MTH 311: Advanced Linear Algebra

Semester 1, 2020-2021

Dr. Prahlad Vaidyanathan

Contents

I. Preliminaries

1. Fields

Throughout this course, we will be talking about "Vector spaces", and "Fields". The definition of a vector space depends on that of a field, so we begin with that.

Example 1.1. Consider $F = \mathbb{R}$, the set of all real numbers. It comes equipped with two operations: Addition and multiplication, which have the following properties:

(i) Addition is *commutative*

$$x + y = y + x$$

for all $x, y \in F$

(ii) Addition is *associative*

$$x + (y + z) = (x + y) + z$$

for all $x, y, z \in F$.

(iii) There is an *additive identity*, 0 (zero) with the property that

$$x + 0 = 0 + x = x$$

for all $x \in F$

(iv) For each $x \in F$, there is an *additive inverse* $(-x) \in F$ which satisfies

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

(v) Multiplication is *commutative*

$$xy = yx$$

for all $x, y \in F$

(vi) Multiplication is *associative*

$$x(yz) = (xy)z$$

for all $x, y, z \in F$

(vii) There is a *multiplicative identity*, 1 (one) with the property that

$$x1 = 1x = x$$

for all $x \in F$

(viii) To each non-zero $x \in F$, there is an *multiplicative inverse* $x^{-1} \in F$ which satisfies

$$xx^{-1} = x^{-1}x = 1$$

(ix) Finally, multiplication *distributes* over addition

$$x(y+z) = xy + xz$$

for all $x, y, z \in F$.

Definition 1.2. A *field* is a set F together with two operations

$$\begin{aligned} \text{Addition} &: (x,y) \mapsto x+y \\ \text{Multiplication} &: (x,y) \mapsto xy \end{aligned}$$

which satisfy all the conditions 1.1-1.9 above. Elements of a field will be termed *scalars*.

Example 1.3. (i) $F = \mathbb{R}$ is a field.

(ii) $F = \mathbb{C}$ is a field with the usual operations

$$\begin{aligned} \text{Addition} &: (a+ib)+(c+id) := (a+c)+i(b+d), \text{ and} \\ \text{Multiplication} &: (a+ib)(c+id) := (ac-bd)+i(ad+bc) \end{aligned}$$

(iii) $F = \mathbb{Q}$, the set of all rational numbers, is also a field. In fact, $\mathbb{Q}$ is a *subfield* of $\mathbb{R}$ (in the sense that it is a subset of $\mathbb{R}$ which also inherits the operations of addition and multiplication from $\mathbb{R}$). Also, $\mathbb{R}$ is a subfield of $\mathbb{C}$.

(iv) $F = \mathbb{Z}$ is not a field, because $2 \in \mathbb{Z}$ does not have a multiplicative inverse.

Standing Assumption: For the rest of this course, all fields will be denoted by F, and will either be $\mathbb{R}$ or $\mathbb{C}$, unless stated otherwise.

2. Matrices and Elementary Row Operations

Definition 2.1. Let F be a field and $n, m \in \mathbb{N}$ be fixed integers. Given m scalars $(y_1, y_2, \ldots, y_m) \in F^m$ and nm elements $\{a_{i,j} : 1 \leq i \leq n, 1 \leq j \leq m\}$, we wish to find n scalars $(x_1, x_2, \ldots, x_n) \in F^n$ which satisfy all the following equations

$$\begin{aligned} a_{1,1}x_1 + a_{1,2}x_2 + \ldots + a_{1,n}x_n &= y_1 \\ a_{2,1}x_1 + a_{2,2}x_2 + \ldots + a_{2,n}x_n &= y_2 \\ &\vdots \\ a_{m,1}x_1 + a_{m,2}x_2 + \ldots + a_{m,n}x_n &= y_m \end{aligned}$$

This problem is called a *system of m linear equations in n unknowns.* A tuple $(x_1, x_2, \ldots, x_n) \in F^n$ that satisfies the above system is called a *solution* of the system. If $y_1 = y_2 = \ldots = y_m = 0$, then the system is called a *homogeneous.*

We may express a system of linear equations more simply in the form

$$AX = Y$$

where

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m,1} & a_{m,2} & \dots & a_{m,n} \end{pmatrix}, X := \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \text{ and } Y := \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}$$

The expression A above is called a *matrix of coefficients* of the system, or just an $m \times n$ matrix over the field F. The term $a_{i,j}$ is called the $(i,j)^{th}$ *entry* of the matrix A. In this notation, X is an $n \times 1$ matrix, and Y is an $m \times 1$ matrix.

In order to solve this system, we employ the method of *row reduction.* You would have seen this in earlier classes on linear algebra, but we now formalize it with definitions and theorems.

Definition 2.2. Let A be an $m \times n$ matrix. An *elementary row operation* associates to A a new $m \times n$ matrix $e(A)$ in one of the following ways:

E_1: Multiplication of one row of A by a non-zero scalar: Choose $1 \leq r \leq m$ and a non-zero scalar c, then

$$e(A)_{i,j} = A_{i,j} \text{ if } i \neq r \text{ and } e(A)_{r,j} = cA_{r,j}$$

E_2: Replacement of the r^{th} row of A by row r plus c times row s, where $c \in F$ is any scalar and $r \neq s$:

$$e(A)_{i,j} = A_{i,j} \text{ if } i \neq r \text{ and } e(A)_{r,j} = A_{r,j} + cA_{s,j}$$

E_3: Interchange of two rows of A:

$$e(A)_{i,j} = A_{i,j} \text{ if } i \notin \{r,s\} \text{ and } e(A)_{r,j} = A_{s,j} \text{ and } e(A)_{s,j} = A_{r,j}$$

The first step in this process is to observe that elementary row operations are *reversible.*

Theorem 2.3. *To every elementary row operation e, there is an operation e_1 of the same type such that*

$$e(e_1(A)) = e_1(e(A)) = A$$

for any $m \times n$ matrix A.

Proof. We prove this for each type of elementary row operation from Definition 2.2.

E_1: Define e_1 by

$$e_1(B)_{i,j} = B_{i,j} \text{ if } i \neq r \text{ and } e_1(B)_{r,j} = c^{-1}B_{r,j}$$

E_2: Define e_1 by

$$e_1(B)_{i,j} = B_{i,j} \text{ if } i \neq r \text{ and } e_1(B)_{r,j} = B_{r,j} - cB_{s,j}$$

E_3: Define e_1 by

$$e_1 = e$$

□

Definition 2.4. Let A and B be two $m \times n$ matrices over a field F. We say that A is *row-equivalent* to B if B can be obtained from A by finitely many elementary row operations.

By Theorem 2.3, this is an equivalence relation on the set $F^{m\times n}$. The reason for the usefulness of this relation is the following result.

Theorem 2.5. *If A and B are row-equivalent, then for any vector $X \in F^n$,*

$$AX = 0 \Leftrightarrow BX = 0$$

Proof. By Theorem 2.3, it suffices to show that $AX = 0 \Rightarrow BX = 0$. Furthermore, we may assume without loss of generality that B is obtained from A by a single elementary row operations. So fix $X = (x_1, x_2, \dots, x_n) \in F^n$ that satisfies $AX = 0$. Then, for each $1 \leq i \leq m$, we have

$$(AX)_i = \sum_{j=1}^n a_{i,j}x_j = 0$$

We wish to show that

$$(BX)_i = \sum_{j=1}^n b_{i,j}x_j = 0$$

We consider the different possible operations as in Definition 2.2

E_1: Here, we have

$$(BX)_i = (AX)_i \text{ if } i \neq r \text{ and } (BX)_r = c(AX)_r$$

E_2: Here, we have

$$(BX)_i = (AX)_i \text{ if } i \neq r \text{ and } (BX)_r = (AX)_r + c(AX)_s$$

E_3: Here, we have

$$(BX)_r = (AX)_i \text{ if } i \notin \{r, s\} \text{ and } (BX)_r = (AX)_s \text{ and } (BX)_s = (AX)_r$$

In all three cases, $BX = 0$ holds. □

Definition 2.6. (i) An $m \times n$ matrix R is said to be *row-reduced* if

(i) The first non-zero entry of each non-zero row of R is equal to 1.

(ii) Each column of R which contains the leading non-zero entry of some row has all other entries zero.

(ii) R is said to be a *row-reduced echelon* matrix if R is row-reduced and further satisfies the following conditions

(i) Every row of R which has all its entries 0 occurs below every non-zero row.

(ii) If $R_1, R_2, \ldots, R_r$ are the non-zero rows of R, and if the leading non-zero entry of R_i occurs in column $k_i, 1 \leq i \leq r$, then

$$k_1 < k_2 < \ldots < k_r$$

Example 2.7. (i) The identity matrix I is an $n \times n$ (square) matrix whose entries are

$$I_{i,j} = \delta_{i,j} = \begin{cases} 1 & : i = j \\ 0 & : i \neq j \end{cases}$$

This is clearly a row-reduced echelon matrix.

(ii)

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

is row-reduced, but not row-reduced echelon.

(iii) The matrix

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

is not row-reduced.

We now give an example to convert a given $m \times n$ matrix to a row-reduced echelon matrix by a sequence of elementary row operations. This will give us the idea to prove the next theorem.

Example 2.8. Set

$$A = \begin{pmatrix} 0 & -1 & 3 & 2 \\ 0 & 0 & 0 & 0 \\ 1 & 4 & 0 & -1 \\ 2 & 6 & -1 & 5 \end{pmatrix}$$

We do this in the following steps, indicating each procedure by the notation from Definition 2.2.

E_3: By interchanging rows 2 and 4, we ensure that the first 3 rows are non-zero, while the last row is zero.

$$\begin{pmatrix}0 & -1 & 3 & 2\\ 1 & 4 & 0 & -1\\ 2 & 6 & -1 & 5\\ 0 & 0 & 0 & 0\end{pmatrix}$$

(i) By interchanging row 1 and 3, we ensure that, for each row R_i, if the first non-zero entry occurs in column k_i, then $k_1 < k_2 < \ldots < k_n$. Here, we get

$$\begin{pmatrix}2 & 6 & -1 & 5\\ 1 & 4 & 0 & -1\\ 0 & -1 & 3 & 2\\ 0 & 0 & 0 & 0\end{pmatrix}$$

E_1: The first non-zero entry of Row 1 is at $a_{1,1} = 2$. We multiply the row by $a_{1,1}^{-1}$ to get

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

E_2: For each following non-zero row, replace row i by (row i + $(-a_{i,1}$ times row 1)). This ensures that the first column has only one non-zeroe entry, at $a_{1,1}$.

$$\begin{pmatrix}1 & 3 & \frac{-1}{2} & \frac{5}{2}\\ 0 & 1 & \frac{1}{2} & \frac{-7}{2}\\ 0 & -1 & 3 & 2\\ 0 & 0 & 0 & 0\end{pmatrix}$$

In the previous two steps, we have ensured that the first non-zero entry of row 1 is 1, and the rest of the column has the entry 0. This process is called *pivoting*, and the element $a_{1,1}$ is called the *pivot*. The column containing this pivot is called the *pivot column* (in this case, that is column 1).

E_1: The first non-zero entry of Row 2 is at $a_{2,2} = 1$. We now pivot at this entry. First, we multiply the row by $a_{2,2}^{-1}$ to get

$$\begin{pmatrix}1 & 3 & \frac{-1}{2} & \frac{5}{2}\\ 0 & 1 & \frac{1}{2} & \frac{-7}{2}\\ 0 & -1 & 3 & 2\\ 0 & 0 & 0 & 0\end{pmatrix}$$

E_2: For each other row, replace row i by (row i + $(-a_{i,2}$ times row 2)). Notice that this does not change the value of the leading 1 in row 1. In this process, every other entry of column 2 other than $a_{2,2}$ becomes zero.

$$\begin{pmatrix} 1 & 0 & -2 & 13 \\ 0 & 1 & \frac{1}{2} & \frac{-7}{2} \\ 0 & 0 & \frac{7}{2} & \frac{-3}{2} \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

E_1: The first non-zero entry of Row 3 is at $a_{3,3} = \frac{7}{2}$. We pivot at this entry. First, we multiply the row by $a_{3,3}^{-1}$ to get

$$\begin{pmatrix} 1 & 0 & -2 & 13 \\ 0 & 1 & \frac{1}{2} & \frac{-7}{2} \\ 0 & 0 & 1 & \frac{-3}{7} \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

E_2: For each other row, replace row i by (row i + $(-a_{i,3}$ times row 3)). Note that this does not change the value of the leading 1's in row 1 and 2. In this process, every other entry of column 3 other than $a_{3,3}$ becomes zero.

$$\begin{pmatrix} 1 & 0 & 0 & \frac{85}{7} \\ 0 & 1 & 0 & \frac{-23}{7} \\ 0 & 0 & 1 & \frac{-3}{7} \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

There are no further non-zero rows, so the process stops. What we are left with a row-reduced echelon matrix.

A formal version of this algorithm will result in a proof. We avoid the gory details, but refer the interested reader to [Hoffman-Kunze, Theorem 4 and 5].

Theorem 2.9. *Every $m \times n$ matrix over a field F is row-equivalent to a row-reduced echelon matrix.*

Lemma 2.10. *Let A be an $m \times n$ matrix with $m < n$. Then the homogeneous equation $AX = 0$ has a non-zero solution.*

Proof. Suppose first that A is a row-reduced echelon matrix. Then A has r non-zero rows, whose non-zero entries occur at the columns $k_1 < k_2 < \ldots < k_r$. Suppose $X = (x_1, x_2, \ldots, x_n)$, then we relabel the $(n-r)$ variables $\{x_j : j \neq k_i\}$ as $u_1, u_2, \ldots, u_{n-r}$.

The equation $AX = 0$ now has the form

$$x_{k_1} + \sum_{j=1}^{(n-r)} c_{1,j}u_j = 0$$
$$x_{k_2} + \sum_{j=1}^{(n-r)} c_{2,j}u_j = 0$$
$$\vdots$$
$$x_{k_r} + \sum_{j=1}^{(n-r)} c_{r,j}u_j = 0$$

Now observe that $r \leq m < n$, so we may choose any values for $u_1, u_2, \ldots, u_{n-r}$, and calculate the $\{x_{k_j} : 1 \leq j \leq r\}$ from the above equations.

For instance, if $c_{1,1} \neq 0$, then take

$$u_1 = 1, u_2 = u_3 = \ldots = u_{n-r} = 0$$

which gives a non-trivial solution to the above system of equations.

Now suppose A is not a row-reduced echelon matrix. Then by Theorem 2.9, A is row-equivalent to a row-reduced echelon matrix B. By hypothesis, the equation $BX = 0$ as a non-zero solution. By Theorem 2.5, the equation $AX = 0$ also has a non-trivial solution. □

Theorem 2.11. *Let A be an $n \times n$ matrix, then A is row-equivalent to the identity matrix if and only if the system of equations $AX = 0$ has only the trivial solution.*

Proof. Suppose A is row-equivalent to the identity matrix, then the equation $IX = 0$ has only the trivial solution, so the equation $AX = 0$ has only the trivial solution by Theorem 2.5.

Conversely, suppose $AX = 0$ has only the trivial solution, then let R denote a row-reduced echelon matrix that is row-equivalent to A. Let r be the number of non-zero rows in R, then by the argument in the previous lemma, $r \geq n$.

But R has n rows, so $r \leq n$, whence $r = n$. Hence, R must have n non-zero rows, each of which has a leading 1. Furthermore, each column has exactly one non-zero entry, so R must be the identity matrix. □

3. Matrix Multiplication

Definition 3.1. Let $A = (a_{i,j})$ be an $m \times n$ matrix over a field F and $B = (b_{k,\ell})$ be an $n \times p$ matrix over F. The product AB is the $m \times p$ matrix C whose $(i,j)^{th}$ entry is given by

$$c_{i,j} := \sum_{k=1}^{n} a_{i,k} b_{k,j}$$

Example 3.2. (i) If

$$A = \begin{pmatrix} 1 & 2 & -4 \\ 3 & 2 & 7 \end{pmatrix}, \text{ and } B = \begin{pmatrix} -1 & 3 \\ 4 & 8 \\ 3 & 1 \end{pmatrix}$$

Then $C := AB$ is a 2×2 matrix given by

$$\begin{aligned} c_{1,1} &= 1(-1) + 2(4) + (-4)(3) = -5 \\ c_{1,2} &= 1(3) + 2(8) + (-4)(1) = 15 \\ c_{2,1} &= 3(-1) + 2(4) + 7(3) = 26 \\ c_{2,2} &= 3(3) + 2(8) + 7(1) = 32 \end{aligned}$$

(ii) The *identity matrix* is

$$I = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix}_{n \times n}$$

If A is any $m \times n$ matrix, then

$$AI = A$$

Similarly, if B is an $n \times p$ matrix, then

$$IB = B$$

Theorem 3.3. *Matrix multiplication is associative.*

Proof. Let A, B, C be $m \times n, n \times k$, and $k \times \ell$ matrices over F respectively. Let $D := BC$

and $E := AB$. Then

$$\begin{aligned}
[A(BC)]_{i,j} = [AD]_{i,j} &= \sum_{s=1}^{n} a_{i,s} d_{s,j} \\
&= \sum_{s=1}^{n} a_{i,s} \left(\sum_{t=1}^{k} b_{s,t} c_{t,j} \right) \\
&= \sum_{s=1}^{n} \sum_{t=1}^{k} a_{i,s} b_{s,t} c_{t,j} \\
&= \sum_{t=1}^{k} \left(\sum_{s=1}^{n} a_{i,s} b_{s,t} \right) c_{t,j} \\
&= \sum_{t=1}^{k} e_{i,t} c_{t,j} \\
&= [EC]_{i,j} = [(AB)C]_{i,j}
\end{aligned}$$

This is true for all $1 \leq i \leq m, 1 \leq j \leq \ell$, so $(AB)C = A(BC)$. □

An $m \times n$ matrix over F is called a *square* matrix if $m = n$.

Definition 3.4. An $m \times m$ matrix is said to be an *elementary* matrix if it is obtained from the $m \times m$ identity matrix by means of a single elementary row operation.

Example 3.5. A 2×2 elementary matrix is one of the following:

E_1:

$$\begin{pmatrix} c & 0 \\ 0 & 1 \end{pmatrix} \text{ or } \begin{pmatrix} 1 & 0 \\ 0 & c \end{pmatrix}$$

for some non-zero $c \in F$.

E_2:

$$\begin{pmatrix} 1 & c \\ 0 & 1 \end{pmatrix} \text{ or } \begin{pmatrix} 1 & 0 \\ c & 1 \end{pmatrix}$$

for some scalar $c \in F$.

E_3:

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

Theorem 3.6. *Let e be an elementary row operation and $E = e(I)$ be the associated $m \times m$ elementary matrix. Then*

$$e(A) = EA$$

for any $m \times n$ matrix A.

Proof. We consider each elementary operation

E_1: Here, the elementary matrix $E = e(I)$ has entries

$$E_{i,j} = \begin{cases} 0 & : i \neq j \\ 1 & : i = j, i \neq r \\ c & : i = j = r \end{cases}$$

And

$$e(A)_{i,j} = A_{i,j} \text{ if } i \neq r \text{ and } e(A)_{r,j} = cA_{r,j}$$

But an easy calculation shows that

$$(EA)_{i,j} = \sum_{k=1}^{m} E_{i,k}A_{k,j} = E_{i,i}A_{i,j} = \begin{cases} A_{i,j} & : i \neq r \\ cA_{i,j} & : i = r \end{cases}$$

Hence, $EA = e(A)$.

E_2: This is similar, and done in [Hoffman-Kunze, Theorem 9].

E_3: We leave this for the reader.

□

The next corollary follows from the definition of row-equivalence and Theorem 3.6.

Corollary 3.7. *Let A and B be two $m \times n$ matrices over a field F. Then B is row-equivalent to A if and only if $B = PA$, where P is a product of $m \times m$ elementary matrices.*

4. Invertible Matrices

Definition 4.1. Let A and B be $n \times n$ square matrices over F. We say that B is a *left inverse* of A if

$$BA = I$$

where I denotes the $n \times n$ identity matrix. Similarly, we say that B is a *right inverse* of A if

$$AB = I$$

If $AB = BA = I$, then we say that B is the *inverse* of A, and that A is *invertible*.

Lemma 4.2. *If A has a left-inverse B and a right-inverse C, then $B = C$.*

Proof.

$$B = BI = B(AC) = (BA)C = IC = C$$

□

In particular, we have shown that if A has an inverse, then that inverse is unique. We denote this inverse by A^{-1}.

Theorem 4.3. *Let A and B be $n \times n$ matrices over F.*

(i) If A is invertible, then so is A^{-1} and $(A^{-1})^{-1} = A$

(ii) If A and B are invertible, then so is AB and $(AB)^{-1} = B^{-1}A^{-1}$. Hence, the product of finitely many invertible matrices is invertible.

Proof. (i) If A is invertible, then there exists B so that $AB = BA = I$. Now $B = A^{-1}$, so since $BA = AB = I$, it follows that B is invertible and $B^{-1} = A$.

(ii) Let $C = A^{-1}$ and $D = B^{-1}$, then

$$(AB)(DC) = A(BD)C = AIC = AC = I$$

Similarly, $(DC)(AB) = I$, whence AB is invertible and $(AB)^{-1} = DC$ as required. □

Theorem 4.4. *An elementary matrix is invertible.*

Proof. Let E be the elementary matrix corresponding to a row operation e. Then by Theorem 2.3, there is an inverse row operation e_1 such that $e_1(e(A)) = e(e_1(A)) = A$. Let B be the elementary matrix corresponding to e_1, then

$$EBA = BEA = A$$

for any matrix A. In particular, $EB = BE = I$, so E is invertible. □

Example 4.5. Consider the 2×2 elementary matrices from Example 3.5. We have

$$\begin{pmatrix} c & 0 \\ 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} c^{-1} & 0 \\ 0 & 1 \end{pmatrix}$$
$$\begin{pmatrix} 1 & 0 \\ 0 & c \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & c^{-1} \end{pmatrix}$$
$$\begin{pmatrix} 1 & c \\ 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & -c \\ 0 & 1 \end{pmatrix}$$
$$\begin{pmatrix} 1 & 0 \\ c & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 0 \\ -c & 1 \end{pmatrix}$$
$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}^{-1} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

Theorem 4.6. *For an $n \times n$ matrix A, the following are equivalent:*

(i) A is invertible.

(ii) A is row-equivalent to the $n \times n$ identity matrix.

(iii) A is a product of elementary matrices.

Proof. We prove $(i) \Rightarrow (ii) \Rightarrow (iii) \Rightarrow (i)$. To begin, we let R be a row-reduced echelon matrix that is row-equivalent to A (by Theorem 2.9). By Theorem 3.6, there is a matrix P that is a product of elementary matrices such that

$$R = PA$$

$(i) \Rightarrow (ii)$: By Theorem 4.4 and Theorem 4.3, it follows that P is invertible. Since A is invertible, it follows that R is invertible. Since R is a row-reduced echelon square matrix, R is invertible if and only if $R = I$. Thus, (ii) holds.

$(ii) \Rightarrow (iii)$: If A is row-equivalent to the identity matrix, then $R = I$ in the above equation. Thus, $A = P^{-1}$. But the inverse of an elementary matrix is again an elementary matrix. Thus, by Theorem 4.3, P^{-1} is also a product of elementary matrices.

$(iii) \Rightarrow (i)$: This follows from Theorem 4.4 and Theorem 4.3.

□

The next corollary follows from Theorem 4.6 and Corollary 3.7.

Corollary 4.7. *Let A and B be $m \times n$ matrices. Then B is row-equivalent to A if and only if $B = PA$ for some invertible matrix P.*

Theorem 4.8. *For an $n \times n$ matrix A, the following are equivalent:*

(i) A is invertible.

(ii) The homogeneous system $AX = 0$ has only the trivial solution $X = 0$.

(iii) For every vector $Y \in F^n$, the system of equations $AX = Y$ has a solution.

Proof. Once again, we prove $(i) \Rightarrow (ii) \Rightarrow (i)$, and $(i) \Rightarrow (iii) \Rightarrow (i)$.

$(i) \Rightarrow (ii)$: Let $B = A^{-1}$ and X be a solution to the homogeneous system $AX = 0$, then

$$X = IX = (BA)X = B(AX) = B(0) = 0$$

Hence, $X = 0$ is the only solution.

$(ii) \Rightarrow (i)$: Suppose $AX = 0$ has only the trivial solution, then A is row-equivalent to the identity matrix by Theorem 2.11. Hence, A is invertible by Theorem 4.6.

$(i) \Rightarrow (iii)$: Given a vector Y, consider $X := A^{-1}Y$, then $AX = Y$ by associativity of matrix multiplication.

$(iii) \Rightarrow (i)$: Let R be a row-reduced echelon matrix that is row-equivalent to A. By Theorem 4.6, it suffices to show that $R = I$. Since R is a row-reduced echelon matrix, it suffices to show that the n^{th} row of R is non-zero. So set

$$Y = (0, 0, \ldots, 1)$$

Then the equation $RX = Y$ has a solution, which must necessarily be non-zero (since $Y \neq 0$). Thus, the last row of R cannot be zero. Hence, $R = I$, whence A is invertible.

□

Corollary 4.9. *A square matrix which is either left or right invertible is invertible.*

Proof. Suppose A is left-invertible, then there exists a matrix B so that $BA = I$. If X is a vector so that $AX = 0$, then $X = B(AX) = (BA)X = 0$. Hence, the equation $AX = 0$ has only the trivial solution. By Theorem 4.8, A is invertible.

Now suppose A is right-invertible, then there exists a matrix B so that $AB = I$. If Y is any vector, then $X := B(Y)$ has the property that $AX = Y$. Hence, by Theorem 4.8, A is invertible. □

Corollary 4.10. *Let $A = A_1A_2 \dots A_k$ where the A_i are $n \times n$ matrices. Then, A is invertible if and only if each A_i is invertible.*

Proof. If each A_i is invertible, then A is invertible by Theorem 4.3. Conversely, suppose A is invertible and X is a vector such that $A_kX = 0$, then

$$AX = (A_1A_2 \dots A_{k-1})A_kX = 0$$

Since A is invertible, this forces $X = 0$. Hence, the only solution to the equation $A_kX = 0$ is the trivial solution. By Theorem 4.8, it follows that A_k is invertible. Hence,

$$A_1A_2 \dots A_{k-1} = AA_k^{-1}$$

is invertible. Now, by induction on k, each A_i is invertible for $1 \leq i \leq k-1$ as well. □

(End of Week 1)

II. Vector Spaces

1. Definition and Examples

Definition 1.1. A vector space V over a field F is a set together with two operations:

$$\begin{aligned} &\text{(Addition)} && + : V \times V \to V \text{ given by } (\alpha, \beta) \mapsto \alpha + \beta \\ &\text{(Scalar Multiplication)} && \cdot : F \times V \to V \text{ given by } (c, \alpha) \mapsto c\alpha \end{aligned}$$

with the following properties:

(i) Addition is commutative

$$\alpha + \beta = \beta + \alpha$$

for all $\alpha, \beta \in V$

(ii) Addition is associative

$$\alpha + (\beta + \gamma) = (\alpha + \beta) + \gamma$$

for all $\alpha, \beta, \gamma \in V$

(iii) There is a unique zero vector $0 \in V$ which satisfies the equation

$$\alpha + 0 = 0 + \alpha = \alpha$$

for all $\alpha \in V$

(iv) For each vector $\alpha \in V$, there is a unique vector $(-\alpha) \in V$ such that

$$\alpha + (-\alpha) = (-\alpha) + \alpha = 0$$

(v) For each $\alpha \in V$,

$$1 \cdot \alpha = \alpha$$

(vi) For every $c_1, c_2 \in F$ and $\alpha \in V$,

$$(c_1 c_2)\alpha = c_1(c_2\alpha)$$

(vii) For every $c \in F$ and $\alpha, \beta \in V$,

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

(viii) For every $c_1, c_2 \in F$ and $\alpha \in V$,

$$(c_1 + c_2)\alpha = c_1\alpha + c_2\alpha$$

An element of the set V is called a *vector*, while an element of F is called a *scalar*.

Technically, a vector space is a tuple $(V, F, +, \cdot)$, but usually, we simply say that V *is a vector space over* F, when the operations $+$ and $\cdot$ are implicit.

Example 1.2. (i) *The n-tuple space F^n*: Let F be any field and V be the set of all n-tuples $\alpha = (x_1, x_2, \ldots, x_n)$ whose entries x_i are in F. If $\beta = (y_1, y_2, \ldots, y_n) \in V$ and $c \in F$, we define addition by

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

and scalar multiplication by

$$c \cdot \alpha := (cx_1, cx_2, \ldots, cx_n)$$

One can then verify that $V = F^n$ satisfies all the conditions of Definition 1.1.

(ii) *The space of $m \times n$ matrices $F^{m\times n}$*: Let F be a field and $m, n \in \mathbb{N}$ be positive integers. Let $F^{m\times n}$ be the set of all $m \times n$ matrices with entries in F. For matrices $A, B \in F^{m\times n}$, we define addition by

$$(A + B)_{i,j} := A_{i,j} + B_{i,j}$$

and scalar multiplication by

$$(cA)_{i,j} := cA_{i,j}$$

for any $c \in F$. [Observe that $F^{1\times n} = F^n$ from the previous example]

(iii) *The space of functions from a set to a field*: Let F be a field and S a non-empty set. Let V denote the set of all functions from S taking values in F. For $f, g \in V$, define

$$(f + g)(s) := f(s) + g(s)$$

where the addition on the right-hand-side is the addition in F. Similarly, scalar multiplication is defined *pointwise* by

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

which the multiplication on the right-hand-side is that of F. Once again, it is easy to verify the axioms (note that zero vector here is zero function).

- If $S = \{1, 2, \ldots, n\}$, then the function $f : S \to F$ may be identified with a tuple $(f(1), f(2), \ldots, f(n))$. Conversely, any n-tuple $(x_1, x_2, \ldots, x_n)$ may be thought of as a function. This identification shows that the first example is a special case of this example.
- Similarly, if $S = \{(i, j) : 1 \leq i \leq m, 1 \leq j \leq n\}$, then any function $f : S \to F$ may be identified with a matrix $A \in F^{m\times n}$ where $A_{i,j} := f(i, j)$. This identification is a bijection between the set of functions from $S \to F$ and the space $F^{m\times n}$. Thus, the second example is also a special case of this one.

(iv) *The space of polynomial functions over a field*: Let F be a field, and V be the set of all functions $f : F \to F$ which are of the form

$$f(x) = c_0 + c_1 x + \ldots + c_n x^n$$

for some scalars $c_0, c_1, \ldots, c_n \in F$. Such a function is called a polynomial function. With addition and scalar multiplication defined exactly as in the previous example, V forms a vector space.

(v) Let $\mathbb{C}$ denote the set of all complex numbers and $F = \mathbb{R}$. Then $\mathbb{C}$ may be thought of as a vector space over $\mathbb{R}$. In fact, $\mathbb{C}$ may be identified with $\mathbb{R}^2$.

Lemma 1.3. *(i) For any $c \in F$,*

$$c0 = 0$$

where $0 \in V$ denotes the zero vector.

(ii) If $c \in F$ is a non-zero scalar and $\alpha \in V$ such that

$$c\alpha = 0$$

Then $\alpha = 0$

(iii) For any $\alpha \in V$

$$(-1)\alpha = -\alpha$$

Proof. (i) For any $c \in F$,

$$0 + c0 = c0 = c(0+0) = c0 + c0$$

Hence, $c0 = 0$

(ii) If $c \in F$ is non-zero and $\alpha \in V$ such that

$$c\alpha = 0$$

Then

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

But

$$c^{-1}(c\alpha) = (c^{-1}c)\alpha = 1\alpha = \alpha$$

Hence, $\alpha = 0$

(iii) For any $\alpha \in V$,

$$\alpha + (-1)\alpha = 1\alpha + (-1)\alpha = (1 + (-1))\alpha = 0\alpha = 0$$

But $\alpha + (-\alpha) = 0$ and $(-\alpha)$ is the unique vector with this property. Hence,

$$(-1)\alpha = (-\alpha)$$

□

Remark 1.4. Since vector space addition is associative, for any vectors $\alpha_1, \alpha_2, \alpha_3, \alpha_4 \in V$, we have

$$\alpha_1 + (\alpha_2 + (\alpha_3 + \alpha_4))$$

can be written in many different ways by moving the parentheses around. For instance,

$$(\alpha_1 + \alpha_2) + (\alpha_3 + \alpha_4)$$

denotes the same vector. Hence, we simply drop all parentheses, and write this vector as

$$\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4$$

The same is true for any finite number of vectors $\alpha_1, \alpha_2, \ldots, \alpha_n \in V$, so the expression

$$\alpha_1 + \alpha_2 + \ldots + \alpha_n$$

denotes the common vector associated to all possible re-arrangements of parentheses.

The next definition is the most fundamental operation in a vector space, and is the reason for defining our axioms the way we have done.

Definition 1.5. Let V be a vector space over a field F, and $\alpha_1, \alpha_2, \ldots, \alpha_n \in V$. A vector $\beta \in V$ is said to be a *linear combination* of $\alpha_1, \alpha_2, \ldots, \alpha_n$ if there exist scalars $c_1, c_2, \ldots, c_n \in F$ such that

$$\beta = c_1\alpha_1 + c_2\alpha_2 + \ldots + c_n\alpha_n$$

When this happens, we write

$$\beta = \sum_{i=1}^n c_i\alpha_i$$

Note that, by the distributivity properties (Properties (vii) and $(viii)$ of Definition 1.1), we have

$$\sum_{i=1}^n c_i\alpha_i + \sum_{j=1}^n d_j\alpha_j = \sum_{k=1}^n (c_k + d_k)\alpha_k$$
$$c\left(\sum_{i=1}^n c_i\alpha_i\right) = \sum_{i=1}^n (cc_i)\alpha_i$$

Exercise: Read the end of [Hoffman-Kunze, Section 2.1] concerning the geometric interpretation of vector spaces, addition, and scalar multiplication.

2. Subspaces

Definition 2.1. Let V be a vector space over a field F. A *subspace* of V is a subset $W \subset V$ which is itself a vector space with the addition and scalar multiplication operations inherited from V.

Remark 2.2. What this definition means is that $W \subset V$ should have the following properties:

(i) If $\alpha, \beta \in W$, then $(\alpha + \beta)$ must be in W.

(ii) If $\alpha \in W$ and $c \in F$, then $c\alpha$ must be in W.

We say that W is *closed* under the operations of addition and scalar multiplication.

Theorem 2.3. *Let V be a vector space over a field F and $W \subset V$ be a non-empty set. Then W is a subspace of V if and only if, for any $\alpha, \beta \in W$ and $c \in F$, the vector $(c\alpha + \beta)$ lies in W.*

Proof. Suppose W is a subspace of V, then W is closed under the operations of scalar multiplication and addition as mentioned above. Hence, if $\alpha, \beta \in W$ and $c \in F$, then $c\alpha \in W$, so $(c\alpha + \beta) \in W$ as well.

Conversely, suppose W satisfies this condition, and we wish to show that W is subspace. In other words, we wish to show that W satisfies the conditions of Definition 1.1. By hypothesis, the addition map $+$ maps $W \times W \to W$, and the scalar multiplication map $\cdot$ maps $F \times W$ to W.

(i) Addition is commutative because it is commutative in V.

(ii) Addition is associative because it is associative in V.

(iii) V has a zero element $0 \in V$. To see that this vector lies in W, observe that W is non-empty, so it contains *some* vector $\alpha \in W$. Then $0\alpha = 0 \in W$ by Lemma 1.3.

(iv) If $\alpha \in W$, then $\alpha \in V$, so there is a unique vector $(-\alpha) \in V$ so that

$$\alpha + (-\alpha) = 0$$

But we know that $0 \in W$, so by Lemma 1.3 once again,

$$(-a) = 0 - (\alpha) = 0 + (-1)\alpha \in W$$

(v) For each $\alpha \in W$, we have $1 \cdot \alpha = \alpha$ in V. But the scalar multiplication on W is the same as that of V, so the same property holds in W as well.

(vi) The remaining three properties of a vector space are satisfied in W because they are satisfied in V (Check!)

□

Example 2.4. (i) Let V be any vector space, then $W := \{0\}$ is a subspace of V. Similarly, $W := V$ is a subspace of V. These are both called the *trivial* subspaces of V.

(ii) Let $V = F^n$ as in Example 1.2. Let

$$W := \{(x_1, x_2, \ldots, x_n) \in V : x_1 = 0\}$$

Note that if $\alpha = (x_1, x_2, \ldots, x_n), \beta = (y_1, y_2, \ldots, y_n) \in W$ and $c \in F$, then

$$x_1 = y_1 = 0 \Rightarrow cx_1 + y_1 = 0$$

Hence, $(c\alpha + \beta) \in W$. Thus W is a subspace by Theorem 2.3.

Note: If $F = \mathbb{R}$ and $n = 2$, then W defines a *line passing through the origin.*

(iii) Let $V = F^n$ as before, and let

$$W = \{(x_1, x_2, \ldots, x_n) \in V : x_1 = 1\}$$

Then W is *not* a subspace.

Note: If $F = \mathbb{R}$ and $n = 2$, then W defines a line that does not pass through the origin.

(iv) Let V denote the set of all functions from F to F, and let W denote the set of all polynomial functions from F to F. Then W is subspace of V.

(v) Let $V = F^{n\times n}$ denote the set of all $n \times n$ matrices over a field F. A matrix $A \in V$ is said to be *symmetric* if

$$A_{i,j} = A_{j,i}$$

for all $1 \leq i, j \leq n$. Let W denote the set of all symmetric matrices, then W is subspace of V (simply verify Theorem 2.3).

