk}}$, then

$$(A|B) = tr(AB^*) = tr(B^*A).$$

This is because

$$\begin{aligned} tr(AB^*) &= \sum_j (AB^*)_{jj} \\ &= \sum_j \sum_k A_{jk} B^*_{kj} \\ &= \sum_j \sum_k A_{jk} \bar{B_{jk}} \end{aligned}$$

5. Let $F^{n\times 1}$, the space of all $n \times 1$ column matrix over F, let Q be an $n \times n$ invertible matrices over F. For $X, Y \in F^{n\times 1}$ set

$$(X|Y) = Y^* Q^* Q X.$$

We are identifying the 1×1 matrix on the right with its single entry. When Q is the identity matrix this inner product is essentially the same as standard inner product.

6. Let V be the vector space of all continuous complex valued functions on the unit interval, $0 \leq t \leq 1$. Let

$$(f|g) = \int_0^1 f(t)\overline{g(t)}dt.$$

When we consider the space of real-valued continuous functions on the unit interval,

$$(f|g) = \int_0^1 f(t)g(t)dt.$$

7. This is a whole class of examples. One may construct new inner products from a given one by the following method. Let V and W be vector spaces over F and suppose $(|)$ is an inner product on W. If T is a non-singular linear transformation from V into W, then the equation

$$pr(\alpha, \beta) = (T\alpha|T\beta)$$

defines an inner product pr on V. The following are two special cases.

(a) Let V be a finite dimensional vector space, and let $\mathscr{B} = \{\alpha_1, \ldots, \alpha_n\}$ be an ordered basis for V. Let $e_1, \ldots, e_n$ be the standard basis vectors in F^n and let T be the linear transformation from V into F^n such that $T\alpha_j = e_j, \quad j = 1, 2, \ldots, n$. In other words let T be the natural isomorphism of V onto F^n that is determined by $\mathscr{B}$. If we take the standard inner product on F^n, then

$$pr(\sum_j x_j \alpha_j, \sum_j y_k \alpha_k) = \sum_{j=1}^{n} x_j \bar{y}_j$$

Thus for any basis for V there is an inner product on V with the property $(\alpha_j|\alpha_k) = \delta_{jk}$ and there exists exactly one such inner product. This method is used to determine the inner product on V corresponding to some basis.

(b) Consider $(f|g) = \int_0^1 f(t)\overline{g(t)}dt$ on the vector space V of all continuous complex valued functions on the unit interval. Let T be the linear operator multiplication by t. That is $(Tf)(t) = tf(t), \quad 0 \leq t \leq 1$.

$Tf = 0 \Rightarrow (Tf)(t) = 0 \quad \forall \quad t \in [0,1]$. That is $tf(t) = 0 \quad \forall \quad t \in [0,1]$. Then $f(t) = 0$ for $t > 0$. Since f is continuous, we have $f(0) = 0$ and therefore $f = 0$. Thus T is non-singular. Now we construct a new inner product on V by setting

$$\begin{aligned} pr(f,g) &= \int_0^1 (Tf)(t)\overline{(Tg)(t)}dt \\ &= \int_0^1 f(t)\overline{g(t)}t^2 dt \end{aligned}$$

Note 4.1. Suppose V is a complex vector space with an inner product. Then for all $\alpha, \beta \in V$,

$$(\alpha|\beta) = Re(\alpha|\beta) + iIm(\alpha|\beta)$$

where $Re(\alpha|\beta)$ is the real part of $(\alpha|\beta)$ and $Im(\alpha|\beta)$ is the imaginary part of $(\alpha|\beta)$.

Since $Im(z) = Re(-iz)$, we have

$$Im(\alpha|\beta) = Re[-i(\alpha|\beta)] = Re(\alpha|i\beta).$$

Thus

$$(\alpha|\beta) = Re(\alpha|\beta) + iRe(\alpha|i\beta) \tag{4.2}$$

Thus inner product is completely determined by its real part.

Definition 4.2. *The* ***quadratic norm*** *determined by the inner product is the function that assigns to each vector α the scalar $||\alpha||^2$.*

Properties of quadratic Norm

1. For α and β

$$||\alpha \pm \beta||^2 = ||\alpha||^2 \pm 2Re(\alpha|\beta) + ||\beta||^2$$

2. Polarization Identities:

$$(\alpha|\beta) = \frac{1}{4}||\alpha + \beta||^2 - \frac{1}{4}||\alpha - \beta||^2.$$

$$(\alpha|\beta) = \frac{1}{4}||\alpha + \beta||^2 - \frac{1}{4}||\alpha - \beta||^2 + \frac{i}{4}||\alpha + i\beta||^2 - \frac{i}{4}||\alpha - i\beta||^2.$$

Matrix of the inner product in the ordered basis $\mathscr{B}$

Suppose V is finite-dimensional, that $\mathscr{B} = \{\alpha_1, \ldots, \alpha_n\}$ is an ordered basis for V. We show that the inner product is completely determined by the values $G_{jk} = (\alpha_k|\alpha_j)$. It assumes on pairs of vectors in $\mathscr{B}$. If $\alpha = x_1\alpha_1 + \ldots + x_n\alpha_n$ and $\beta = y_1\alpha_1 + \ldots + y_n\alpha_n$, then

$$\begin{aligned}(\alpha|\beta) &= (x_1\alpha_1 + \ldots + x_n\alpha_n|\beta) \\ &= x_1(\alpha_1|\beta) + \ldots + x_n(\alpha_n|\beta)\end{aligned}$$