(vi) Let $V = \mathbb{C}^{n\times n}$ denote the set of all $n \times n$ matrices over the field $\mathbb{C}$ of complex numbers. A matrix $A \in V$ is said to be *Hermitian* (or *self-adjoint*) if

$$A_{k,\ell} = \overline{A_{\ell,k}}$$

for all $1 \leq k, \ell \leq n$. Then W is *not* a subspace of V because if $A \in W$, and $i := \sqrt{-1}$, then

$$(iA)_{k,\ell} = iA_{k,\ell}$$

while

$$\overline{iA)_{\ell,k}} = \overline{iA_{\ell,k}} = -i\overline{A_{\ell,k}} = -iA_{k,\ell}$$

Hence if A a non-zero hermitian matrix, then iA is *not* hermitian.

(vii) *The solution space of a system of homogeneous equations*: Let A be an $m \times n$ matrix over a field F, and let $V = F^n$, and set

$$W := \{X \in V : AX = 0\}$$

Then W is a subspace of V by the following lemma, because if $X, Y \in W$ and $c \in F$, then

$$A(cX + Y) = c(AX) + (AY) = 0 + 0 = 0$$

so $cX + Y \in W$.

The next lemma says that *matrix multiplication is linear.*

Lemma 2.5. *Let A be an $m \times n$ matrix over a field F, and B, C both be $n \times p$ matrices. For any scalar $d \in F$, we have*

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

Proof. For any $1 \leq i \leq m$ and $1 \leq j \leq p$, we have

$$\begin{aligned}
[A(dB + C)]_{i,j} &= \sum_{k=1}^{n} A_{i,k}[(dB + C)]_{k,j} \\
&= \sum_{k=1}^{n} A_{i,k}(dB_{k,j} + C_{k,j}) \\
&= \sum_{k=1}^{n} dA_{i,k}B_{k,j} + A_{i,k}C_{k,j} \\
&= d\left(\sum_{k=1}^{n} A_{i,k}B_{k,j}\right) + \sum_{k=1}^{n} A_{i,k}C_{k,j} \\
&= d[AB]_{i,j} + [AC]_{i,j}
\end{aligned}$$

Hence the result. □

Theorem 2.6. *Let V be a vector space, and $\{W_a : a \in A\}$ be a collection of subspaces of V. Then*

$$W := \bigcap_{a \in A} W_a$$

is a subspace of V.

Proof. We verify Theorem 2.3. If $\alpha, \beta \in W$ and $c \in F$, then we wish to show that

$$c\alpha + \beta \in W$$

Fix $a \in A$. Then $\alpha, \beta \in W_a$. Since W_a is subspace

$$c\alpha + \beta \in W_a$$

This is true for any $a \in A$, so

$$c\alpha + \beta \in W$$

as required. □

Note: If V is a vector space, and $S \subset V$ is any set, then consider the collection

$$\mathcal{F} := \{W : W \text{ is a subspace of } V, \text{ and } S \subset W\}$$

of all subspaces of V that contain S. Note that $\mathcal{F}$ is a non-empty set because $V \in \mathcal{F}$. Hence, it makes sense to take the intersection of all members of $\mathcal{F}$. By Theorem 2.6, this intersection is once again a subspace.

Definition 2.7. Let V be a vector space and $S \subset V$ be any subset. The subspace *spanned by* S is the intersection of all subspaces of V containing S.

Note that this intersection is once again a subspace of V. Furthermore, if this intersection is denoted by W, then W is the smallest subspace of V containing S. In other words, if W' is another subspace of V such that $S \subset W'$, then it follows that $W \subset W'$.

Theorem 2.8. *The subspace spanned by a set S is the set of all linear combinations of vectors in S.*

Proof. Define

$$W := \{c_1\alpha_1 + c_2\alpha_2 + \ldots + c_n\alpha_n : c_i \in F, \alpha_i \in S\}$$

In other words, $\beta \in W$ if and only if there exist $\alpha_1, \alpha_2, \ldots, \alpha_n \in S$ and scalars $c_1, c_2, \ldots, c_n \in F$ such that

$$\beta = \sum_{i=1}^n c_i\alpha_i \tag{II.1}$$

Then

(i) W is a subspace of V

Proof. If $\alpha, \beta \in W$ and $c \in F$, then write

$$\alpha = \sum_{i=1}^n c_i\alpha_i$$

for some $c_i \in F$ and $\alpha_i \in S$. Similarly,

$$\beta = \sum_{j=1}^m d_j\beta_j$$

for some $d_j \in F$ and $\beta_j \in S$. Then

$$c\alpha + \beta = \sum_{i=1}^n (cc_i)\alpha_i + \sum_{j=1}^m d_j\beta_j$$

Thus, $c\alpha + \beta$ is also of the form in Equation II.7, and so $c\alpha + \beta \in W$. So, by Theorem 2.3, W is a subspace of V. □

(ii) If L is any other subspace of V containing S, then $W \subset L$.

Proof. If $\beta \in W$, then there exists $c_i \in F$ and $\alpha_i \in S$ such that

$$\beta = \sum_{i=1}^{n} c_i \alpha_i$$

Since L is a subspace containing S, $\alpha_i \in L$ for all $1 \leq i \leq n$. Hence, $\sum_{i=1}^{n} c_i \alpha_i \in L$. Thus, $W \subset L$ as required. □

By (i) and (ii), W is the smallest subspace containing S. Hence, W is the subspace spanned by S. □

Example 2.9. (i) Let $F = \mathbb{R}, V = \mathbb{R}^3$ and $S = \{(1,0,1),(2,0,3)\}$. Then the subspace W spanned by S has the form

$$W = \{c(1,0,1) + d(2,0,3) : c, d \in \mathbb{R}\}$$

Hence, $\alpha = (a_1, a_2, a_3) \in W$ if and only if there exist $c, d \in \mathbb{R}$ such that

$$\alpha = c(1,0,1) + d(2,0,3) = (c + 2d, 0, c + 3d)$$

Replacing $x \leftrightarrow c + 2d, y \leftrightarrow c + 3d$, we get

$$\alpha = (x, 0, y)$$

Hence,

$$W = \{(x,0,y) : x, y \in \mathbb{R}\}$$

Thus, $(2,0,5) \in W$ but $(1,1,1) \notin W$.

(ii) Let V be the space of all functions from F to F and W be the subspace of all polynomial functions. For $n \geq 0$, define $f_n \in V$ by

$$f_n(x) = x^n$$

Then, W is the subspace spanned by the set $\{f_0, f_1, f_2, \ldots\}$

Definition 2.10. Let $S_1, S_2, \ldots, S_k$ be k subsets of a vector space V. Define

$$S_1 + S_2 + \ldots + S_k$$

to be the set consisting of all vectors of the form

$$\alpha_1 + \alpha_2 + \ldots + \alpha_k$$

where $\alpha_i \in S_i$ for all $1 \leq i \leq k$.

Remark 2.11. If $W_1, W_2, \ldots, W_k$ are k subspaces of a vector space V, then

$$W := W_1 + W_2 + \ldots + W_k$$

is a subspace of V (Check!)

3. Bases and Dimension

Definition 3.1. Let V be a vector space over a field F and $S \subset V$ be a subset of V. We say that S is *linearly dependent* if there exist distinct vectors $\{\alpha_1, a_2, \ldots, \alpha_n\} \subset S$ and scalars $\{c_1, c_2, \ldots, c_n\} \subset F$, not all of which are zero, such that

$$\sum_{i=1}^{n} c_i\alpha_i = 0$$

A set which is not linearly dependent is said to be *linearly independent.*

Remark 3.2. (i) Any set which contains a linearly dependent set is linearly dependent.

(ii) Any subset of a linearly independent set is linearly independent.

(iii) The set $\{0\}$ is linearly dependent. So, if S contains 0, then S is linearly dependent.

(iv) If $S = \{\alpha\}$ where $\alpha \neq 0$, then S is linearly independent.

(v) A set $S = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ is linearly independent if and only if, whenever $c_1, c_2, \ldots, c_n \in F$ are scalars such that

$$\sum_{i=1}^{n} c_i\alpha_i = 0$$

then $c_i = 0$ for all $1 \leq i \leq n$.

(vi) Let S be an infinite set such that every finite subset of S is linearly independent, then S is linearly independent.

Example 3.3. (i) If $S = \{\alpha_1, \alpha_2\}$, then S is linearly dependent if and only if there exists a non-zero scalar $c \in F$ such that

$$\alpha_2 = c\alpha_1$$

In other words, α_2 lies on the *line* containing α_1.

(ii) If $S = \{\alpha_1, \alpha_2, \alpha_3\}$ is linearly dependent, then choose scalars $c_1, c_2, c_3 \in F$ not all zero such that

$$c_1\alpha_1 + c_2\alpha_2 + c_3\alpha_3 = 0$$

Suppose that $c_1 \neq 0$, then dividing by c_1, we get an expression

$$\alpha_1 = d_2\alpha_2 + d_3\alpha_3$$

In other words, α_1 lies on the *plane* generated by $\{\alpha_2, \alpha_3\}$.

(iii) Let $V = \mathbb{R}^3$ and $S = \{\alpha_1, \alpha_2, \alpha_3\}$ where

$$\begin{aligned}\alpha_1 &:= (1, 1, 0)\\ \alpha_2 &:= (0, 1, 0)\\ \alpha_3 &:= (1, 2, 0)\end{aligned}$$

Then S is linearly dependent because

$$\alpha_3 = \alpha_1 + \alpha_2$$

(iv) Let $V = F^n$, and define

$$\begin{aligned}\epsilon_1 &:= (1, 0, 0, \ldots, 0)\\ \epsilon_2 &:= (0, 1, 0, \ldots, 0)\\ &\vdots\\ \epsilon_n &:= (0, 0, 0, \ldots, 1)\end{aligned}$$

Suppose $c_1, c_2, \ldots, c_n \in F$ are scalars such that

$$\sum_{i=1}^n c_i\epsilon_i = 0$$

Then,

$$(c_1, c_2, \ldots, c_n) = 0 \Rightarrow c_i = 0 \quad \forall 1 \leq i \leq n$$

Hence, $\{\epsilon_1, \epsilon_2, \ldots, \epsilon_n\}$ is linearly independent.

Definition 3.4. A *basis* for V is a linearly independent spanning set. If V has a finite basis, then we say that V is *finite dimensional.*

Example 3.5. (i) If $V = F^n$ and $S = \{\epsilon_1, \epsilon_2, \ldots, \epsilon_n\}$ from Example 3.3, then S is a basis for V. Hence, V is finite dimensional. S is called the *standard basis* for F^n.

(ii) Let $V = F^n$ and P be an invertible $n \times n$ matrix. Let $P_1, P_2, \ldots, P_n$ denote the columns of P. Then, we claim that $S = \{P_1, P_2, \ldots, P_n\}$ is a basis for V.

Proof. (i) S is linearly independent: To see this, suppose $c_1, c_2, \ldots, c_n \in F$ are such that

$$c_1P_1 + c_2P_2 + \ldots + c_nP_n = 0$$

Let $X = (c_1, c_2, \ldots, c_n) \in V$, then it follows that

$$PX = 0$$

But this implies $X = IX = P^{-1}(PX) = P^{-1}(0) = 0$. Hence, $c_i = 0$ for all $1 \leq i \leq n$.

(ii) S is a spanning set for V: To see this, suppose $Y = (x_1, x_2, \ldots, x_n) \in V$, then consider

$$X := P^{-1}Y$$

so that $PX = Y$. It follows that, if $X = (c_1, c_2, \ldots, c_n)$, then

$$c_1P_1 + c_2P_2 + \ldots + c_nP_n = Y$$

Hence the claim.

□

(iii) Let V be the space of all polynomial functions from F to F. For $n \geq 0$, define $f_n \in V$ by

$$f_n(x) = x^n$$

Then, as we saw in Example 2.9, $S := \{f_0, f_1, f_2, \ldots\}$ is a spanning set. Also, if $c_0, c_2, \ldots, c_k \in F$ are scalars such that

$$\sum_{i=0}^{n} c_i f_i = 0$$

Then, it follows that the polynomial

$$c_0 + c_1 x + c_2 x^2 + \ldots + c_k x^k$$

is the zero polynomial. Since a non-zero polynomial can only have finitely many roots, it follows that $c_i = 0$ for all $0 \leq i \leq k$. Thus, every finite subset of S is linearly independent, and so S is linearly independent. Hence, S is a basis for V.

(iv) Let V be the space of all continuous functions from F to F, and let S be as in the previous example. Then, we claim that S is not a basis for V.

(i) S remains linearly independent in V

(ii) S does not span V: To see this, let $f \in V$ be any function that is non-zero, but is zero on an infinite set (for instance, $f(x) = \sin(x)$). Then f cannot be expressed as a polynomial, and so is not in the span of S.

Remark 3.6. Note that, even if a vector space has an infinite basis, there is no such thing as an infinite linear combination. In other words, a set S is a basis for a vector space V if and only if

(i) Every finite subset of S is linearly independent.

(ii) For every $\alpha \in V$, there exist finitely many vectors $\alpha_1, \alpha_2, \ldots, \alpha_n$ in S and scalars $c_1, c_2, \ldots, c_n \in F$ such that

$$\alpha = \sum_{i=1}^{n} c_i \alpha_i$$

Hence, the symbols

$$\text{``}\sum_{n=1}^{\infty} c_n \alpha_n\text{''}$$

does not make sense.

Theorem 3.7. *Let V be a vector space which is spanned by a set $\{\beta_1, \beta_2, \ldots, \beta_m\}$. Then, any linearly independent set of vectors in V is finite, and contains no more than m elements.*

Proof. Let S be a set with more than m elements. Choose $\{\alpha_1, \alpha_2, \ldots, \alpha_n\} \subset S$ where $n > m$. Since $\{\beta_1, \beta_2, \ldots, \beta_m\}$ is a spanning set, there exist scalars $\{A_{i,j} : 1 \leq i \leq m, 1 \leq j \leq n\}$ such that

$$\alpha_j = \sum_{i=1}^{m} A_{i,j}\beta_i$$

Let $A = (A_{i,j})$ be the corresponding matrix, then A is an $m \times n$ matrix, where $m < n$. By Lemma I.2.10, there is a vector $X = (x_1, x_2, \ldots, x_n)$ such that $X \neq 0$ and

$$AX = 0$$

Now consider

$$\begin{aligned}
x_1\alpha_1 + x_2\alpha_2 + \ldots + x_n\alpha_n &= \sum_{j=1}^{n} x_j\alpha_j \\
&= \sum_{j=1}^{n} x_j \left(\sum_{i=1}^{m} A_{i,j}\beta_i\right) \\
&= \sum_{i=1}^{m}\sum_{j=1}^{n} x_j A_{i,j}\beta_i \\
&= \sum_{i=1}^{m} \left(\sum_{j=1}^{n} A_{i,j}x_j\right) \beta_i \\
&= \sum_{i=1}^{m} (AX)_i\beta_i \\
&= 0
\end{aligned}$$

Hence, the set $\{\alpha_1, \alpha_2, \ldots, a_n\}$ is *not* linearly independent, and so S cannot be linearly independent. This proves our theorem. □

Corollary 3.8. *If V is a finite dimensional vector space, then any two bases of V have the same (finite) cardinality.*

Proof. By hypothesis, V has a basis S consisting of finitely many elements, say $m := |S|$. Let T be any any other basis of V. By Theorem 3.7, since S is a spanning set, and T is linearly independent, it follows that T is finite, and

$$|T| \leq m$$

But by applying Theorem 3.7 again (in reverse), we see that

$$|S| \leq |T|$$

Hence, $|S| = |T|$. Thus, any other basis is finite and has cardinality m. □

This corollary now allows us to make the following definition, which is independent of the choice of basis.

Definition 3.9. Let V be a finite dimensional vector space. Then, the *dimension* of V is the cardinality if any basis of V. We denote this number by

$$\dim(V)$$

Note that $V = \{0\}$, then V does not contain a linearly independent set, so we simply set

$$\dim(\{0\}) := 0$$

The next corollary is essentially a restatement of Theorem 3.7.

Corollary 3.10. *Let V be a finite dimensional vector space and $n := \dim(V)$. Then*

(i) Any subset of V which contains more than n vectors is linearly dependent.

(ii) Any subset of V which is a spanning set must contain at least n elements.

Example 3.11. (i) Let F be a field and $V := F^n$, then the standard basis $\{\epsilon_1, \epsilon_2, \ldots, \epsilon_n\}$ has cardinality n. Therefore,

$$\dim(F^n) = n$$

(ii) Let F be a field and $V := F^{m\times n}$ be the space of $m \times n$ matrices over F. For $1 \leq i \leq m, 1 \leq j \leq n$, let $B^{i,j}$ denote the matrix whose entries are all zero, except the $(i,j)^{th}$ entry, which is 1. Then (Check!) that

$$S := \{B^{i,j} : 1 \leq i \leq m, 1 \leq j \leq n\}$$

is a basis for V. Hence,

$$\dim(F^{m\times n}) = mn$$

(iii) Let A be an $m \times n$ matrix, and consider the subspace

$$W := \{X \in F^n : AX = 0\}$$

Let R be a row-reduced echelon marix that is row equivalent to A. Let r denote the number of non-zero rows in R, then (as in Lemma I.2.10), the subspace

$$\{X \in F^n : RX = 0\}$$

has dimension $(n - r)$ (Check!). Hence, $\dim(W) = (n - r)$.

Lemma 3.12. *Let S be a linearly independent subset of a vector space V. Let $\beta \in V$ be a vector which is not in the subspace spanned by S. Then $S \cup \{\beta\}$ is a linearly independent set.*

Proof. Let $\{\alpha_1, \alpha_2, \dots, \alpha_m\} \subset S$ and $c_1, c_2, \dots, c_m, c_{m+1} \in F$ are scalars such that

$$c_1\alpha_1 + c_2\alpha_2 + \ldots + c_m\alpha_m + c_{m+1}\beta = 0$$

Suppose $c_{m+1} \neq 0$, then we may rewrite the above equation as

$$\beta = \frac{-c_1}{c_{m+1}}\alpha_1 + \frac{-c_2}{c_{m+1}}\alpha_2 + \ldots + \frac{-c_m}{c_{m+1}}\alpha_m$$

Hence, β is in the subspace spanned by S - a contradiction. Hence, it must happen that

$$c_{m+1} = 0$$

Then, the above equation reduces to

$$c_1\alpha_1 + c_2\alpha_2 + \ldots + c_m\alpha_m = 0$$

But S is linearly independent, so $c_i = 0$ for all $1 \leq i \leq m$. So we conclude that $S \cup \{\beta\}$ is linearly independent. □

Theorem 3.13. *Let W be a subspace of a finite dimensional vector space V, then every linearly independent subset of W is finite, and is contained in a (finite) basis of V.*

Proof. Let $S_0 \subset W$ be a linearly independent set. If S is a linearly independent subset of W containing S_0, then S is also linearly independent in V. Since V is finite dimensional,

$$|S| \leq n := \dim(V)$$

Now, we extend S_0 to form a basis of W: If S_0 spans W, there is nothing to do, since S_0 is a linearly independent set. If S_0 does not span W, then there exists a $\beta_1 \in W$ which does not belong to the subspace spanned by S_0. By Lemma 3.12,

$$S_1 := S_0 \cup \{\beta_1\}$$

is a linearly independent set. Once again, if S_1 spans W, then we stop the process.

If not, we continue as above to take a vector $\beta_2 \in W$ so that

$$S_2 := S_1 \cup \{\beta_2\}$$

is linearly independent. Thus proceeding, we obtain (after finitely many such steps), a set

$$S_m = S_0 \cup \{\beta_1, \beta_2, \dots, \beta_m\}$$

which is linearly independent, and must span W. □

Corollary 3.14. *If W is a proper subspace of a finite dimensional vector space V, then W is finite dimensional, and*

$$\dim(W) < \dim(V)$$

Proof. Since $W \neq \{0\}$, there is a non-zero vector $\alpha \in W$. Let

$$S_0 := \{\alpha\}$$

Then S_0 is linearly independent. By Theorem 3.13, there is a finite basis S of W containing S_0. Furthermore, by the previous proof, we have that

$$|S| \leq \dim(V)$$

Hence,

$$\dim(W) \leq \dim(V)$$

Since $W \neq V$, there is a vector $\beta \in V$ which is not in W. Hence, $T = S \cup \{\beta\}$ is a linearly independent set. So by Corollary 3.10, we have

$$|S \cup \{\beta\}| \leq \dim(V)$$

Hence,

$$\dim(W) = |S| < \dim(V)$$

□

Corollary 3.15. *Let V be a finite dimensional vector space and $S \subset V$ be a linearly independent set. Then, there exists a basis B of V such that $S \subset B$.*

Proof. Let W be the subspace spanned by S. Now apply Theorem 3.13. □

Corollary 3.16. *Let A be an $n \times n$ matrix over a field F such that the row vectors of A form a linearly independent set of vectors in F^n. Then, A is invertible.*

Proof. Let $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ be the row vectors of A. By Corollary 3.14, this set is a basis for F^n (Why?). Let ϵ_i denote the i^{th} standard basis vector, then there exist scalars $\{B_{i,j} : 1 \leq j \leq n\}$ such that

$$\epsilon_i = \sum_{j=1}^{n} B_{i,j}\alpha_j$$

This is true for each $1 \leq i \leq n$, so we get a matrix $B = (B_{i,j})$ such that

$$BA = I$$

By Corollary I.4.9, A is invertible. □

Theorem 3.17. *Let W_1 and W_2 be two subspaces of a vector space V, then*

$$\dim(W_1 + W_2) = \dim(W_1) + \dim(W_2) - \dim(W_1 \cap W_2)$$

Proof. Let $\{\alpha_1, \alpha_2, \ldots, \alpha_k\}$ be a basis for the subspace $W_1 \cap W_2$. By Theorem 3.13, there is a basis

$$B_1 = \{\alpha_1, \alpha_2, \ldots, \alpha_k, \beta_1, \beta_2, \ldots, \beta_n\}$$

of W_1, and a basis

$$B_2 = \{\alpha_1, \alpha_2, \ldots, \alpha_k, \gamma_1, \gamma_2, \ldots, \gamma_m\}$$

of W_2. Consider

$$B = \{\alpha_1, \alpha_2, \ldots, \alpha_k, \beta_1, \beta_2, \ldots, \beta_n, \gamma_1, \gamma_2, \ldots, \gamma_m\}$$

We claim that B is a basis for $W_1 + W_2$.

(i) B is linearly independent: If we have scalars $c_i, d_j, e_s \in F$ such that

$$\sum_{i=1}^{k} c_i\alpha_i + \sum_{j=1}^{n} d_j\beta_j + \sum_{s=1}^{m} e_s\gamma_s = 0$$

Consider the vector

$$\delta := \sum_{s=1}^{m} e_s\gamma_s \tag{II.2}$$

Then $\delta \in W_2$ since $B_2 \subset W_2$. Furthermore,

$$\delta = -\left(\sum_{i=1}^{k} c_i\alpha_i + \sum_{j=1}^{n} d_j\beta_j\right) \tag{II.3}$$

so $\delta \in W_1$ as well. Hence, $\delta \in W_1 \cap W_2$, so there exist scalars f_ℓ such that

$$\delta = \sum_{\ell=1}^{k} f_\ell\alpha_\ell$$

By Equation II.2, we see that

$$\sum_{\ell=1}^{k} f_\ell\alpha_\ell + \sum_{s=1}^{m} (-e_s)\gamma_s = 0$$

But the set B_2 is linearly independent, so we conclude that

$$f_\ell = e_s = 0$$

for all $1 \leq \ell \leq k$ and $1 \leq s \leq m$. From this and Equation II.2, we conclude that $\delta = 0$. Hence, from Equation II.3, we have

$$\sum_{i=1}^{k} c_i\alpha_i + \sum_{j=1}^{n} d_j\beta_j = 0$$

But the set B_1 is linearly independent, so we conclude that

$$c_i = d_j = 0$$

for all $1 \leq i \leq k, 1 \leq j \leq n$. Thus, B is linearly independent as well.

(ii) B spans $W_1 + W_2$: Let $\alpha \in W_1 + W_2$, then there exist $\alpha_1 \in W_1$ and $\alpha_2 \in W_2$ such that

$$\alpha = \alpha_1 + \alpha_2$$

Since B_1 is a basis for W_1, there are scalars c_i, d_j such that

$$\alpha_1 = \sum_{i=1}^{k} c_i\alpha_i + \sum_{j=1}^{n} d_j\beta_j$$

Similarly, there are scalars $e_s, f_\ell \in F$ such that

$$\alpha_2 = \sum_{s=1}^{k} e_s\alpha_s + \sum_{\ell=1}^{m} f_\ell\gamma_\ell$$

Combining the like terms in these equations, we get

$$\alpha = \sum_{i=1}^{k}(c_i + e_i)\alpha_i + \sum_{j=1}^{n} d_j\beta_j + \sum_{\ell=1}^{m} f_\ell\gamma_\ell$$

Thus, B spans $W_1 + W_2$.

Hence, we conclude that B is a basis for $W_1 + W_2$, so that

$$\dim(W_1+W_2) = |B| = k+m+n = |B_1|+|B_2|-k = \dim(W_1)+\dim(W_2)-\dim(W_1\cap W_2)$$

□

4. Coordinates

Remark 4.1. Let V be a vector space and $\mathcal{B} = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ be a basis for V. Given a vector $\alpha \in V$, we may express it in the form

$$\alpha = \sum_{i=1}^{n} c_i\alpha_i \qquad \text{(II.4)}$$

for some scalars $c_i \in F$. Furthermore, this expression is *unique*. If $d_j \in F$ are any other scalars such that

$$\alpha = \sum_{j=1}^{n} d_j\alpha_j$$

then $c_i = d_i$ for all $1 \leq i \leq n$ (Why?). Hence, the scalars $\{c_1, c_2, \ldots, c_n\}$ are uniquely determined by α, and also uniquely determine α. Therefore, we would like to assocate to α the tuple

$$(c_1, c_2, \ldots, c_n)$$

and say that c_i is the i^{th} coordinate of α. However, this only makes sense if we fix the *order* in which the basis elements appear in $\mathcal{B}$ (remember, a set has no ordering).

Definition 4.2. Let V be a finite dimensional vector space. An *ordered basis* of V is a finite sequence of vectors $\alpha_1, \alpha_2, \ldots, \alpha_n$ which together form a basis of V.

In other words, we are imposing an order on the basis $\mathcal{B} = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ by saying that α_1 is the first vector, α_2 is the second, and so on. Now, given an ordered basis $\mathcal{B}$ as above, and a vector $\alpha \in V$, we may associate to α the tuple

$$[\alpha]_{\mathcal{B}} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}$$

provided Equation II.4 is satisfied.

Example 4.3. Let F be a field and $V = F^n$. If $\mathcal{B} = \{\epsilon_1, \epsilon_2, \ldots, \epsilon_n\}$ is the standard ordered basis, then for a vector $\alpha = (x_1, x_2, \ldots, x_n) \in V$, we have

$$[\alpha]_{\mathcal{B}} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

However, if we take $\mathcal{B}' = \{\epsilon_n, \epsilon_1, \epsilon_1, \ldots, \epsilon_{n-1}\}$ as the same basis ordered differently (by a cyclic permutation), then

$$[\alpha]_{\mathcal{B}'} = \begin{pmatrix} x_n \\ x_1 \\ x_2 \\ \vdots \\ x_{n-1} \end{pmatrix}$$

Remark 4.4. Now suppose we are given two ordered bases $\mathcal{B} = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ and $\mathcal{B}' = \{\beta_1, \beta_2, \ldots, \beta_n\}$ of V (Note that these two sets have the same cardinality). Given a vector $\alpha \in V$, we have two expressions associated to α

$$[\alpha]_{\mathcal{B}} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix} \text{ and } [\alpha]_{\mathcal{B}'} = \begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{pmatrix}$$

The question is, How are these two column vectors related to each other?

Observe that, since $\mathcal{B}$ is a basis, for each $1 \leq i \leq n$, there are scalars $P_{j,i} \in F$ such that

$$\beta_i = \sum_{j=1}^{n} P_{j,i} \alpha_j \tag{II.5}$$

Now observe that

$$\begin{aligned}\alpha &= \sum_{i=1}^n d_i\beta_i \\ &= \sum_{i=1}^n d_i\left(\sum_{j=1}^n P_{j,i}\alpha_j\right) \\ &= \sum_{j=1}^n \left(\sum_{i=1}^n d_iP_{j,i}\right)\alpha_j\end{aligned}$$

However,

$$\alpha = \sum_{i=1}^n c_j\alpha_j$$

so by the uniqueness of these scalars, we see that

$$c_j = \sum_{i=1}^n P_{j,i}d_i$$

for each $1 \leq j \leq n$. Hence, we conclude that

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

where $P = (P_{j,i})$.

Now consider the expression in Equation II.5. Reversing the roles of $\mathcal{B}$ and $\mathcal{B}'$, we obtain scalars $Q_{i,k} \in F$ such that

$$\alpha_k = \sum_{i=1}^n Q_{i,k}\beta_i$$

Combining this with Equation II.5, we see that

$$\alpha_k = \sum_{i=1}^n Q_{i,k}\left(\sum_{j=1}^n P_{j,i}\alpha_j\right) = \sum_{j=1}^n\left(\sum_{i=1}^n P_{j,i}Q_{i,k}\right)\alpha_j$$

But the $\{\alpha_j\}$ are a basis, so we conclude that

$$\sum_{i=1}^n P_{j,i}Q_{i,k} = \begin{cases} 1 & : k = j \\ 0 & : k \neq j\end{cases}$$

Thus, if $Q = (Q_{i,j})$, then we conclude that

$$PQ = I$$

Hence the matrix P chosen above is invertible and $Q = P^{-1}$. The following theorem is the conclusion of this discussion.

Theorem 4.5. *Let V be an n-dimensional vector space and $\mathcal{B}$ and $\mathcal{B}'$ be two ordered bases of V. Then, there is a unique $n \times n$ invertible matrix P such that, for any $\alpha in V$, we have*

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

and

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

Furthermore, the columns of P are given by

$$P_j = [\beta_j]_{\mathcal{B}}$$

Definition 4.6. The matrix P constructed in the above theorem is called a *change of basis* matrix.

(End of Week 2)

The next theorem is a converse to Theorem 4.5.

Theorem 4.7. *Let P be an $n \times n$ invertible matrix over F. Let V be an n-dimensional vector space over F and let $\mathcal{B}$ be an ordered basis of V. Then there is a unique ordered basis $\mathcal{B}'$ of V such that, for any vector $\alpha \in V$, we have*

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

and

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

Proof. We write $\mathcal{B} = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ and set $P = (P_{j,i})$. We define

$$\beta_i = \sum_{j=1}^{n} P_{j,i}\alpha_j \tag{II.6}$$

Then we claim that $\mathcal{B}' = \{\beta_1, \beta_2, \ldots, \beta_n\}$ is a basis for V.

(i) $\mathcal{B}'$ is linearly independent: If we have scalars $c_i \in F$ such that

$$\sum_{i=1}^{n} c_i\beta_i = 0$$

Then we get

$$\sum_{i=1}^{n}\sum_{j=1}^{n} c_i P_{j,i}\alpha_j = 0$$

Rewriting the above expression, and using the linear independence of $\mathcal{B}$, we conclude that

$$\sum_{i=1}^{n} P_{j,i}c_i = 0$$

for each $1 \leq j \leq n$. If $X = (c_1, c_2, \ldots, c_n) \in F^n$, then we conclude that

$$PX = 0$$

However, P is invertible, so $X = 0$, whence $c_i = 0$ for all $1 \leq i \leq n$.

(ii) $\mathcal{B}'$ spans V: If $\alpha \in V$, then there are scalars $d_i \in F$ such that

$$\alpha = \sum_{i=1}^n d_i\alpha_i \tag{II.7}$$

Now let $Q = (Q_{i,j}) = P^{-1}$, then we have $PQ = I$, so

$$\sum_{k=1}^n P_{j,k}Q_{k,i} = \begin{cases} 0 & : i \neq j \\ 1 & : i = j \end{cases}$$

Thus, from Equation II.6, we get

$$\begin{aligned} \sum_{k=1}^n Q_{k,i}\beta_k &= \sum_{k=1}^n \sum_{j=1}^n Q_{k,i}P_{j,k}\alpha_j \\ &= \sum_{j=1}^n \left(\sum_{k=1}^n P_{j,k}Q_{k,i} \right) \alpha_j \\ &= \alpha_i \end{aligned}$$

Hence, if $\alpha \in V$ as above, we have

$$\alpha = \sum_{i=1}^n d_i\alpha_i = \sum_{i=1}^n \sum_{k=1}^n Q_{k,i}d_i\beta_k = \sum_{k=1}^n \left(\sum_{i=1}^n d_iQ_{k,i} \right) \beta_k \tag{II.8}$$

Thus, every vector $\alpha \in V$ is in the subspace spanned by $\mathcal{B}'$, whence $\mathcal{B}'$ is a basis for V.

Finally, if $\alpha \in V$ and suppose

$$[\alpha]_{\mathcal{B}} = \begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{pmatrix} \text{ and } [\alpha]_{\mathcal{B}'} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}$$

so that Equation II.7 holds, then by Equation II.8, we see that

$$c_k = \sum_{i=1}^n Q_{k,i}d_i$$

Hence,

$$[\alpha]_{\mathcal{B}'} = Q[\alpha]_{\mathcal{B}} = P^{-1}[\alpha]_{\mathcal{B}}$$

By symmetry, it follows that

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

This completes the proof. □

Example 4.8. Let $F = \mathbb{R}$ and $V = \mathbb{R}^2$ and $\mathcal{B} = \{\epsilon_1, \epsilon_2\}$ the standard ordered basis. For a fixed $\theta \in \mathbb{R}$, let

$$P := \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$$

Then P is an invertible matrix and

$$P^{-1} = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix}$$

Then $\mathcal{B}' = \{(\cos(\theta), \sin(\theta)), (-\sin(\theta), \cos(\theta))\}$ is a basis for V. It is, geometrically, the usual pair of axes rotated by an angle θ. In this basis, for $\alpha = (x_1, x_2) \in V$, we have

$$[\alpha]_{\mathcal{B}'} = \begin{pmatrix} x_1\cos(\theta) + x_2\sin(\theta) \\ -x_1\sin(\theta) + x_2\cos(\theta) \end{pmatrix}$$

5. Summary of Row Equivalence

Definition 5.1. Let A be an $m \times n$ matrix over a field F, and write its rows as vectors $\{\alpha_1, \alpha_2, \ldots, \alpha_m\} \subset F^n$.

(i) The *row space* of A is the subspace of F^n spanned by this set.

(ii) The *row rank* of A is the dimension of the row space of A.

Theorem 5.2. *Row equivalent matrices have the same row space.*

Proof. If A and B are two row-equivalent $m \times n$ matrices, then there is an invertible matrix P such that

$$B = PA$$

If $\{\alpha_1, \alpha_2, \ldots, \alpha_m\}$ are the row vectors A and $\{\beta_1, \beta_2, \ldots, \beta_m\}$ are the row vectors of B, then

$$\beta_i = \sum_{j=1}^{m} P_{i,j}\alpha_j$$

If W_A and W_B are the row spaces of A and B respectively, then we see that

$$\{\beta_1, \beta_2, \ldots, \beta_m\} \subset W_A$$

Since W_B is the smallest subspace containing this set, we conclude that

$$W_B \subset W_A$$

Since row equivalence is an equivalence relation, we have $W_A \subset W_B$ as well. □

Theorem 5.3. *Let R be a row-reduced echelon matrix, then the non-zero rows of R form a basis for the row space of R.*

Proof. Let $\rho_1, \rho_2, \ldots, \rho_r$ be the non-zero rows of R and write

$$\rho_i = (R_{i,1}, R_{i,2}, \ldots, R_{i,n})$$

By definition, the set $\{\rho_1, \rho_2, \ldots, \rho_r\}$ spans the row space W_R of R. Therefore, it suffices to check that this set is linearly independent. Since R is a row-reduced echelon matrix, there are positive integers $k_1 < k_2 < \ldots < k_r$ such that, for all $i \leq r$

(i) $R(i,j) = 0$ if $j < k_i$

(ii) $R(i,k_j) = \delta_{i,j}$

Hence, if there are scalars $c_i \in F$ such that

$$\sum_{i=1}^{r} c_i \rho_i = 0$$

Then consider the k_j^{th} entry of the vector in the LHS, and we have

$$\begin{aligned} 0 &= \left[\sum_{i=1}^{r} c_i \rho_i\right]_{k_j} \\ &= \sum_{i=1}^{r} c_i [\rho_i]_{k_j} \\ &= \sum_{i=1}^{r} c_i R(i,k_j) \\ &= \sum_{i=1}^{r} c_i \delta_{i,j} = c_j \end{aligned}$$

Hence each $c_j = 0$, whence $\{\rho_1, \rho_2, \ldots, \rho_r\}$ is a linearly independent set. □

Theorem 5.4. *Let F be a field and $m, n \in \mathbb{N}$ be positive integers. Given a subspace $W < F^n$ with $\dim(W) \leq m$, there is a unique $m \times n$ row reduced echelon matrix R whose row space is W.*

Proof.

(i) Existence: Since $\dim(W) \leq m$, there is a spanning set of W consisting of m vectors. Let A be the $m \times n$ matrix whose rows are these vectors. Then the row space of A is W. Let R be a row-reduced echelon matrix that is row-equivalent to A. Then by Theorem 5.2, the row space of R is also W.

(ii) Uniqueness: Let R and S be two row-reduced echelon matrices with the same row space W. Let $\rho_1, \rho_2, \ldots, \rho_r$ be the non-zero row vectors of R. Write $\rho_i = (R_{i,1}, R_{i,2}, \ldots, R_{i,n})$. Since R is row-reduced, there are integers $k_1, k_2, \ldots, k_r$ such that, for $i \leq r$,

(i) $R(i,j) = 0$ if $j < k_i$

(ii) $R(i,k_j) = \delta_{i,j}$

(iii) $k_1 < k_2 < \ldots < k_r$

By Theorem 5.3, the set $\{\rho_1, \rho_2, \ldots, \rho_r\}$ forms a basis for W. Hence, S has exactly r non-zero rows, which we enumerate as $\eta_1, \eta_2, \ldots, \eta_r$. Furthermore, there are integers $\ell_1, \ell_2, \ldots, \ell_r$ such that, for $i \leq r$

(i) $S(i,j) = 0$ if $j < \ell_i$

(ii) $S(i,\ell_j) = \delta_{i,j}$

(iii) $\ell_1 < \ell_2 < \ldots < \ell_r$

Write $\eta_1 = (b_1, b_2, \ldots, b_n)$. Then, there exist scalars $c_1, c_2, \ldots, c_r \in F$ such that

$$\eta_1 = \sum_{i=1}^r c_i \rho_i$$

Then observe that

$$\begin{aligned} b_{k_j} &= \sum_{i=1}^r c_i R(i,k_j) \\ &= \sum_{i=1}^r c_i \delta_{i,j} \\ &= c_j \end{aligned}$$

Hence,

$$\eta_1 = \sum_{i=1}^r b_{k_i} \rho_i \tag{II.9}$$

It now follows from the conditions on $\{R(i,j)\}$ listed above that the first non-zero entry of η_1 occurs $b_{k_{s_1}}$ for some $1 \leq s_1 \leq r$. It follows that

$$\ell_1 = k_{s_1}$$

Thus proceeding, for each $1 \leq i \leq r$, there is some $1 \leq s_i \leq r$ such that

$$\ell_i = k_{s_i}$$

Since both sets of integers are strictly increasing, it follows that

$$\ell_i = k_i \quad \forall 1 \leq i \leq r$$

Now consider the expression in Equation II.9, and observe that

$$b_{k_i} = S(1, k_i) = 0 \text{ if } i \geq 2$$

Hence, $\eta_1 = \rho_1$. Thus proceeding, we may conclude that $\eta_i = \rho_i$ for all i, whence $S = R$.

□

Corollary 5.5. *Every $m \times n$ matrix A is row-equivalent to one and only one row-reduced echelon matrix.*

Proof. We know that A is row-equivalent to one row-reduced echelon matrix from Theorem I.2.9. If A is row-equivalent to two row-reduced echelon matrices R and S, then by Theorem 5.3, both R and S have the same row space. By Theorem 5.4, $R = S$. □

Corollary 5.6. *Let A and B be two $m \times n$ matrices over a field F. Then A is row-equivalent to B if and only if they have the same row space.*

Proof. We know from Theorem 5.2 that if A and B are row-equivalent, then they have the same row space.

Conversely, suppose A and B have the same row space. By Theorem I.2.9, A and B are both row-equivalent to row-reduced echelon matrices R and S respectively. By Theorem 5.2, R and S have the same row space. By Theorem 5.4, $R = S$. Hence, A and B are row-equivalent to each other. □

III. Linear Transformations

1. Linear Transformations

Definition 1.1. Let V and W be two vector spaces over a common field F. A function $T : V \to W$ is called a *linear transformation* if, for any two vectors $\alpha, \beta \in V$ and any scalar $c \in F$, we have

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

Example 1.2.

(i) Let V be any vector space and $I : V \to V$ be the identity map. Then I is linear.

(ii) Similarly, the zero map $0 : V \to V$ is a linear map.

(iii) Let $V = F^n$ and $W = F^m$, and $A \in F^{m\times n}$ be an $m \times n$ matrix with entries in F. Then $T : V \to W$ given by

$$T(X) := AX$$

is a linear transformation by Lemma II.2.5.

(iv) Let V be the space of all polynomials over F. Define $D : V \to V$ be the 'derivative' map, defined by the rule: If

$$f(x) = c_0 + c_1x + c_2x^2 + \ldots + c_nx^n$$

Then

$$(Df)(x) = c_1 + 2c_2x + \ldots + nc_nx^{n-1}$$

(v) Let $F = \mathbb{R}$ and V be the space of all functions $f : \mathbb{R} \to \mathbb{R}$ that are continuous (Note that V is, indeed, a vector space with the point-wise operations as in Example II.1.2). Define $T : V \to V$ by

$$T(f)(x) := \int_0^x f(t)dt$$

(vi) With V as in the previous example and $W = \mathbb{R}$, we may also define $T : V \to W$ by

$$T(f) := \int_0^1 f(t)dt$$

Remark 1.3. If $T : V \to W$ is a linear transformation

(i) $T(0) = 0$ because if $\alpha := T(0)$, then

$$2\alpha = \alpha + \alpha = T(0) + T(0) = T(0 + 0) = T(0) = \alpha$$

Hence, $\alpha = 0$ by Lemma II.1.3.

(ii) If β is a linear combination of vectors $\{\alpha_1, \alpha_2, \ldots, \alpha_m\}$, then we may write

$$\beta = \sum_{i=1}^n c_i\alpha_i$$

for some scalars $c_1, c_2, \ldots, c_n \in F$. Then it follows that

$$T(\beta) = \sum_{i=1}^n c_iT(\alpha_i)$$

Theorem 1.4. *Let V be a finite dimensional vector space over a field F and let $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ be an ordered basis of V. Let W be another vector space over F and $\{\beta_1, \beta_2, \ldots, \beta_n\}$ be any set of n vectors in W. Then, there is a unique linear transformation $T : V \to W$ such that*

$$T(\alpha_i) = \beta_i \quad \forall 1 \leq i \leq n$$

Proof.

(i) Existence: Given a vector $\alpha \in V$, there is a unique expression of the form

$$\alpha = \sum_{i=1}^n c_i\alpha_i$$

We define $T : V \to W$ by

$$T(\alpha) := \sum_{i=1}^n c_i\beta_i$$

Since the above expression is uniquely associated to α, this map is well-defined. Now we check linearity: If

$$\beta = \sum_{i=1}^n d_i\alpha_i$$

and $c \in F$ a scalar, then we have

$$c\alpha + \beta = \sum_{i=1}^n (cc_i + d_i)\alpha_i$$

So by definition

$$T(c\alpha + \beta) = \sum_{i=1}^n (cc_i + d_i)\beta_i$$

Now consider

$$\begin{aligned} cT(\alpha) + T(\beta) &= c\left(\sum_{i=1}^{n} c_i\beta_i\right) + \sum_{i=1}^{n} d_i\beta_i \\ &= \sum_{i=1}^{n}(cc_i + d_i)\beta_i \\ &= T(c\alpha + \beta) \end{aligned}$$

Hence, T is linear as required.

(ii) Uniqueness: If $S : V \to W$ is another linear transformation such that

$$S(\alpha_i) = \beta_i \quad \forall 1 \leq i \leq n$$

Then for any $\alpha \in V$, we write

$$\alpha = \sum_{i=1}^{n} c_i\alpha_i$$

So that, by linearity,

$$S(\alpha) = \sum_{i=1}^{n} c_i\beta_i = T(\alpha)$$

Hence, $T(\alpha) = S(\alpha)$ for all $\alpha \in V$, so $T = S$.

□

Example 1.5.

(i) Let $\alpha_1 = (1, 2), \alpha_2 = (3, 4)$. Then the set $\{\alpha_1, \alpha_2\}$ is a basis for $\mathbb{R}^2$ (Check!). Hence, there is a unique linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^3$ such that

$$T(\alpha_1) = (3, 2, 1) \text{ and } T(\alpha_2) = (6, 5, 4)$$

We find $T(\epsilon_2)$: To do that, we write

$$\epsilon_2 = c_1\alpha_1 + c_2\alpha_2 = (c_1 + 3c_2, 2c_1 + 4c_2) = (1, 0)$$

Hence,

$$c_1 = -2, c_1 = 1$$

So that

$$T(\epsilon_2) = -2T(\alpha_1) + T(\alpha_2) = -2(3, 2, 1) + (6, 5, 4) = (0, 1, 2)$$

(ii) If $T : F^m \to F^n$ is a linear transformation, then T is uniquely determined by the vectors

$$\beta_i = T(\epsilon_i), 1 \leq i \leq m$$

If $\alpha = (x_1, x_2, \ldots, x_m) \in F^m$, then it follows that

$$T(\alpha) = \sum_{i=1}^{m} x_i\beta_i$$

So if we write B for the matrix whose row vectors are $\beta_1, \beta_2, \ldots, \beta_m$, then

$$T(\alpha) = \alpha B$$

Hence, a linear tranformation $T : F^m \to F^n$ is given by multiplication by an $m \times n$ matrix.

Definition 1.6. Let $T : V \to W$ be a linear transformation.

(i) The *range of* T is the set
$$R_T := \{T(\alpha) : \alpha \in V\}$$

(ii) The *kernel of* T (or the *nullspace of* T) is the set
$$\ker(T) = \{\alpha \in V : T(\alpha) = 0\}$$

Lemma 1.7. *If $T : V \to W$ is a linear transformation, then*

(i) *R_T is a subspace of W*

(ii) $\ker(T)$ *is a subspace of V.*

Proof. Exercise. (Verify Theorem II.2.3) □

Definition 1.8. Let V be a finite dimensional vector space and $T : V \to W$ a linear transformation.

(i) The *rank of* T is $\dim(R_T)$, and is denoted by $\mathrm{rank}(T)$

(ii) The *nullity of* T is $\dim(\ker(T))$, and is denoted by $\mathrm{nullity}(T)$.

The next result is an important theorem, and is called the *Rank-Nullity Theorem*

Theorem 1.9. *Let V be a finite dimensional vector space and $T : V \to W$ a linear transformation. Then*
$$rank(T) + nullity(T) = \dim(V)$$

Proof. Let $\{\alpha_1, \alpha_2, \ldots, \alpha_k\}$ be a basis of $\ker(T)$. Then, by Corollary II.3.15, we can extend it to form a basis

$$\mathcal{B} := \{\alpha_1, \alpha_2, \ldots, \alpha_k, \alpha_{k+1}, \alpha_{k+2}, \ldots, \alpha_n\}$$

of V. Consider the set

$$S := \{T(\alpha_{k+1}), T(\alpha_{k+2}), \ldots, T(\alpha_n)\} \subset R_T$$

We claim that this set is a basis.

(i) S is linearly independent: If $c_{k+1}, c_{k+2}, \ldots, c_n \in F$ are scalars such that

$$\sum_{i=k+1}^{n} c_i T(\alpha_i) = 0$$

By linearity

$$T\left(\sum_{i=k+1}^{n} c_i \alpha_i\right) = 0 \Rightarrow \sum_{i=k+1}^{n} c_i \alpha_i \in \ker(T)$$

Hence, there exist scalars $d_1, d_2, \ldots, d_k \in F$ such that

$$\sum_{i=k+1}^{n} c_i \alpha_i = \sum_{j=1}^{k} d_j \alpha_j$$

Since the set $\mathcal{B}$ is linearly independent, we conclude that

$$c_i = 0 = d_j$$

for all $1 \leq j \leq k, k+1 \leq i \leq n$. Hence, we conclude that S is linearly independent.

(ii) S spans R_T: If $\beta \in R(T)$, then there exists $\alpha \in V$ such that $\beta = T(\alpha)$. Since $\mathcal{B}$ is a basis for V, there exist scalars $c_1, c_2, \ldots, c_n \in F$ such that

$$\alpha = \sum_{i=1}^{n} c_i \alpha_i$$

Hence,

$$\beta = T(\alpha) = \sum_{i=1}^{n} c_i T(\alpha_i)$$

But $T(\alpha_i) = 0$ for all $1 \leq i \leq k$. Hence,

$$\beta = \sum_{i=k+1}^{n} c_i T(\alpha_i)$$

This proves the theorem.

□

2. The Algebra of Linear Transformations

Lemma 2.1. *Let V and W be two vector spaces over a common field F. Let $U, T : V \to W$ be two linear transformations, and $c \in F$ a scalar.*

(i) Define $(T + U) : V \to W$ by

$$(T + U)(\alpha) = T(\alpha) + U(\alpha)$$

(ii) Define $(cT) : V \to W$ by

$$(cT)(\alpha) := cT(\alpha)$$

Then $(T+U)$ and cT are both linear transformations.

Proof. We prove that $(T+U)$ is a linear transformation. The proof for (cT) is similar. Fix $\alpha, \beta \in V$ and $d \in F$ a scalar, and consider

$$\begin{aligned}(T+U)(d\alpha+\beta) &= T(d\alpha+\beta) + U(d\alpha+\beta) \\ &= dT(\alpha) + T(\beta) + dU(\alpha) + U(\beta) \\ &= d\left(T(\alpha)+U(\alpha)\right) + \left(T(\beta)+U(\beta)\right) \\ &= d(T+U)(\alpha) + (T+U)(\beta)\end{aligned}$$

Hence, $(T+U)$ is linear. □

Definition 2.2. Let V and W be two vector spaces over a common field F. Let $L(V,W)$ be the space of all linear transformations from V to W.

Theorem 2.3. *Under the operations defined in Lemma 2.1, $L(V,W)$ is a vector space.*

Proof. By Lemma 2.1, the operations

$$+ : L(V,W) \times L(V,W) \to L(V,W)$$

and

$$\cdot : F \times L(V,W) \to L(V,W)$$

are well-defined operations. We now need to verify all the axioms of Definition II.1.1. For convenience, we simply verify a few of them, and leave the rest for you.

(i) Addition is commutative: If $T, U \in L(V,W)$, we need to check that $(T+U) = (U+T)$. Hence, we need to check that, for any $\alpha \in V$,

$$(U+T)(\alpha) = (T+U)(\alpha)$$

But this follows from the fact that addition in W is commutative, and so

$$(T+U)(\alpha) = T(\alpha) + U(\alpha) = U(\alpha) + T(\alpha) = (U+T)(\alpha)$$

(ii) Observe that the zero linear transformation $0 : V \to W$ is the zero element in $L(V,W)$.

(iii) Let $d \in F$ and $T, U \in L(V,W)$, then we verify that

$$d(T+U) = dT + dU$$

So fix $\alpha \in V$, then

$$\begin{aligned}\left[d(T+U)\right](\alpha) &= d(T+U)(\alpha) \\ &= d\left(T(\alpha)+U(\alpha)\right) \\ &= dT(\alpha) + dU(\alpha) \\ &= (dT+dU)(\alpha)\end{aligned}$$

This is true for every $\alpha in V$, so $d(T+U) = dT + dU$.

The other axioms are verified in a similar fashion. □

Theorem 2.4. *Let V and W be two finite dimensional vector spaces over F. Then $L(V,W)$ is finite dimensional, and*

$$\dim(L(V,W)) = \dim(V)\dim(W)$$

Proof. Let

$$\mathcal{B} := \{\alpha_1, \alpha_2, \dots, \alpha_n\} \text{ and } \mathcal{B}' := \{\beta_1, \beta_2, \dots, \beta_m\}$$

be bases of V and W respectively. Then, we wish to show that

$$\dim(L(V,W)) = mn$$

For each $1 \leq p \leq m, 1 \leq q \leq n$, by Theorem 1.4, there is a unique $E^{p,q} \in L(V,W)$ such that

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

We claim that

$$S := \{E^{p,q} : 1 \leq p \leq m, 1 \leq q \leq n\}$$

forms a basis for $L(V,W)$.

(i) S is linearly independent: Suppose $c_{p,q} \in F$ are scalars such that

$$\sum_{p=1}^{m}\sum_{q=1}^{n} c_{p,q}E^{p,q} = 0$$

Then evaluating this expression on α_i gives

$$\sum_{p=1}^{m} c_{p,i}\beta_p = 0$$

But $\mathcal{B}'$ is a linearly independent set in W, so

$$c_{p,i} = 0 \quad \forall 1 \leq p \leq m$$

This is true for each $1 \leq i \leq n$, proving that S is linearly independent.

(ii) S spans $L(V,W)$: Let $T \in L(V,W)$, then for each $1 \leq i \leq n$,

$$T(\alpha_i) \in W$$

so it can be expessed as a linear combination of elements of $\mathcal{B}'$ in a unique way. So we write

$$T(\alpha_i) = \sum_{p=1}^{m} a_{p,i}\beta_p$$

We define $S \in L(V, W)$ by

$$S = \sum_{p=1}^{m} \sum_{q=1}^{n} a_{p,q} E^{p,q}$$

and we claim that $S = T$. By Theorem 1.4, it suffices to verify that

$$S(\alpha_i) = T(\alpha_i) \quad \forall 1 \leq i \leq n$$

so consider

$$\begin{aligned} S(\alpha_i) &= \sum_{p=1}^{m} \sum_{q=1}^{n} a_{p,q} E^{p,q}(\alpha_i) \\ &= \sum_{q=1}^{n} a_{p,i} \beta_p \\ &= T(\alpha_i) \end{aligned}$$

This proves that $S = T$ as required. Hence, S spans $L(V, W)$.

□

Theorem 2.5. *Let V, W and Z be three vector spaces over a common field F. Let $T \in L(V, W)$ and $U \in L(W, Z)$. Then define $UT : V \to Z$ by*

$$(UT)(\alpha) := U(T(\alpha))$$

Then $(UT) \in L(V, Z)$

Proof. Fix $\alpha, \beta \in V$ and $c \in F$, and note that

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

Hence, (UT) is linear. □

Definition 2.6. A *linear operator* is a linear transformation from a vector space V to itself.

Note that $L(V, V)$ now has a 'multiplication' operation, given by composition of linear operators. We let $I \in L(V, V)$ denote the identity linear operator. For $T \in L(V, V)$, we may now write

$$T^2 = TT$$

and similarly, T^n makes sense for all $n \in \mathbb{N}$. We simply define $T^0 = I$ for convenience.

Lemma 2.7. *Let $U, T_1, T_2 \in L(V, V)$ and $c \in F$. Then*

(i) $IU = UI = U$

(ii) $U(T_1 + T_2) = UT_1 + UT_2$

(iii) $(T_1 + T_2)U = T_1U + T_2U$

(iv) $c(UT_1) = (cU)T_1 = U(cT_1)$

Proof.

(i) This is obvious

(ii) Fix $\alpha \in V$ and consider

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

This is true for every $\alpha \in V$, so

$$U(T_1 + T_2) = UT_1 + UT_2$$

(iii) This is similar to part (ii) [See [Hoffman-Kunze, Page 77]]

(iv) Fix $\alpha \in V$ and consider

$$\begin{aligned}[c(UT_1)](\alpha) &= c[(UT_1)(\alpha)] \\ &= c[U(T_1(\alpha))] \\ &= (cU)(T_1(\alpha)) \\ &= [(cU)T_1](\alpha)\end{aligned}$$

This is true for every $\alpha \in V$, so

$$c(UT_1) = (cU)T_1$$

The other equality is proved similarly.

□

Example 2.8.

(i) Let $A \in F^{m\times n}$ and $B \in F^{p\times m}$ be two matrices. Let $V = F^n, W = F^m$, and $Z = F^p$, and define $T \in L(V, W)$ and $U \in L(W, Z)$ by

$$T(X) = AX \text{ and } U(Y) = BY$$

by matrix multiplication. Then, by Lemma II.2.5,

$$(UT)(X) = U(T(X)) = U(AX) = B(AX) = (BA)(X)$$

Hence, (UT) is given by multiplication by (BA).

(ii) Let V be the vector space of polynomials over F. Define $D : V \to V$ by the 'derivative' operator (See Example 1.2). If $f \in V$ is given by

$$f(x) = c_0 + c_1x + c_2x^2 + \ldots + c_nx^n$$

Then $D(f) \in V$ is the function

$$D(f)(x) = c_1 + 2c_2x + 3c_3x^2 \ldots + nc_nx^{n-1}$$

Let $T : V \to V$ be the linear transformation

$$T(f)(x) = xf(x)$$

Then, if $f_n(x) := x^n$ and $n \geq 1$, then

$$\begin{aligned}(DT - TD)(f_n)(x) = DT(f_n)(x) - TD(f_n)(x) \\ &= D(xf_n(x)) - T(nx^{n-1}) \\ &= D(x^{n+1}) - nx^n \\ &= (n+1)x^n - nx^n \\ &= x^n \\ &= f_n(x)\end{aligned}$$

By Theorem 1.4,

$$DT - TD = I$$

In particular, $DT \neq TD$. Hence, composition of operators is not necessarily a commutative operation.

(iii) Let $\mathcal{B} = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ be an ordered basis of a vector space V. For $1 \leq p, q \leq n$, let $E^{p,q} \in L(V, V)$ be the unique operator such that

$$E^{p,q}(\alpha_i) = \delta_{i,q}\alpha_p$$

The n^2 operators $\{E^{p,q} : 1 \leq p, q \leq n\}$ forms a basis for $L(V, V)$ by Theorem 2.4. Now consider

$$E^{p,q}E^{r,s}$$

For a fixed $1 \leq i \leq n$, we have

$$\begin{aligned}E^{p,q}E^{r,s}(\alpha_i) &= E^{p,q}(\delta_{i,s}\alpha_r) \\ &= \delta_{i,s}E^{p,q}(\alpha_r) \\ &= \delta_{i,s}\delta_{r,q}\alpha_p\end{aligned}$$

Hence,

$$E^{p,q}E^{r,s} = \delta_{r,q}E^{p,s}$$

Now suppose $T, U \in L(V, V)$ are two operators. Then, by Theorem 2.4, there are scalars $(a_{i,j})$ and $(b_{i,j})$ such that

$$T = \sum_{p=1}^n \sum_{q=1}^n a_{p,q} E^{p,q} \text{ and } U = \sum_{r=1}^n \sum_{s=1}^n b_{r,s} E^{r,s}$$

Now consider $UT \in L(V, V)$, and using the above relation, we calculate

$$\begin{aligned} TU &= \sum_{p=1}^n \sum_{q=1}^n \sum_{r=1}^n \sum_{s=1}^n a_{p,q} b_{r,s} E^{p,q} E^{r,s} \\ &= \sum_{p=1}^n \sum_{q=1}^n \sum_{r=1}^n \sum_{s=1}^n a_{p,q} b_{r,s} \delta_{r,q} E^{p,s} \\ &= \sum_{p=1}^n \sum_{q=1}^n \sum_{s=1}^n a_{p,q} b_{q,s} E^{p,s} \end{aligned}$$

Hence, if we associate

$$T \mapsto A := (a_{i,j}) \text{ and } U \mapsto B := (b_{i,j})$$

Then

$$UT \mapsto AB$$

(End of Week 3)

Definition 2.9. A linear transformation $T : V \to W$ is said to be *invertible* if there is a linear transformtion $S : W \to V$ such that

$$ST = I_V \text{ and } TS = I_W$$

Definition 2.10. A function $f : S \to T$ between two sets is said to be

(i) *injective* if f is one-to-one. In other words, if $x, y \in S$ and $f(x) = f(y)$, then $x = y$

(ii) *surjective* if f is onto. In other words, for any $z \in T$, there exists $x \in S$ such that $f(x) = z$.

(iii) *bijective* if f is both injective and surjective.

Theorem 2.11. *Let $T : V \to W$ be a linear transformation. Then T is invertible if and only if T is bijective.*

Proof. (i) If T is invertible, then there is a linear transformation $S : V \to W$ as above.

(i) S is injective: If $\alpha, \beta \in V$ are such that $T(\alpha) = T(\beta)$, Then

$$ST(\alpha) = ST(\beta)$$

But $ST = I_V$, so $\alpha = \beta$.

(ii) S is surjective: If $\beta \in W$, then $S(\beta) \in V$, and

$$T(S(\beta)) = (TS)(\beta) = I_W(\beta) = \beta$$

(ii) Conversely, suppose T is bijective. Then, by usual set theory, there is a function $S : W \to V$ such that

$$ST = I_V \text{ and } TS = I_W$$

We claim S is also a linear map. To this end, fix $c \in F$ and $\alpha, \beta \in W$. Then we wish to show that

$$S(c\alpha + \beta) = cS(\alpha) + S(\beta)$$

Since T is injective, it suffices to show that

$$T\left(S(c\alpha + \beta)\right) = T\left(cS(\alpha) + S(\beta)\right)$$

Bu this follows from the 'linearity of composition' (Lemma 2.7). Hence, S is linear, and thus, T is invertible.

□

Definition 2.12. A linear transformation $T : V \to W$ is said to be *non-singular* if, for any $\alpha \in V$

$$T(\alpha) = 0 \Rightarrow \alpha = 0$$

Equivalently, T is non-singular if $\ker(T) = \{0_V\}$

Theorem 2.13. *Let $T : V \to W$ be a non-singular matrix. If S is a linearly independent subset of V, then $T(S) = \{T(\alpha) : \alpha \in S\}$ is a linearly independent subset of W.*

Proof. Suppose $\{b_1, \beta_2, \ldots, \beta_n\} \subset T(S)$ are vectors and $c_1, c_2, \ldots, c_n \in F$ are scalars such that

$$\sum_{i=1}^n c_i\beta_i = 0$$

Then for each $1 \leq i \leq n$, there exists $\alpha_i \in S$ such that $\beta_i = T(\alpha_i)$, so that

$$\sum_{i=1}^n c_iT(\alpha_i) = 0$$

Using linearity, we see that

$$T\left(\sum_{i=1}^n c_i\alpha_i\right) = 0$$

Since T is non-singular, it follows that

$$\sum_{i=1}^n c_i\alpha_i = 0$$

Since S is linearly independent, it follows that $c_i = 0$ for all $1 \leq i \leq n$. Hence, $T(S)$ is linearly independent. □

Example 2.14. (i) Let $T : \mathbb{R}^2 \to \mathbb{R}^3$ be the linear map

$$T(x, y) := (x, y, 0)$$

Then T is clearly non-singular, and is not surjective.

(ii) Let V be the space of polynomials over a field F. Define $D : V \to V$ to be the 'derivative' operator from earlier. Define $E : V \to V$ be the 'integral' operator, described as follows: If $f \in V$ is given by

$$f(x) = c_0 + c_1x + c_2x^2 + \ldots + c_nx^n$$

Then define

$$(Ef)(x) = c_0x + c_1\frac{x^2}{2} + c_2\frac{x^3}{3} + \ldots + c_n\frac{x^{n+1}}{n+1}$$

Then it is clear that

$$DE = I_V$$

However, $ED \neq I_V$ because ED is zero on constant functions. Furthermore, E is not surjective because constant functions are not in the range of E.

Hence, it is possible for an operator to be non-singular, but not invertible. This, however, is not possible for an operator on a finite dimensional vector space.

Theorem 2.15. *Let V and W be finite dimensional vector spaces over a common field F such that*

$$\dim(V) = \dim(W)$$

For a linear transformation $T : V \to W$, the following are equivalent:

(i) T is invertible.

(ii) T is non-singular.

(iii) T is surjective.

(iv) If $\mathcal{B} = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ is a basis of V, then $T(\mathcal{B}) = \{T(\alpha_1), T(\alpha_2), \ldots, T(\alpha_n)\}$ is a basis of W.

(v) There is some basis $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ of V such that $\{T(\alpha_1), T(\alpha_2), \ldots, T(\alpha_n)\}$ is a basis for W.

Proof.

(i) ⇒ *(ii)*: If T is invertible, then T is bijective. Hence, if $\alpha \in V$ is such that $T(\alpha) = 0$, then since $T(0) = 0$, it must follow that $\alpha = 0$. Hence, T is non-singular.

(ii) ⇒ *(iii)*: If T is non-singular, then $\text{nullity}(T) = 0$, so by the Rank-Nullity theorem, we know that

$$\text{rank}(T) = \text{rank}(T) + \text{nullity}(T) = \dim(V) = \dim(W)$$

But R_T is a subspace of W, and so by Corollary II.3.14, it follows that $R_T = W$. Hence, T is surjective.

$(iii) \Rightarrow (i)$: If T is surjective, then $R_T = W$. By the Rank-Nullity theorem, it follows that nullity$(T) = 0$. We claim that T is injective. To see this, suppose $\alpha, \beta \in V$ are such that $T(\alpha) = T(\beta)$, then $T(\alpha - \beta) = 0$. Hence,

$$\alpha = \beta = 0 \Rightarrow \alpha = \beta$$

Thus, T is injective, and hence bijective. So by Theorem 2.11, T is invertible.

$(i) \Rightarrow (iv)$: If $\mathcal{B}$ is a basis of V and T is invertible, then T is non-singular by the earlier steps. Hence, by Theorem 2.13, $T(\mathcal{B})$ is a linearly independent set in W. Since

$$\dim(W) = n$$

it follows that this set is a basis for W.

$(iv) \Rightarrow (v)$: Trivial.

$(v) \Rightarrow (iii)$: Suppose $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ is a basis for V such that $\{T(\alpha_1), T(\alpha_2), \ldots, T(\alpha_n)\}$ is a basis for W, then if $\beta \in W$, then there exist scalars $c_1, c_2, \ldots, c_n \in F$ such that

$$\beta = \sum_{i=1}^n c_i T(\alpha_i)$$

Hence, if

$$\alpha = \sum_{i=1}^n c_n \alpha_i \in V$$

Then $\beta = T(\alpha)$. So T is surjective as required.

□

3. Isomorphism

Definition 3.1. An *isomorphism* between two vector spaces V and W is a bijective linear transformation $T : V \to W$. If such an isomorphism exists, we say that V and W are *isomorphic*, and we write $V \cong W$.

Note that if $T : V \to W$ is an isomorphism, then so is T^{-1} (by Theorem 2.11). Similarly, if $T : V \to W$ and $S : W \to Z$ are both isomorphisms, then so is $ST : V \to Z$. Hence, the notion of isomorphism is an equivalence relation on the set of all vector spaces.

Theorem 3.2. *Any n dimensional vector space over a field F is isomorphic to F^n.*

Proof. Fix a basis $\mathcal{B} := \{\alpha_1, \alpha_2, \ldots, \alpha_n\} \subset V$, and define $T : F^n \to V$ by

$$T(x_1, x_2, \ldots, x_n) := \sum_{i=1}^n x_i \alpha_i$$

Note that T sends the standard basis of F^n to the basis $\mathcal{B}$. By Theorem 2.15, T is an isomorphism. □

4. Representation of Transformations by Matrices

Let V and W be two vector spaces, and fix two ordered bases $\mathcal{B} = \{\alpha_1, \alpha_2, \dots, \alpha_n\}$ and $\mathcal{B}' = \{\beta_1, \beta_2, \dots, \beta_m\}$ of V and W respectively. Let $T : V \to W$ be a linear transformation. For any $1 \leq j \leq n$, the vector $T(\alpha_j)$ can be expressed as a linear combination

$$T(\alpha_j) = \sum_{i=1}^{m} A_{i,j}\beta_i$$

By the notation of section 4, this means

$$[T(\alpha_j)]_{\mathcal{B}'} = \begin{pmatrix} A_{1,j} \\ A_{2,j} \\ \vdots \\ A_{m,j} \end{pmatrix}$$

Since the basis $\mathcal{B}$ is also ordered, we may now associate to T the $m \times n$ matrix

$$A = \begin{pmatrix} A_{1,1} & A_{1,2} & \dots & A_{1,j} & \dots & A_{1,n} \\ A_{2,1} & A_{2,2} & \dots & A_{2,j} & \dots & A_{2,n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ A_{m,1} & A_{m,2} & \dots & A_{m,j} & \dots & A_{m,n} \end{pmatrix}$$

In other words, the j^{th} column of A is $[T(\alpha_j)]_{\mathcal{B}'}$.

Definition 4.1. The matrix defined above is called the *matrix associated to T* and is denoted by

$$[T]_{\mathcal{B}'}^{\mathcal{B}}$$

Now suppose $\alpha \in V$, then write

$$\alpha = \sum_{j=1}^{n} x_j\alpha_j \Rightarrow [\alpha]_{\mathcal{B}} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

Then

$$\begin{aligned} T(\alpha) &= \sum_{j=1}^{n} x_j T(\alpha_j) \\ &= \sum_{j=1}^{n} x_j \left(\sum_{i=1}^{m} A_{i,j}\beta_i \right) \\ &= \sum_{i=1}^{n} \sum_{j=1}^{m} (A_{i,j}x_j)\,\beta_i \end{aligned}$$

Hence,

$$[T(\alpha)]_{\mathcal{B}'} = \begin{pmatrix} \sum_{j=1}^n A_{1,j}x_j \\ \sum_{j=1}^n A_{2,j}x_j \\ \vdots \\ \sum_{j=1}^n A_{m,j}x_j \end{pmatrix} = A[\alpha]_{\mathcal{B}}$$

Hence, we obtain the following result

Theorem 4.2. *Let $V, W, \mathcal{B}, \mathcal{B}'$ be as above. For each linear transformation $T : V \to W$, there is an $m \times n$ matrix $A = [T]_{\mathcal{B}'}^{\mathcal{B}}$ in $F^{m\times n}$ such that, for any vector $\alpha \in V$,*

$$[T(\alpha)]_{\mathcal{B}'} = A[\alpha]_{\mathcal{B}}$$

Furthermore, the map

$$\Theta : L(V, W) \to F^{m\times n}$$

given by

$$T \to [T]_{\mathcal{B}'}^{\mathcal{B}}$$

is a linear isomorphism of F-vector spaces.

Proof. Using the construction as before, we have that Θ is a well-defined map.

(i) Θ is linear: If $T, S \in L(V, W)$, then write $A := [T]_{\mathcal{B}'}^{\mathcal{B}}$ and $B = [S]_{\mathcal{B}'}^{\mathcal{B}}$. Then the j^{th} columns of A and B respectively are

$$[T(\alpha_j)]_{\mathcal{B}'} \text{ and } [S(\alpha_j)]_{\mathcal{B}'}$$

Hence, the j^{th} column of $[T + S]_{\mathcal{B}'}^{\mathcal{B}}$ is

$$[(T + S)(\alpha_j)]_{\mathcal{B}'} = [T(\alpha_j) + S(\alpha_j)]_{\mathcal{B}'} = [T(\alpha_j)]_{\mathcal{B}'} + [S(\alpha_j)]_{\mathcal{B}'}$$

Hence, $\Theta(T + S) = \Theta(T) + \Theta(S)$.

Similarly, if $T \in L(V, W)$ and $c \in F$, then $\Theta(cT) = c\Theta(T)$, so Θ is linear.

(ii) Θ is injective: If $T, S \in L(V, W)$ such that $[T]_{\mathcal{B}'}^{\mathcal{B}} = [S]_{\mathcal{B}'}^{\mathcal{B}}$, then, for each $1 \leq j \leq n$, we have

$$[T(\alpha_j)]_{\mathcal{B}'} = [S(\alpha_j)]_{\mathcal{B}'}$$

Hence, $T(\alpha_j) = S(\alpha_j)$ for all $1 \leq j \leq n$, whence $S = T$ by Theorem 1.4.

(iii) Θ is surjective: Note that T is an injective function, and

$$\dim(L(V, W)) = nm = \dim(F^{m\times n})$$

by Theorem 2.4 and Example II.3.11. Hence, T is an isomorphism by Theorem 2.15.

□

Definition 4.3. Let V be a finite dimensional vector space over a field F, and $\mathcal{B}$ be an ordered basis of V. For a linear operator $T \in L(V, V)$, we write

$$[T]_{\mathcal{B}} := [T]_{\mathcal{B}}^{\mathcal{B}}$$

This is called the *matrix of T relative to the ordered basis $\mathcal{B}$*.

Note that, if $\alpha \in V$, then, by this notation,

$$[T(\alpha)]_{\mathcal{B}} = [T]_{\mathcal{B}}[\alpha]_{\mathcal{B}}$$

Example 4.4.

(i) Let $V = F^n, W = F^m$ and $A \in F^{m\times n}$. Define $T : V \to W$ by

$$T(X) = AX$$

If $\mathcal{B} = \{\epsilon_1, \epsilon_2, \ldots, \epsilon_n\}$ and $\mathcal{B}' = \{\beta_1, \beta_2, \ldots, \beta_m\}$ be the standard bases of V and W respectively, then

$$T(\epsilon_j) = A_{1,j}\beta_1 + A_{2,j}\beta_2 + \ldots + A_{m,j}\beta_m$$

Hence,

$$[T(\epsilon_j)]_{\mathcal{B}'} = \begin{pmatrix} A_{1,j} \\ A_{2,j} \\ \vdots \\ A_{m,j} \end{pmatrix}$$

Hence,

$$[T]_{\mathcal{B}'}^{\mathcal{B}} = A$$

(ii) Let $V = W = \mathbb{R}^2$ and $T(X) = AX$ where

$$A = \begin{pmatrix} 3 & 1 \\ 0 & 2 \end{pmatrix}$$

If $\mathcal{B} = \{\epsilon_1, \epsilon_2\}$, then $[T]_{\mathcal{B}} = A$, but if $\mathcal{B} = \{\epsilon_2, \epsilon_1\}$, then

$$T(\epsilon_2) = (1, 2) = 2\epsilon_2 + 1\epsilon_1 \Rightarrow [T(\epsilon_2)]_{\mathcal{B}} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

Similarly,

$$[T(\epsilon_1)]_{\mathcal{B}} = \begin{pmatrix} 0 \\ 3 \end{pmatrix}$$

Hence,

$$[T]_{\mathcal{B}} = \begin{pmatrix} 2 & 0 \\ 1 & 3 \end{pmatrix}$$

Hence, the matrix $[T]_{\mathcal{B}}$ very much depends on the basis.