$$
\begin{aligned}
&= x_1(\alpha_1|y_1\alpha_1+\ldots+y_n\alpha_n)+\ldots+x_n(\alpha_n|y_1\alpha_1+\ldots+y_n\alpha_n)\\
&= x_1\bar{y_1}(\alpha_1|\alpha_1)+x_1\bar{y_2}(\alpha_1|\alpha_2)+\ldots+x_n\bar{y_1}(\alpha_n|\alpha_1)+\ldots+x_n\bar{y_n}(\alpha_n|\alpha_n)\\
&= \sum_k x_k\sum_j \bar{y_j}(\alpha_k|\alpha_j)\\
&= \sum_{j,k}\bar{y_j}G_{jk}x_k\\
&= Y^*GX
\end{aligned}
$$

where X, Y are the co-ordinate matrices of α and β in the ordered basis $\mathscr{B}$, G is the matrix with entries $G_{jk} = (\alpha_k|\alpha_j)$. We call G the matrix of the inner product in the ordered basis $\mathscr{B}$. Also note that G is an invertible Hermitian matrix satisfying the condition

$$X^*GX > 0, X \neq 0. \tag{4.3}$$

If G is any $n \times n$ matrix which satisfies 4.3 and the condition $G = G^*$, then G is the matrix in the ordered basis $\mathscr{B}$ of an inner product on V. This inner product is given by the equation

$$(\alpha|\beta) = Y^*GX$$

where X, Y are the co-ordinate matrices of α and β in the ordered basis $\mathscr{B}$.

Exercises

1. Let V be a vector space and $(\ |\ )$ an inner product on V.
   (a) Show that $(0|\beta) = 0$ for all $\beta \in V$.
   (b) Show that if $(\alpha|\beta) = 0$ for all $\beta \in V$, then $\alpha = 0$.
2. Let V be a vector space over F. Show that the sum of two inner products of V is an inner product on V. Is the difference of two inner products of V an inner product? Show that a positive multiple of an inner product is an inner product.
3. Describe explicitly all inner products on R^1 and on C^1.
4. Verify that standard inner product on F^n is an inner product.
5. Let $(|)$ be the standard inner product on R^2.
   (a) Let $\alpha = (1, 2)$, $\beta = (-1, 1)$. If γ is a vector such that $(\alpha|\gamma) = -1$ and $(\beta|\gamma) = 3$, find γ.
   (b) Show that for any $\alpha \in R^2$ we have $\alpha = (\alpha|e_1)e_1 + (\alpha|e_2)e_2$.

4.2 Inner Product Spaces

Definition 4.3. *An inner product space is a real or complex vector space together with a specified inner product on that space.*

Note that a finite dimensional real inner product space is often called a **Euclidean Space**. A complex inner product space is called a **Unitary Space**.

Theorem 4.1. *If V is an inner product space, then for any vectors $\alpha, \beta \in V$, and any scalar c,*

(i) $||c\alpha|| = |c|\ ||\alpha||$;

(ii) $||\alpha|| > 0$ *for* $\alpha \neq 0$;

(iii) $|(\alpha|\beta)| \leq ||\alpha||\ ||\beta||$;*(Cauchy-Schwarz inequality.)*

(iv) $||\alpha + \beta|| \leq ||\alpha|| + ||\beta||$.

Proof. We have $||\alpha||^2 = (\alpha|\alpha)$. Thus (i) and (ii) can be easily verified using the definition of norm and inner product.

Proof of (iii):
It is true when $\alpha = 0$. Assume $\alpha \neq 0$. Then $||\alpha|| \neq 0$. Define

$$\gamma = \beta - \frac{(\beta|\alpha)}{||\alpha||^2}\,\alpha.$$

Then $(\gamma|\alpha) = 0$ and $||\gamma||^2 > 0$. But

$$\begin{aligned}
||\gamma||^2 &= \left(\beta - \frac{(\beta|\alpha)}{||\alpha||^2}\,\alpha \quad | \quad \beta - \frac{(\beta|\alpha)}{||\alpha||^2}\,\alpha\right) \\
&= (\beta|\beta) - \frac{\overline{(\beta|\alpha)}(\beta|\alpha)}{||\alpha||^2} - \frac{(\beta|\alpha)(\alpha|\beta)}{||\alpha||^2} + \frac{(\beta|\alpha)(\beta|\alpha)||\alpha||^2}{||\alpha||^2||\alpha||^2} \\
&= (\beta|\beta) - \frac{(\beta|\alpha)(\alpha|\beta)}{||\alpha||^2} \\
&= ||\beta||^2 - \frac{|(\alpha|\beta)|^2}{||\alpha||^2}
\end{aligned}$$

Hence $|(\alpha|\beta)| \leq ||\alpha||\ ||\beta||$. This inequality is called Cauchy Schwarz Inequality.

Proof of (iv)

$$
\begin{aligned}
||\alpha+\beta||^2 = ||\alpha||^2 + (\alpha|\beta) + \overline{(\alpha|\beta)} + ||\beta||^2 \\
&= ||\alpha||^2 + 2Re(\alpha|\beta) + ||\beta||^2 \\
&\leq ||\alpha||^2 + 2||\alpha||\ ||\beta|| + ||\beta||^2 \\
&= (||\alpha|| + ||\beta||)^2.
\end{aligned}
$$

Hence We get $||\alpha+\beta|| \leq ||\alpha|| + ||\beta||$. □

Remark 19. Cauchy -Schwarz inequality has a wide variety of applications. The proof shows that if α is non zero then $|(\alpha|\beta)| \leq ||\alpha||\ ||\beta||$ unless

$$\beta = \frac{(\beta|\alpha)}{||\alpha||^2}\alpha.$$

Thus equality occurs in (iii) if and only if α and β are linearly dependent.

Definition 4.4. *Let α and β be vectors in an inner product space V. Then α is orthogonal to β if $(\alpha|\beta) = 0$*

Since $(\alpha|\beta) = 0$ implies that $(\beta|\alpha) = 0$. That is β is orthogonal to α. Hence we can simply say that α and β are orthogonal instead of saying α is orthogonal to β.