(iii) Let $V = F^2 = W$ and $T : V \to W$ be the map $T(x,y) := (x,0)$. If $\mathcal{B}$ denotes the standard basis of V, then

$$[T]_{\mathcal{B}} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

(iv) Let V be the space of all polynomials of degree ≤ 3 and $D : V \to V$ be the 'derivative' operator. Let $\mathcal{B} = \{\alpha_0, \alpha_1, \alpha_2, \alpha_3\}$ be the basis given by

$$\alpha_i(x) := x^i$$

Then $D(\alpha_0) = 0$, and for $i \geq 1$,

$$D(\alpha_i)(x) = ix^{i-1} \Rightarrow D(\alpha_i) = i\alpha_{i-1}$$

Hence,

$$[D]_{\mathcal{B}} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Let $T : V \to W$ and $S : W \to Z$ be two linear transformations, and let $\mathcal{B} = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}, \mathcal{B}' = \{\beta_1, \beta_2, \ldots, \beta_m\}$, and $\mathcal{B}'' = \{\gamma_1, \gamma_2, \ldots, \gamma_p\}$ be fixed ordered bases of V, W, and Z respectively. Suppose further that

$$A := [T]_{\mathcal{B}'}^{\mathcal{B}} = (a_{i,j}) \text{ and } B := [S]_{\mathcal{B}''}^{\mathcal{B}'} = (b_{s,t})$$

Set $C := [ST]_{\mathcal{B}''}^{\mathcal{B}}$, and observe that, for each $1 \leq j \leq n$, the j^{th} column of C is

$$[ST(\alpha_j)]_{\mathcal{B}''}$$

Now note that

$$\begin{aligned} ST(\alpha_j) &= S(T(\alpha_j)) \\ &= S\left(\sum_{k=1}^m a_{k,j}\beta_k\right) \\ &= \sum_{k=1}^m a_{k,j} S(\beta_k) \\ &= \sum_{k=1}^m a_{k,j}\left(\sum_{i=1}^p b_{i,k}\gamma_i\right) \\ &= \sum_{i=1}^p \left(\sum_{k=1}^m b_{i,k}a_{k,j}\right)\gamma_i \end{aligned}$$

Hence,

$$[ST(\alpha_j)]_{\mathcal{B}''} = \begin{pmatrix} (\sum_{k=1}^m b_{1,k}a_{k,j}) \\ (\sum_{k=1}^m b_{2,k}a_{k,j}) \\ \vdots \\ (\sum_{k=1}^m b_{p,k}a_{k,j}) \end{pmatrix}$$

By definition, this means

$$c_{i,j} = \sum_{k=1}^{m} b_{i,k}a_{k,j}$$

Hence, we get

Theorem 4.5. *Let $T : V \to W$ and $S : W \to Z$ as above. Then*

$$[ST]_{\mathcal{B}''}^{\mathcal{B}} = [S]_{\mathcal{B}''}^{\mathcal{B}'}[T]_{\mathcal{B}'}^{\mathcal{B}}$$

Remark 4.6.

(i) This above calculation gives us a simple proof that matrix multiplication is associative (because composition of functions is clearly associative).

(ii) If $T, U \in L(V, V)$, then the above theorem implies that

$$[UT]_{\mathcal{B}} = [U]_{\mathcal{B}}[T]_{\mathcal{B}}$$

(iii) Hence, if $T \in L(V, V)$ is invertible, with inverse U, then

$$[U]_{\mathcal{B}}[T]_{\mathcal{B}} = [T]_{\mathcal{B}}[U]_{\mathcal{B}} = I$$

where I denotes the $n \times n$ identity matrix (Here, $n = \dim(V)$). So if T is invertible as a linear transformation, then $[T]_{\mathcal{B}}$ is an invertible matrix. Furthermore,

$$[T^{-1}]_{\mathcal{B}} = [T]_{\mathcal{B}}^{-1}$$

Conversely, if $[T]_{\mathcal{B}}$ is an invertible matrix with inverse B, then by Theorem 4.2, there is a linear map $U \in L(V, V)$ such that

$$[S]_{\mathcal{B}} = B$$

Hence, it follows that

$$[ST]_{\mathcal{B}} = [TS]_{\mathcal{B}} = [I]_{\mathcal{B}}$$

where I denotes the identity linear map on V. Since the map $T \mapsto [T]_{\mathcal{B}}$ is injective (again by Theorem 4.2), it follows that

$$ST = TS = I$$

so T is invertible.

Let $T \in L(V, V)$ be a linear operator and suppose we have two ordered bases

$$\mathcal{B} = \{\alpha_1, \alpha_2, \ldots, \alpha_n\} \text{ and } \mathcal{B}' = \{\beta_1, \beta_2, \ldots, \beta_n\}$$

of V. We would like to know how the matrices

$$[T]_{\mathcal{B}} \text{ and } [T]_{\mathcal{B}'}$$

are related.

By Theorem II.4.5, there is an invertible $n \times n$ matrix P such that, for any $\alpha \in V$,

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

Hence, if $\alpha \in V$, then

$$[T(\alpha)]_{\mathcal{B}} = P[T(\alpha)]_{\mathcal{B}'} = P[T]_{\mathcal{B}'}[\alpha]_{\mathcal{B}'}$$

But

$$[T(\alpha)]_{\mathcal{B}} = [T]_{\mathcal{B}}[\alpha]_{\mathcal{B}} = [T]_{\mathcal{B}}P[\alpha]_{\mathcal{B}'}$$

Equating these two, we get

$$[T]_{\mathcal{B}}P = P[T]_{\mathcal{B}'}$$

(since the above equations hold for all $\alpha \in V$). Since P is invertible, we conclude that

$$[T]_{\mathcal{B}'} = P^{-1}[T]_{\mathcal{B}}P$$

Remark 4.7. Let $U \in L(V, V)$ be the unique linear operator such that

$$U(\alpha_j) = \beta_j$$

for lal $1 \leq j \leq n$, then U is invertible since it maps one basis of V to another (by Theorem 2.15). Furthermore, if P is the change of basis matrix as above, then

$$\beta_j = \sum_{i=1}^{n} P_{i,j}\alpha_i$$

Since $U(\alpha_j) = \beta_j$, we conclude that

$$P = [U]_{\mathcal{B}}$$

Hence, we get the following theorem.

Theorem 4.8. *Let V be a finite dimensional vector space over a field F, and let*

$$\mathcal{B} = \{\alpha_1, \alpha_2, \ldots, \alpha_n\} \text{ and } \mathcal{B}' = \{\beta_1, \beta_2, \ldots, \beta_n\}$$

be two ordered bases of V. If $T \in L(V, V)$ and P is the change of basis matrix (as in Theorem II.4.5) whose j^{th} column is

$$P_j = [\beta_j]_{\mathcal{B}}$$

Then

$$[T]_{\mathcal{B}'} = P^{-1}[T]_{\mathcal{B}}P$$

Equivalently, if $U \in L(V, V)$ is the invertible operator defined by $U(\alpha_j) = \beta_j$ for all $1 \leq j \leq n$, then

$$[T]_{\mathcal{B}'} = [U^{-1}]_{\mathcal{B}}[T]_{\mathcal{B}}[U]_{\mathcal{B}}$$

Example 4.9.

(i) Let $V = \mathbb{R}^2$, $\mathcal{B} = \{\epsilon_1, \epsilon_2\}$ and $\mathcal{B}' = \{\beta_1, \beta_2\}$, where

$$\beta_1 = \epsilon_1 + \epsilon_2 \text{ and } \beta_2 = 2\epsilon_1 + \epsilon_2$$

Then the change of basis matrix as above is

$$P = \begin{pmatrix} 1 & 2 \\ 1 & 1 \end{pmatrix}$$

Hence,

$$P^{-1} = \begin{pmatrix} -1 & 2 \\ 1 & -1 \end{pmatrix}$$

Hence, if $T \in L(V, V)$ is the linear operator given by $T(x, y) := (x, 0)$, then observe that

(i) $T(\beta_1) = T(1, 1) = (1, 0) = -\beta_2 + \beta_1$, while $T(\beta_2) = (2, 0) = -2\beta_1 + 2\beta_2$, so that

$$[T]_{\mathcal{B}'} = \begin{pmatrix} -1 & -2 \\ 1 & 2 \end{pmatrix}$$

(ii) Now note that

$$[T]_{\mathcal{B}} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

so that

$$\begin{aligned} P^{-1}[T]_{\mathcal{B}}P &= \begin{pmatrix} -1 & 2 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 1 & 1 \end{pmatrix} \\ &= \begin{pmatrix} -1 & 2 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} -1 & -2 \\ 1 & 2 \end{pmatrix} \end{aligned}$$

which agrees with Theorem 4.8.

(ii) Let V be the space of all real polynomials of degree ≤ 3, and let $\mathcal{B} = \{\alpha_0, \alpha_1, \alpha_2, \alpha_3\}$ be the basis given by

$$\alpha_0(x) := x^i$$

Define a new basis $\{\beta_0, \beta_1, \beta_2, \beta_3\}$ by

$$\beta_i(x) := (x + 2)^i$$

Then

$$\begin{aligned} \beta_0 &= \alpha_0 \\ \beta_1 &= 2\alpha_0 + \alpha_1 \\ \beta_2 &= 4\alpha_0 + 2\alpha_1 + \alpha_2 \\ \beta_3 &= 8\alpha_0 + 12\alpha_1 + 6\alpha_3 + \alpha_4 \end{aligned}$$

Hence, the change of basis matrix is

$$P = \begin{pmatrix} 1 & 2 & 4 & 8 \\ 0 & 1 & 4 & 12 \\ 0 & 0 & 1 & 6 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Hence,

$$P^{-1} = \begin{pmatrix} 1 & -2 & 4 & -8 \\ 0 & 1 & -4 & 12 \\ 0 & 0 & 1 & -6 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Now let $D : V \to V$ be the derivative operator. Then,

(i)

$$\begin{aligned} D(\beta_0) &= 0 \\ D(\beta_1) &= \beta_0 \\ D(\beta_2) &= 2\beta_1 \\ D(\beta_3) &= 3\beta_2 \end{aligned}$$

Hence,

$$[D]_{\mathcal{B}'} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

(ii) Now we saw in Example 4.4 that

$$[D]_{\mathcal{B}} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

So note that

$$\begin{aligned} P^{-1}[D]_{\mathcal{B}}P &= \begin{pmatrix} 1 & -2 & 4 & -8 \\ 0 & 1 & -4 & 12 \\ 0 & 0 & 1 & -6 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 2 & 4 & 8 \\ 0 & 1 & 4 & 12 \\ 0 & 0 & 1 & 6 \\ 0 & 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 & -2 & 4 & -8 \\ 0 & 1 & -4 & 12 \\ 0 & 0 & 1 & -6 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 & 4 & 12 \\ 0 & 0 & 2 & 12 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix} \end{aligned}$$

which agrees with Theorem 4.8.

This leads to the following definition for matrices.

Definition 4.10. Let A and B be two $n \times n$ matrices over a field F. We say that A is *similar to* B if there exists an invertible $n \times n$ matrix P such that

$$B = P^{-1}AP$$

Remark 4.11. Note that the notion of similarity is an equivalence relation on the set of all $n \times n$ matrices (Check!). Furthermore, if A is similar to the zero matrix, then A must be the zero matrix, and if A is similar to the identity matrix, then $A = I$.

Finally, we have the following corollaries, the first of which follows directly from Theorem 4.8.

Corollary 4.12. *Let V be a finite dimensional vector space with two ordered bases $\mathcal{B}$ and $\mathcal{B}'$. Let $T \in L(V, V)$, then the matrices $[T]_{\mathcal{B}}$ and $[T]_{\mathcal{B}'}$ are similar.*

Corollary 4.13. *Let $V = F^n$ and A and B be two $n \times n$ matrices. Define $T : V \to V$ be the linear operator*

$$T(X) = AX$$

Then, B is similar to A if and only if there is a basis $\mathcal{B}'$ of V such that

$$[T]_{\mathcal{B}'} = B$$

Proof. By Example 4.4, if $\mathcal{B}$ denotes the standard basis of V, then

$$[T]_{\mathcal{B}} = A$$

Hence if $\mathcal{B}'$ is another basis such that $[T]_{\mathcal{B}'} = B$, then A and B are similar by Theorem 4.8.

Conversely, if A and B are similar, then there exists an invertible matrix P such that

$$B = P^{-1}AP$$

Let $\mathcal{B}' = \{\beta_1, \beta_2, \ldots, \beta_n\}$ be given by the formula

$$\beta_j = \sum_{i=1}^{n} P_{i,j}\epsilon_i$$

Then, since P is invertible, it follows from Theorem 2.15, that $\mathcal{B}'$ is a basis of V. Now one can verify (please check!) that

$$[T]_{\mathcal{B}'} = B$$

□

5. Linear Functionals

Definition 5.1. Let V be a vector space over a field F. A *linear functional* on V is a linear transformation $L : V \to F$.

Example 5.2.

(i) Let $V = F^n$ and fix an n tuple $(a_1, a_2, \ldots, a_n) \in F^n$. We define $L : V \to F$ by

$$L(x_1, x_2, \ldots, x_n) := \sum_{i=1}^{n} a_i x_i$$

Then L is a linear functional.

(ii) Conversely, if $L : F^n \to F$ is a linear functional, and we set $a_j := L(\epsilon_j)$, then, for any $\alpha = (x_1, x_2, \ldots, x_n) \in V$, we have

$$L(\alpha) = L\left(\sum_{i=1}^{n} x_i \epsilon_i\right) = \sum_{i=1}^{n} x_i L(\epsilon_i) = \sum_{i=1}^{n} x_i a_i$$

Hence, L is associated to the tuple $(a_1, a_2, \ldots, a_n)$. In fact, if $\mathcal{B} = \{\epsilon_1, \epsilon_2, \ldots, \epsilon_n\}$ is the standard basis for F^n and $\mathcal{B}' := \{1\}$ is taken as a basis for F, then

$$[L]_{\mathcal{B}'}^{\mathcal{B}} = (a_1, a_2, \ldots, a_n)$$

in the notation of the previous section.

(iii) Let $V = F^{n\times n}$ be the vector space of $n \times n$ matrices over a field F. Define $L : V \to F$ by

$$L(A) = \text{trace}(A) = \sum_{i=1}^{n} A_{i,i}$$

Then, L is a linear functional (Check!)

(iv) Let V be the space of all polynomials over a field F, and let $t \in F$. Define $L_t : V \to F$ by

$$L_t(f) := f(t)$$

obtained by 'evaluating a polynomial at t'. This is a linear functional (Check!)

(v) Let $V = C([a, b])$ denote the vector space of all continuous functions $f : [a, b] \to F$, and define $L : V \to F$ by

$$L(f) := \int_a^b f(t)dt$$

Then L is a linear functional.

Definition 5.3. Let V be a vector space over a field F. The *dual space* of V is the space

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

Remark 5.4. Let V be a finite dimensional vector space and $\mathcal{B} = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ be a basis for V. By Theorem 2.4, we have

$$\dim(V^*) = \dim(V) = n$$

Note that $\mathcal{B}' = \{1\}$ is a basis for F. Hence, by Theorem 1.4, for each $1 \leq i \leq n$, there is a unique linear functional f_i such that

$$f_i(\alpha_j) = \delta_{i,j}$$

Now observe that the set $\mathcal{B}^* := \{f_1, f_2, \ldots, f_n\}$ is a linearly independent set, because if $c_i \in F$ are scalars such that

$$\sum_{i=1}^n c_i f_i = 0$$

Then for a fixed $1 \leq j \leq n$, we get

$$\left(\sum_{i=1}^n c_i f_i\right)(\alpha_j) = 0 \Rightarrow c_j = 0$$

Hence, it follows that $\mathcal{B}^*$ is a basis for V^*.

Theorem 5.5. *Let V be a finite dimensional vector space over a field F and $\mathcal{B} = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ be a basis for V. Then there is a basis $\mathcal{B}^* = \{f_1, f_2, \ldots, f_n\}$ of V^* which satisfies*

$$f_i(\alpha_j) = \delta_{i,j}$$

for all $1 \leq i, j \leq n$. Furthermore, for each $f \in V^*$, we have

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

and for each $\alpha \in V$, we have

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

Proof. (i) We have just proved above that such a basis exists.

(ii) Now suppose $f \in V^*$, then consider the linear functional given by

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

Evaluating at α_j, we see that

$$g(\alpha_j) = \left(\sum_{i=1}^n f(\alpha_i f_i\right)(\alpha_j) = f(\alpha_j)$$

By the uniqueness of Theorem 1.4, we have that $f = g$ as required.

(iii) Finally, if $\alpha \in V$, then we write

$$\alpha = \sum_{i=1}^n c_i \alpha_i$$

Applying f_j to both sides, we see that

$$c_j = f_j(\alpha)$$

as required.

□

Definition 5.6. The basis constructed above is called the *dual basis* of $\mathcal{B}$.

Remark 5.7. If V is a finite dimensional vector space and $\mathcal{B} = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ is an ordered basis for V, then the dual basis $\mathcal{B}^* = \{f_1, f_2, \ldots, f_n\}$ allows us to recover the coordinates of a vector in the basis $\mathcal{B}$. In other words, if $\alpha \in V$, then

$$[\alpha]_{\mathcal{B}} = \begin{pmatrix} f_1(\alpha) \\ f_2(\alpha) \\ \vdots \\ f_n(\alpha) \end{pmatrix}$$

Example 5.8. Let V be the space of polynomials over $\mathbb{R}$ of degree ≤ 2. Fix three distinct real numbers $t_1, t_2, t_3 \in \mathbb{R}$ and define $L_i \in V^*$ by

$$L_i(p) := p(t_i)$$

We claim that the set $\mathcal{S} := \{L_1, L_2, L_3\}$ is a basis for V^*. Since $\dim(V^*) = \dim(V) = 3$, it suffices to show that $\mathcal{S}$ is linearly independent. To see this, fix scalars $c_i \in \mathbb{R}$ such that

$$\sum_{i=1}^3 c_i L_i = 0$$

Evaluating at the 'standard basis' $\mathcal{B} = \{\alpha_0, \alpha_1, \alpha_2\}$ of V (where $\alpha_i(x) = x^i$), we get three equations

$$\begin{aligned} c_1 + c_2 + c_3 &= 0 \\ t_1 c_1 + t_2 c_2 + t_3 c_3 &= 0 \\ t_1^2 c_1 + t_2^2 c_2 + t_3^2 c_3 &= 0 \end{aligned}$$

But the matrix

$$\begin{pmatrix} 1 & 1 & 1 \\ t_1 & t_2 & t_3 \\ t_1^2 & t_2^2 & t_3^2 \end{pmatrix}$$

is an invertible matrix when t_1, t_2, t_3 are three distinct numbers. Hence, we conclude that

$$c_1 = c_2 = c_3 = 0$$

Hence, $\mathcal{S}$ forms a basis for V^*. We wish to find a basis $\mathcal{B}' = \{p_1, p_2, p_3\}$ of V such that $\mathcal{S}$ is the dual basis of $\mathcal{B}'$. In other words, we wish to find polynomials p_1, p_2, and p_3 such that

$$p_j(t_i) = \delta_{i,j}$$

One can do this by hand, by taking

$$\begin{aligned}
p_1(x) &= \frac{(x - t_2)(x - t_3)}{(t_1 - t_2)(t_1 - t_3)} \\
p_2(x) &= \frac{(x - t_1)(x - t_3)}{(t_2 - t_1)(t_2 - t_3)} \\
p_3(x) &= \frac{(x - t_2)(x - t_1)}{(t_3 - t_2)(t_3 - t_1)}
\end{aligned}$$

Remark 5.9. Let V be a n-dimensional vector space and $f \in V^*$ be a non-zero linear functional. Then the rank of f is 1 (Why?). So by the rank-nullity theorem,

$$\dim(N_f) = n - 1$$

where N_f denotes the null space of f.

Definition 5.10. If V is a vector space of dimension n, then a subspace of dimension $(n - 1)$ is called a *hyperspace.*

We wish to know if every hyperspace is the kernel of a non-zero linear functional. To do that, we need a definition.

Definition 5.11. Let V be a vector space and $S \subset V$ be a subset of V. The set

$$S^0 := \{f \in V^* : f(\alpha) = 0 \quad \forall \alpha \in S\}$$

is called the *annihilator* of S.

Now the following facts are easy to prove (Check!)

Example 5.12. Let V be a finite dimensional vector space.

(i) For any set $S \subset V$, S^0 is a subspace of V^*.

(ii) If $S = \{0\}$, then $S^0 = V^*$.

(iii) If $S = V$, then $S^0 = \{0\}$.

(iv) If $S_1 \subset S_2$, then $S_2^0 \subset S_1^0$.

(v) For any subset $S \subset V$, if $W = \text{span}(S)$, then

$$S^0 = W^0$$

Theorem 5.13. *Let V be a finite dimensional vector space over a field F, and let W be a subspace of V. Then*

$$\dim(W) + \dim(W^0) = \dim(V)$$

Proof. Suppose $\dim(W) = k$ and $\mathcal{S} = \{\alpha_1, \alpha_2, \ldots, \alpha_k\}$ be a basis for W. Choose vectors $\{\alpha_{k+1}, \alpha_{k+2}, \ldots, \alpha_n\} \subset V$ such that $\mathcal{B} = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ is a basis for V. Let $\mathcal{B}^* = \{f_1, f_2, \ldots, f_n\}$ be the dual basis of $\mathcal{B}$, so that, for any $1 \leq i, j \leq n$, we have

$$f_i(\alpha_j) = \delta_{i,j}$$

So if $k+1 \leq i \leq n$, then, for any $1 \leq j \leq k$, we have

$$f_i(\alpha_j) = 0$$

Since $\mathcal{S}$ is a basis for W, it follows (Why?) that

$$f_i \in W^0$$

Hence, $\mathcal{T} := \{f_{k+1}, f_{k+2}, \ldots, f_n\} \subset W^0$. Since $\mathcal{B}^*$ is a linearly independent set, so is $\mathcal{T}$. We claim that $\mathcal{T}$ is a basis for W^0. To see this, fix $f \in W^0$, then $f \in V^*$, so by Theorem 5.5, we have

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

But $f(\alpha_i) = 0$ for all $1 \leq i \leq k$, so

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

Hence, $\mathcal{T}$ spans W^0 as required.

We now conclude that

$$\dim(V) = n = k + (n-k) = \dim(W) + \dim(W^0)$$

□

Corollary 5.14. *If W is a k-dimensional subspace of an n-dimensional vector space V, then there exist $(n-k)$ hyperspaces $W_1, W_2, \ldots, W_{n-k}$ such that*

$$W = \bigcap_{i=1}^{n-k} W_i$$

Proof. Consider the proof of Theorem 5.13. We constructed a basis $\mathcal{T} := \{f_{k+1}, f_{k+2}, \ldots, f_n\}$ of W^0. Set

$$V_i := \ker(f_i)$$

and set

$$X := \bigcap_{i=k+1}^n V_i$$

We claim that $W = X$ proving the result.

(i) If $\alpha \in W$, then, since $\mathcal{T} \subset W^0$, we have

$$\alpha \in \ker(f_i)$$

for all $k+1 \leq i \leq n$. Hence, $\alpha \in X$.

(ii) Conversely, if $\alpha \in X$, then $\alpha \in V$, so we write

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

by Theorem 5.5. But $\alpha \in \ker(f_i)$ for all $k+1 \leq i \leq n$, so

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

But $\mathcal{S} = \{\alpha_1, \alpha_2, \ldots, \alpha_k\} \subset W$, so $\alpha \in W$.

□

Corollary 5.15. *Let W_1 and W_2 be two subspaces of a finite dimensional vector space V. Then $W_1 = W_2$ if and only if $W_1^0 = W_2^0$.*

Proof. Clearly, if $W_1 = W_2$, then $W_1^0 = W_2^0$.

Conversely, suppose $W_1 \neq W_2$, then we may assume without loss of generality, that there is a vector $\alpha \in W_2 \setminus W_1$. Let $\mathcal{S} = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ be a basis for W_1, then the set $\mathcal{S} \cup \{\alpha\}$ is also linearly independent. So by Theorem II.3.13, there is a basis $\mathcal{B}$ containing $S \cup \{\alpha\}$. Hence, by the proof of Theorem 5.13, there is a linear functional $f \in \mathcal{B}^*$ such that

$$f(\alpha_i) = 0$$

for all $1 \leq i \leq k$, but

$$f(\alpha) = 1$$

Hence, $f \in W_1^0$, but $f \notin W_2^0$. Thus, $W_1^0 \neq W_2^0$ as required. □

(End of Week 4)

6. The Double Dual

Definition 6.1. Let V be a vector space over a field F and let $\alpha \in V$. Define $L_\alpha : V^* \to F$ by

$$L_\alpha(f) := f(\alpha)$$

The proof of the next lemma is an easy exercise.

Lemma 6.2. *For each $\alpha \in V, L_\alpha$ is a linear functional on V^**

Definition 6.3. The *double dual* of V is the vector space $V^{**} := (V^*)^*$

Theorem 6.4. *Let V be a finite dimensional vector space. The map $\Theta : V \to V^{**}$ given by*

$$\Theta(\alpha) := L_\alpha$$

is a linear isomorphism.

Proof.

(i) Θ is well-defined because $L_\alpha \in V^{**}$ for each $\alpha \in V$ by the previous lemma.

(ii) Θ is linear: If $\alpha, \beta \in V$, then, for any $f \in V^*$

$$L_{\alpha+\beta}(f) = f(\alpha+\beta) = f(\alpha) + f(\beta) = L_\alpha(f) + L_\beta(f) = (L_\alpha + L_\beta)(f)$$

Hence,

$$L_{\alpha+\beta} = L_\alpha + L_\beta$$

Hence, Θ is additive. Similarly, $L_{c\alpha} = cL_\alpha$ for any $c \in F$, so Θ is linear.

(iii) Θ is injective: If $\alpha \in V$ is a non-zero vector, then consider $W_1 := \text{span}(\alpha)$ and $W_2 = \{0\}$. Since $W_1 \neq W_2$,

$$W_1^0 \neq W_2^0$$

by Corollary 5.15. Since $W_2^0 = V^*$, it follows that there is a linear functional $f \in V^*$ such that

$$f(\alpha) \neq 0$$

Hence, $L_\alpha \neq 0$. Thus, (Why?)

$$\Theta(\alpha) = 0 \Rightarrow \alpha = 0$$

so Θ is injective.

(iv) Now note that

$$\dim(V) = \dim(V^*) = \dim(V^{**})$$

so Θ is surjective as well.

□

Corollary 6.5. *If $L \in V^{**}$, then there exists $\alpha \in V$ such that*

$$L(f) = f(\alpha) \quad \forall f \in V^*$$

Corollary 6.6. *If $\mathcal{B}$ is a basis of V^*, then there exists a basis $\mathcal{B}'$ of V such that $\mathcal{B}$ is the dual basis of $\mathcal{B}'$.*

Proof. Write $\mathcal{B} = \{f_1, f_2, \ldots, f_n\}$. By Theorem 5.5, there is a basis $\mathcal{S} = \{L_1, L_2, \ldots, L_n\}$ of V^{**} such that

$$L_i(f_j) = \delta_{i,j}$$

For each $1 \leq i \leq n$, there exists $\alpha_i \in V$ such that $L_i = L_{\alpha_i}$. In other words,

$$f_j(\alpha_i) = \delta_{i,j}$$

for all $1 \leq i, j \leq n$. Now set $\mathcal{S} = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$. Then

(i) $\mathcal{S}$ is linearly independent: If $c_i \in F$ such that

$$c_i\alpha_i = 0$$

Applying f_j to this expression, we have

$$c_j = 0$$

This is true for each $1 \leq j \leq n$, so $\mathcal{S}$ is linearly independent.

(ii) Since $\dim(V) = n$, it follows that $\mathcal{S}$ is a basis for V.

□

Recall that, if $S \subset V$, we write

$$S^0 = \{f \in V^* : f(\alpha) = 0 \quad \forall \alpha \in S\}$$

Definition 6.7. If $S \subset V^*$, we write

$$S^0 = \{\alpha \in V : f(\alpha) = 0 \quad \forall f \in S\}$$

Note that the two definitions agree if we identify V with V^{**} via Θ.

Theorem 6.8. *If S is any subset of a finite dimensional vector space V, then*

$$(S^0)^0 = \textit{span}(S)$$

Proof. Let $W := \text{span}(S)$, then, by Example 5.12,

$$S^0 = W^0$$

Therefore, we wish to show that $W^0 = (W^0)^0$.

(i) Observe that, if $\alpha \in W$ and $f \in W^0$, then $f(\alpha) = 0$. Hence, $\alpha \in (W^0)^0$. Thus,

$$W \subset (W^0)^0$$

(ii) Now note that W an $(W^0)^0$ are both subspaces over V. By Theorem 5.13, we have

$$\dim(W) + \dim(W^0) = \dim(V)$$

and

$$\dim(W^0) + \dim((W^0)^0) = \dim(V^*) = \dim(V)$$

Hence, $\dim(W) = \dim((W^0)^0)$. By Corollary II.3.14, we conclude that

$$W = (W^0)^0$$

□

We now wish to prove Corollary 5.14 for vector spaces that are not finite dimensional.

Definition 6.9. Let V be a vector space. A subspace $W \subset V$ is called a *hyperspace* if

(i) $W \neq V$ (ie. W is a proper subspace)

(ii) For any subspace N of V such that

$$W \subset N \subset V$$

we must have $W = N$ or $N = V$.

In other words, a hyperspace is a *maximal* proper subspace.

Theorem 6.10. *Let V be a vector space over a field F.*

(i) If $f \in V^*$ is non-zero, then $\ker(f)$ *is a hyperspace.*

(ii) If $W \subset V$ is a hyperspace, then there exists $f \in V^*$ such that $W =$ $\ker(f)$

Proof.

(i) If $f \in V^*$ is non-zero, then $W := \ker(f)$ is a subspace of V. Furthermore, since $f \neq 0$, it follows that $\ker(f) \neq V$. Now suppose N is a subspace of V such that

$$W \subset N \subset V$$

We wish to conclude that $W = N$ or $N = V$. Suppose $W \neq N$, then we will show that $N = V$.

Since $W \neq N$ and $W \subset N$, there is a vector $\alpha \in N$ such that $\alpha \notin W$. Hence,

$$f(\alpha) \neq 0$$

Fix $\beta \in V$, then we wish to show that $\beta \in N$.

- If $f(\beta) = 0$, then $\beta \in W \subset N$

- If $f(\beta) \neq 0$, then set

  $$\gamma := \beta - \frac{f(\beta)}{f(\alpha)}\alpha$$

  Then

  $$f(\gamma) = f(\beta) - f(\beta) = 0$$

  Hence, $\gamma \in W \subset N$, so

  $$\beta = \gamma + \frac{f(\beta)}{f(\alpha)}\alpha \in N$$

Either way, we conclude that $\beta \in N$. This is true for any $\beta \in V$, so $V = N$ as required.

(ii) Now let $W \subset V$ be a hyperspace. Since $W \neq V$, choose $\alpha \notin W$, so that

$$W + \text{span}(\alpha)$$

is a subspace of V by Remark II.2.11. Since $\alpha \notin W$, this subspace is not W. Since W is a hyperspace, it follows that

$$V = W + \text{span}(\alpha)$$

Hence, for each $\beta \in V$, there exists $\gamma \in W$ and $c \in F$ such that

$$\beta = \gamma + c\alpha$$

We claim that this expression is unique: If

$$\beta = \gamma' + c'\alpha$$

Then

$$(\gamma - \gamma') = (c' - c)\alpha$$

But $(\gamma - \gamma') \in W$ and $\alpha \notin W$. So we conclude (Why?) that

$$c = c'$$

Hence, $\gamma = \gamma'$ as well. Thus, we define $g : V \to F$ by

$$g(\beta) = c$$

Then, (Check!) that g is a linear functional. It is now clear that

$$\ker(g) = W$$

as required.

□

Lemma 6.11. *Let $f, g \in V^*$ be two linear functionals. Then*

$$\ker(f) \subset \ker(g)$$

if and only if there is a scalar $c \in F$ such that $g = cf$

Proof. Clearly, if $g = cf$ for some $c \in F$, then $\ker(f) \subset \ker(g)$.

Conversely, suppose $\ker(f) \subset \ker(g)$. If $g \equiv 0$, then take $c = 0$. Otherwise, f must also be non-zero, so that $\ker(f)$ is a hyperspace. Since $\ker(g) \neq V$, we conclude that

$$\ker(f) = \ker(g)$$

Now choose a vector $\alpha \in V$ such that $f(\alpha) \neq 0$. Consider

$$c := \frac{g(\alpha)}{f(\alpha)} \in F$$

Then we claim that $g = cf$. So set

$$h := g - cf$$

and we wish to show that $h \equiv 0$. Note that, if $\alpha \in \ker(f) = \ker(g)$, then $h(\alpha) = 0$, so

$$\ker(f) \subset \ker(h)$$

Furthermore, by construction,

$$h(\alpha) = 0$$

Since $\alpha \notin \ker(f)$, it follows that $\ker(h)$ is a subspace of V that is strictly larger than $\ker(f)$. But $\ker(f)$ is a hyperspace, so

$$\ker(h) = V$$

whence $h \equiv 0$ as required. □

We now extend this lemma to a finite family of linear functionals.

Theorem 6.12. *Let $f_1, f_2, \ldots, f_n, g \in V^*$. Then, g is a linear combination of $\{f_1, f_2, \ldots, f_n\}$ if and only if*

$$\bigcap_{i=1}^n \ker(f_i) \subset \ker(g)$$

Proof.

(i) If there are scalars $c_i \in F$ such that

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

Then if $\alpha \in \ker(f_i)$ for all $1 \leq i \leq n$, then $g(\alpha) = 0$. So

$$\bigcap_{i=1}^n \ker(f_i) \subset \ker(g)$$

(ii) Conversely, suppose

$$\bigcap_{i=1}^n \ker(f_i) \subset \ker(g)$$

holds, then we proceed by induction on n.

- If $n = 1$, then this is Lemma 6.11.
- Suppose the theorem is true for $n = k-1$, and suppose $n = k$. So set $W := \ker(f_k)$, and restrict $g, f_1, f_2, \ldots, f_{k-1}$ to W to obtain linear functionals $g', f_1', f_2', \ldots, f_{k-1}'$. Now, if $\alpha \in W$ such that

  $$f_i'(\alpha) = 0 \quad \forall 1 \leq i \leq k-1$$

  Then, by definition,

  $$\alpha \in \bigcap_{i=1}^k \ker(f_i)$$

  Therefore, $g(\alpha) = 0$. Hence, $g'(\alpha) = 0$, so, by induction hypothesis,

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

  for some scalars $c_i \in F$. Now consider $h \in V^*$ given by

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

  Then, $h \equiv 0$ on $W = \ker(f_k)$. Hence,

  $$\ker(f_k) \subset \ker(h)$$

  By Lemma 6.11, there is a scalar $c \in F$ such that $h = cf_k$, whence

  $$g = c_1 f_1 + c_2 f_2 + \ldots + c_{k-1} f_{k-1} + c f_k$$

  as required.

□

7. The Transpose of a Linear Transformation

Let $T : V \to W$ be a linear transformation. Given $g \in W^*$, we define $f \in V^*$ by the formula

$$f(\alpha) := g(T(\alpha)) \tag{III.1}$$

Note that f is, indeed, a linear functional. Thus, we get an association

$$W^* \to V^*$$

which sends $g \to f$.

Theorem 7.1. *Given a linear transformation $T : V \to W$, there is a unique linear transformation*

$$T^t : W^* \to V^*$$

given by the formula

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

for each $g \in W^*$ and $\alpha \in V$. The map T^t is called the transpose *of T.*

Proof.

(i) We have just explained that T^t is well-defined. ie. if $g \in W^*$, then $T^t(g) \in V^*$.

(ii) Now suppose $g_1, g_2 \in W^*$, and set $f_1 := T^t(g_1), f_2 = T^*(g_2)$ and $f_3 = T^t(g_1 + g_2)$. Then, for any $\alpha \in V$, we have

$$f_3(\alpha) = (g_1 + g_2)(T(\alpha)) = g_1(T(\alpha)) + g_2(T(\alpha)) = f_1(\alpha) + f_2(\alpha)$$

Hence,

$$T^t(g_1 + g_2) = T^t(g_1) + T^t(g_2)$$

Similarly, $T^t(cg) = cT^t(g)$ for $c \in F, g \in W^*$. Hence, T^t is linear.

□

Theorem 7.2. *Let $T : V \to W$ be a linear transformation between finite dimensional vector spaces. Then*

(i) $\ker(T^t) = \mathit{Range}(T)^0$

(ii) $\mathit{rank}(T^t) = \mathit{rank}(T)$

(iii) $\mathit{Range}(T^t) = \ker(T)^0$

Proof.

(i) For any $g \in W^*$,

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

So if $g \in \ker(T^t)$, then $g(T(\alpha)) = 0$ for all $\alpha \in V$, whence $g(\beta) = 0$ for all $\beta \in \text{Range}(T)$, so

$$g \in \text{Range}(T)^0$$

Hence, $\ker(T^t) \subset \text{Range}(T)^0$. Conversely, if $g \in \text{Range}(T)^0$, then

$$g(T(\alpha)) = 0 \quad \forall \alpha \in V$$

whence, $T^t(g) = 0$. So the reverse containment also holds.