*A set S of vectors in V is said to be **orthogonal** if all pairs of distinct vectors in S are orthogonal.*

*An **orthonormal set** is an orthogonal set with the additional property that $||\alpha = 1||$ for every $\alpha \in S$.*

Since $(0|\alpha) = 0$, the zero vector is orthogonal to every vector in V and is the only vector with this property. It is appropriate to think of an orthonormal set as a set of mutually perpendicular vectors, each having length 1.

Example 43. The standard basis of either R^n or C^n is an orthonormal set with respect to the standard inner product.

Example 44. The vector $(x, y) \in \quad R^1$ is orthogonal to $(-y, x)$ with respect to the standard inner product, for $((x, y)|(-y, x)) = -xy + yx = 0$.

Example 45. For $\alpha = (x_1, x_2)$ and $\beta = (y_1, y_2)$ in R^2, let $(\alpha|\beta) = x_1y_1 - x_2y_1 - x_1y_2 + 4x_2y_2$ In this inner product (x, y) and $(-y, x)$ are orthogonal if and only if

$$y^2 - x^2 + 3xy = 0 \Rightarrow y = 1/2.(-3 \pm \sqrt{13})x.$$

Example 46. Let V be the space of complex $n \times n$ matrices, and let E^{pq} be the matrix whose only non-zero entry is a 1 in row p and column q. Then the set of all such matrices E^{pq} is orthonormal with respect to the inner product given by $(A|B) = tr(AB*)$.

For,

$$\begin{aligned} (E^{pq}|E^{rs}) &= tr(E^{pq}E^{rs}) \\ &= \delta_{qs}tr(E^{pr}) \\ &= \delta_{qs}\delta_{pr}. \end{aligned}$$

Example 47. Let V be the space of continuous complex – valued (or real-valued) functions on the interval $0 \leq x \leq 1$ with the inner product

$$(f|g) = \int_0^1 f(x)\overline{g(x)}dx.$$

Suppose $f_n(x) = \sqrt{2}\cos 2\pi nx$ and that $g_n(x) = \sqrt{2}\sin 2\pi nx$, for any positive integer n. Then $\{1, f_1, g_1, f_2, g_2, \ldots\}$ is an infinite orthonormal set. In the complex case, we may also form the linear combinations $1\sqrt{2}(f_n + ig_n)$, $n = 1, 2, \ldots$. In this way we get a new orthonormal set S which consists of all functions of the form $h_n(x) = e^{2\pi inx}$, $n = \pm 1, \pm 2, \ldots$ The set S' obtained from S by adjoining the constant function 1 is also orthonormal.

Theorem 4.2. *An orthogonal set of non-zero vectors is linearly independent.*

Proof. Let S be a finite or infinite dimensional set of non-zero vectors in a given inner product space. Suppose $\alpha_1, \alpha_2, \ldots, \alpha_m$ are distinct vectors in S. Then $(\alpha_j|\alpha_k) = 0$ for $j \neq k$. Let

$$\beta = c_1\alpha_1 + \ldots + c_m\alpha_m.$$

Then

$$\begin{aligned}
(\beta|\alpha_k) &= (\sum_j c_j\alpha_j|\alpha_k) \\
&= \sum_j c_j(\alpha_j|\alpha_k) \\
&= c_k(\alpha_k|\alpha_k) \quad (\text{since } (\alpha_j|\alpha_k) = 0 \text{ for} j \neq k)
\end{aligned}$$

Since $(\alpha_k|\alpha_k) \neq 0$, we have

$$c_k = \frac{(\beta|\alpha_k)}{||\alpha_k||^2} \quad 1 \leq k \leq m.$$

Thus when $\beta = 0$, each $c_k = 0$, so S is an independent set. □

Corollary 4.1. *If a vector β is a linear combination of an orthogonal sequence of non-zero vectors $\alpha_1, \ldots, \alpha_m$ then β is the particular linear combination*

$$\beta = \sum_{k=1}^{m} \frac{(\beta|\alpha_k)}{||\alpha_k||^2}\alpha_k$$

Proof. Suppose $\alpha_1, \ldots, \alpha_m$, be an orthogonal sequence of non-zero vectors. Then

$$\beta = c_1\alpha_1 + \ldots + c_m\alpha_m.$$

Then

$$(\beta|\alpha_k) = (\sum_j c_j\alpha_j|\alpha_k)$$

$$
\begin{aligned}
&= \sum_j c_j(\alpha_j|\alpha_k) \\
&= c_k(\alpha_k|\alpha_k) \\
&= c_k||\alpha_k||^2
\end{aligned}
$$

This implies that

$$c_k = \frac{(\beta|\alpha_k)}{||\alpha_k||^2} \quad 1 \leq k \leq m.$$

Hence

$$\beta = \sum_{k=1}^{m} \frac{(\beta|\alpha_k)}{||\alpha_k||^2}\alpha_k$$

□

Corollary 4.2. *: If $\{\alpha_1, \ldots, \alpha_m\}$ is an orthogonal set of non-zero vectors in a finite-dimensional inner product space V, then $m \leq dimV$.*

Proof. If $\{\alpha_1, \ldots, \alpha_m\}$ is an orthogonal set of non-zero vectors in V, then they are linearly independent. Hence $m \leq dimV$. □

This says that the number of mutually orthogonal directions in V cannot exceed the algebraically defined dimension of V. The maximum number of mutually orthogonal directions in V is what one would regard as the geometric dimension of V, and we have just seen that this is not greater than the algebraic dimension.

Theorem 4.3. *Let V be an inner product space and let $\{\beta_1, \ldots, \beta_n\}$ be any independent vectors in V. Then one may construct orthogonal vectors $\{\alpha_1, \ldots, \alpha_n\}$ in V such that for each $k = 1, 2, \ldots, n$, the set $\{\alpha_1, \ldots, \alpha_k\}$ is a basis for the subspace spanned by $\beta_1, \ldots, \beta_k$.*

Proof. The vectors will be obtained by means of a construction known as the Gram-Schmidt Orthogonalization process. First let $\alpha_1 = \beta_1$. The other vectors are then given inductively as follows: Suppose $\alpha_1, \ldots, \alpha_m$, $1 \leq m < n$ have been chosen so that for every k,

$$\{\alpha_1, \ldots, \alpha_k\}, \quad 1 \leq k \leq m$$

is an orthogonal basis for the subspace of V that is spanned by $\beta_1, \ldots, \beta_k$. To construct the next vector α_{m+1}, let

$$\alpha_{m+1} = \beta_{m+1} - \sum_{k=1}^{m} \frac{(\beta_{m+1}|\alpha_k)}{||\alpha_k||^2}\alpha_k.$$

Then $\alpha_{m+1} \neq 0$, for otherwise β_{m+1} is a linear combination of $\alpha_1, \ldots, \alpha_k$ and hence a linear combination of $\beta_1, \ldots, \beta_m$. Further more, if $1 \leq j \leq m$, then

$$\begin{aligned}
(\alpha_{m+1}|\alpha_j) &= (\beta_{m+1}|\alpha_j) - \sum_{k=1}^{m} \frac{(\beta_{m+1}|\alpha_k)}{||\alpha_k||^2}(\alpha_k|\alpha_j) \\
&= (\beta_{m+1}|\alpha_j) - (\beta_{m+1}|\alpha_j) \\
&= 0.
\end{aligned}$$

Therefore $\{\alpha_1, \ldots, \alpha_{m+1}\}$ is an orthogonal set consisting of $m+1$ non-zero vectors in the subspace spanned by $\beta_1, \beta_2, \ldots, \beta_{m+1}$. Since an orthogonal set of non-zero vectors is linearly independent, it is a basis for this subspace. $\square$

In particular when $n = 4$, we have

$$\alpha_1 = \beta_1 \tag{4.4}$$

$$\alpha_2 = \beta_2 - \frac{(\beta_2|\alpha_1)}{||\alpha_1||^2}\alpha_1 \tag{4.5}$$

$$\alpha_3 = \beta_3 - \frac{(\beta_3|\alpha_1)}{||\alpha_1||^2}\alpha_1 - \frac{(\beta_3|\alpha_2)}{||\alpha_2||^2}\alpha_2 \tag{4.6}$$

$$\alpha_4 = \beta_4 - \frac{(\beta_4|\alpha_1)}{||\alpha_1||^2}\alpha_1 - \frac{(\beta_4|\alpha_2)}{||\alpha_2||^2}\alpha_2 - \frac{(\beta_4|\alpha_3)}{||\alpha_3||^2}\alpha_3. \tag{4.7}$$

Corollary 4.3. *Every finite-dimensional inner product space has an orthonormal basis.*

Proof. Let V be a finite-dimensional inner product space and $\{\beta_1, \beta_2, \ldots, \beta_n\}$ is a basis for V. They by using Gram-Schmidt process we can construct an orthogonal basis $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$. Then

$$||(\frac{\alpha_i}{||\alpha_i||})|| = 1 \quad i = 1, 2, \ldots, n.$$

Hence to obtain an orthonormal basis, replace each vector α_i by $\frac{\alpha_i}{||\alpha_i||}$, $i = 1, 2, \dots, n$. Thus

$$\{\frac{\alpha_1}{||\alpha_1||}, \frac{\alpha_2}{||\alpha_2||}, \dots, \frac{\alpha_n}{||\alpha_n||}\}$$

is an orthonormal basis for V. □

Remark 20.

1. Computations involving coordinates are simpler when we use orthonormal bases instead of arbitrary bases.

2. Suppose V is a finite dimensional inner product space. Then by using the equation $G_{jk} = (\alpha_k|\alpha_j)$ we can associate a matrix G with every ordered basis $\mathscr{B} = \{\alpha_1, \alpha_2, \dots, \alpha_n\}$ of V. Using this matrix we may compute inner products in terms of co-ordinates. If $\mathscr{B}$ is an orthonormal basis, then G is the identity matrix, and for any scalars x_j and y_k,

$$(\sum_j x_j\alpha_j|\sum_k y_k\alpha_k) = \sum_j x_j\bar{y_j}.$$

Thus in terms of an orthonormal basis, the inner product in V looks like the standard inner product in F^n.

Remark 21. Gram-Schmidt process is also used to test linear dependence of vectors. For suppose $\beta_1, \dots, \beta_n$ are linearly dependent vectors in an inner product space V. To exclude a trivial case, assume that $\beta_1 \neq 0$. Let m be the largest integer for which $\beta_1, \dots, \beta_m$ are independent. Then $1 \leq m < n$. Let $\alpha_1, \dots, \alpha_m$ be the vectors obtained by applying the orthogonalization process to $\beta_1, \dots, \beta_m$. Then the vector

$$\alpha_{m+1} = \beta_{m+1} - \sum_{k=1}^{m} \frac{(\beta_{m+1}|\alpha_k)}{||\alpha_k||^2}\alpha_k.$$

is necessarily 0. For β_{m+1} is in the subspace spanned by $\alpha_1, \dots, \alpha_m$ and orthogonal to each of these vectors; hence it is 0 (If a vector β is a linear combination of an orthonormal sequence of non-zero vectors $\alpha_1, \ldots, \alpha_m$ then $\beta = \sum_{k=1}^{m} \frac{(\beta|\alpha_k)}{||\alpha_k||^2}\alpha_k)$. Conversely, if $\alpha_1, \ldots, \alpha_m$ are different from 0 and $\alpha_{m+1} = 0$, then $\beta_1, \ldots, \beta_m$ are linearly dependent.