(ii) Let $r := \text{rank}(T)$ and $m = \dim(W)$, then by Theorem 5.13, we have

$$r + \dim(\text{Range}(T)^0) = m \Rightarrow \dim(\text{Range}(T)^0) = m - r$$

But $T^t : W^* \to V^*$ is a linear transformation with $m = \dim(W^*)$. So by Rank-Nullity, we have

$$\text{nullity}(T^t) + \text{rank}(T^t) = m$$

But by the first part, we have

$$\text{nullity}(T^t) = \dim(\text{Range}(T)^0) = m - r$$

so that $\text{rank}(T^t) = r = \text{rank}(T)$.

(iii) Now if $\alpha \in \ker(T)$, then, for any $g \in W^*$,

$$T^t(g)(\alpha) = g(T(\alpha)) = g(0) = 0$$

Hence, $T^t(g) \in \ker(T)^0$ for all $g \in W^*$, so that

$$\text{Range}(T^t) \subset \ker(T)^0$$

But if $n = \dim(V)$, then by Rank-Nullity,

$$\text{rank}(T^t) = \text{rank}(T) = n - \text{nullity}(T)$$

and by Theorem 5.13, we have

$$\dim(\ker(T)^0) + \text{nullity}(T) = n$$

Hence,

$$\dim(\text{Range}(T^t)) = \dim(\ker(T)^0)$$

so by Corollary II.3.14, we have

$$\text{Range}(T^t) = \ker(T)^0$$

□

Theorem 7.3. *Let $T : V \to W$ be a linear transformation between two finite dimensional vector spaces, and fix two ordered bases $\mathcal{B}$ and $\mathcal{B}'$ of V and W respectively. Then, we consider the matrix*

$$A = [T]_{\mathcal{B}'}^{\mathcal{B}}$$

Now consider $T^t : W^* \to V^*$, and the ordered bases $(\mathcal{B}')^*$ of W^* and $\mathcal{B}^*$ of V^*. We again have a matrix

$$B = [T^t]_{\mathcal{B}^*}^{(\mathcal{B}')^*}$$

Then,

$$B_{i,j} = A_{j,i}$$

Proof. Write

$$\begin{aligned}
\mathcal{B} &= \{\alpha_1, \alpha_2, \dots, \alpha_n\} \\
\mathcal{B}' &= \{\beta_1, \beta_2, \dots, \beta_m\} \\
\mathcal{B}^* &= \{f_1, f_2, \dots, f_n\} \\
(\mathcal{B}')^* &= \{g_1, g_2, \dots, g_m\} \text{ so that} \\
f_i(\alpha_j) &= \delta_{i,j}, \quad \forall 1 \leq i, j \leq n, \text{ and} \\
g_i(\beta_j) &= \delta_{i,j} \quad \forall 1 \leq i, j \leq m
\end{aligned}$$

Furthermore, we have the expressions

$$\begin{aligned}
T(\alpha_i) &= \sum_{k=1}^m A_{k,i}\beta_k, \quad \forall 1 \leq i \leq n \\
T^t(g_j) &= \sum_{i=1}^m B_{i,j} f_i, \quad \forall 1 \leq j \leq m
\end{aligned}$$

But by definition,

$$\begin{aligned}
T^t(g_j)(\alpha_i) &= g_j(T(\alpha_i)) \\
&= g_j\left(\sum_{k=1}^m A_{k,i}\beta_k\right) \\
&= \sum_{k=1}^m A_{k,i} g_j(\beta_k) \\
&= A_{j,i}
\end{aligned}$$

But for the linear functional $f = T^t(g_j)$, we have the formula

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

by Theorem 5.5. Hence,

$$T^t(g_j) = \sum_{i=1}^m T^t(g)(\alpha_i) f_i = \sum_{i=1}^m A_{j,i} f_i$$

By the uniqueness of the expression

$$T^t(g_j) = \sum_{i=1}^m B_{i,j} f_i$$

we conclude that $B_{i,j} = A_{j,i}$ as required. □

Definition 7.4. Let A be an $m \times n$ matrix over a field F, then the *transpose* of A is the $n \times m$ matrix B whose $(i,j)^{th}$ entry is given by

$$B_{i,j} = A_{j,i}$$

Therefore, the above theorem says that, once we fix (coherent, ordered) bases for V, W, V^*, and W^*, then the matrix of T^t is the transpose of the matrix of T.

Definition 7.5. Let A be an $m \times n$ matrix over a field F.

(i) The *column space* of A is the subspace of F^m spanned by the n columns of A.

(ii) The *column rank* of A is the dimension of the column space of A.

Theorem 7.6. *Let A be an $m \times n$ matrix over a field F, then*

$$\textit{row rank}(A) = \textit{column rank}(A)$$

Proof. Define $T : F^n \to F^m$ by

$$T(X) := AX$$

Let $\mathcal{B}$ and $\mathcal{B}'$ denote the standard bases of F^n and F^m respectively, so that

$$[T]_{\mathcal{B}'}^{\mathcal{B}} = A$$

by Example 4.4. Then, by Theorem 7.3,

$$[T^t]_{\mathcal{B}^*}^{(\mathcal{B}')^*} = A^t$$

Now note that the columns of T are the images of T under $\mathcal{B}$. Hence,

$$\text{column rank}(A) = \text{rank}(T)$$

Similarly,

$$\text{row rank}(A) = \text{rank}(T^t)$$

The result now follows from Theorem 7.2. □

IV. Polynomials

1. Algebras

Definition 1.1. A *linear algebra over a field F* is a vector space $\mathcal{A}$ together with a multiplication map

$$\times : \mathcal{A} \times \mathcal{A} \to \mathcal{A}$$

denoted by

$$(\alpha, \beta) \mapsto \alpha\beta$$

which satisfies the following axioms:

(i) Multiplication is associative:
$$\alpha(\beta\gamma) = (\alpha\beta)\gamma$$

(ii) Multiplication distributes over addition
$$\alpha(\beta + \gamma) = \alpha\beta + \alpha\gamma \text{ and } (\alpha + \beta)\gamma = \alpha\gamma + \beta\gamma$$

(iii) For each scalar $c \in F$,
$$c(\alpha\beta) = (c\alpha)\beta = \alpha(c\beta)$$

If there is an element $1_{\mathcal{A}} \in \mathcal{A}$ such that

$$1_{\mathcal{A}}\alpha = \alpha = \alpha 1_{\mathcal{A}}$$

for all $\alpha \in \mathcal{A}$, then $1_{\mathcal{A}}$ is called the *identity of $\mathcal{A}$*, and $\mathcal{A}$ is said to be a *linear algebra with identity.* Furthermore, $\mathcal{A}$ is said to be *commutative* if

$$\alpha\beta = \beta\alpha$$

for all $\alpha, \beta \in \mathcal{A}$.

Example 1.2.

(i) Any field is an algebra over itself, which has an identity, and is commutative.

(ii) Let $\mathcal{A} = M_n(F)$ be the space of all $n \times n$ matrices over the field F. With multiplication given by matrix multiplication, $\mathcal{A}$ is a linear algebra. Furthermore, the identity matrix I_n is the identity of $\mathcal{A}$. If $n \geq 2$, then $\mathcal{A}$ is not commutative.

(iii) Let V be any vector space and $\mathcal{A} = L(V, V)$ be the space of all linear operators on V. With multiplication given by composition of operators, $\mathcal{A}$ is a linear algebra. Furthermore, the identity operator is the identity of $\mathcal{A}$. Once again, $\mathcal{A}$ is not commutative unless $\dim(V) = 1$.

(iv) Let $\mathcal{A} = C[0, 1]$ be the space of continuous F-valued functions on $[0, 1]$. With multiplication defined pointwise,

$$(f \cdot g)(x) := f(x)g(x)$$

Then $\mathcal{A}$ is a linear algebra. The constant function 1 is the identity of $\mathcal{A}$, and it is commutative.

(v) Let $\mathcal{A} = C_c(\mathbb{R})$ be the space of real-valued continuous functions on $\mathbb{R}$ which have compact support (A function $f : \mathbb{R} \to \mathbb{R}$ is said to have compact support if the set $\{x \in \mathbb{R} : f(x) \neq 0\}$ has compact closure). This is a linear algebra over $\mathbb{R}$ which is commutative, but does not have an identity (This requires a proof, which I will leave as an exercise).

We will now construct an important example. Fix a field F. Define

$$F^\infty := \{(f_0, f_1, f_2, \ldots, f_n, \ldots) : f_i \in F\}$$

be the set of all sequences from F. We now define operations on F^∞ as follows:

(i) Addition: Given two sequences $f = (f_i), g = (g_i)$, we write

$$f + g := (f_i + g_i)$$

(ii) Scalar multiplication: Given $f = (f_i) \in F^\infty$ and $c \in F$, define

$$c \cdot f := (cf_i)$$

(iii) Vector multiplication: This is the *Cauchy product.* Given $f = (f_i), g = (g_i) \in F^\infty$, we define the sequence $f \cdot g = (x_i) \in F^\infty$ by

$$x_n := \sum_{i=0}^{n} f_i g_{n-i} \tag{IV.1}$$

Thus,

$$fg = (f_0g_0, f_0g_1 + g_0f_1, f_0g_2 + f_1g_1 + g_0f_2, \ldots)$$

Now, one has to verify that F^∞ is, indeed, an algebra over F (See [Hoffman-Kunze, Page 118]). Furthermore, it is clear that,

$$fg = gf$$

so F^∞ is a commutative algebra. Furthermore, the element

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

plays the role of the identity of F^∞. Now define

$$x := (0, 1, 0, 0, \ldots)$$

Then note that

$$x^2 = (0, 0, 1, 0, 0, \ldots), x^3 = (0, 0, 0, 1, 0, \ldots), \ldots$$

In other words, if we set $x^0 = 1$, then for each integer $k \geq 0$, we have

$$(x^k)_i = \delta_{i,k}$$

Now, one can check that the set

$$\{1, x, x^2, x^3, \ldots\}$$

is an infinite linearly independent set in F^∞.

Definition 1.3. The algebra F^∞ is called the *algebra of formal power series over F*. An element $\overline{f} = (f_0, f_1, f_2, \ldots) \in F^\infty$ is written as a formal expression

$$f = \sum_{n=0}^{\infty} f_n x^n$$

Note that the above expression is only a formal expression - there is no series convergence involved, as there is no metric.

2. Algebra of Polynomials

Definition 2.1. Let $F[x]$ be the subspace of F^∞ spanned by the vectors $\{1, x, x^2, \ldots\}$. An element of $F[x]$ is called a *polynomial over F*

Remark 2.2.

(i) Any $f \in F[x]$ is of the form

$$f = f_0 + f_1 x + f_2 x^2 + \ldots + f_n x^n$$

(ii) If $f_n \neq 0$ and $f_k = 0$ for all $k \geq n$, then we say that f has *degree n*, denoted by $\deg(f)$.

(iii) If $f = 0$ is the zero polynomial, then we simply define $\deg(0) = 0$.

(iv) The scalars $f_0, f_1, \ldots, f_n$ are called the *coefficients* of f.

(v) If $f = cx^0$, then f is called a *scalar polynomial.*

(vi) If $f_n = 1$, then f is called a *monic polynomial.*

Theorem 2.3. *Let $f, g \in F[x]$ be non-zero polynomials. Then*

(i) fg is a non-zero polynomial.

(ii) $\deg(fg) = \deg(f) + \deg(g)$

(iii) If both f and g are monic, then fg is monic.

(iv) fg is a scalar polynomial if and only if both f and g are scalar polynomials.

(v) If $f + g \neq 0$, then

$$\deg(f+g) \leq \max\{deg(f), \deg(g)\}$$

Proof. Write

$$f = \sum_{i=0}^{n} f_i x^i \text{ and } g = \sum_{j=0}^{m} g_j x^j$$

with $f_n \neq 0$ and $g_m \neq 0$. Then, by Equation IV.1, we have

$$(fg)_k = \sum_{i=0}^{n} f_i g_{k-i}$$

Now note that, if $0 \leq i \leq n$, and $k - i > m$, then $g_{k-i} = 0$. Hence, In particular,

$$(fg)_k = 0 \text{ if } k - n > m$$

Hence, $\deg(fg) \leq n + m$. But

$$(fg)_{n+m} = f_n g_m \neq 0$$

so $\deg(fg) = n+m$. Thus proves (i), (ii), (iii) and (iv). We leave (v) as an exercise. □

Corollary 2.4. *For any field F, $F[x]$ is a commutative linear algebra with identity over F.*

Corollary 2.5. *Let $f, g, h \in F[x]$ such that $fg = fh$. If $f \neq 0$, then $g = h$.*

Proof. Note that $f(g - h) = 0$. Since $f \neq 0$, by part (i) of Theorem 2.3, we conclude that $(g - h) = 0$. □

Remark 2.6. If $f, g \in F[x]$ are expressed as

$$f = \sum_{i=0}^{m} f_i x^i \text{ and } g = \sum_{j=0}^{n} g_j x^j$$

Then

$$fg = \sum_{s=0}^{m+n} \left(\sum_{r=0}^{s} f_r g_{s-r} \right) x^s$$

In the case $f = cx^m$ and $g = dx^n$, we have

$$(cx^m)(dx^n) = (cd)x^{m+n}$$

Hence, by distributivity of addition in $F[x]$, we get

$$fg = \sum_{i=0}^{m} \sum_{j=0}^{n} f_i g_j x^{i+j}$$

Definition 2.7. Let $\mathcal{A}$ be a linear algebra with identity over a field F. We write $1 = 1_{\mathcal{A}}$, and for each $\alpha \in \mathcal{A}$, we write $\alpha^0 = 1$. Then, given a polynomial

$$f = \sum_{i=0}^{n} f_i x^i$$

in $F[x]$, and $\alpha \in \mathcal{A}$, we define $f(\alpha) \in \mathcal{A}$ by

$$f(\alpha) = \sum_{i=0}^{n} f_i \alpha^i$$

Example 2.8. Let $f \in \mathbb{C}[x]$ be the polynomial $f = 2 + x^2$.

(i) If $\mathcal{A} = \mathbb{C}$ and $\alpha = 2 \in \mathbb{C}$, then

$$f(\alpha) = 2^2 + 2 = 6$$

(ii) If $\mathcal{A} = \mathbb{C}$ and $\alpha = \frac{1+i}{1-i} \in \mathcal{A}$, then

$$f(\alpha) = 1$$

(iii) If $\mathcal{A} = M_2(\mathbb{C})$ is the algebra of 2×2 matrices over $\mathbb{C}$, and

$$B = \begin{pmatrix} 1 & 0 \\ -1 & 2 \end{pmatrix}$$

Then

$$f(B) = 2\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ -1 & 2 \end{pmatrix}^2 = \begin{pmatrix} 3 & 0 \\ -3 & 6 \end{pmatrix}$$

(iv) If $\mathcal{A} = L(V, V)$ where $V = \mathbb{C}^3$, and $T \in \mathcal{A}$ is the linear operator given by

$$T(x_1, x_2, x_3) = (i\sqrt{2}x_1, x_2, i\sqrt{2}x_3)$$

Then $f(T) \in \mathcal{A}$ is the operator

$$f(T)(x_1, x_2, x_3) = (0, 3x_2, 0)$$

(v) If $\mathcal{A} = \mathbb{C}[x]$ and $g = x^4 + 3i$, then

$$f(g) = -7 + 6ix^4 + x^8$$

Theorem 2.9. *Let $\mathcal{A}$ be a linear algebra with identity over F, let $f, g \in F[x]$ be two fixed polynomials, let $\alpha \in \mathcal{A}$, and $c \in F$ be a scalar. Then*

(i) $(cf + g)(\alpha) = cf(\alpha) + g(\alpha)$

(ii) $(fg)(\alpha) = f(\alpha)g(\alpha)$

Proof. We prove (ii) since (i) is easy. Write

$$f = \sum_{i=0}^{n} f_i x^i \text{ and } g = \sum_{j=0}^{m} g_j x^j$$

So that

$$fg = \sum_{i=0}^{n}\sum_{j=0}^{m} f_i g_j x^{i+j}$$

Hence,

$$(fg)(\alpha) = \sum_{i=0}^{n}\sum_{j=0}^{m} f_i g_j \alpha^{i+j} = \left(\sum_{i=0}^{n} f_i \alpha^i\right)\left(\sum_{j=0}^{m} g_j \alpha^j\right) = f(\alpha)g(\alpha)$$

□

(End of Week 5)

3. Lagrange Interpolation

Theorem 3.1. *Let F be a field and $t_0, t_2, \ldots, t_n$ be $(n+1)$ distint elements of F. Let V be the subspace of $F[x]$ consisting of polynomials of degree $\leq n$. Define $L_i : V \to F$ by*

$$L_i(f) := f(t_i)$$

Then $\mathcal{S} := \{L_0, L_2, \ldots, L_n\}$ is a basis for V^*.

Proof. It suffices to show that $\mathcal{S}$ is the dual basis to a basis $\mathcal{B}$ of V. For $0 \leq i \leq n$, define $P_i \in V$ by

$$P_i = \prod_{j \neq i} \frac{(x - t_i)}{t_i - t_j} = \frac{(x - t_0)(x - t_1)\ldots(x - t_{i-1})(x - t_{i+1})\ldots(x - t_n)}{(t_i - t_0)(t_i - t_1)\ldots(t_i - t_{i-1})(t_i - t_{i+1})\ldots(t_i - t_n)}$$

Then we claim that $\mathcal{B} = \{P_0, P_1, \ldots, P_n\}$ is a basis for V. Note that $\{1, x, x^2, \ldots, x^n\}$ is a basis for V, so $\dim(V) = n+1$. Hence, it suffices to show that $\mathcal{B}$ is linearly independent. So suppose $c_i \in F$ are such that

$$\sum_{i=0}^{n} c_i P_i = 0$$

where the right hand side denotes the zero polynomial. Now applying L_j to this expression, and note that

$$L_j(P_i) = \delta_{i,j}$$

Hence,

$$c_j = L_j\left(\sum_{i=0}^{n} c_i P_i\right) = 0$$

This is true for all $0 \leq j \leq n$, so $\mathcal{B}$ is a basis for V, and so $\mathcal{S}$ is a basis for V^*. □

Remark 3.2.

(i) By Theorem III.5.5, any $f \in V$ may be expressed in the form

$$f = \sum_{i=0}^{n} f(t_i) P_i$$

This is called the *Lagrange Interpolation formula*

(ii) If $f = x^j$, then we obtain

$$f = \sum_{i=0}^{n} t_i^j P_i$$

Since the collection $\{1, x, x^2, \ldots, x^n\}$ forms a basis for V, it follows that the matrix

$$\begin{pmatrix} 1 & t_0 & t_0^2 & \ldots & t_0^n \\ 1 & t_1 & t_1^2 & \ldots & t_1^n \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & t_n & t_n^2 & \ldots & t_n^n \end{pmatrix}$$

is invertible by Theorem II.4.5. This is called a *Vandermonde matrix.*

Definition 3.3. Let V be as above and let W be the space of all polynomial functions on F of degree $\leq n$. For each $f \in V$, define $\tilde{f} \in W$ by

$$\tilde{f}(t) := f(t)$$

In other words, we send each *formal* polynomial to the corresponding polynomial *function.*

Theorem 3.4. *The map $V \to W$ given by $f \mapsto \tilde{f}$ defined above is an isomorphism of vector spaces.*

Proof. Let $T(f) := \tilde{f}$, then T is linear by Theorem 2.9. Since $\dim(V) = \dim(W) = n+1$, it suffices to show that T is injective. If $\tilde{f} = 0$, then $f(t) = 0$ for all $t \in F$. In particular, if we choose $(n+1)$ distinct elements $\{t_0, t_1, \ldots, t_n\} \subset F$, then

$$f(t_i) = 0 \quad \forall 0 \leq i \leq n$$

By Theorem 3.1, it follows that $f = 0$. Hence, $\ker(T) = \{0\}$ so T is injective. □

Definition 3.5. Let F be a field and $\mathcal{A}$ and $\widetilde{\mathcal{A}}$ be two linear algebras over F. A map $T : \mathcal{A} \to \widetilde{\mathcal{A}}$ is said to be an *isomorphism* if

(i) T is bijective.

(ii) T is linear.

(iii) $T(\alpha\beta) = T(\alpha)T(\beta)$ for all $\alpha, \beta \in \mathcal{A}$

If such an isomorphism exists, then we say that $\mathcal{A}$ and $\widetilde{\mathcal{A}}$ are *isomorphic.*

Theorem 3.6. *Let $\mathcal{A} = F[x]$ and $\widetilde{\mathcal{A}}$ denote the algebra of all polynomial functions on F, then the map*

$$f \mapsto \tilde{f}$$

induces an isomorphism $\mathcal{A} \to \widetilde{\mathcal{A}}$ of algebras.

Proof. Once again, by Theorem 2.9, T is a morphism of algebras. Also, T is injective as in Theorem 3.4. Since T is clearly surjective, T is an isomorphism. □

Example 3.7. Let V be an n-dimensional vector space over F and $\mathcal{B}$ be an ordered basis for V. By Theorem III.4.2, we have a linear isomorphism

$$\Theta : L(V, V) \to F^{n\times n} \text{ given by } T \mapsto [T]_{\mathcal{B}}$$

Furthermore, by Theorem III.4.5, Θ is multiplicative, and hence an isomorphism of algebras. Now if

$$f = \sum_{i=0}^{n} c_i x^i$$

is a polynomial in $F[x]$, and $T \in L(V, V)$, then we may associate two polynomials to it:

$$f(T) = \sum_{i=0}^{n} c_i T^i \text{ and } f([T]_{\mathcal{B}}) = \sum_{i=0}^{n} c_i [T]_{\mathcal{B}}^i$$

where $f(T) \in L(V, V)$ and $f([T]_{\mathcal{B}}) \in F^{n\times n}$. Since Θ is linear and multiplicative, it follows that

$$[f(T)]_{\mathcal{B}} = f([T]_{\mathcal{B}})$$

4. Polynomial Ideals

Lemma 4.1. *Let $f, d \in F[x]$ such that $\deg(d) \leq \deg(f)$. Then there exists $g \in F[x]$ such that either*

$$f = dg \textit{ or } \deg(f - dg) < \deg(f)$$

Proof. Write

$$f = a_m x^m + \sum_{i=0}^{m-1} a_i x^i$$

$$d = b_n x^n + \sum_{j=0}^{n-1} b_j x^j$$

with $a_m \neq 0$ and $b_n \neq 0$. Since $m \geq n$, take

$$g = \frac{a_m}{b_n} x^{m-n}$$

Then this g works. □

Theorem 4.2 (Euclidean Division). *Let $f, d \in F[x]$ with $d \neq 0$. Then there exist polynomials $q, r \in F[x]$ such that*

(i) $f = dq + r$

(ii) Either $r = 0$ *or* $\deg(r) < \deg(d)$

The polynomials q, r satisfying (i) and (ii) are unique.

Proof.

(i) Uniqueness: Suppose q_1, r_1 are another pair of polynomials satisfying (i) and (ii) in addition to q, r. Then

$$d(q_1 - q) = r - r_1$$

Furthermore, if $r - r_1 \neq 0$, then by Theorem 2.3,

$$\deg(r - r_1) \leq \max\{\deg(r), \deg(r_1)\} < \deg(d)$$

But

$$\deg(d(q_1 - q)) = \deg(d) + \deg(q - q_1) \geq \deg(d)$$

This is impossible, so $r = r_1$, and so $q = q_1$ as well.

(ii) Existence:

(i) If $\deg(f) < \deg(d)$, we may take $q = 0$ and $r = f$.

(ii) If $f = 0$, then we take $q = 0 = r$.

(iii) So suppose $f \neq 0$ and

$$\deg(d) \leq \deg(f)$$

We now induct on $\deg(f)$.

- If $\deg(f) = 0$, then $f = c$ is a constant, so that d is also a constant. Since $d \neq 0$, we take

$$q = \frac{c}{d} \in F$$

and $r = 0$.

- Now suppose $\deg(f) > 0$ and that the theorem is true for any polynomail h such that $\deg(h) < \deg(f)$. Since $\deg(d) \leq \deg(f)$, by the previous lemma, we may choose $g \in F[x]$ such that either

  $$f = dg \text{ or } \deg(f - dg) < \deg(f)$$

  If $f = dg$, then we take $q = g$ and $r = 0$ and we are done. If not, then take

  $$h := f - dg$$

  By induction hypothesis, there exists $q_2, r_2 \in F[x]$ such that

  $$h = dq_2 + r_2$$

  with either $r_2 = 0$ or $\deg(r_2) < \deg(h)$. Hence,

  $$f = d(g + q_2) + r_2$$

  with the required conditions satisfied.

□

Definition 4.3. Let $d \in F[x]$ be non-zero, and $f \in F[x]$ be any polynomial. Write

$$f = dq + r \text{ with } r = 0 \text{ or } \deg(r) < \deg(d)$$

(i) The element q is called the *quotient* and r is called the *remainder*.

(ii) If $r = 0$, then we say that d *divides* f, or that f is *divisible* by d. In symbols, we write $d \mid f$. If this happens, we also write $q = f/d$.

(iii) If $r \neq 0$, then we say that d *does not divide* f and we write $d \nmid f$.

Corollary 4.4. *Let $f \in F[x]$ and $c \in F$. Then $(x - c) \mid f$ if and only if $f(c) = 0$.*

If this happens, we say that c is a *root* of f (or a *zero* of f).

Proof. Take $d := (x - c)$, then $\deg(d) = 1$, so if $f = qd + r$, then either $r = 0$ or $\deg(r) = 0$. So write $r \in F$, then

$$f = q(x - c) + r$$

Evaluating at c by Theorem 2.9, we see that

$$f(c) = 0 + r$$

Hence,

$$f = q(x - c) + f(c)$$

Thus, $(x - c) \mid f$ if and only if $f(c) = 0$. □

Corollary 4.5. *Let $f \in F[x]$ is non-zero, then f has atmost $\deg(f)$ roots in F.*

Proof. We induct on $\deg(f)$.

- If $\deg(f) = 0$, then $f \in F$ is non-zero, so f has no roots.
- Suppose $\deg(f) > 0$, and assume that the theorem is true for any polynomial g with $\deg(g) < \deg(f)$. If f has no roots in F, then we are done. Suppose f has a root at $c \in F$, then by Corollary 4.4, write

  $$f = q(x-c)$$

  Note that $\deg(f) = \deg(q) + \deg(x-c)$, so

  $$\deg(q) < \deg(f)$$

  By induction hypothesis, q has atmost $\deg(q)$ roots. Furthermore, for any $b \in F$,

  $$f(b) = q(b)(b-c)$$

  So if $b \in F$ is a root of f and $b \neq c$, then it must follow that b is a root of f. Hence,

  $$\{\text{Roots of } f\} = \{c\} \cup \{\text{Roots of } q\}$$

  Thus,

  $$|\{\text{Roots of } f\}| \leq 1 + |\{\text{Roots of } q\}| \leq 1 + \deg(q) \leq 1 + \deg(f) - 1 = \deg(f)$$

□

Definition 4.6. Let $f \in F[x]$ be the polynomial

$$f = \sum_{i=0}^{n} c_i x^i$$

Then the *derivative* of f is the polynomial

$$Df = \sum_{i=1}^{n} i c_i x^{i-1}$$

This defines a linear operator $D : F[x] \to F[x]$. The higher derivatives of f are denoted by $D^2 f, D^3 f, \ldots$.

Theorem 4.7 (Taylor's Formula). *Let $f \in F[x]$ be a polynomial with $\deg(f) \leq n$, then*

$$f = \sum_{k=0}^{n} \frac{(D^k f)(c)}{k!}(x-c)^k$$

Proof. By the binomial theorem, we have

$$\begin{aligned} x^m &= [c + (x-c)]^m \\ &= \sum_{k=0}^{m} \binom{m}{k} c^{m-k}(x-c)^k \end{aligned}$$

Hence, Taylor's formula holds when $f = x^m$. Now by linearity, if

$$f = \sum_{m=0}^{n} a_m x^m$$

then by linearity of D^k and the evaluation map (Theorem 2.9), we have

$$D^k f(c) = \sum_{m=0}^{n} a_m (D^k x^m)(c)$$

Hence,

$$\begin{aligned} \sum_{k=0}^{n} \frac{D^k f(c)}{k!}(x-c)^k &= \sum_{k=0}^{n}\sum_{m=0}^{n} a_m \frac{(D^k x^m)(c)}{k!}(x-c)^k \\ &= \sum_{m=0}^{n} a_m \left(\sum_{k=0}^{m} \frac{D^k f(c)}{k!}(x-c)^k \right) \\ &= \sum_{m=0}^{n} a_m x^m \\ &= f \end{aligned}$$

□

Definition 4.8. Let $f \in F[x]$ and $c \in F$, then we say that c *is a root of* f *of multiplicity* r if

(i) $(x-c)^r \mid f$

(ii) $(x-c)^{r+1} \nmid f$

Lemma 4.9. *Let $S = \{f_1, f_2, \ldots, f_n\} \subset F[x]$ be a set of non-zero polynomials such that no two elements of S have the same degree. Then S is a linearly independent set.*

Proof. We may enumerate S so that, if $k_i := \deg(f_i)$, then

$$k_1 < k_2 < \ldots < k_n$$

We proceed by induction on $n := |S|$. If $n = 1$, then the theorem is true because $f_1 \neq 0$. So suppose the theorem is true for any set T as above such that $|T| \leq n-1$. Now suppose $c_i \in F$ such that

$$\sum_{i=1}^{n} c_i f_i = 0$$

Let $m := k_n - 1$, then observe that

$$D^m f_i = 0 \quad \forall 1 \leq i \leq n-1$$

and that $D^m f_n \neq 0$. Applying D^m to the above equation, we see that

$$c_n D^m f_n = 0$$

Since $D^m f_n \neq 0$, it follows that $c_n = 0$ (by Corollary 2.5). Hence,

$$\sum_{i=1}^{n-1} c_i f_i = 0$$

By induction, it follows that $c_i = 0$ for all $1 \leq i \leq n$. □

Proposition 4.10. *Let V be the vector subspace of $F[x]$ of all polynomials of degree $\leq n$. For any $c \in F$, the set*

$$\{1, (x-c), (x-c)^2, \ldots, (x-c)^n\}$$

forms a basis for V.

Proof. The set is linearly independent by Lemma 4.9 and spans V by Theorem 4.7. □

Theorem 4.11. *Let $f \in F[x]$ be a polynomial with* $\deg(f) \leq n$*, and $c \in F$. Then c is a root of f of multiplicity $r \in \mathbb{N}$ if and only if*

(i) $D^k f(c) = 0$ *for all* $0 \leq k \leq r-1$

(ii) $D^r f(c) \neq 0$.

Proof.

(i) Suppose c is a root of f of multiplicity r, then we may write

$$f = (x-c)^r g$$

since $(x-c)^r \mid f$. Furthermore, if $g(c) = 0$, then $(x-c) \mid g$ by Corollary 4.4, so that would imply that

$$(x-c)^{r+1} \mid f$$

This is not possible, so $g(c) \neq 0$. Since $\deg(g) \leq n - r$, by applying Taylor's formula to g, we see that

$$\begin{aligned} f &= (x-c)^r \left[\sum_{m=0}^{n-r} \frac{D^m g(c)}{m!} (x-c)^m \right] \\ &= \sum_{m=0}^{n-r} \frac{D^m g(c)}{m!} (x-c)^{m+r} \end{aligned}$$

By Taylor's formula,

$$f = \sum_{k=0}^{n} \frac{D^k f(c)}{k!}(x-c)^k$$

The set $\{1, (x-c), (x-c)^2, \ldots, (x-c)^n\}$ forms a basis for the space of polynomials of degree $\leq n$ by Proposition 4.10, so there is only one way to express f as a linear combination of these polynomials. Hence,

$$\frac{D^k f(c)}{k!} = \begin{cases} 0 & : 0 \leq k \leq r-1 \\ \frac{D^{k-r} g(c)}{(k-r)!} & : r \leq k \leq n \end{cases}$$

Hence, $D^k f(c) = 0$ for all $0 \leq k \leq r-1$ and

$$D^r f(c) = g(c) r! \neq 0$$

since $g(c) \neq 0$

(ii) Conversely, suppose conditions (i) and (ii) are satisfied, then Taylor's formula gives

$$f = \sum_{k=r}^{n} \frac{D^k f(c)}{k!}(x-c)^k = (x-c)^r g$$

where

$$g = \sum_{m=0}^{n-r} \frac{D^{m+r} f(c)}{(m+r)!}(x-c)^m$$

Thus, $(x-c)^r \mid f$. Furthermore,

$$g(c) = \frac{D^r f(c)}{r!} \neq 0$$

so that $(x-c)^{r+1} \nmid f$, so that c is a root of f of multiplicity r.

□

Definition 4.12. A subset $M \subset F[x]$ is called an *ideal* of $F[x]$ if M is a subspace of $F[x]$ such that, for any $f \in M$ and $g \in F[x]$, the product fg lies in M.

Example 4.13.

(i) If $d \in F[x]$ is a polynomial, set

$$M := \{dg : g \in F[x]\}$$

Then M is an ideal of $F[x]$. This is called the *principal ideal of* $F[x]$ generated by d. We denote this ideal by

$$M = dF[x]$$

Note that the generator d may not be unique (See below)

(ii) If $d_1, d_2, \ldots, d_n \in F[x]$, the define

$$M = \{d_1g_1 + d_2g_2 + \ldots + d_ng_n : g_i \in F[x]\}$$

Then M is an ideal of $F[x]$, called the *ideal generated by* $d_1, d_2, \ldots, d_n$.

Theorem 4.14. *If $M \subset F[x]$ is a non-zero ideal of $F[x]$, then there is a unique monic polynomial $d \in F[x]$ such that M is the principal ideal generated by d.*

Proof.

(i) Existence: Since M is non-zero, we may define

$$S = \{\deg(g) : g \in M, g \neq 0\}$$

Then S is a non-empty subset of $\mathbb{N} \cup \{0\}$, so it contains a minimal element. Hence, there exists $d \in M$ such that

$$\deg(d) \leq \deg(g) \quad \forall g \in M$$

Since M is a subspace, we may multiply d by a scalar (if necessary) to ensure that d is monic. Set

$$M' := dF[x]$$

Since M is an ideal, $M' \subset M$. To show the reverse containment, let $f \in M$. By Euclidean division Theorem 4.2, we may write

$$f = dq + r$$

where $r = 0$ or $\deg(r) < \deg(d)$. Since $r = f - dq$, we conclude that $r \in M$. Since

$$\deg(d) \leq \deg(g) \quad \forall g \in M$$

it must happen that $r = 0$. Hence, $f = dq \in M'$. Thus, $M = M'$ as required.

(ii) Uniqueness: If $d_1, d_2 \in M$ are two monic polynomials such that

$$M = d_1F[x] = d_2F[x]$$

Then $d_1 \in M$, so there exists $g \in F[x]$ such that $d_1 = d_2g$. Thus,

$$\deg(d_1) \geq \deg(d_2)$$

By symmetry, it follows that $\deg(d_2) = \deg(d_1)$, so that $g \in F$. Since both d_1 and d_2 are monic, we conclude that

$$g = 1$$

so that $d_1 = d_2$ as required.

□

Corollary 4.15. *Let $p_1, p_2, \ldots, p_n \in F[x]$ where not all the p_i are zero. Then, there exists a unique monic polynomial $d \in F[x]$ such that*

(i) $d \mid p_i$ for all $1 \leq i \leq n$.

(ii) If $f \in F[x]$ is any polynomial such that $f \mid p_i$ for all $1 \leq i \leq n$, then $f \mid d$.

Proof.

(i) Existence: Let M be the ideal generated by $p_1, p_2, \ldots, p_n$, and let d be its unique monic generator (by Theorem 4.14). Then, for each $1 \leq i \leq n$, we have $p_i \in M = dF[x]$, so

$$d \mid p_i$$

Furthermore, if $f \mid p_i$ for all $1 \leq i \leq n$, then there exist $g_i \in F[x]$ such that

$$p_i = fg_i$$

Since d is in the ideal M, there exist $f_i \in F[x]$ such that

$$d = \sum_{i=1}^{n} f_i p_i = \sum_{i=1}^{n} f_i f g_i = \left(\sum_{i=1}^{n} f_i g_i \right) f$$

So that $d \mid f$.

(ii) Uniqueness: If $d_1, d_2 \in F[x]$ are two polynomials satisfying both (i) and (ii), then $d_1 \mid p_i$ for all $1 \leq i \leq n$ implies that $d_1 \mid d_2$. By symmetry, we have $d_2 \mid d_1$. This implies (Why?) that

$$d_1 = cd_2$$

for some constant $c \in F$. Since both d_1 and d_2 are monic, we conclude that $d_1 = d_2$.

□

Definition 4.16. Let $p_1, p_2, \ldots, p_n \in F[x]$ be polynomials (not all zero).

(i) The monic polynomial $d \in F[x]$ satisfying the conditions of Corollary 4.15 is called the *greatest common divisor (gcd)* of $p_1, p_2, \ldots, p_n$. If this happens, we write

$$d = (p_1, p_2, \ldots, p_n)$$

(ii) We say that $p_1, p_2, \ldots, p_n$ are *relatively prime* if $(p_1, p_2, \ldots, p_n) = 1$

Example 4.17.

(i) Let $p_1 = (x+2)$ and $p_2 = (x^2 + 8x + 16) \in \mathbb{R}[x]$, and let M be the ideal generated by $\{p_1, p_2\}$. Then

$$(x^2 + 8x + 16) = (x+6)(x+2) + 4$$

Hence, $4 \in M$, so M contains scalar polynomials. Therefore, $1 \in M$, whence

$$(p_1, p_2) = 1$$

(ii) If $p_1, p_2, \ldots, p_n \in F[x]$ are relatively prime, then the ideal M generated by them is all of $F[x]$. Hence, there exist $f_1, f_2, \ldots, f_n \in F[x]$ such that

$$\sum_{i=1}^{n} f_i p_i = 1$$

5. Prime Factorization of a Polynomial

Definition 5.1. Let F be a field and $f \in F[x]$. f is said to be

(i) *reducible* if there exist two polynomials $g, h \in F[x]$ such that $\deg(g), \deg(h) \geq 1$ and

$$f = gh$$

(ii) *irreducible* or *prime* if f is not a scalar, and it is not reducible.