Example 48. 1. Consider the vectors Let $\beta_1 = (3, 0, 4)$ $\beta_2 = (-1, 0, 7)$ $\beta_3 = (2, 9, 11)$ in R^3 equipped with the standard inner product. Applying the Gram-Schmidt process to $\beta_1, \beta_2, \beta_3$, we obtain the following vectors. $\alpha_1 = (3, 0, 4)$

$$\begin{aligned}
\alpha_2 &= \beta_2 - \frac{(\beta_2|\alpha_1)}{||\alpha_1||^2} \\
&= (-1,0,7) - \frac{(-1,0,7)|(3,0,4)}{25}(3,0,4) \\
&= (-1,0,7) - (3,0,4) \\
&= (-4,0,3)
\end{aligned}$$

$$\begin{aligned}
\alpha_3 &= \beta_3 - \frac{\beta_3|\alpha_1}{||\alpha_1||^2}\alpha_1 - \frac{\beta_3|\alpha_2}{||\alpha_2||^2}\alpha_2 \\
&= (2,9,11) - \frac{(2,9,11)|(3,0,4)}{25}(3,0,4) - \frac{(2,9,11)|(-4,0,3)}{25}(-4,0,3) \\
&= (2,9,11) - 2(3,0,4) - (-4,0,3) \\
&= (0,9,0).
\end{aligned}$$

These vectors are non-zero and mutually orthogonal. Hence $\{\alpha_1, \alpha_2, \alpha_3\}$ is an orthogonal basis for R^3. Then any vector (x_1, x_2, x_3) in R^3 can be written as a linear combination of $\alpha_1, \alpha_2, \alpha_3$.

$$(x_1, x_2, x_3) = \frac{3x_1 + 4x_3}{25}\alpha_1 + \frac{-4x_1 + 3x_3}{25}\alpha_2 + \frac{x_2}{9}\alpha_3.$$

We can say the same in a dual point of view. That is, if $\{f_1, f_2, f_3\}$ is the basis of $(R^3)^*$ which is dual to the basis $\{\alpha_1, \alpha_2, \alpha_3\}$ is defined explicitly by the equations

$$f_1(x_1,x_2,x_3) = \frac{3x_1+4x_3}{25}\alpha_1$$
$$f_2(x_1,x_2,x_3) = \frac{-4x_1+3x_3}{25}\alpha_2$$
$$f_3(x_1,x_2,x_3) = \frac{x_2}{9}\alpha_3.$$

And these equations can be more generally written as

$$f_j(x_1,x_2,x_3) = \frac{((x_1,x_2,x_3)|\alpha_j)}{||\alpha_j||^2}.$$

Also note that $\{\frac{1}{5}(3,0,4), \frac{1}{5}(-4,0,3), (0,1,0)\}$ is an orthonormal basis for R^3.

2. : Let
$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$
where a,b,c,d are complex numbers. Set $\beta_1 = (a,b)$, $\beta_2 = (c,d)$, and suppose that $\beta_1 \neq 0$. If we apply the orthogonalization process to β_1, β_2 using the standard inner product in C^2, we obtain the following vectors: $\alpha_1 = \beta_1 = (a,b)$

$$\begin{aligned}\alpha_2 &= (c,d) - \frac{(c,d)(a,b)}{|a|^2+|b|^2}(a,b) \\ &= (c,d) - \frac{c\bar{a}+d\bar{b}}{|a|^2+|b|^2}(a,b) \\ &= (\frac{cb\bar{b}-d\bar{b}a}{|a|^2+|b|^2}, \frac{d\bar{a}a-c\bar{a}b}{|a|^2+|b|^2}) = \frac{detA}{|a|^2+|b|^2}(-\bar{b},\bar{a})\end{aligned}$$

Now the general theory tells us that $\alpha_2 \neq 0$ if and only if β_1, β_2 are linearly independent. On the other hand, the formula for α_2 shows that this is the case if and only if $detA \neq 0$.

Note 4.2. Suppose W is a subspace of an inner product space V, and let β be an arbitrary vector in V. The problem is to find a best possible approximation to β by vectors in W. This means, we want to find a vector α for which $||\beta - \alpha||$ is as small as possible subject to the restriction that α should belong to W.

Definition 4.5. *Suppose W is a subspace of an inner product space V, and let β be an arbitrary vector in V. A best approximation to β by vectors in W is a vector α in W such that $||\beta - \alpha|| \leq ||\beta - \gamma||$ every vector γ in W.*

Theorem 4.4. *Let W be a subspace of an inner product space V and β let be a vector in V.*

1. *The vector α in W is a best approximation to β by vectors in W if and only if $\beta - \alpha$ is orthogonal to every vector in W.*
2. *If a best approximation to β by vectors in W exists, it is unique.*
3. *If W is finite-dimensional and $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ is any orthogonal basis for W, then the vector*

$$\alpha = \sum_{k=1}^{m} \frac{(\beta|\alpha_k)}{||\alpha_k||^2}\alpha_k$$

is the (unique) best approximation to β by vectors in W.

Proof. First note that if γ is any vector in V, then $\beta - \gamma = \beta - \alpha + \alpha - \gamma$, and

$$||\beta - \gamma||^2 = ||\beta - \alpha||^2 + 2\,Re(\beta - \alpha|\alpha - \gamma) + ||\alpha - \gamma||^2. \tag{4.8}$$

Now suppose $\beta - \alpha$ is orthogonal to every vector in W, that γ is in W and that $\gamma \neq \alpha$. Then since $\alpha - \gamma$ is in W, it follows that

$$||\beta - \gamma||^2 = ||\beta - \alpha||^2 + ||\alpha - \gamma||^2 > ||\beta - \alpha||^2. \tag{4.9}$$

Hence α is a best approximation to β by vectors in W.

Conversely suppose that $||\beta - \gamma|| \geq ||\beta - \alpha||$ is a best approximation to β by vectors in W. Then $||\beta - \gamma|| \geq ||\beta - \alpha||$ for every γ in W. Hence from Equation 4.8, we get $2\,Re(\beta - \alpha|\alpha - \gamma)$ for all γ in W. Since every vector in W may be expressed in the form $\alpha - \gamma$ with γ in W, we see that

$$2\,Re(\beta - \alpha|\tau) + ||\tau||^2 \geq 0$$

for every τ in W. In particular, if γ is in W and $\gamma \neq \alpha$, we may take

$$\tau = -\frac{(\beta - \alpha|\alpha - \gamma)}{||\alpha - \gamma||^2}(\alpha - \gamma).$$

Then the inequality reduces to the statement

$$-2\frac{|(\beta - \alpha|\alpha - \gamma)|^2}{||\alpha - \gamma||^2} + \frac{|(\beta - \alpha|\alpha - \gamma)|^2}{||\alpha - \gamma||^2} \geq 0.$$

This holds if and only if $(\beta - \alpha|\alpha - \gamma) = 0$. Therefore, $\beta - \alpha$ is orthogonal to every vector in W. This completes the proof of the equivalence of the two conditions on α.

ii). Suppose α_1 and α_2 are two best approximation to β by vectors in W. Then $\beta - \alpha_1$ is orthogonal to every vector in W and $\beta - \alpha_2$ also orthogonal to every vector in W. Then

$$\begin{aligned}
(\alpha_1 - \alpha_2|\alpha) &= (\alpha_1 - \beta + \beta - \alpha_2|\alpha) \\
&= (\alpha_1 - \beta|\alpha) + (\beta - \alpha_2|\alpha) \\
&= 0 + 0 = 0.
\end{aligned}$$

Then $\alpha_1 = \alpha_2$. Hence a best approximation to β by vectors in W, if it exists is unique.

iii) Suppose W is finite dimensional subspace of V. Then by Corollary 4.3, W has an orthonormal basis. Let $\{\alpha_1, \ldots, \alpha_n\}$ be a basis for W. Then every element of W is a linear combination of the vectors $\alpha_1, \ldots, \alpha_n$. Now to prove that

$$\alpha = \sum_{k=1}^{m} \frac{(\beta|\alpha_k)}{||\alpha_k||^2}\alpha_k$$

is the best approximation to β by vectors in W, it is enough to prove that $\beta - \alpha$ is orthogonal to every vector in W.

By the computation in the proof of Theorem 4.3, $\beta - \alpha$ is orthogonal to each of the vectors α_k ($\beta - \alpha$ is the vector obtained at the last stage when the orthogonalization process is applied to $\alpha_1, \ldots, \alpha_n, \beta$). Thus $\beta - \alpha$ is orthogonal to every linear combination of the vectors $\alpha_1, \ldots, \alpha_n$ and hence to every vector in W. If γ is in W and $\gamma \neq \alpha$, it follows that $||\beta - \gamma|| \geq ||\beta - \alpha||$. Therefore α is the best approximation of β that lies in W. □

Definition 4.6. *Let V be an inner product space and S any set of vectors in V. The orthogonal complement of S is the set $S^\perp$ (can be read as S perp) of all vectors in V which are orthogonal to every vector is S.*

The orthogonal complement of V is the zero subspace, and conversely $0^\perp = V$.

Theorem 4.5. *If S is any subset of an inner product space V, then its orthogonal complement $S^\perp$ is a subspace of V.*

Proof: Clearly $S^\perp$ is non-empty, since it contains 0. Let α and β be in $S^\perp$. Then by definition, $(\alpha|\gamma) = 0$ for all γ in S; $(\beta|\gamma) = 0$ for all γ in S Now for any scalar c

$$\begin{aligned} (c\alpha + \beta|\gamma) &= (c\alpha|\gamma) + (\beta|\gamma) \\ &= c.0 + 0 \\ &= 0 \text{ for every } \gamma \text{ in S.} \end{aligned}$$

Thus $c\alpha + \beta$ also lies in S. Hence $S^\perp$ is a subspace of V.

Whenever the vector α in 4.4 exists it is called orthogonal projection of β on W.

Definition 4.7. *Let W be a subspace of an inner product space V and let β be a vector in V and α in W is the best approximation to β by vectors in W. Then α, if it exists is called the orthogonal projection of β on W. If every vector in V has an orthogonal projection on W, the mapping that assigns to each vector in V its orthogonal projection on W is called the orthogonal projection of V on W.*

If W is finite dimensional and $\{\alpha_1, \alpha_2, \dots, \alpha_n\}$ is any orthogonal basis for W, then we know that

$$\alpha = \sum_{k=1}^{m} \frac{(\beta|\alpha_k)}{||\alpha_k||^2}\alpha_k$$

is the unique best approximation to β by vectors in W. Hence the orthogonal projection of an inner product space on a finite-dimensional subspace always exists. Theorem 4.4 also implies the following result.

Corollary 4.4. *Let V be an inner product space, W and finite dimensional subspace, and E the orthogonal projection of V on W. Then the mapping*

$$\beta \to \beta - E\beta$$

is the orthogonal projection of V on $W^{\perp}$.