Example 5.2.

(i) If $f \in F[x]$ has degree 1, then f is prime because $\deg(gh) = \deg(g) + \deg(h)$.

(ii) $f = x^2 + 1$ is reducible in $\mathbb{C}[x]$ because $f = (x+i)(x-i)$

(iii) $f = x^2 + 1$ is irreducible in $\mathbb{R}[x]$ because if $f = gh$ with $\deg(g), \deg(h) \geq 1$, then since

$$\deg(g) + \deg(h) = \deg(gh) = \deg(f) = 2$$

it follows that $\deg(g) = \deg(h) = 1$, so we may write

$$g = ax + b, h = cx + d$$

Multiplying, we get equations

$$\begin{aligned} ac &= 1 \\ ad + bc &= 0 \\ bd &= 1 \end{aligned}$$

This implies (Check!) that $a^2 + b^2 = 0$ which is not possible for $a, b \in \mathbb{R}$ unless $a = b = 0$. Thus, f is irreducible.

Remark 5.3.

(i) If $p \in F[x]$ is prime and $d \mid p$, then either $d = cp$ for some constant $c \in F$ or $d \in F$.

(ii) If $p, q \in F[x]$ are both primes and $p \mid q$, then $p = cq$ (because a prime polynomial cannot be a scalar).

Theorem 5.4 (Euclid's Lemma). *Let $p, f, g \in F[x]$ where p is prime. If $p \mid (fg)$, then either $p \mid f$ or $p \mid g$*

Proof. Assume without loss of generality that p is monic. Let $d = (f, p)$. Since $d \mid p$ is monic, and p is irreducible, we must have that either

$$d = p \text{ or } d = 1$$

If $d = p$, then $p \mid f$, and we are done.

If not, then $d = 1$, so by Example 4.17, there exist $f_0, p_0 \in F[x]$ such that

$$f_0 f + p_0 p = 1$$

So that

$$g = f_0 f g + p_0 p g$$

Now observe that $p \mid fg$ and $p \mid p$, so

$$p \mid g$$

□

The proof of the next corollary follows by induction (Check!).

Corollary 5.5. *Let $p, f_1, f_2, \ldots, f_n \in F[x]$ where p is prime. If*

$$p \mid (f_1 f_2 \ldots f_n)$$

Then there exists $1 \leq i \leq n$ such that $p \mid f_i$

Theorem 5.6 (Prime Factorization)**.** *Let F be a field and $f \in F[x]$ be a non-scalar monic polynomial. Then, f can be expressed as a product of finitely many monic prime polynomials. Furthermore, this expression is unique (upto rearrangement).*

Proof.

(i) Existence: We induct on $\deg(f)$. Note that by assumption, $\deg(f) \geq 1$.

- If $\deg(f) = 1$, then f is prime by Example 5.2.
- Now assume $\deg(f) \geq 2$ and that every polynomial h with $\deg(h) < \deg(f)$ has a prime factorization. Then, if f is itself prime, there is nothing to prove. So suppose f is not prime. Then, by definition, there exist $g, h \in F[x]$ of degree ≥ 1 such that

  $$f = gh$$

  Since f is monic, we may arrange it so that g and h are both monic as well. But then $\deg(g), \deg(h) < \deg(f)$, so by induction hypothesis, both g and h can be expressed a product of primes. Thus, f can also be expressed as a product of primes.

(ii) Uniqueness: Suppose that

$$f = p_1 p_2 \ldots p_n = q_1 q_2 \ldots q_m$$

where $p_i, q_j \in F[x]$ are monic primes. We wish to show that $n = m$ and that (upto reordering), $p_i = q_i$ for all $1 \leq i \leq n$. We induct on n.

- If $n = 1$, then
$$f = p_1 = q_1q_2\ldots q_m$$
Since p_1 is prime, there exists $1 \leq j \leq m$ such that $p_1 \mid q_j$. But both p_1 and q_j are monic primes, so $p_1 = q_j$ by Remark 5.3. Assume WLOG that $j = 1$, so by Corollary 2.5, it follows that
$$q_2q_3\ldots q_m = 1$$
Since each q_j is prime (and so has degree ≥ 2), this cannot happen. Hence, $m = 1$ must hold.

- Now suppose $n \geq 2$, and we assume that the uniqueness of prime factorization holds for any monic polynomial h that is expressed as a product of $(n-1)$ primes. Then, we have
$$p_1 \mid q_1q_2\ldots q_m$$
So by Corollary 5.5, there exists $1 \leq j \leq m$ such that $p_1 \mid q_j$. Assume WLOG that $j = 1$, then (as before),
$$p_1 = q_1$$
must hold. Hence, by Corollary 2.5, we have
$$p_2p_3\ldots p_n = q_2q_3\ldots q_m$$
By induction hypothesis, we have $(n-1) = (m-1)$ and $p_j = q_j$ for all $2 \leq j \leq m$ (upto rearrangement). Thus,
$$n = m \text{ and } p_i = q_i \quad \forall 1 \leq i \leq n$$
as required.

□

Definition 5.7. Let $f \in F[x]$ be a non-scalar monic polynomial, and write
$$f - q_1q_2\ldots q_m$$
where each q_i is prime. Combining like terms, we get
$$f = p_1^{n_1}p_2^{n_2}\ldots p_r^{n_r}$$
where the p_i are *distinct* primes. This is called the *primary decomposition* of f (and it is also unique).

Remark 5.8. Suppose that $f, g \in F[x]$ are monic polynomials with primary decomposition
$$f = p_1^{n_1}p_2^{n_2}\ldots p_r^{n_r} \text{ and } g = p_1^{m_1}p_2^{m_2}\ldots p_r^{m_r}$$
(where some of the n_i, m_j may also be zero). Then (Check!) that the g.c.d. of f and g is given by
$$(f,g) = p_1^{\min\{n_1,m_1\}}p_2^{\min\{n_2,m_2\}}\ldots p_r^{\min\{n_r,m_r\}}$$

The proof of the next theorem is now an easy corollary of this remark.

Theorem 5.9. *Let $f \in F[x]$ be a non-scalar polynomial with primary decomposition*

$$f = p_1^{n_1} p_2^{n_2} \dots p_r^{n_r}$$

Write

$$f_j := f/p_j^{n_j} = \prod_{i \neq j} p_i^{n_i}$$

then the polynomials $f_1, f_2, \dots, f_r$ are relatively prime

Theorem 5.10. *A polynomial $f \in F[x]$ is a product of distinct irreducible polynomials over F if and only if $(f, Df) = 1$*

Proof.

(i) Suppose that f is a product of distinct irreducible polynomials. So the prime decomposition of f has the form

$$f = p_1 p_2 \dots p_r$$

where the p_j are mutually distinct primes, and let $d = (f, Df)$. If $d \neq 1$, then there is a monic prime q such that

$$q \mid d$$

Hence, $q \mid f$, so by Corollary 5.5, there exists $1 \leq i \leq m$ such that $q \mid p_i$. Since q and p_i are both monic primes, it follows that

$$q = p_i$$

Hence, we assume WLOG that $i = 1$ so that $p_1 \mid Df$. Now we write

$$f_j = f/p_j$$

so by Leibnitz' rule, we have

$$Df = D(p_1)f_1 + D(p_2)f_2 + \ldots + D(p_r)f_r$$

Now observe that $p_1 \mid f_j$ for all $j \geq 2$. Since $p_1 \mid Df$, it follows that

$$p_1 \mid D(p_1)f_1$$

Now $D(p_1)$ is a polynomial whose degree is $< \deg(p_1)$. So $p_1 \nmid D(p_1)$. So by Euclid's Lemma Theorem 5.4, it follows that

$$p_1 \mid f_1$$

But $f_1 = p_2 p_3 \dots p_n$, so by Corollary 5.5, it follows that there exists $2 \leq j \leq n$ such that

$$p_1 \mid p_j$$

Since both p_1 and p_j are monic primes, it follows that $p_1 = p_j$. This contradicts the fact that the p_j are all mutually distinct.

(ii) Conversely, suppose f has a prime decomposition with at least one prime occuring multiple times. Then, we may write

$$f = p^2 g$$

for some prime $p \in F[x]$ and some other $g \in F[x]$. Then, by Leibnitz' rule,

$$Df = 2pD(p)g + p^2 D(g)$$

Hence, $p \mid D(f)$, so $(f, Df) \neq 1$.

□

Recall that, if $p \in F[x]$ has degree 1, then p is irreducible.

Definition 5.11. A field F is said to be *algebraically closed* if every irreducible polynomial over F has degree 1.

Example 5.12.

(i) If $F = \mathbb{R}$, then $x^2 + 1 \in F[x]$ is irreducible by Example 5.2. Hence, $\mathbb{R}$ is not algebraically closed.

(ii) By the Fundamental Theorem of Algebra, $\mathbb{C}$ is algebraically closed.

Remark 5.13. If F is algebraically closed, then any non-zero $f \in F[x]$ can be expressed in form

$$f = c(x - a_1)(x - a_2)\ldots(x - a_n)$$

for some scalars $c, a_1, a_2, \ldots, a_n \in F$.

(End of Week 6)

V. Determinants

1. Commutative Rings

Definition 1.1. A *ring* is a set K together with two operations $\times : K \times K \to K$ and $+ : K \times K \to K$ satisfying the following conditions:

(i) $(K, +)$ is a commutative group.

(ii) $(xy)z = x(yz)$ for all $x, y, z \in K$

(iii) $x(y+z) = xy + xz$ and $(y+z)x = yx + zx$ for all $x, y, z \in K$

Furthermore, we say that K is *commutative* if $xy = yx$ for all $x, y \in K$. An element $1 \in K$ is said to be a *unit* if $1x = x = x1$ for all $x \in K$. If such an element exists, then K is said to be a *ring with identity.*

Example 1.2.

(i) Every field is a commutative ring with identity.

(ii) $\mathbb{Z}$ is a ring that is not a field.

Note: The important distinction between a commutative ring with identity and a field is that, if F is a field and $x \in F$ is non-zero, then there exists $y \in F$ such that $xy = yx = 1$. In a ring, it is not necessary that every non-zero element has a multiplicative inverse.

(iii) If F is a field, then $F[x]$ is a ring.

Definition 1.3. Let K be a commutative ring with identity. An $m \times n$ *matrix* over K is a function A from the set $\{(i,j) : 1 \leq i \leq m, 1 \leq j \leq n\}$ to K. We write $K^{m\times n}$ for the set of all $m \times n$ matrices over K.

As usual, we represent such a function the same way we do for matrices over fields. Given two matrices $A, B \in K^{m\times n}$, we define addition and mutliplication in the usual way. The basic algebraic identities still hold. For instance,

$$A(B+C) = AB + AC, (AB)C = A(BC), \ldots$$

Many of our earlier results about matrices over a field also hold for matrices over a ring, except those that may involve 'dividing by elements of K'.

2. Determinant Functions

Standing Assumption: Throughout the remainder of the chapter, K will denote a commutative ring with identity.
Given an $n \times n$ matrix A over a ring K, we think of A as a tuple of rows

$$A \leftrightarrow (\alpha_1, \alpha_2, \ldots, \alpha_n)$$

We wish to define the determinant of A axiomatically so that the final formula we arrive at will not be mysterious. One of the advantages of this approach is that it is more conceptual and less computational.

Definition 2.1. Given a function $D : K^{n\times n} \to K$, we say that D is *n-linear* if, given any matrix $A = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ as above, and each $1 \leq j \leq n$, the function $D^j : K^n \to K$ defined by

$$D^j(\cdot) := D(\alpha_1, \alpha_2, \ldots, \alpha_{j-1}, \cdot, \alpha_{j+1}, \ldots, \alpha_n)$$

is linear. (ie. If $\beta_1, \beta_2 \in K^n$ and $c \in K$, then $D^j(c\beta_1 + \beta_2) = cD(\beta_1) + D(\beta_2)$)

Example 2.2.

(i) Fix integers $k_1, k_2, \ldots, k_n$ such that $1 \leq k_i \leq n$ and fix $a \in K$. Define $D : K^{n\times n} \to K$ by

$$D(A) = aA(1, k_1)A(2, k_2) \ldots A(n, k_n)$$

Then, for any fixed $1 \leq j \leq n$, the map $D^j : K^n \to K$ has the form

$$D^j(\beta) = cA(j, k_j) = c\beta$$

Hence, each D^j is linear, so D is n-linear.

(ii) As a special case of the previous example, the map $D : K^{n\times n} \to K$ by

$$D(A) = A(1, 1)A(2, 2) \ldots A(n, n)$$

is n-linear. (ie. D maps a matrix to the product of its diagonal entries).

(iii) Let $n = 2$ and $D : K^{2\times 2} \to K$ be any n-linear function. Write $\{\epsilon_1, \epsilon_2\}$ denote the rows of the 2×2 identity matrix. For $A \in K^{2\times 2}$, we have

$$\begin{aligned} D(A) &= D(A_{1,1}\epsilon_1 + A_{1,2}\epsilon_2, A_{2,1}\epsilon_1 + A_{2,2}\epsilon_2) \\ &= A_{1,1}D(\epsilon_1, A_{2,1}\epsilon_1 + A_{2,2}\epsilon_2) + A_{1,2}D(\epsilon_2, A_{2,1}\epsilon_1 + A_{2,2}\epsilon_2) \\ &= A_{1,1}A_{2,1}D(\epsilon_1, \epsilon_1) + A_{1,2}A_{2,1}D(\epsilon_2, \epsilon_1) + A_{1,1}A_{2,2}D(\epsilon_1, \epsilon_2) + A_{1,2}A_{2,2}D(\epsilon_2, \epsilon_2) \end{aligned}$$

Hence, D is completely determined by four scalars

$$\begin{aligned} a &:= D(\epsilon_1, \epsilon_1), \quad b := D(\epsilon_1, \epsilon_2) \\ c &:= D(\epsilon_2, \epsilon_1), \quad d := D(\epsilon_2, \epsilon_2) \end{aligned}$$

Conversely, given any four scalars a, b, c, d, if we define $D : K^{2\times 2} \to K$ by

$$D(A) = A_{1,1}A_{2,1}a + A_{1,2}A_{2,1}b + A_{1,1}A_{2,2}c + A_{1,2}A_{2,2}d$$

Then D defines a 2-linear map on $K^{2\times 2}$.

(iv) For instance, we may define $D : K^{2\times 2} \to K$ by

$$D(A) = A_{1,1}A_{2,2} - A_{1,2}A_{2,1}$$

This is 2-linear and is the family 'determinant' of a 2×2 matrix.

Lemma 2.3. *A linear combination of n-linear functions is n-linear.*

Proof. Let D_1 and D_2 be two n-linear functions and $c \in K$ be a scalar, then we wish to show that

$$D_3 := cD_1 + D_2$$

is n-linear. So fix $A = (\alpha_1, \alpha_2, \dots, \alpha_n) \in K^{n\times n}$ and $1 \leq j \leq n$, and consider the map

$$D_3^j : K^n \to K$$

defined by

$$D_3^j(\beta) := D_3(\alpha_1, \alpha_2, \dots, \alpha_{j-1}, \beta, \alpha_{j+1}, \dots, \alpha_n)$$

It is clear that

$$D_3^j = cD_1^j + D_2^j$$

and each of D_1^j and D_2^j are linear. Hence, D_3^j is linear, so D is n-linear as required. □

We now wish to isolate the kind of function that will eventually lead to the definition of a 'determinant'.

Definition 2.4. An n-linear function D is said to be *alternating* (or *alternate*) if both the following conditions are satisfied:

(i) $D(A) = 0$ whenever two rows of A are equal.

(ii) If A' is obtained from A by exchanging two rows of A, then $D(A') = -D(A)$.

Remark 2.5.

(i) We will show that the condition (i) implies the condition (ii) above.

(ii) It is not, in general, true that condition (ii) implies condition (i). It is true, however, if $1 + 1 \neq 0$ in K (Check!)

Definition 2.6. An n-linear function D is said to be a *determinant function* if D is alternating, and satisfies

$$D(I) = 1$$

Example 2.7.

(i) Let D be a 2-linear determinant function. As mentioned above, D has the form

$$D(A) = A_{1,1}A_{2,1}a + A_{1,2}A_{2,1}b + A_{1,1}A_{2,2}c + A_{1,2}A_{2,2}d$$

where

$$\begin{aligned} a &:= D(\epsilon_1, \epsilon_1), \quad b := D(\epsilon_1, \epsilon_2) \\ c &:= D(\epsilon_2, \epsilon_1), \quad d := D(\epsilon_2, \epsilon_2) \end{aligned}$$

Since D is alternating,

$$D(\epsilon_1, \epsilon_1) = D(\epsilon_2, \epsilon_2) = 0$$

Furthermore,

$$D(\epsilon_1, \epsilon_2) = -D(\epsilon_2, \epsilon_1)$$

Hence,

$$D(A) = c(A_{1,1}A_{2,2} - A_{1,2}A_{2,1})$$

But since $D(I) = 1$, we conclude that $c = 1$, so that

$$D(A) = A_{1,1}A_{2,2} - A_{1,2}A_{2,1}$$

Hence, there is only one 2-linear determinant function.

(ii) Let F be a field and $K = F[x]$ be the polynomial ring over F. Let D be any 3-linear determinant function on K, and let

$$A = \begin{pmatrix} x & 0 & -x^2 \\ 0 & 1 & 0 \\ 1 & 0 & x^3 \end{pmatrix}$$

Then

$$\begin{aligned} D(A) &= D(x\epsilon_1 - x^2\epsilon_3, \epsilon_2, \epsilon_1 + x^3\epsilon_3) \\ &= xD(\epsilon_1, \epsilon_2, \epsilon_1 + x^2\epsilon_3) - x^2D(\epsilon_3, \epsilon_2, \epsilon_1 + x^3\epsilon_3) \\ &= xD(\epsilon_1, \epsilon_2, \epsilon_1) + x^4D(\epsilon_1, \epsilon_2, \epsilon_3) - x^2D(\epsilon_3, \epsilon_2, \epsilon_1) - x^5D(\epsilon_3, \epsilon_2, \epsilon_3) \end{aligned}$$

Since D is alternating,

$$D(\epsilon_1, \epsilon_2, \epsilon_1) = D(\epsilon_3, \epsilon_2, \epsilon_3) = 0$$

and

$$D(\epsilon_3, \epsilon_2, \epsilon_1) = -D(\epsilon_1, \epsilon_2, \epsilon_3)$$

Hence,

$$D(A) = (x^4 + x^2)D(\epsilon_1, \epsilon_2, \epsilon_3) = x^4 + x^2$$

where the last equality holds because $D(I) = 1$.

Lemma 2.8. *Let D be a 2-linear function with the property that $D(A) = 0$ for any 2×2 matrix A over K having equal rows. Then D is alternating.*

Proof. Let A be a fixed 2×2 matrix and A' be obtained from A by interchanging two rows. We wish to prove that

$$D(A') = -D(A)$$

So write $A = (\alpha, \beta)$ as before, then we wish to show that

$$D(\beta, \alpha) = -D(\alpha, \beta)$$

But consider

$$D(\alpha + \beta, \alpha + \beta) = D(\alpha, \alpha) + D(\beta, \alpha) + D(\alpha, \beta) + D(\beta, \beta)$$

and note that, by hypothesis,

$$D(\alpha + \beta, \alpha + \beta) = D(\alpha, \alpha) = D(\beta, \beta) = 0$$

Hence,

$$D(\alpha, \beta) + D(\beta, \alpha) = 0$$

as required. □

Lemma 2.9. *Let D be an n-linear function on $K^{n\times n}$ with the property that $D(A) = 0$ whenever two adjacent rows of A are equal. Then, D is alternating.*

Proof. We have to verify both conditions of Definition 2.4. Namely,

- $D(A) = 0$ whenever two rows of A are equal.
- If A' is obtained from A by exchanging two rows of A, then $D(A') = -D(A)$.

We first verify condition (ii) and then verify (i).

(i) Suppose first that A' is obtained from A by interchanging two adjacent rows of A. Then, we assume without loss of generality that the rows α_1 and α_2 are interchanged. In other words,

$$A = (\alpha_1, \alpha_2, \ldots, \alpha_n) \text{ and } A' = (\alpha_2, \alpha_1, \ldots, \alpha_n)$$

But the same logic as in the previous lemma shows that

$$D(A') = -D(A)$$

(ii) Now suppose that A' is obtained from A by interchanging row i with row j where $i < j$. Then consider the matrix B_1 obtained from A by successively interchanging rows

$$\begin{aligned} i &\leftrightarrow (i+1) \\ (i+1) &\leftrightarrow (i+2) \\ &\vdots \\ (j-1) &\leftrightarrow j \end{aligned}$$

This requires $k := (j-i)$ interchanges of adjacent rows, so that

$$D(B_1) = (-1)^k D(A)$$

Now A' is obtained from B_1 by successively interchangning rows

$$\begin{aligned}(j-1) &\leftrightarrow (j-2)\\ (j-2) &\leftrightarrow (j-3)\\ &\vdots\\ (i+1)&i\end{aligned}$$

This requires $(k-1)$ interchanges of adjacent rows, so that

$$D(A') = (-1)^{k-1}D(B_1) = (-1)^{2k-1}D(A) = -D(A)$$

(iii) Finally, suppose A is any matrix in which two rows are equal, say $\alpha_i = \alpha_j$. If $j = i+1$, then A has two adjacent rows that are equal, so $D(A) = 0$ by hypothesis. If $j > i+1$, then we interchange rows $(i+1) \leftrightarrow j$, to obtain a matrix B which has two adjacent rows equal. Therefore, by hypothesis, $D(B) = 0$. But by step (ii), we have

$$D(A) = -D(B)$$

so that $D(A) = 0$ as well.

□

Definition 2.10.

(i) Let $A \in K^{n\times n}$ be a matrix and $1 \leq i, j \leq n$. Then $A(i \mid j)$ is the $(n-1)\times(n-1)$ matrix obtained by deleting the i^{th} row and the j^{th} column of A.

(ii) If D is an $(n-1)$-linear function and A is an $n \times n$ matrix, then we define $D_{i,j}$ as

$$D_{i,j}(A) := D(A(i \mid j))$$

Theorem 2.11. *LeT $n > 1$ and D be an $(n-1)$-linear function. For each $1 \leq j \leq n$, define a function $E_j : K^{n\times n} \to K$ by*

$$E_j(A) := \sum_{i=1}^n (-1)^{i+j} A_{i,j} D_{i,j}(A)$$

Then E_j is an alternating function on $K^{n\times n}$. Furthermore, if D is a determinant function, then so is E_j.

Proof.

(i) If A is an $n \times n$ matrix, then the scalar $D_{i,j}(A)$ is independent of the i^{th} row of A. Furthermore, since D is $(n-1)$-linear, it follows that $D_{i,j}$ is linear in all rows except the i^{th} row. Hence,

$$A \mapsto A_{i,j}D_{i,j}(A)$$

is n-linear. By Lemma 2.3, it follows that E_j is n-linear.

(ii) To prove that E_j is alternating, it suffices to show (by Lemma 2.9) that $E_j(A) = 0$ whenever any two adjacent rows of A are equal. So suppose $A = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ and $\alpha_k = \alpha_{k+1}$. If $i \notin \{k, k+1\}$, then the matrix $A(i \mid j)$ has row equal rows, so that $D_{i,j}(A) = 0$. Therefore,

$$E_j(A) = (-1)^{k+j}A_{k,j}D_{k,j}(A) + (-1)^{k+1+j}A_{k+1,j}D_{k+1,j}(A)$$

Since $\alpha_k = \alpha_{k+1}$,

$$A_{k,j} = A_{k+1,j} \text{ and } A(k \mid j) = A(k+1 \mid j)$$

Hence, $E_j(A) = 0$. Thus E is alternating.

(iii) Now suppose D is a determinant function and $I^{(n)}$ denotes the $n \times n$ identity matrix, then $I^{(n)}(j \mid j)$ is the $(n-1) \times (n-1)$ identity matrix $I^{(n-1)}$. Since $I^{(n)}_{i,j} = \delta_{i,j}$, we have

$$E_j(I^{(n)}) = D_{j,j}(I^{(n)}) = D(I^{(n-1)}) = 1$$

so that E_j is a determinant function.

□

Corollary 2.12. *Let K be a commutative ring with identity and let $n \in \mathbb{N}$. Then, there exists at least one determinant function on $K^{n\times n}$.*

Proof. For $n = 1$, we simply define $D([a]) = a$.

For $n > 1$, we assume by induction that we have constructed a determinant function on $K^{(n-1)\times(n-1)}$. By Theorem 2.11, we may construct a determinant function on $K^{n\times n}$. □

Example 2.13. We have already seen that any determinant function $D : K^{2\times 2} \to K$ must be of the form

$$D(B) = B_{1,1}B_{2,2} - B_{1,2}B_{2,2} =: |B|$$

Let $A \in K^{3\times 3}$, then we define E_j as in Theorem 2.11. Then

$$\begin{aligned} E_1(A) &= \sum_{i=1}^{3} (-1)^{i+1} A_{i,1}D_{i,1}(A) \\ &= A_{1,1}D_{1,1}(A) - A_{2,1}D_{2,1}(A) + A_{3,1}D_{3,1}(A) \\ &= A_{1,1}\left|\begin{pmatrix} A_{2,2} & A_{2,3} \\ A_{3,2} & A_{3,3} \end{pmatrix}\right| - A_{2,1}\left|\begin{pmatrix} A_{1,2} & A_{1,3} \\ A_{3,2} & A_{3,3} \end{pmatrix}\right| + A_{3,1}\left|\begin{pmatrix} A_{1,2} & A_{1,3} \\ A_{2,2} & A_{2,3} \end{pmatrix}\right| \end{aligned}$$

Similarly, we may calculate that

$$\begin{aligned}
E_2(A) &= -A_{1,2}\left|\begin{pmatrix} A_{2,1} & A_{2,3} \\ A_{3,1} & A_{3,3} \end{pmatrix}\right| + A_{2,2}\left|\begin{pmatrix} A_{1,1} & A_{1,3} \\ A_{3,1} & A_{3,3} \end{pmatrix}\right| - A_{3,2}\left|\begin{pmatrix} A_{1,1} & A_{1,3} \\ A_{2,1} & A_{2,3} \end{pmatrix}\right|, \text{ and} \\
E_3(A) &= A_{1,3}\left|\begin{pmatrix} A_{2,1} & A_{2,2} \\ A_{3,1} & A_{3,2} \end{pmatrix}\right| - A_{2,3}\left|\begin{pmatrix} A_{1,1} & A_{1,2} \\ A_{3,1} & A_{3,2} \end{pmatrix}\right| + A_{3,3}\left|\begin{pmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{pmatrix}\right|
\end{aligned}$$

It follows from Theorem 2.11 that E_1, E_2, and E_3 are all determinant functions. In fact, we will soon prove (and you can verify if you like) that

$$E_1 = E_2 = E_3$$

We take one example to describe this phenomenon: Let $K = \mathbb{R}[x]$ and

$$A = \begin{pmatrix} (x-1) & x^2 & x^3 \\ 0 & x-2 & 1 \\ 0 & 0 & x-3 \end{pmatrix}$$

Then

$$\begin{aligned}
E_1(A) &= (x-1)\left|\begin{pmatrix} x-2 & 1 \\ 0 & x-3 \end{pmatrix}\right| - x^2\left|\begin{pmatrix} 0 & 1 \\ 0 & x-3 \end{pmatrix}\right| + x^3\left|\begin{pmatrix} 0 & x-2 \\ 0 & 0 \end{pmatrix}\right| \\
&= (x-1)(x-2)(x-3) \\
E_2(A) &= -x^2\left|\begin{pmatrix} 0 & 1 \\ 0 & x-3 \end{pmatrix}\right| + (x-2)\left|\begin{pmatrix} x-1 & x^3 \\ 0 & x-3 \end{pmatrix}\right| - 1\left|\begin{pmatrix} x-1 & x^2 \\ 0 & 0 \end{pmatrix}\right| \\
&= (x-1)(x-2)(x-3)
\end{aligned}$$

Similarly, one can check that

$$E_3(A) = (x-1)(x-2)(x-3)$$

as well.

3. Permutations and Uniqueness of Determinants

Remark 3.1. (i) Let D be an n-linear function, and $A \in K^{n\times n}$ be a matrix with rows $\alpha_1, \alpha_2, \ldots, \alpha_n$. Then,

$$\alpha_i = \sum_{j=1}^{n} A(i,j)\epsilon_j$$

where $\{\epsilon_1, \epsilon_2, \dots, \epsilon_n\}$ denote the rows of the identity matrix I. Since D is n-linear, we see that

$$\begin{aligned} D(A) &= \sum_j A(1,j)D(\epsilon_j, \alpha_2, \dots, \alpha_n) \\ &= \sum_j \sum_k A(1,j)A(2,k)D(\epsilon_j, \epsilon_k, \alpha_3, \dots \alpha_n) \\ &= \dots \\ &= \sum_{k_1,k_2,\dots,k_n} A(1,k_1)A(2,k_2)\dots A(n,k_n)D(\epsilon_{k_1}, \epsilon_{k_2}, \dots, \epsilon_{k_n}) \end{aligned}$$

(ii) Now suppose that D is also alternating. Then,

$$D(\epsilon_{k_1}, \epsilon_{k_2}, \dots, \epsilon_{k_n}) = 0$$

for any tuple $(k_1, k_2, \dots, k_n)$ such that any two k_i coincide. Hence, we conclude that

$$D(A) = \sum_{k_1,k_2,\dots,k_n} A(1,k_1)A(2,k_2)\dots A(n,k_n)D(\epsilon_{k_1}, \epsilon_{k_2}, \dots, \epsilon_{k_n})$$

where the sum is taken over all tuples $(k_1, k_2, \dots, k_n)$ such that the $\{k_i\}$ are mutually distinct integers with $1 \leq k_i \leq n$.

Definition 3.2.

(i) Let X be a set. A *permutation* of X is a bijective function $\sigma : X \to X$.

(ii) If $X = \{1, 2, \dots, n\}$, then a permutation of X is called a *permutation of degree n*.

(iii) We write S_n for the set of all permutations of degree n. Note that, since $X := \{1, 2, \dots, n\}$ is a finite set, a function $\sigma : X \to X$ is bijective if and only if it is either injective or surjective.

Now, in the earlier remark, a tuple $(k_1, k_2, \dots, k_n)$ is equivalent to a function

$$\sigma : X \to X \text{ given by } \sigma(i) = k_i$$

To say that the $\{k_i\}$ are mutually distinct is equivalent to saying that σ is injective (and hence bijective). We now conclude the following fact.

Lemma 3.3. *Let D be an n-linear alternating function. Then, for any $A \in K^{n\times n}$, we have*

$$D(A) = \sum_{\sigma \in S_n} A(1,\sigma(1))A(2,\sigma(2))\dots A(n,\sigma(n))D(\epsilon_{\sigma(1)}, \epsilon_{\sigma(2)}, \dots, \epsilon_{\sigma(n)})$$

Remark 3.4. If $\sigma_1, \sigma_2 \in S_n$, then the composition $\sigma_1 \circ \sigma_2$ is also a bijection of X, and hence we get a binary operation

$$\circ : S_n \times S_n \to S_n$$

Observe that

(i) Composition is associative:

$$\sigma_1 \circ (\sigma_2 \circ \sigma_3) = (\sigma_1 \circ \sigma_2) \circ \sigma_3$$

(ii) If $\tau = \mathrm{id}_X$, then

$$\sigma \circ \tau = \tau \circ \sigma$$

holds for all $\sigma \in S_n$.

(iii) If $\sigma \in S_n$, there is an inverse function $\sigma^{-1} : X \to X$, which is also bijective, and satisfies

$$\sigma \circ \sigma^{-1} = \tau = \sigma^{-1} \circ \sigma$$

Hence, the pair $(S_n, \circ)$ is a *group*, and is called the *symmetric group of degree n.*

Definition 3.5. A *transposition* is an element $\sigma \in S_n$ such that, there exist $1 \leq i, j \leq n$ with $i \neq j$ such that

$$\sigma(k) = \begin{cases} k & : k \notin \{i, j\} \\ j & : k = i \\ i & : k = j \end{cases}$$

We will need the following fact, which we will not prove (it will hopefully be proved in MTH301 - if not, you can look up [Conrad]).

Theorem 3.6. *Every $\sigma \in S_n$ can be expressed in the form*

$$\sigma = \tau_1 \tau_2 \dots \tau_k$$

where each τ_i is a transposition. This expression is not necessarily unique, but if

$$\sigma = \eta_1 \eta_2 \dots \eta_{k'}$$

is another such expression where each η_j is a transposition, then

$$k = k' \mod (2)$$

Definition 3.7. Let $\sigma \in S_n$

(i) We say that σ is an *even* permutation if it can be expressed as an even number of transpositions. If it can be expressed as an odd number of transpositions, then we say that σ is *odd*. Note that this definition makes sense (ie. an odd permutation cannot also be even) because of the previous theorem.

(ii) The *sign function* is the map $\mathrm{sgn} : S_n \to \{\pm 1\}$ given by

$$\mathrm{sgn}(\sigma) = \begin{cases} 1 & : \sigma \text{ is even} \\ -1 & : \sigma \text{ is odd} \end{cases}$$

Example 3.8.

(i) If τ is a transposition, then $\text{sgn}(\tau) = -1$.

(ii) If $\sigma \in S_4$ is the permutation given by

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{pmatrix}$$

Then, σ can be expressed as a product of two transpositions (Check!), so

$$\text{sgn}(\sigma) = +1$$

Lemma 3.9. *Let D be an n-linear alternating function, and $\sigma \in S_n$. Then*

$$D(\epsilon_{\sigma(1)}, \epsilon_{\sigma(2)}, \ldots, \epsilon_{\sigma(n)}) = sgn(\sigma)D(I)$$

Proof.

(i) Suppose first that σ is a transposition

$$\sigma(k) = \begin{cases} l & : k \notin \{i, j\} \\ j & : k = i \\ i & : k = j \end{cases}$$

Then the matrix

$$A = (\epsilon_{\sigma(1)}, \epsilon_{\sigma(2)}, \ldots, \epsilon_{\sigma(n)})$$

is obtained from the identity matrix by interchanging the i^{th} row with the j^{th} row (and keeping all other rows fixed). Hence,

$$D(A) = (-1)D(I) = \text{sgn}(\sigma)D(I)$$

(ii) Now suppose σ is any other permutation, then write σ as a product of transpositions

$$\sigma = \tau_1\tau_2 \ldots \tau_k$$

Thus, if we pass from

$$(1, 2, \ldots, n) \to (\sigma(1), \sigma(2), \ldots, \sigma(n))$$

there are exactly k interchanges if pairs. Since D is alternating, each such interchange results in a multiplication by (-1). Hence,

$$D(\epsilon_{\sigma(1)}, \epsilon_{\sigma(2)}, \ldots, \epsilon_{\sigma(n)}) = (-1)^k D(I) = \text{sgn}(\sigma)D(I)$$

□

The next theorem is thus a consequence of Lemma 3.3 and Lemma 3.9.

Theorem 3.10. *Let D be an n-linear alternating function and $A \in K^{n\times n}$. Then*

$$D(A) = \left[\sum_{\sigma \in S_n} A(1,\sigma(1))A(2,\sigma(2))\dots A(n,\sigma(n))sgn(\sigma)\right] D(I)$$

Recall that a determinant function is one that is n-linear, alternating, and satisfies $D(I) = 1$. From the above theorem, we thus conclude

Corollary 3.11. *There is exactly one determinant function*

$$\det : K^{n\times n} \to K$$

Furthermore, if $A \in K^{n\times n}$, then

$$\det(A) = \sum_{\sigma\in S_n} sgn(\sigma)A(1,\sigma(1))A(2,\sigma(2))\dots A(n,\sigma(n))$$

Furthermore, if $D : K^{n\times n} \to K$ is any n-linear alternating function, then

$$D(A) = \det(A)D(I)$$

for any $A \in K^{n\times n}$

Theorem 3.12. *Let K be a commutative ring with identity, and $A, B \in K^{n\times n}$. Then*

$$\det(AB) = \det(A)\det(B)$$

Proof. Fix B, and define $D : K^{n\times n} \to K$ by

$$D(A) := \det(AB)$$

Then

(i) D is n-linear: If $C = (\alpha_1, \alpha_2, \dots, \alpha_n)$, we write

$$D(C) = D(\alpha_1, \alpha_2, \dots, \alpha_n)$$

Then observe that

$$D(\alpha_1, \alpha_2, \dots, \alpha_n) = \det(\alpha_1 B, \alpha_2 B, \dots, \alpha_n)$$

where $\alpha_j B$ denotes the $1 \times n$ matrix obtained by multiplying the $1 \times n$ matrix α_j by the $n \times n$ matrix B. Now note that

$$(\alpha_i + c\alpha_i')B = \alpha_i B + c\alpha_i' B$$

Since det is n-linear, it now follows that

$$\begin{aligned}
&D(\alpha_1, \alpha_2, \dots, \alpha_{i-1}, \alpha_i + c\alpha_i', \alpha_{i+1}, \dots, \alpha_n)\\
&= \det(\alpha_1 B, \alpha_2 B, \dots, \alpha_{i-1}B, (\alpha_i + c\alpha_i')B, \alpha_{i+1}B, \dots, \alpha_n B)\\
&= \det(\alpha_1 B, \alpha_2 B, \dots, \alpha_{i-1}B, \alpha_i B, \alpha_{i+1}B, \dots, \alpha_n B)\\
&+ c\det(\alpha_1 B, \alpha_2 B, \dots, \alpha_{i-1}B, \alpha_i' B, \alpha_{i+1}B, \dots, \alpha_n B)\\
&= D(\alpha_1, \alpha_2, \dots, \alpha_{i-1}, \alpha_i, \alpha_{i+1}, \dots, \alpha_n) + cD(\alpha_1, \alpha_2, \dots, \alpha_{i-1}, \alpha_i', \alpha_{i+1}, \dots, \alpha_n)
\end{aligned}$$

This is true for each $1 \le i \le n$. Hence, D is n-linear.