Proof. Let β be an arbitrary vector in V. Then $E\beta$ is the unique best approximation to β by vectors in W. Hence $\beta - E\beta$ is orthogonal to every vector in W, i.e. $\beta - E\beta \in W^{\perp}$. Also for any vector γ in $W^{\perp}$, $\beta - \gamma = E\beta + (\beta - E\beta - \gamma)$. Since $E\beta$ is in W and $(\beta - E\beta - \gamma)$ is in $W^{\perp}$, it follows that

$$\begin{aligned} ||\beta - \gamma||^2 &= ||E\beta||^2 + ||\beta - E\beta - \gamma||^2 \\ &\geq ||\beta - (\beta - E\beta)||^2. \end{aligned}$$

with strict inequality when $\gamma \neq \beta - E\beta$. Therefore, $\beta - E\beta$ is the best approximation to β by vectors in $W^{\perp}$. $\square$

Example 49. Consider R^3 with standard inner product. Then the orthogonal projection of $(-10, 2, 8)$ on the subspace W that is spanned by $(3, 12, -1)$ is the vector

$$\begin{aligned}
\alpha &= \frac{((-10,2,8)|(3,12,-1))}{9+144+1}(3,12,-1) && (4.10)\\
&= \frac{-30+24+-8}{154}(3,12,-1) && (4.11)\\
&= \frac{-14}{154}(3,12,-1). && (4.12)
\end{aligned}$$

The orthogonal projection of R^3 on W is the linear transformation E defined by

$$(x_1, x_2, x_3) \to (\frac{3x_1 + 12x_2 - x_3}{154})(3, 12, -1).$$

The rank of E is 1, therefore nullity is 2. We have

$$\begin{aligned}
(x_1, x_2, x_3) \in N(E) &\Leftrightarrow E(x_1, x_2, x_3) = 0\\
&\Leftrightarrow (\frac{3x_1 + 12x_2 - x_3}{154})(3, 12, -1) = 0\\
&\Leftrightarrow 3x_1 + 12x_2 - x_3 = 0\\
&\Leftrightarrow (x_1, x_2, x_3) \in W^{\perp}
\end{aligned}$$

Therefore $N(E) = W^{\perp}$. Dimension of $W^{\perp} = 2$. On computing

$$(x_1, x_2, x_3) - (\frac{3x_1 + 12x_2 - x_3}{154})(3, 12, -1),$$

we can see that the orthogonal projection of R^3 on $W^{\perp}$ is the linear transformation $I - E$ that maps the vector (x_1, x_2, x_3) onto the vector

$$\frac{1}{154}(145x_1 - 36x_2 + 3x_3, -36x_1 + 10x_2 + 12x_3, 3x_1 + 12x_2 + 153x_3).$$

Theorem 4.6. *Let W be a finite-dimensional subspace of an inner product space V and let E be the orthogonal projection of V on W. Then E is an idempotent linear transformation of V onto W, $W^{\perp}$ is the null space of E, and $V = W \oplus W^{\perp}$.*

Proof. (i) E is idempotent.

Let β be an arbitrary vector in V. Then $E\beta$ is the best approximation to β that lies in W. In particular, $E\beta = \beta$ when β is in W. Therefore, $E(E\beta) = E\beta$ for every β in V; that is, E is idempotent: $E^2 = E$.

(ii)E is linear

Let α and β be any two vectors in V and c an arbitrary scalar. Then $E\alpha$ and $E\beta$ are the best approximations to α and β by vectors in E. Hence, $\alpha - E\alpha$ and $\beta - E\beta$ are each orthogonal to every vector in W. Hence the vector

$$c(\alpha - E\alpha) + (\beta - E\beta) = c(\alpha + \beta) - (cE\alpha + E\beta)$$

also belongs to $W^\perp$. Since $cE\alpha + E\beta$ is a vector in W as W is a subspace of V, it follows from Theorem 4.4 that

$$E(c\alpha + \beta) = cE\alpha + E\beta.$$

Hence E is linear.

(iii) $W^\perp$ is the null space of E.

Let N be the null space of E. To prove that $N = W^\perp$. Let β be arbitrary vector in N. Then $E\beta = 0$ Now

$$\begin{aligned}
\beta \in N \quad &\Rightarrow \quad \beta \in V \\
&\Rightarrow \quad E\beta \text{ is the unique vector in } W \text{ such that } \beta - E\beta \text{ is in } W^\perp \\
&\Rightarrow \quad 0 \text{ is the unique vector in } W \text{ such that } \beta - 0 \text{ is in } W^\perp \\
&\Rightarrow \quad \beta \in W^\perp.
\end{aligned}$$

Hence N is a subset of $W^\perp$.

Conversely suppose that $\beta \in W^\perp$. Then $(\beta|\alpha) = 0$ for all α in W. Then $(\beta - 0|\alpha) = 0$ for all α in W. Hence 0 is the best approximation to β by vectors in W. Thus $E\beta = 0$ and hence β is in N. So $W^\perp$ is a subset of N. Therefore $N = W^\perp$.

(iv). $V = W + W^\perp$

Let β be any vector in V. We have $E\beta \in W$ and $(\beta - E\beta) \in W^\perp$. Then $\beta = E\beta + (\beta - E\beta)$. Then $V = W + W^\perp$. We have $E\beta \in W$ Also $W \cap W^\perp = \{0\}$. For, if α is a vector in $W \cap W^\perp$ then $\alpha \in W$ and $\alpha \in W^\perp$. Then $(\alpha|\alpha) = 0$. Therefore, $\alpha = 0$, and hence V is the direct sum of W and $W^\perp$. □

Corollary 4.5. *Let W be a finite-dimensional subspace of an inner product space V and let E be the orthogonal projection of V on W. Then $I - E$ is the orthogonal projection of V on $W^\perp$. It is an idempotent linear transformation of V onto $W^\perp$ with null space W.*

Proof. We know that the mapping $\beta \to \beta - E\beta$ is the orthogonal projection of V on $W^\perp$. Since E is a linear transformation, this projection on $W^\perp$ is the linear transformation $I - E$. Also

$$\begin{aligned}(I-E)(I-E) &= I - E - E + E^2 \\ &= I - E \text{ since } (E^2 = E)\end{aligned}$$

Moreover, $(I - E)\beta = 0$ if and only if $\beta = E\beta$, and this is the case if and only if β is in W. Therefore W is the null space of $I - E$. □

Above theorem implies another result known as Bessel's Inequality.