(ii) D is alternating: If $\alpha_i = \alpha_j$ for some $i \neq j$, then

$$\alpha_i B = \alpha_j B$$

Since det is alternating,

$$\det(\alpha_1 B, \alpha_2 B, \ldots, \alpha_n B) = 0$$

and so $D(\alpha_1, \alpha_2, \ldots, \alpha_n) = 0$. Thus, D is alternating by Lemma 2.9.

Thus, by Corollary 3.11, it follows that

$$D(A) = \det(A) D(I)$$

Now observe that $D(I) = \det(IB) = \det(B)$. This completes the proof. □

(End of Week 7)

4. Additional Properties of Determinants

Theorem 4.1. *Let K be a commutative ring with identity, and let $A \in K^{n\times n}$ be an $n \times n$ matrix over K. Then*

$$\det(A^t) = \det(A)$$

Proof. Let $\sigma \in S_n$ be a permutation of degree n, then

$$A^t(i, \sigma(i)) = A(\sigma(i), i)$$

Hence, by Corollary 3.11, we have

$$det(A^t) = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) A(\sigma(1), 1) A(\sigma(2), 2) \ldots A(\sigma(n), n)$$

But if $\sigma(i) = j$, then $\sigma^{-1}(j) = i$, so $A(\sigma(i), i) = A(j, \sigma^{-1}(j))$. Multiplying, we get

$$A(\sigma(1), 1) A(\sigma(2), 2) \ldots A(\sigma(n), n) = A(1, \sigma^{-1}(1)) A(2, \sigma^{-1}(2)) \ldots A(n, \sigma^{-1}(n))$$

Furthermore,

$$\sigma\sigma^{-1} = 1 \Rightarrow \operatorname{sgn}(\sigma)\operatorname{sgn}(\sigma^{-1}) = \operatorname{sgn}(1) = 1 \Rightarrow \operatorname{sgn}(\sigma) = \operatorname{sgn}(\sigma^{-1})$$

The map $\sigma \mapsto \sigma^{-1}$ is a permutation of S_n, so we get

$$\begin{aligned}
\det(A^t) &= \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) A(\sigma(1), 1) A(\sigma(2), 2) \ldots A(\sigma(n), n) \\
&= \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma^{-1}) A(1, \sigma^{-1}(1)) A(2, \sigma^{-1}(2)) \ldots A(n, \sigma^{-1}(n)) \\
&= \sum_{\tau \in S_n} \operatorname{sgn}(\tau) A(1, \tau(1)) A(2, \tau(2)) \ldots A(n, \tau(n)) \\
&= \det(A)
\end{aligned}$$

□

Lemma 4.2. *Let $A, B \in K^{n\times n}$. Suppose that B is obtained from A by adding a multiple of one row of A to another (or a multiple of one column of A to another), then* $\det(A) = \det(B)$.

Proof.

(i) Suppose the rows of A are $\alpha_1, \alpha_2, \ldots, \alpha_n$, and assume that the rows of B are $\alpha_1, \alpha_2 + c\alpha_1, \alpha_3, \ldots \alpha_n$. Then, using the fact that det is n-linear, we have

$$\begin{aligned}\det(B) &= \det(\alpha_1, \alpha_2 + c\alpha_1, \alpha_3, \ldots, \alpha_n)\\ &= \det(\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_n) + c\det(\alpha_1, \alpha_1, \alpha_3, \ldots, \alpha_n)\\ &= \det(A) + c(0) = \det(A)\end{aligned}$$

(ii) Now suppose B is obtained by replacing a column of A by a multiple of another, then B^t is obtained from A^t by replacing a row by a multiple of another row. Hence, by the first part

$$\det(B^t) = \det(A^t)$$

The result now follows from Theorem 4.1.

□

Theorem 4.3. *Let $A \in K^{r\times r}, B \in K^{r\times s}$, and $C \in K^{s\times s}$, then*

$$\det\begin{pmatrix} A & B \\ 0 & C\end{pmatrix} = \det(A)\det(C)$$

where 0 *denotes the $s \times r$ zero matrix over K.*

Similarly, if $D \in K^{s\times r}$, then

$$\det\begin{pmatrix} A & 0 \\ D & C\end{pmatrix} = \det(A)\det(C)$$

where 0 *denotes the $r \times s$ zero matrix over K.*

Proof. Note that the second formula follows from the first by taking adjoints, so we only prove the first one.

(i) Define a function $D : K^{s\times s} \to K$ by

$$D(C) := \det\begin{pmatrix} A & B \\ 0 & C\end{pmatrix}$$

Then, it is clear that D is s-linear and alternating. So by Corollary 3.11, we conclude that

$$D(C) = \det(C)D(I)$$

(ii) Now consider

$$D(I) = \det \begin{pmatrix} A & B \\ 0 & I \end{pmatrix}$$

Fix an entry $B_{i,j}$ of B, and consider the matrix U obtained by subtracting $B_{i,j}$ times the $(s+j)^{th}$ row of the given matrix $\begin{pmatrix} A & B \\ 0 & I \end{pmatrix}$ from itself. Then, U is of the form

$$\begin{pmatrix} A & B' \\ 0 & I \end{pmatrix}$$

where $B'_{i,j} = 0$. Furthermore, by Lemma 4.2, we have

$$\det \begin{pmatrix} A & B \\ 0 & I \end{pmatrix} = \det \begin{pmatrix} A & B' \\ 0 & I \end{pmatrix}$$

Repeating this process finitely many times, we conclude that

$$D(I) = \det \begin{pmatrix} A & B \\ 0 & I \end{pmatrix} = \det \begin{pmatrix} A & 0 \\ 0 & I \end{pmatrix}$$

(iii) Finally, consider the function $\widetilde{D} : K^{r\times r} \to K$ by

$$\widetilde{D}(A) = \det \begin{pmatrix} A & 0 \\ 0 & I \end{pmatrix}$$

Then $\widetilde{D}$ is an n-linear and alternating function, so by Corollary 3.11, we have

$$\widetilde{D}(A) = \det(A)\widetilde{D}(I)$$

However, $\widetilde{D}(I) = 1$, so by part (i), we have

$$\det \begin{pmatrix} A & B \\ 0 & C \end{pmatrix} = \det(C)\det(A)$$

□

Example 4.4. Consider $K = \mathbb{Q}$ and

$$A = \begin{pmatrix} 1 & -1 & 2 & 3 \\ 2 & 2 & 0 & 2 \\ 4 & 1 & -1 & -1 \\ 1 & 2 & 3 & 0 \end{pmatrix}$$

Label the rows of A as $\alpha_1, \alpha_2, \alpha_3$, and α_4. Replacing α_2 by $\alpha_2 - 2\alpha_1$, we get a new matrix

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

Note that this new matrix has the same determinant as that of A by Lemma 4.2. Doing the same for α_3 and α_4, we get

$$\begin{pmatrix} 1 & -1 & 2 & 3 \\ 0 & 4 & -4 & -4 \\ 0 & 5 & -9 & -13 \\ 0 & 3 & 1 & -3 \end{pmatrix}$$

Now label these rows $\beta_1, \beta_2, \beta_3$, and β_4, we may replace β_3 by $\beta_3 - \frac{5}{4}\beta_2$ to get

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

Also replacing β_4 by $\beta_4 - \frac{3}{4}\beta_2$, we get

$$B := \begin{pmatrix} 1 & -1 & 2 & 3 \\ 0 & 4 & -4 & -4 \\ 0 & 0 & -4 & -8 \\ 0 & 0 & 4 & 0 \end{pmatrix}$$

Now note that $\det(B) = \det(A)$, and by Theorem 4.3, we have

$$\det(A) = \det\begin{pmatrix} 1 & -1 \\ 0 & 4 \end{pmatrix} \det\begin{pmatrix} -4 & -8 \\ 4 & 0 \end{pmatrix} = (4)(32) = 128$$

Recall that, in Theorem 2.11, we had defined, for each $1 \leq j \leq n$,

$$E_j(A) = \sum_{i=1}^n (-1)^{i+j} A_{i,j} \det(A(i|j))$$

and shown that this function is also a determinant function. By Corollary 3.11, we have that

$$\det(A) = \sum_{i=1}^n (-1)^{i+j} A_{i,j} \det(A(i|j)) \tag{V.1}$$

Definition 4.5. Let $A \in K^{n\times n}$

(i) For $1 \leq i,j \leq n$, the (i,j) *cofactor* of A is

$$C_{i,j} := (-1)^{i+j} \det(A(i|j))$$

(ii) The *(classical) adjoint* of A is the matrix $\text{adj}(A)$ whose $(i,j)^{th}$ entry is $C_{j,i}$.

Theorem 4.6. *For any $A \in K^{n\times n}$, then*

$$adj(A)A = A\,adj(A) = \det(A)I$$

Proof. Fix $A \in K^{n\times n}$ and write $\text{adj}(A) = (C_{j,i})$ as above.

(i) Observe that, by our formula (Equation V.1), we have

$$\det(A) = \sum_{i=1}^{n} A_{i,j}C_{i,j}$$

(ii) Now fix $1 \leq k \leq n$, and supopse B is the matrix obtained by replacing the j^{th} column of A by the k^{th} column, then two columns of B are the same, so

$$\det(B) = 0$$

Furthermore, $B(i|j) = A(i|j)$, so by Equation V.1,

$$\begin{aligned} 0 &= \sum_{i=1}^{n}(-1)^{i+j}B_{i,j}\det(B(i|j)) \\ &= \sum_{i=1}^{n}(-1)^{i+j}A_{i,k}\det(A(i|j)) \\ &= \sum_{i=1}^{n} A_{i,k}C_{i,j} \end{aligned}$$

Hence, we conclude that

$$\sum_{i=1}^{n} A_{i,k}C_{i,j} = \delta_{j,k}\det(A)$$

Since $\text{adj}(A)_{i,j} = C_{j,i}$, we conclude that

$$\text{adj}(A)A = \det(A)I$$

(iii) Now consider the matrix A^t, and observe that $A^t(i|j) = A(j|i)^t$. Hence, we have

$$(-1)^{i+j}\det(A^t(i|j)) = (-1)^{j+i}\det(A(j|i))$$

Thus, the $(i,j)^{th}$ cofactor of A is the $(j,i)^{th}$ cofactor of A^t. Hence,

$$\text{adj}(A^t) = \text{adj}(A)^t$$

(iv) Now applying part (ii) to A^t, we have

$$\text{adj}(A^t)A^t = \det(A^t)I$$

Transposing this equation, we get

$$A\text{adj}(A) = A\text{adj}(A^t)^t = \det(A^t)I = \det(A)I$$

by Theorem 4.1.

□

Note that, throughout this discussion, we have only used the fact that K is a commutative ring with identity, and we have not needed K to be a field. We can thus conclude the following theorem.

Theorem 4.7. *Lert K be a commutative ring with identity, and $A \in K^{n\times n}$. Then, A is invertible over K if and only if* $\det(A)$ *is invertible in K. If this happens, then A has a unique inverse given by*

$$A^{-1} = \det(A)^{-1} \mathit{adj}(A)$$

In particular, if K is a field, then A is invertible if and only if $\det(A) \neq 0$.

Example 4.8.

(i) Let $K = \mathbb{Z}$ denote the ring of integers, and

$$A := \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$$

Then $\det(A) = -2$ which is not invertible in K, so A is not invertible over K.

(ii) However, if you think of A as a 2×2 matrix over $\mathbb{Q}$, then A is invertible. Since

$$\text{adj}(A) = \begin{pmatrix} 4 & -2 \\ -3 & 1 \end{pmatrix}$$

we have

$$A^{-1} = \frac{-1}{2}\begin{pmatrix} 4 & -2 \\ -3 & 1 \end{pmatrix}$$

(iii) Let $K := \mathbb{R}[x]$, then the invertible elements of K are precisely the non-zero scalar polynomials (Why?). Consider

$$A := \begin{pmatrix} x^2+x & x+1 \\ x-1 & 1 \end{pmatrix} \text{ and } B := \begin{pmatrix} x^2-1 & x+2 \\ x^2-2x+3 & x \end{pmatrix}$$

Then,

$$\det(A) = x+1 \text{ and } \det(B) = -6$$

Hence, B is invertible, but A is not. Furthermore,

$$\text{adj}(B) = \begin{pmatrix} x & -x-2 \\ -x^2+2x-3 & x^2-1 \end{pmatrix}$$

Hence,

$$B^{-1} = \frac{-1}{6}\begin{pmatrix} x & -x-2 \\ -x^2+2x-3 & x^2-1 \end{pmatrix}$$

Remark 4.9. Let $T \in L(V)$ be a linear operator on a finite dimensional vector space V, and let $\mathcal{B}$ and $\mathcal{B}'$ be two ordered bases of V. Then, there is an invertible matrix P such that

$$[T]_{\mathcal{B}} = P^{-1}[T]_{\mathcal{B}'}P$$

Since det is multiplicative, we see that

$$\det([T]_{\mathcal{B}}) = \det([T]_{\mathcal{B}'})$$

We may thus define this common number to be the determinant of T, since it does not depend on the choice of ordered basis.

Theorem 4.10 (Cramer's Rule)**.** *Let A be an invertible matrix over a field F, and suppose $Y \in F^n$ is given. Then, the unique solution to the system of linear equations*

$$AX = Y$$

is given by $X = (x_j)$ where

$$x_j = \frac{\det(B_j)}{\det(A)}$$

where B_j is the matrix obtained by replacing the j^{th} column of A by Y.

Proof. Note that if $AX = Y$, then

$$\begin{aligned}
\operatorname{adj}(A)AX &= \operatorname{adj}(A)Y \\
\Rightarrow \det(A)X &= \operatorname{adj}(A)Y \\
\Rightarrow \det(A)x_j &= \sum_{i=1}^{n} \operatorname{adj}(A)_{j,i}y_i \\
&= \sum_{i=1}^{n} (-1)^{i+j}\det(A(i|j))y_i \\
&= \det(B_j)
\end{aligned}$$

(Check the last line!). This completes the proof. □

VI. Elementary Canonical Forms

1. Introduction

Our goal in this section is to study a fixed linear operator T on a finite dimensional vector space V. We know that, given an ordered basis $\mathcal{B}$, we may represent T as a matrix

$$[T]_{\mathcal{B}}$$

Depending on the basis $\mathcal{B}$, we may get different matrices. We would like to know

(i) What is the 'simplest' such matrix A that represents T.

(ii) Can we find the ordered basis $\mathcal{B}$ such that $[T]_{\mathcal{B}} = A$.

The meaning of the term 'simplest' will change as we encounter more difficulties. For now, we will understand 'simple' to mean a *diagonal matrix*, ie. one in the form

$$D = \begin{pmatrix} c_1 & 0 & 0 & \dots & 0 \\ 0 & c_2 & 0 & \dots & 0 \\ 0 & 0 & c_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & c_n \end{pmatrix}$$

Now, if $\mathcal{B} = \{\alpha_1, \alpha_2, \dots, \alpha_n\}$ is an ordered basis such that $[T]_{\mathcal{B}} = D$, then, for each $1 \leq i \leq n$, we have

$$T(\alpha_i) = c_i\alpha_i$$

This leads to the next section.

2. Characteristic Values

Definition 2.1. Let T be a linear operator on a vector space V.

(i) A scalar $c \in F$ is called a *characteristic value* (or *eigen value*) of T if, there exists a non-zero vector $\alpha \in V$ such that

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

(ii) If this happens, then α is called a *characteristic vector* (or *eigen vector*).

(iii) Given $c \in F$, the space

$$\{\alpha \in V : T(\alpha) = c\alpha\}$$

is called the *characteristic space* (or *eigen space*) of T associated with c. (Note that this is a subspace of V)

Note that, if $T \in L(V)$, we may write $(T - cI)$ for the linear operator $\alpha \mapsto T(\alpha) - c\alpha$. Now c is a characteristic value if and only if this operator has a non-zero kernel. Hence,

Theorem 2.2. *Let $T \in L(V)$ where V is a finite dimensional vector space, and $c \in F$. Then, TFAE:*

(i) c is a characteristic value of T

(ii) $(T - cI)$ is non-singular (ie. not invertible)

(iii) $det(T - cI) = 0$

Recall that $\det(T-cI)$ is defined in terms of matrices, so we make the following definition.

Definition 2.3. Let A be an $n \times n$ matrix over a field F.

(i) A *characteristic value* of A is a scalar $c \in F$ such that

$$\det(A - cI) = 0$$

(ii) The *characteristic polynomial* of A is $f \in F[x]$ defined by

$$f(x) = \det(xI - A)$$

Note that $\deg(f) = n$, and that f is monic.

Lemma 2.4. *Similar matrices have the same characteristic polynomial.*

Proof. If $B = P^{-1}AP$, then

$$\det(xI - B) = \det(P^{-1}(xI - A)P) = \det(xI - A)$$

□

Remark 2.5.

(i) The previous lemma allows us to make sense of the characteristic polynomial of a linear operator: Let $T \in L(V)$, then the characteristic polynomial of T is simply

$$f(x) = \det(xI - A)$$

where A is any matrix that represents the operator (as in Theorem III.4.2).

Hence, $\deg(f) = \dim(V) =: n$, so T has atmost n characteristic values by Corollary IV.4.5.

Example 2.6.

(i) Let $T \in L(\mathbb{R}^2)$ be the linear operator given by

$$T(x_1, x_2) = (-x_2, x_1)$$

If $\mathcal{B}$ denotes the standard basis, then

$$A := [T]_{\mathcal{B}} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

Hence, the characteristic polynomial of T is given by

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

Hence, T has no characteristic values.

(ii) However, if we think of the same operator in $L(\mathbb{C}^2)$, then T has two characteristic values $\{i, -i\}$. Now we wish to find the corresponding characteristic vectors.

(i) If $c = i$, then consider the matrix

$$(A - iI) = \begin{pmatrix} -i & -1 \\ 1 & -i \end{pmatrix}$$

It is clear that, if $\alpha_1 = (1, -i)$, then $\alpha_1 \in \ker(A - iI)$. Furthermore, by row reducing this matrix, we arrive at

$$B = \begin{pmatrix} -i & -1 \\ 0 & 0 \end{pmatrix}$$

so the rank of $(A - iI)$ is 1, and so it has nullity 1 as well. Thus,

$$\ker(T - iI) = \operatorname{span}\{\alpha_1\}$$

(ii) If $c = -i$, then consider the matrix

$$(A + iI) = \begin{pmatrix} i & -1 \\ 1 & i \end{pmatrix}$$

Now, $\alpha_2 = (1, i)$ is a characteristic vector. Once again, row reducing this matrix yields

$$C = \begin{pmatrix} i & -1 \\ 0 & 0 \end{pmatrix}$$

so by the same argument, $\dim(\ker(A + iI)) = 1$. Hence,

$$\ker(T + iI) = \operatorname{span}\{\alpha_2\}$$

(iii) Let A be the 3×3 real matrix given by

$$A = \begin{pmatrix} 3 & 1 & -1 \\ 2 & 2 & -1 \\ 2 & 2 & 0 \end{pmatrix}$$

Then the characteristic polynomial of A is

$$f(x) = \det \begin{pmatrix} x-3 & -1 & 1 \\ -2 & x-2 & 1 \\ -2 & -2 & x \end{pmatrix} = x^3 - 5x^2 + 8x - 4 = (x-1)(x-2)^2$$

So A has two characteristic values, 1 and 2.

(iv) Let $T \in L(\mathbb{R}^3)$ be the linear operator which is represented in the standard basis by A. We wish to find characteristic vectors associated to these two characteristic values:

(i) Consider $c = 1$ and the matrix

$$A - I = \begin{pmatrix} 2 & 1 & -1 \\ 2 & 1 & -1 \\ 2 & 2 & -1 \end{pmatrix}$$

Row reducing this matrix results in

$$B = \begin{pmatrix} 2 & 1 & -1 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

From this, it is clear that $(A - I)$ has rank 2, and so has nullity 1 (by Rank-Nullity). It is also clear that $\alpha_1 = (1, 0, 2) \in \ker(A - I)$. Hence, α_1 is a characteristic vector, and the characteristic space is given by

$$\ker(A - I) = \text{span}\{\alpha_1\}$$

(ii) Consider $c = 2$ and the matrix

$$A - 2I = \begin{pmatrix} 1 & 1 & -1 \\ 2 & 0 & -1 \\ 2 & 2 & -2 \end{pmatrix}$$

Row reducing this matrix results in

$$C = \begin{pmatrix} 1 & 1 & -1 \\ 0 & -2 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

Hence, $(A - 2I)$ has rank 2, so has nullity 1, once again. Also, it is clear that $\alpha_2 = (1, 1, 2) \in \ker(A - 2I)$. Hence, α_2 is a characteristic vector, and the characteristic space is given by

$$\ker(A - 2I) = \text{span}\{\alpha_2\}$$

Definition 2.7. An operator $T \in L(V)$ is said to be *diagonalizable* if there is a basis for V each vector of which is a characteristic vector of T.

Remark 2.8.

(i) If $\mathcal{B}$ is a basis satisfying the requirement of this definition, then

$$[T]_{\mathcal{B}}$$

is a diagonal matrix.

(ii) Note that there is no requirement that the diagonal entries be distinct. They may all be equal too!

(iii) We may as well require (in this definition) that V has a spanning set consisting of characteristic vectors. This is because any spanning set will contain a basis.

Example 2.9.

(i) In Example 2.6 (i), T is not diagonalizable because it has no characteristic values.

(ii) In Example 2.6 (ii), T is diagonalizable, because we have found two characteristic vectors $\mathcal{B} := \{\alpha_1, \alpha_2\}$ where $\alpha_1 = (1, -i)$ and $\alpha_2 = (1, i)$ which are characteristic vectors. Since these vectors are not scalar multiples of each other, they are linearly independent. Since $\dim(\mathbb{C}^2) = 2$, they form a basis. In this basis $\mathcal{B}$, we may represent T as

$$[T]_{\mathcal{B}} = \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix}$$

(iii) In Example 2.6 (iv), we have found that T has two linearly independent characteristic vectors $\{\alpha_1, \alpha_2\}$ where $\alpha_1 = (1, 0, 2)$ and $\alpha_2 = (1, 1, 2)$. However, there are no other characteristic vectors (other than scalar multiples of these). Since $\dim(\mathbb{R}^3) = 3$, this operator does not have a basis consisting of characteristic vectors. Hence, T is *not* diagonalizable.

Lemma 2.10. *Suppose $T \in L(V)$ is a diagonalizable operator, with distinct characteristic values $c_1, c_2, \ldots, c_k$. Then the characteristic polynomial of T has the form*

$$f(x) = (x - c_1)^{d_1}(x - c_2)^{d_2} \ldots (x - c_k)^{d_k}$$

Furthermore, the multiplicity d_i is the dimension of the characteristict space $\ker(T - c_i I)$.

Proof. Since T is diagonalizable, there is a ordered basis $\mathcal{B} = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ such that

$$A := [T]_{\mathcal{B}} = \begin{pmatrix} c_1 I_1 & 0 & 0 & \ldots & 0 \\ 0 & c_2 I_2 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & c_k I_k \end{pmatrix}$$

where $I_1, I_2, \dots I_k$ are the identity matrices of dimension $e_i = \dim(\ker(A - c_i I))$. Now the characteristic polynomial of T is the same as that of A, whose characteristic polynomial has the form

$$f(x) = \det(xI - A) = (x - c_1)^{d_1}(x - c_2)^{d_2} \dots (x - c_k)^{d_k}$$

Furthermore,

$$A - c_1 I = \begin{pmatrix} 0I_1 & 0 & 0 & \dots & 0 \\ 0 & (c_2 - c_1)I_2 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & (c_k - c_1)I_k \end{pmatrix}$$

Since $c_i \neq c_1$ for all $i \geq 2$, we have

$$d_1 = \dim(\ker(A - c_1 I))$$

and similarly, $d_i = \dim(\ker(A - c_i I))$ as required. Hence the result. □

Lemma 2.11. *Let $T \in L(V), c \in f$ and $\alpha \in V$ such that $T\alpha = c\alpha$. Then, for any polynomial $f \in F[x]$, we have*

$$f(T)\alpha = f(c)\alpha$$

Proof. Exercise. □

Lemma 2.12. *Let $T \in L(V)$ and $c_1, c_2, \dots, c_k$ be the distinct characteristic values of T, and let $W_i := \ker(T - c_i I)$ denote the corresponding characteristic spaces. If*

$$W := W_1 + W_2 + \dots + W_k$$

Then

$$\dim(W) = \dim(W_1) + \dim(W_2) + \dots + \dim(W_k)$$

In fact, if $\mathcal{B}_i, 1 \leq i \leq k$ are bases for the W_i, then $\mathcal{B} = \sqcup_{i=1}^n \mathcal{B}_i$ is a basis for W.

Proof.

(i) If $i \neq j$ and suppose $\alpha \in W_i \cap W_j$, then $T\alpha = c_i\alpha = c_j\alpha$. Since $c_i \neq c_j$, it follows that $\alpha = 0$. Hence,

$$W_i \cap W_j = \emptyset$$

In particular,

$$\mathcal{B} = \cup_{i=1}^n \mathcal{B}_i = \sqcup_{i=1}^n \mathcal{B}_i$$

(ii) Now suppose $\beta_i \in W_i$ are vectors such that

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

We claim that $\beta_i = 0$ for all $1 \leq i \leq k$. To see this, note that $T\beta_i = c_i\beta_i$. Hence, if $f \in F[x]$, then by Lemma 2.11,

$$0 = f(T)(\beta_1 + \beta_2 + \ldots + \beta_k) = \sum_{i=1}^k f(c_i)\beta_i$$

Since $\{c_1, c_2, \ldots, c_k\}$ are distinct scalars, there exist polynomials $\{f_1, f_2, \ldots, f_k\}$ such that

$$f_j(c_i) = \delta_{i,j}$$

(by Lagrange interpolation - Theorem IV.3.1). Applying f_j in the above equation, we conclude that

$$0 = \sum_{i=1}^k f_j(c_i)\beta_i = \beta_j$$

This is true for every $1 \leq j \leq k$. Hence the claim.

(iii) Now we claim that $\mathcal{B}$ is a basis for W. Since $\mathcal{B}$ clearly spans W, it suffices to show that it is linearly independent. So suppose there exist scalars d_i and vectors $\gamma_j \in \mathcal{B}$ such that

$$\sum_{j=1}^{\ell} d_j\gamma_j = 0$$

Then, by separating out the terms from each $\mathcal{B}_i$, we obtain an expression of the form

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

where $\beta_i \in W_i$ for each $1 \leq i \leq k$. By the previous step, we conclude that $\beta_i = 0$. However, β_i is itself a linear combination of the vectors in $\mathcal{B}_i$, which is linearly independent. Hence, each γ_j must be zero.

Thus, $\mathcal{B}$ is linearly independent, and hence a basis for W. This proves the result. □

Theorem 2.13. *Let $T \in L(V)$, and $c_1, c_2, \ldots, c_k \in F$ be the distinct characteristic values of T, and let $W_i := \ker(T - c_iI)$. Then, TFAE:*

(i) T is diagonalizable.

(ii) The characteristic polynomial of T is

$$f = (x - c_1)^{d_1}(x - c_2)^{d_2}\ldots(x - c_k)^{d_k}$$

where $d_i = \dim(W_i)$ for all $1 \leq i \leq k$.

(iii) $\dim(W_1) + \dim(W_2) + \ldots + \dim(W_k) = \dim(V)$.

Proof.

$(i) \Rightarrow (ii)$: This is the content of Lemma 2.10.

$(ii) \Rightarrow (iii)$: If (ii) holds, then

$$\deg(f) = d_1 + d_2 + \ldots + d_k$$

But $\deg(f) = \dim(V)$ because f is the characteristic polynomial of T. Hence, (iii) holds.

$(iii) \Rightarrow (i)$: By Lemma 2.12, it follows that

$$V = W_1 + W_2 + \ldots + W_k$$

Hence, the basis $\mathcal{B}$ from Lemma 2.12 is a basis for V consisting of characteristic vectors T. Thus, T is diagonalizable.

□

Remark 2.14. Let $A \in F^{n\times n}$ be a square matrix and $c_1, c_2, \ldots, c_k$ be the distinct characteristic values of A. Let $W_i \subset F^n$ be the subspace

$$W_i = \{X \in F^n : (A - c_i I)X = 0\}$$

and let $\mathcal{B}_i = \{\alpha_{i,1}, \alpha_{i,2}, \ldots, \alpha_{i,n_i}\}$ be an ordered basis for W_i. Placing these basis vectors in columns, we get a matrix

$$P = [\alpha_{1,1}\alpha_{1,2}\ldots\alpha_{1,n_1}\alpha_{2,1}\alpha_{2,2}\ldots\alpha_{2,n_2}\ldots\alpha_{k,1}\alpha_{k,2}\ldots\alpha_{k,n_k}]$$

Then, the set $\mathcal{B} = \sqcup_{i=1}^k \mathcal{B}_i$ is a basis for F^n if and only if P is a square matrix. In that case, P is a invertible matrix, and $P^{-1}AP$ is diagonal.

Example 2.15. Let $T \in L(\mathbb{R}^3)$ be the linear operator represented in the standard basis by the matrix

$$A = \begin{pmatrix} 5 & -6 & -6 \\ -1 & 4 & 2 \\ 3 & -6 & -4 \end{pmatrix}$$

(i) We first compute the characteristic polynomial of A: $f = \det(xI = A)$

$$\det \begin{pmatrix} (x-5) & 6 & 6 \\ 1 & x-4 & -2 \\ -3 & 6 & x+4 \end{pmatrix}$$

Subtracting column 3 from column 2 gives us a new matrix with the same determinant by Lemma V.4.2. Hence,

$$f = \det \begin{pmatrix} x-5 & 0 & 6 \\ 1 & x-2 & -2 \\ -3 & 2-x & x+4 \end{pmatrix} = (x-2)\det \begin{pmatrix} x-5 & 0 & 6 \\ 1 & 1 & -2 \\ -3 & -1 & x+4 \end{pmatrix}$$

Now adding row 2 to row 3 does not change the determinant, so

$$f = (x-2)\det\begin{pmatrix} x-5 & 0 & 6 \\ 1 & 1 & -2 \\ -2 & 0 & x+2 \end{pmatrix}$$

Expanding along column 2 now gives

$$f = (x-2)\det\begin{pmatrix} x-5 & 6 \\ -2 & x+2 \end{pmatrix} = (x-2)(x^2-3x+2) = (x-2)^2(x-1)$$

(ii) Hence, the characteristic values of T are 1 and 2.

(iii) We wish to determine the dimensions of the characteristic spaces, W_1 and W_2.

  (i) Consider the case $c = 1$, and the matrix

$$(A-I) = \begin{pmatrix} 4 & -6 & -6 \\ -1 & 3 & 2 \\ 3 & -6 & -5 \end{pmatrix}$$

  Row reducing this matrix gives

$$\begin{pmatrix} 4 & -6 & -6 \\ 0 & \frac{3}{2} & \frac{1}{2} \\ 0 & \frac{-3}{2} & \frac{-1}{2} \end{pmatrix} \mapsto \begin{pmatrix} 4 & -6 & -6 \\ 0 & \frac{3}{2} & \frac{1}{2} \\ 0 & 0 & 0 \end{pmatrix} =: B$$

  Hence, $\text{rank}(A-I) = 2$

  (ii) Consider the case $c = 2$: We know that

$$\text{rank}(A-2I) \leq 2$$

  since $(A-2I)$ is a singular matrix. Furthermore, we know that

$$\text{rank}(A-I) + \text{rank}(A-2I) \leq \dim(\mathbb{R}^3) = 3$$

  So it follows that $\text{rank}(A-2I) = 1$ and equality holds in the previous equation.

Hence, we conclude by Theorem 2.13 that T is diagonalizable.

(iv) We now determine a basis consisting of characteristic vectors:

  (i) Consider the case $c = 1$: Using the matrix B above, we solve the system of linear equations $BX = 0$. This gives a solution

$$\alpha_1 = (3, -1, 3)$$

  (ii) Consider the case $c = 2$, and the matrix

$$(A-2I) = \begin{pmatrix} 3 & -6 & -6 \\ -1 & 2 & 2 \\ 3 & -6 & -6 \end{pmatrix}$$

Row reducing gives

$$C = \begin{pmatrix} 3 & -6 & -6 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

So solving the system $CX = 0$ gives two solutions

$$\alpha_2 = (2,1,0) \text{ and } \alpha_3 = (2,0,1)$$

(v) Thus, we get an ordered basis

$$\mathcal{B} = \{(3,-1,3),(2,1,0),(2,0,1)\}$$

consisting of characteristic vectors of T. Furthermore,

$$[T]_{\mathcal{B}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix} := D$$

(vi) Furthermore, if P is the matrix

$$P = \begin{pmatrix} 3 & 2 & 2 \\ -1 & 1 & 0 \\ 3 & 0 & 1 \end{pmatrix}$$

Then

$$P^{-1}AP = D$$

(End of Week 8)

3. Annihilating Polynomials

Recall that: If V is a vector space, then $L(V)$ is a (non-commutative) linear algebra with unit. Hence, if $f \in F[x]$ is a polynomial and $T \in L(V)$, then $f(T) \in L(V)$ makes sense (See Definition IV.2.7). Furthermore, the map $f \mapsto f(T)$ is an algebra homomorphism (Theorem IV.2.9). In other words, if $f, g \in F[x]$ and $c \in F$, then

$$(f+cg)(T) = f(T) + cg(T) \text{ and } (fg)(T) = f(T)g(T) \tag{VI.1}$$

Definition 3.1. Let $T \in L(V)$ be a linear operator on a finite dimensional vector space V. Define the *annihilator* of T to be

$$\text{Ann}(T) := \{f \in F[x] : f(T) = 0\}$$

Lemma 3.2. *If V is a finite dimensional vector space and $T \in L(V)$, then $Ann(T)$ is a non-zero ideal of $F[x]$.*

Proof.

(i) If $f, g \in \text{Ann}(T)$, then $f(T) = g(T) = 0$. Hence, if $c \in F$, then

$$(f + cg)(T) = 0$$

by Equation VI.1, so $(f + cg) \in \text{Ann}(T)$. Thus, $\text{Ann}(T)$ is a vector subspace of $F[x]$.

(ii) If $f \in \text{Ann}(T)$ and $g \in F[x]$, then $f(T) = 0$, so by Equation VI.1, we have

$$(fg)(T) = f(T)g(T) = 0$$

so $fg \in \text{Ann}(T)$. Thus, $\text{Ann}(T)$ is an ideal of $F[x]$.

(iii) Suppose $n := \dim(V)$, then $\dim(L(V)) = n^2$ by Theorem III.2.4. By Corollary II.3.10, the set

$$\{I, T, T^2, \ldots, T^{n^2}\}$$

is linearly dependent. Hence, there exist scalars $a_i \in F, 0 \leq i \leq n^2$ (not all zero) such that

$$\sum_{i=0}^{n^2} a_i T^i = 0$$

Hence, if $f \in F[x]$ is the (non-zero) polynomial

$$f = \sum_{i=0}^{n^2} a_i x^i$$

Then $f \in \text{Ann}(T)$. Thus, $\text{Ann}(T) \neq \{0\}$.

□

Definition 3.3. Let $T \in L(V)$ be a linear operator on a finite dimensional vector space V. The *minimal polynomial* of T is the unique monic generator of $\text{Ann}(T)$.

Remark 3.4. Note that the minimal polynomial $p \in F[x]$ of T has the following properties:

(i) p is a monic, non-zero polynomial.

(ii) $p(T) = 0$

(iii) If $f \in F[x]$ is another polynomial such that $f(T) = 0$, then $p \mid f$.

(iv) In particular, if $f \in F[x]$ is any other polynomial such that $f(T) = 0$, then

$$\deg(f) \geq \deg(p)$$

The same argument that we had above also applies to the algebra $F^{n\times n}$.

Definition 3.5. If $A \in F^{n\times n}$ is a square matrix, then we define the *annihilator* of A to be the set

$$\mathrm{Ann}(A) = \{f \in F[x] : f(A) = 0\}$$

Once again, this is a non-zero ideal of $F[x]$. So we may define the *minimal polynomial* of A the same way as above.

Remark 3.6.

(i) If $T \in L(V)$ is a linear operator and $\mathcal{B}$ is an ordered basis of V. We write

$$A := [T]_{\mathcal{B}}$$

Then, by Example IV.3.7, for any $f \in F[x]$, we have

$$[f(T)]_{\mathcal{B}} = f(A)$$

Since the map $S \mapsto [S]_{\mathcal{B}}$ is an isomorphism from $L(V)$ to $F^{n\times n}$ (by Theorem III.4.2), it follows that

$$f(A) = 0 \text{ in } F^{n\times n} \Leftrightarrow f(T) = 0 \text{ in } L(V)$$

Hence,

$$\mathrm{Ann}(T) = \mathrm{Ann}([T]_{\mathcal{B}})$$

In particular, T and $[T]_{\mathcal{B}}$ have the same minimal polynomial. Hence, in what follows, whatever we say for linear operators also holds for matrices.