Corollary 4.6. ***(Bessel's Inequality)***

Let $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ be an orthogonal set of non-zero vectors in an inner product space V. If β is any vector in V, then

$$\sum_k \frac{|(\beta|\alpha_k)|^2}{||\alpha_k||^2} \leq ||\beta||^2$$

and equality holds if and only if

$$\beta = \sum_k \frac{(\beta|\alpha_k)}{||\alpha_k||^2}\alpha_k.$$

Proof. Let

$$\gamma = \beta - \sum_k [\frac{(\beta|\alpha_k)}{||\alpha_k||^2}]\alpha_k.$$

Let

$$\gamma = \beta - \sum_k c_k\alpha_k,$$

where

$$c_k = \frac{(\beta|\alpha_k)}{||\alpha_k||^2}$$

Then $||\gamma||^2 \geq 0$. We have

$$\begin{aligned}
||\gamma||^2 &= (\gamma|\gamma) \\
&= (\beta - \sum_k c_k\alpha_k|\beta - \sum_k c_k\alpha_k) \\
&= ||\beta||^2 - \sum_k(\beta|c_k\alpha_k) - \sum_k(c_k\alpha_k|\beta) + (\sum_k c_k\alpha_k|\sum_k c_k\alpha_k) \\
&= ||\beta||^2 - \sum_k \bar{c_k}(\beta|\alpha_k) - \sum_k c_k(\alpha_k|\beta) + \sum_k c_k\bar{c_k}(\alpha_k|\alpha_k) \\
&= ||\beta||^2 - \sum_k \bar{c_k}(\beta|\alpha_k) - \sum_k c_k\overline{(\beta|\alpha_k)} + \sum_k c_k\bar{c_k}||\alpha_k||^2 \\
&= ||\beta||^2 - \sum_k \frac{\overline{(\beta|\alpha_k)}}{||\alpha_k||^2}(\beta|\alpha_k) - \sum_k \frac{(\beta|\alpha_k)}{||\alpha_k||^2}\overline{(\beta|\alpha_k)} + \sum_k \frac{(\beta|\alpha_k)}{||\alpha_k||^2}\frac{\overline{(\beta|\alpha_k)}}{||\alpha_k||^2}||\alpha_k||^2 \\
&= ||\beta||^2 - 2\sum_k \frac{\overline{(\beta|\alpha_k)}}{||\alpha_k||^2}(\beta|\alpha_k) + \sum_k \frac{(\beta|\alpha_k)}{||\alpha_k||^2}\overline{(\beta|\alpha_k)} \\
&= ||\beta||^2 - \sum_k \frac{|(\beta|\alpha_k)|^2}{||\alpha_k||^2}.
\end{aligned}$$

Now

$$||\gamma||^2 \geq 0 \Rightarrow ||\beta||^2 - \sum_k \frac{|(\beta|\alpha_k)|^2}{||\alpha_k||^2} \geq 0$$

$$\Rightarrow ||\beta||^2 \geq \sum_k \frac{|(\beta|\alpha_k)|^2}{||\alpha_k||^2}.$$

Also equality holds if and only if

$$\beta = \frac{(\beta|\alpha_k)}{||\alpha_k||^2}\alpha_k.$$

□

Remark 22. 1. If $\{\alpha_1, \ldots, \alpha_n\}$, is an orthonormal set, $||\alpha_k|| = 1$ for each k. Then Bessel's inequality says that

$$||\beta||^2 \geq \sum_k |(\beta|\alpha_k)|^2$$

2. If β is in the subspace spanned by $\{\beta_1, \ldots, \beta_n\}$ if and only if

$$\beta = \sum_k (\beta|\alpha_k)\alpha_k.$$

or if and only if Bessel's inequality is actually an equality. Of course, in the event that V is finite dimensional and $\{\alpha_1, \ldots, \alpha_n\}$ is an orthogonal basis for V, the above formula holds for every vector β in V. In other words, if $\{\alpha_1, \ldots, \alpha_n\}$ is an orthonormal basis for V, the k-th co-ordinate of β in the ordered basis $\{\alpha_1, \ldots, \alpha_n\}$ is $(\beta|\alpha_k)$.

Exercises

1. Consider R^4 with the standard inner product.Let W be the subspace of R^4 consisting of all vectors which are orthogonal to both $\alpha = (1, 0, -1, 1)$ and $\beta = (2, 3, -1, 2)$. Find a basis for W.
2. Apply the Gram-Schmidt process to the vectors $\beta_1 = (1, 0, 1)$, $\beta_2 = (1, 0, -1)$, $\beta_3 = (0, 3, 4)$ to obtain an orthonormal basis for R^3 with the standard inner product.
3. Find an inner product on R^2 such that $(e_1|e_2) = 2$.

4. Let V be an inner product space, and let α, β be vectors in V. Show that $\alpha = \beta$ if and only if $(\alpha|\gamma) = (\beta|\gamma)$ for every $\gamma \in V$.

5. Let W be a finite dimensional subspace of an inner product space V and let E be the orthogonal projection of V on W. Prove that $(E\alpha|\beta) = (\alpha|E\beta)$ for all $\alpha, \beta \in V$.