(ii) Now we consider the following situation: Consider $\mathbb{R} \subset \mathbb{C}$ as a subfield, and let $A \in M_n(\mathbb{R})$ be an $n \times n$ square matrix with entries in $\mathbb{R}$. Let $p_1 \in \mathbb{R}[x]$ be the minimal polynomial of T with respect to $\mathbb{R}$. In other words, p_1 is the monic generator of the ideal

$$\mathrm{Ann}_{\mathbb{R}}(T) = \{f \in \mathbb{R}[x] : f(T) = 0\}$$

Similarly, let $p \in \mathbb{C}[x]$ be the minimal polynomial of T with respect to F. In other words, p is the monic generator of the ideal

$$\mathrm{Ann}_{\mathbb{C}}(T) = \{g \in \mathbb{C}[x] : g(T) = 0\}$$

Note that, since $\mathbb{R} \subset \mathbb{C}$, we have

$$\mathrm{Ann}_{\mathbb{R}}(T) \subset \mathrm{Ann}_{\mathbb{C}}(T)$$

In particular, $p_1 \in \mathrm{Ann}_{\mathbb{C}}(T)$, so that

$$p \mid p_1 \text{ in } \mathbb{C}[x]$$

Conversely, $p \in \mathbb{C}[x]$ can be expressed in the form $p = h + ig$ for some $h, g \in \mathbb{R}[x]$, so that

$$0 = p(A) = h(A) + ig(A)$$

so that $h(A) = g(A) = 0$. Hence, $p_1 \mid h$ and $p_1 \mid g$ in $\mathbb{R}[x]$. Thus,

$$p_1 \mid p \text{ in } \mathbb{C}[x]$$

We conclude that $p_1 = p$ (because they are both monic). Thus, if $A \in M_n(\mathbb{R})$, then the minimal polynomial of A with respect to $\mathbb{R}$ is the same as the minimal polynomial of A with respect to $\mathbb{C}$.

Theorem 3.7. *Let $T \in L(V)$ where V is a finite dimensional vector space. Then, the characteristic polynomial of T and the minimal polynomial of T have the same roots (except for multiplicities).*

Proof. Let $p \in F[x]$ denote the minimal polynomial of T and let $f \in F[x]$ denote the characteristic polynomial of T, given by

$$f = \det(xI - T)$$

(i) If $c \in F$ is a root of p, then by Corollary IV.4.4, $(x - c) \mid p$, so there exists $q \in F[x]$ such that

$$p = (x - c)q$$

Since $\deg(q) < \deg(p)$ it follows that $q(T) \neq 0$ by Remark 3.4. Hence, there exists $\beta \in V$ such that

$$\alpha := q(T)(\beta) \neq 0$$

Then, since $p(T) = 0$, we have

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

Hence, c is a characteristic value of T, so $f(c) = 0$.

(ii) Conversely, suppose $c \in F$ is such that $f(c) = 0$, then c is a characteristic value of T, so there exists $\alpha \in V$ non-zero such that

$$T\alpha = c\alpha$$

By Lemma 2.11, we have

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

But $p(T) = 0$, so

$$p(c)\alpha = 0$$

Since $\alpha \neq 0$, it follows that $p(c) = 0$ by Lemma II.1.3.

□

Remark 3.8. Suppose T is a diagonalizable operator with distinct characteristic values $c_1, c_2, \ldots, c_k$. Then, by Lemma 2.10, the characteristic polynomial of T has the form

$$f = (x - c_1)^{d_1}(x - c_2)^{d_2} \ldots (x - c_k)^{d_k}$$

where d_i is the dimension of the characteristic subspace associated to c_i. Let $q \in F[x]$ be the polynomial

$$q = (x - c_1)(x - c_2) \ldots (x - c_k)$$

If $\alpha \in V$ is a characteristic vector of T associated to the characteristic value c_i, then

$$q(T)\alpha = (T - c_1 I)(T - c_2 I) \ldots (T - c_k I)\alpha = 0$$

since $(T - c_i I)$ and $(T - c_j I$ commute for all $1 \leq i, j \leq k$. Since V has a basis consisting of characteristic vectors of T, it follows that

$$q(T) = 0$$

Let $p \in F[x]$ denote the minimal polynomial of T, then p and f share the same roots by Theorem 3.7. Thus, by Corollary IV.4.4, it follows that

$$(x - c_i) \mid p$$

for all $1 \leq i \leq k$. Hence, $q \mid p$. But $q(T) = 0$, so $p \mid q$ by Remark 3.4. Since both p and q are monic, we conclude that

$$p = q = (x - c_1)(x - c_2) \ldots (x - c_k)$$

Thus, if T is diagonalizable, then the minimal polynomial of T is the product of distinct linear factors.

We now consider the following examples, the first two of which are from Example 2.6.

Example 3.9.

(i) Let $T \in L(\mathbb{R}^2)$ be the linear operator given by

$$T(x_1, x_2) = (-x_2, x_1)$$

In the standard basis $\mathcal{B}$, the associated matrix of T is given by

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

So, the characteristic polynomial of T is

$$f = x^2 + 1$$

Considering $T \in L(\mathbb{C}^2)$, the characteristic polynomial of T is

$$f = (x - i)(x + i)$$

By Theorem 3.7, the minimal polynomial of T in $\mathbb{C}[x]$ is

$$p = (x - i)(x + i) = x^2 + 1$$

By Remark 3.6, the minimal polynomial of T in $\mathbb{R}[x]$ is also

$$p = x^2 + 1$$

(ii) Let A be the 3×3 real matrix given by

$$A = \begin{pmatrix} 3 & 1 & -1 \\ 2 & 2 & -1 \\ 2 & 2 & 0 \end{pmatrix}$$

Then the characteristic polynomial of A is

$$f = (x - 1)(x - 2)^2$$

By Theorem 3.7, the minimal polynomial of A has the form

$$p = (x - 1)^i (x - 2)^j$$

Observe that

$$(A - I)(A - 2I) = \begin{pmatrix} 2 & 1 & -1 \\ 2 & 1 & -1 \\ 2 & 2 & -1 \end{pmatrix} \begin{pmatrix} 1 & 1 & -1 \\ 2 & 0 & -1 \\ 2 & 2 & -2 \end{pmatrix} = \begin{pmatrix} 2 & 0 & -1 \\ 2 & 0 & -1 \\ 4 & 0 & -2 \end{pmatrix}$$

Therefore, $p \neq (x - 1)(x - 2)$. One can then verify that

$$(A - I)(A - 2I)^2 = 0$$

Thus, the minimal polynomial of A is

$$p = (x - 1)(x - 2)^2 = f$$

(iii) Let $T \in L(\mathbb{R}^3)$ be the linear operator (from Example 2.15) represented in the standard basis by the matrix

$$A = \begin{pmatrix} 5 & -6 & -6 \\ -1 & 4 & 2 \\ 3 & -6 & -4 \end{pmatrix}$$

Then the characteristic polynomial of T is

$$f = (x - 1)(x - 2)^2$$

Hence, by Theorem 3.7, the minimal polynomial of T has the form

$$p = (x - 1)^i (x - 2)^j$$

But by Remark 3.8, the minimal polynomial is a product of distinct linear factors. Hence,

$$p = (x - 1)(x - 2)$$

Theorem 3.10. *[Cayley-Hamilton] Let T be a linear operator on a finite dimensional vector space and $f \in F[x]$ be the characteristic polynomial of T. Then*

$$f(T) = 0$$

In particular, the minimal polynomial of T divides the characteristic polynomial of T.

Proof.

(i) Set

$$K := \{g(T) : g \in F[x]\}$$

Since the map $g \mapsto g(T)$ is a homomorphism of rings, it follows that K is a commutative ring with identity.

(ii) Let $\mathcal{B} = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ be an ordered basis for V, and set $A := [T]_{\mathcal{B}}$. Then, for each $1 \leq i \leq n$, we have

$$T\alpha_i = \sum_{j=1}^{n} A_{j,i}\alpha_j = \sum_{j=1}^{n} \delta_{i,j} T\alpha_j$$

Hence, if $B \in K^{n\times n}$ is the matrix whose $(i,j)^{th}$ entry is given by

$$B_{i,j} = \delta_{i,j}T - A_{j,i}I$$

then we have

$$B\alpha_j = 0$$

for all $1 \leq j \leq n$.

(iii) Observe that, for $n = 2$,

$$B = \begin{pmatrix} T - A_{1,1}I & -A_{2,1}I \\ -A_{1,2}I & T - A_{2,2}I \end{pmatrix}$$

Hence,

$$\det(B) = (T - A_{1,1}I)(T - A_{2,2}I) - A_{1,2}A_{2,1}I = f(T)$$

More generally, f is the polynomial given by $f = \det(xI - A) = \det(xI - A^t)$ (by Theorem V.4.1), and

$$(xI - A^t)_{i,j} = \delta_{i,j}x - A_{j,i}$$

Hence,

$$f(T) = \det(B)$$

Hence, to prove that $f(T)$ is the zero operator, it suffices to prove that

$$\det(B)\alpha_j = 0$$

for all $1 \leq j \leq n$.

(iv) Let $\widetilde{B} := \operatorname{adj}(B)$ be the adjoint of B. By Theorem V.4.6, we have

$$\widetilde{B}B = \det(B)I$$

Since $B\alpha_j = 0$, it follows that

$$\det(B)\alpha_j = 0$$

This completes the proof.

□

The advantage of the Cayley-Hamilton theorem is that it is easier to determine the minimal polynomial now.

Example 3.11. Let $A \in M_4(\mathbb{Q})$ be the 4×4 rational matrix given by

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

One can compute (do it!) that

$$A^2 = \begin{pmatrix} 2 & 0 & 2 & 0 \\ 0 & 2 & 0 & 2 \\ 2 & 0 & 2 & 0 \\ 0 & 2 & 0 & 2 \end{pmatrix}, \text{ and}$$
$$A^3 = \begin{pmatrix} 0 & 4 & 0 & 4 \\ 4 & 0 & 4 & 0 \\ 0 & 4 & 0 & 4 \\ 4 & 0 & 4 & 0 \end{pmatrix}$$

Thus,

$$A^3 = 4A$$

Hence, if $f \in \mathbb{Q}[x]$ is the polynomial

$$f = x^3 - 4x = x(x+2)(x-2)$$

Then, the minimal polynomial $p \in \mathbb{Q}[x]$ of A must divide f.

(i) Note that, since A is not a scalar polynomial,

$$\deg(p) \geq 2$$

Hence, there are four possible candidates for p

$$x(x+2), x(x-2), (x+2)(x-2), \text{ or } f$$

(ii) It is clear from the above calculation that

$$A^2 \neq -2A$$

Hence, $p \neq x(x+2)$.

(iii) Similarly,

$$A^2 \neq 2A$$

Hence, $p \neq x(x+2)$

(iv) Similarly,

$$A^2 \neq 4I$$

Hence, $p \neq (x-2)(x+2)$

Hence, the minimal polynomial of A is

$$p = f = x(x+2)(x-2)$$

Now, the characteristic polynomial $g \in \mathbb{Q}[x]$ is a polynomial of degree 4, $p \mid g$ and g has the same roots as p. Thus, there are three options for g, namely

$$x^2(x+2)(x-2), x(x+2)^2(x-2), \text{ or } x(x+2)(x-2)^2$$

We now row reduce A to get

$$\begin{aligned} A &\mapsto \begin{pmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \end{pmatrix} \\ &\mapsto \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \end{pmatrix} \\ &\mapsto \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \end{aligned}$$

Hence,

$$\text{rank}(A) = 2$$

Thus, nullity$(A) = 2$. Thus, the characteristic value 0 has a 2-dimensional characteristic space. Thus, it must occur with multiplicity 2 in the characteristic polynomial of T (by Lemma 2.10). Hence,

$$g = x^2(x-2)(x+2)$$

4. Invariant Subspaces

Definition 4.1. Let $T \in L(V)$ be a linear operator on a vector space V, and let $W \subset V$ be a subspace. We say that W is an *invariant subspace* for T if, for all $\alpha \in W$, then $T(\alpha) \in W$. In other words, $T(W) \subset W$.

Example 4.2.

(i) For any T, the subspaces V and $\{0\}$ are invariant. These are called the *trivial* invariant subspaces.

(ii) The null space of T is invariant under T, and so is the range of T.

(iii) Let F be a field and $D : F[x] \to F[x]$ be the derivative operator (See Example III.1.2). Let W be the subspace of all polynomials of degree ≤ 3, then W is invariant under D because D lowers the degree of a polynomial.

(iv) Let $T \in L(V)$ and $U \in L(V)$ be an operator such that

$$TU = UT$$

Then $W := U(V)$, the range of U is invariant under T: If $\beta = U(\alpha)$ for some $\alpha \in V$, then

$$U\beta = TU(\alpha) = U(T(\alpha)) \in U(V) = W$$

Similarly, if $N := \ker(U)$, then N is invariant undeer T: If $\beta \in N$, then $U(\beta) = 0$, so

$$U(T(\beta)) = TU(\beta) = T(0) = 0$$

so $T(\beta) \in N$ as well.

(v) A special case of the above example: Let $g \in F[x]$ be any polynomial, and $U := g(T)$, then

$$UT = TU$$

(vi) An even more special case: Let $g = x - c$, then $U = g(T) = (T - cI)$. If $c \in F$ is a characteristic value of T, then

$$N := \ker(U)$$

is the characteristic subspace associated to c. This is an invariant subspace for T.

(vii) Let $T \in L(\mathbb{R}^2)$ be the linear operator given in the standard basis by the matrix

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

We claim that T has no non-trivial invariant subspace. Suppose $W \subset \mathbb{R}^2$ is a non-trivial invariant subspace, then

$$\dim(W) = 1$$

So if $\alpha \in W$ is non-zero, then $T(\alpha) = c\alpha$ for some scalar $c \in F$. Thus, α is a characteristic vector of T. However, T has no characteristic values (See Example 2.6).

Definition 4.3. Let $T \in L(V)$ and $W \subset V$ be a T-invariant subspace. Then $T_W \in L(W)$ denotes the restriction of T to W. ie. For $\alpha \in W, T_W(\alpha) = T(\alpha)$.

Note that this is well-defined precisely because W is T-invariant.

Remark 4.4. Suppose $W \subset V$ is a T-invariant subspace, and let $\mathcal{B}' = \{\alpha_1, \alpha_2, \ldots, \alpha_r\}$ be an ordered basis for W. Let $\mathcal{B} = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ be an ordered basis for V extending $\mathcal{B}'$ (See Theorem II.3.13). Let $A = [T]_{\mathcal{B}}$, so that, for each $1 \leq j \leq n$, we have

$$T\alpha_j = \sum_{i=1}^{n} A_{i,j}\alpha_i$$

Now, since W is T-invariant, it follows that, for each $1 \leq j \leq r$, we have

$$T\alpha_j \in W$$

can be expressed a linear combination of the elements of $\mathcal{B}'$. By the uniqueness of this expansion, we have

$$T\alpha_j = \sum_{i=1}^{r} A_{i,j}\alpha_i$$

Hence, if $j \leq r$ and $i > r$, we have

$$A_{i,j} = 0$$

Thus, A can be written in block form

$$A = \begin{pmatrix} B & C \\ 0 & D \end{pmatrix}$$

where B is an $r \times r$ matrix, D is an $(n-r) \times (n-r)$ matrix, and C is an $r \times (n-r)$ matrix. Note that

$$[T_W]_{\mathcal{B}'} = B$$

in this expression as well.

Lemma 4.5. *Let $T \in L(V)$ and $W \subset V$ be a T-invariant subspace, and T_W denote the restriction of T to W. Then,*

(i) The characteristic polynomial of T_W divides the characteristic polynomial of T.

(ii) The minimal polynomial of T_W divides the minimal polynomial of T.

Proof.

(i) Consider the characteristic polynomial of T, denoted by f, and the characteristic polynomial of T_W, denoted by g. Let $\mathcal{B}'$ be an ordered basis for W, and $\mathcal{B}$ be an ordered basis for V containing $\mathcal{B}'$. Write

$$A := [T]_{\mathcal{B}'} \text{ and } B := [T_W]_{\mathcal{B}'}$$

So that A has a block form (as above) given by

$$A = \begin{pmatrix} B & C \\ 0 & D \end{pmatrix}$$

Then, by Remark 2.5, we have

$$f = \det(xI - A) \text{ and } g = \det(xI - B)$$

However, by Theorem V.4.3, we have

$$\det(xI - A) = \det(xI - B)\det(xI - D)$$

Hence, $g \mid f$

(ii) Consider the minimal polynomial of T, denoted by p, and the minimal polynomial of T_W, denoted by q. Then, by Example IV.3.7, we have

$$p(A) = 0$$

However, for any $k \in \mathbb{N}$, A^k has the form

$$A^k = \begin{pmatrix} B^k & C_k \\ 0 & D^k \end{pmatrix}$$

for some $r \times (n-r)$ matrix C_k. Hence, for any polynomial $h \in F[x]$,

$$h(A) = 0 \Rightarrow h(B) = 0$$

In particular, $p(B) = 0$. But q is the minimal polynomial for B, so $q \mid p$ by definition.

□

Remark 4.6. Let $T \in L(V)$, and let $c_1, c_2, \ldots, c_k$ denote the distinct characteristic values of T. Let $W_i := \ker(T - c_iI)$ denote the corresponding characteristic spaces, and let $\mathcal{B}_i$ denote an ordered basis for W_i.

(i) Set

$$W := W_1 + W_2 + \ldots + W_k \text{ and } \mathcal{B} := \sqcup_{i=1}^k \mathcal{B}_i$$

By Lemma 2.12, $\mathcal{B}$ is an ordered basis for W and

$$\dim(W) = \dim(W_1) + \dim(W_2) + \ldots + \dim(W_k)$$

Now, we write

$$\mathcal{B}_i = \{\alpha_{i,1}, \alpha_{i,2}, \ldots, \alpha_{i,n_i}\}$$

and

$$\mathcal{B} = \{\alpha_{1,1}, \alpha_{1,2}, \ldots, \alpha_{1,n_1}, \alpha_{2,1}, \alpha_{2,2}, \ldots, \alpha_{2,n_2}, \ldots, \alpha_{k,1}, \alpha_{k,2}, \ldots, \alpha_{k,n_k}\}$$

Then, we have

$$T\alpha_{i,j} = c_i\alpha_{i,j}$$

for all $1 \leq i \leq k, 1 \leq j \leq n_i$. Hence, W is a T-invariant subspace, and if $B := [T_W]_{\mathcal{B}}$, we have

$$B = \begin{pmatrix} t_1 & 0 & 0 & \ldots & 0 \\ 0 & t_2 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & t_k \end{pmatrix}$$

where each t_i denotes a $n_i \times n_i$ block matrix $t_iI_{n_i}$.

(ii) Thus, the characteristic polynomial of T_W is

$$g = (x-c_1)^{n_1}(x-c_2)^{n_2}\ldots(x-c_k)^{n_k}$$

Let f denote the characteristic polynomial of T, then by Lemma 4.5, we have that $g \mid f$. Hence, the multiplicity of c_i as a root of f is at least $n_i = \dim(W_i)$.

(iii) Furthermore, it is now clear (as we proved in Theorem 2.13) that T is diagonalizable if and only if

$$n = \dim(W) = n_1 + n_2 + \ldots + n_k$$

Definition 4.7. An operator $T \in L(V)$ is said to be *triangulable* if there exists an ordered basis $\mathcal{B}$ of V such that $[T]_{\mathcal{B}}$ is an upper triangular matrix. ie.

$$[T]_{\mathcal{B}} = \begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} & \ldots & a_{1,n} \\ 0 & a_{2,2} & a_{2,3} & \ldots & a_{2,n} \\ 0 & 0 & a_{3,3} & \ldots & a_{3,n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & a_{n,n} \end{pmatrix}$$

Remark 4.8.

(i) Observe that, if $\mathcal{B} = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ is an ordered basis, then the matrix $[T]_{\mathcal{B}}$ is upper triangular if and only if, for each $1 \leq i \leq n$,

$$T\alpha_i \in \operatorname{span}\{\alpha_1, \alpha_2, \ldots, \alpha_i\}$$

(ii) If T is a triangulable matrix, and $A := [T]_{\mathcal{B}}$ is an upper triangular matrix as above, then the characteristic polynomial of T is

$$f = \det(xI - A)$$

By expanding along the first column, we get

$$f = (x - a_{1,1})\det\begin{pmatrix} a_{2,2} & a_{2,3} & \ldots & a_{2,n} \\ 0 & a_{3,3} & \ldots & a_{3,n} \\ 0 & 0 & \ldots & a_{n,n} \end{pmatrix}$$

By induction, we may prove that

$$f = (x - a_{1,1})(x - a_{2,2}) \dots (x - a_{n,n})$$

Hence, by combining the like terms, we see that the characteristic polynomial has the form

$$f = (x - c_1)^{d_1}(x - c_2)^{d_2} \dots (x - c_k)^{d_k}$$

where the $c_1, c_2, \dots, c_k$ are the distinct characteristic values of T. In particular, if T is triangulable, then the characteristic polynomial can be factored as a product of linear terms.

(iii) Since the minimal polynomial divides the characteristic polynomial (by Cayley-Hamilton - Theorem 3.10), it follows that, if T is triangulable, then the minimal polynomial can be factored as a product of linear terms.

We will now show that this last condition is also sufficient to prove that T is triangulable. For this, we need a definition.

Definition 4.9. Let $T \in L(V), W \subset V$ be a T-invariant subspace, and let $\alpha \in V$ be a fixed vector.

(i) The *T-conductor of α into W* is the set

$$S(\alpha, W) := S_T(\alpha, W) = \{g \in F[x] : g(T)\alpha \in W\}$$

(ii) If $W = \{0\}$, then the T-conductor of α into W is called the *T-annihilator of α*.

Example 4.10.

(i) If W is any T-invariant subspace and $\alpha \in W$, then

$$S(\alpha, W) = F[x]$$

Conversely, if $\alpha \notin W$, then $S(\alpha, W) \neq F[x]$.

(ii) If $W_1 \subset W_2$, then

$$S(\alpha, W_1) \subset S(\alpha, W_2)$$

(iii) For any T-invariant subspace W and any $\alpha \in V$, we have

$$\text{Ann}(T) \subset S(\alpha, W)$$

In particular, if $p \in F[x]$ is the minimal polynomial of T, then $p \in S(\alpha, W)$.

Lemma 4.11. *Let $T \in L(V)$ and $W \subset V$ be an invariant subspace of T. Then,*

(i) W is invariant under $f(T)$ for any polynomial $f \in F[x]$.

(ii) For each $\alpha \in V$, the T-conductor of α into W, $S(\alpha, W)$ is an ideal in $F[x]$

Proof.

(i) If $\beta \in W$, then $T\beta \in W$. Since W is T-invariant, we have

$$T^2\beta = T(T\beta) \in W$$

Thus proceeding, we conclude that

$$T^n\beta \in W$$

for all $n \geq 0$. By linearity, we conclude that

$$f(T)\beta \in W$$

for all $f \in F[x]$

(ii) We prove both conditions of Definition IV.4.12.

- $S(\alpha, W)$ is a subspace of $F[x]$: If $f, g \in S(\alpha, W)$ and $c \in F$, then, by definition

$$f(T)\alpha \in W \text{ and } g(T)\alpha \in W$$

  Since W is a subspace,

$$(f + cg)(T)\alpha = f(T)\alpha + cg(T)\alpha \in W$$

- If $g \in S(\alpha, W)$ and $f \in F[x]$, then by definition

$$\beta := g(T)\alpha \in W$$

  By part (i), we conclude that

$$(fg)(T)\alpha = f(T)g(T)\alpha = f(T)\beta \in W$$

□

Remark 4.12. We conclude that there is a unique monic polynomial $p_{\alpha,W} \in F[x]$ such that, for any $f \in F[x]$,

$$f \in S(\alpha, W) \Leftrightarrow p_{\alpha,W} \mid f$$

We will often conflate the ideal and the polynomial, and simply say that $p_{\alpha,W}$ *is* the T-conductor of $\alpha \in W$.

By Example 4.10 (iii), every T-conductor divides the minimal polynomial of T.

Example 4.13.

(i) If $\alpha \in V$ is a characteristic vector of T with characteristic value c, and $W \subset V$ is any T-invariant subspace of V, then

$$S(\alpha, W) = \begin{cases} 1 & : \alpha \in W \\ (x - c) & : \alpha \notin W \end{cases}$$

(ii) Let $T \in L(\mathbb{R}^3)$ be the operator which is expressed in the standard basis $\mathcal{B} = \{\epsilon_1, \epsilon_2, \epsilon_3\}$ by the matrix

$$A = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 4 \end{pmatrix}$$

and let $W = \text{span}\{\epsilon_1\}$.

- If $\alpha = \epsilon_1$, then $S(\alpha, W) = F[x]$
- If $\alpha = \epsilon_2$: Observe that the minimal polynomial of A is (Check!)

$$p = (x-2)(x-3)(x-4)$$

Since $\alpha \notin W$ and $p_{\alpha,W} \mid p$, we have many options for $p_{\alpha,W}$, namely

$$(x-2), (x-3), (x-4), (x-2)(x-3), (x-3)(x-4), (x-2)(x-4), \text{ and } p$$

Note that

$$(A - 3I)\epsilon_2 = \begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

Hence, $S(\epsilon_2, W) = (x-3)$

- If $\alpha = \epsilon_3$: Since ϵ_3 is a characteristic vector with characteristic value 4, it follows from part (i) that

$$S(\alpha, W) = (x-4)$$

(End of Week 9)

Lemma 4.14. *Let $T \in L(V)$ be an operator on a finite dimensional vector space V such that the minimal polynomial of T is a product of linear factors*

$$p = (x-c_1)^{r_1}(x-c_2)^{r_2}\ldots(x-c_k^{r_k}$$

Let $W \neq V$ be a proper T-invariant subspace of V. Then, there exists $\alpha \in V$ such that

(i) $\alpha \notin W$

(ii) $(T - cI)\alpha \in W$ for some characteristic value $c \in F$.

In other words, the T-conductor of α into W is of the form $(x-c)$ for some characteristic value $c \in F$.

Proof. Let $\beta \in V$ be any vector such that $\beta \notin W$. Let $g = p_{\beta,W}$ be the T-conductor of β in W. Then, since $\beta \notin W$,

$$g \neq 1$$

Furthermore, by Remark 4.12, we have that

$$g \mid p$$

Hence, there exist $0 \leq e_i \leq r_i$ such that

$$g = (x - c_1)^{e_1}(x - c_2)^{e_2} \dots (x - c_k)^{e_k}$$

In other words, g is a product of linear terms. Since $g \neq 1$, it follows that there exists $1 \leq i \leq k$ such that $e_i > 0$. Hence,

$$(x - c_i) \mid g$$

So combining the other terms, we write

$$g = (x - c_i)h$$

for some polynomial $h \in F[x]$ with $\deg(h) < \deg(g)$. Since g is the monic generator of $S(\beta, W)$ we have that

$$\alpha := h(T)\beta \notin W$$

However,

$$(T - c_i)\alpha = g(T)\beta \in W$$

as required. □

Theorem 4.15. *Let V be a finite dimensional vector space and $T \in L(V)$. Then T is triangulable if and only if the minimal polynomial of T is a product of linear polynomials over F.*

Proof. If T is triangulable, then the minimal polynomial is a product of linear factors by Remark 4.8 (iii).

Conversely, suppose the minimal polynomial $p \in F[x]$ has the form

$$p = (x - c_1)^{r_1}(x - c_2)^{r_2} \dots (x - c_k)^{r_k}$$

We now construct a basis $\mathcal{B} = \{\alpha_1, \alpha_2, \dots, \alpha_n\}$ such that, for each $1 \leq i \leq n$,

$$\{\alpha_1, \alpha_2, \dots, \alpha_i\} \text{ is linearly independent, and } T\alpha_i \in \{\alpha_1, \alpha_2, \dots, \alpha_i\} \tag{VI.2}$$

Observe that the c_i are precisely the characteristic values of T. We proceed by induction:

- Set $W_0 = \{0\}$. By Lemma 4.14, there exists $\alpha_1 \in V$ such that $\alpha_1 \neq 0$ and, there exists $1 \leq i \leq k$ such that
$$(T - c_1 I)\alpha_1 = 0$$
- Suppose that we have construction $\alpha_1, \alpha_2, \dots, \alpha_i$ such that Equation VI.2. We now construct α_{i+1}. Let
$$W_i := \text{span}\{\alpha_1, \alpha_2, \dots, \alpha_i\}$$

If $W_i = n$, then we are done, so suppose $W_i \neq n$. Applying Lemma 4.14, there exists $\alpha_{i+1} \notin W_i$, and a characteristic value $c \in F$ such that

$$(T - cI)\alpha_{i+1} \in W_i$$

Hence, $\{\alpha_1, \alpha_2, \ldots, \alpha_{i+1}\}$ is linearly independent, and

$$T\alpha_{i+1} = \text{span}\{\alpha_1, \alpha_2, \ldots, \alpha_{i+1}\}$$

as required.

□

Recall that a field F is said to be algebraically closed if any polynomial over F can be factored as a product of linear terms (See Definition IV.5.11).

Corollary 4.16. *Let F be an algebraically closed field, and $A \in F^{n\times n}$ be an $n \times n$ matrix over F. Then, A is similar to an upper triangular matrix over F.*

The next theorem gives us an easy (verifiable) way to determine if a linear operator is diagonalizable or not.

Theorem 4.17. *Let V be a finite dimensional vector space over a field F and $T \in L(V)$. Then T is diagonalizable if and only if the minimal polynomial of T is a product of distinct linear terms.*

Proof.

(i) Suppose T is diagonalizable, there is an ordered basis $\mathcal{B}$ such that the matrix $A = [T]_{\mathcal{B}}$ is diagonal.

$$A = \begin{pmatrix} c_1 I_1 & 0 & 0 & \ldots & 0 \\ 0 & c_2 I_2 & 0 & \ldots & 0 \\ 0 & 0 & c_3 I_3 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & c_k I_k \end{pmatrix}$$

where the c_i are all distinct. The minimal polynomial of A is clearly

$$p = (x - c_1)(x - c_2)\ldots(x - c_k)$$

(See also Remark 3.8).

(ii) Conversely, suppose the minimal polynomial of T has this form, we let

$$W_i := \ker(T - c_i I)$$

denote the characteristic space associated to the characteristic value c_i, and let

$$W = W_1 + W_2 + \ldots + W_k$$

To show that T is diagonalizable, it suffices to show that $W = V$ (See Theorem 2.13).

- Note that, if $\beta \in W$, then

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

  for some $\beta_i \in W_i$. Hence,

  $$T\beta = c_1\beta_1 + c_2\beta_2 + \ldots + c_k\beta_k \in W_1 + W_2 + \ldots + W_k = W$$

  So that W is T-invariant.

- Furthermore, if $h \in F[x]$ is a polynomial, then the above calculation shows that

  $$h(T)\beta = h(c_1)\beta_1 + h(c_2)\beta_2 + \ldots + h(c_k)\beta_k$$

- Now, if $W \neq V$, then by Lemma 4.14, there exists $\alpha \in V$ such that $\alpha notin W$, but there exists $1 \leq i \leq k$ such that

  $$\beta := (T - c_iI)\alpha \in W$$

  If we set

  $$q = \prod_{j \neq i}(x - c_j)$$

  Then, $p = (x - c_i)q$, and $q(c_i) \neq 0$.

- Now observe that the polynomial $q - q(c_i)$ has a root at c_i. So by Corollary IV.4.4, there is a polynomial $h \in F[x]$ such that

  $$q - q(c_i) = (x - c_i)h$$

  Hence,

  $$q(T)\alpha - q(c_i)\alpha = (T - c_i)h(T)\alpha = h(T)(T - c_i)\alpha = h(T)\beta$$

  Furthermore,

  $$0 = p(T)\alpha = (T - c_iI)q(T)\alpha$$

  Hence, $q(T)\alpha$ is a characteristic vector with characteristic value c_i. In particular,

  $$q(T)\alpha \in W$$

  Thus,

  $$q(c_i)\alpha = q(T)\alpha - h(T)\beta \in W$$

  Since $q(c_i) \neq 0$, we conclude that $\alpha \in W$.

This is a contradiction, which proves the $W = V$ must hold. Hence, T must be diagonalizable.

□

5. Direct-Sum Decomposition

Given an operator T on a finite dimensional vector space V, our goal in this chapter was to find an ordered basis $\mathcal{B}$ of V such that the matrix representation

$$[T]_{\mathcal{B}}$$

is of a particularly 'simple' form (diagonal, triangular, etc.). Now, we will phrase the question in a slightly more sophisticated manner. We wish to decompose the underlying space V as a sum of invariant subspaces, such that the restriction of T to each such invariant subspace has a simple form.

We will also see how this relates to the results of the earlier sections.

Definition 5.1. Let $W_1, W_2, \ldots, W_k$ be subspaces of a vector space V. We say that $W_1, W_2, \ldots, W_k$ are *independent* if, for any vectors $\alpha_1, \alpha_2, \ldots, \alpha_k$ with $\alpha_i \in W_i$ for all $1 \leq i \leq k$, if

$$\alpha_1 + \alpha_2 + \ldots + \alpha_k = 0$$

then $\alpha_i = 0$ for all $1 \leq i \leq k$.

Remark 5.2.

If $k = 2$, then two subspaces W_1 and W_2 are independent if and only if $W_1 \cap W_2 = \{0\}$ (Check!)

For an example of this phenomenon, look at Lemma 2.12.

If $W_1, W_2, \ldots, W_k$ are independent, then any vector

$$\alpha \in W := W_1 + W_2 + \ldots + W_k$$

can be expressed uniquely as a sum of elements $\alpha_i \in W_i$

$$\alpha = \alpha_1 + \alpha_2 + \ldots + \alpha_k$$

Lemma 5.3. *Let V be a finite dimensional vector space and $W_1, W_2, \ldots, W_k$ be subspaces of V, and let $W := W_1 + W_2 + \ldots + W_k$. Then, TFAE:*

(i) $W_1, W_2, \ldots, W_k$ *are independent.*

(ii) For each $2 \leq j \leq k$, we have

$$W_j \cap (W_1 + W_2 + \ldots + W_{j-1}) = \{0\}$$

(iii) If $\mathcal{B}_i$ is an ordered basis for W_i, then $\mathcal{B} := \sqcup_{i=1}^n \mathcal{B}_i$ is a basis for W.

Proof.

$(i) \Rightarrow (ii)$: Suppose

$$\alpha \in W_j \cap (W_1 + W_2 + \ldots + W_{j-1})$$

Then write $\alpha = \alpha_1 + \alpha_2 + \ldots + \alpha_{j-1}$ for $\alpha_i \in W_i, 1 \leq i \leq j-1$. Then,

$$\alpha_1 + \alpha_2 + \ldots + \alpha_{j-1} + (-\alpha) = 0$$

Since the W_i are independent, it follows that $\alpha_i = 0$ for all $1 \leq i \leq j-1$ and $\alpha = 0$, as required.

$(ii) \Rightarrow (iii)$: Let $\mathcal{B}_i$ be a basis for W_i, then, for any $i \neq j$, we claim that

$$\mathcal{B}_i \cap \mathcal{B}_j = \emptyset$$

If not, then suppose $\alpha \in \mathcal{B}_i \cap \mathcal{B}_j$ for some pair with $i \neq j$, we may assume that $i > j$, then

$$\alpha \in W_i \cap (W_1 + W_2 + \ldots + W_j + W_{j+1} + \ldots + W_{i-1})$$

This implies $\alpha = 0$, but 0 cannot belong to any linearly independent set. Now, $\mathcal{B} = \sqcup_{i=1}^k \mathcal{B}_i$ is clearly a spanning set for W, so it suffices to show that it is linearly independent. So suppose $\mathcal{B}_i = \{\beta_{i,1}, \beta_{i,2}, \ldots, \beta_{i,n_i}\}$ and $c_{i,j}$ are scalars such that

$$\sum_{i=1}^k \sum_{j=1}^{n_i} c_{i,j}\beta_{i,j} = 0$$

Write $\alpha_i := \sum_{j=1}^{n_i} c_{i,j}\beta_{i,j}$, then $\alpha_i \in W_i$ and

$$\alpha_1 + \alpha_2 + \ldots + \alpha_k = 0$$

Since $W_1, W_2, \ldots, W_k$ are independent, we conclude that $\alpha_i = 0$ for all $1 \leq i \leq k$. But then, since $\mathcal{B}_i$ is linearly independent, it must happen that $c_{i,j} = 0$ for all $1 \leq j \leq n_i$ as well.

$(iii) \Rightarrow (i)$: Suppose $\mathcal{B}_i := \{\beta_{i,1}, \beta_{i,2}, \ldots, \beta_{i,n_i}\}$ is a basis for W_i and $\mathcal{B} := \sqcup_{i=1}^k \mathcal{B}_i$ is a basis for W. Now suppose $\alpha_i \in W_i$ such that

$$\alpha_1 + \alpha_2 + \ldots + \alpha_k = 0$$

Express each α_i as a linear combinations of vectors from $\mathcal{B}_i$ by

$$\alpha_i = \sum_{j=1}^{n_i} c_{i,j}\beta_{i,j}$$

Adding these up, we get

$$\sum_{i=1}^k \sum_{j=1}^{n_i} c_{i,j}\beta_{i,j} = 0$$

Since the collection $\mathcal{B}$ is linearly independent, we conclude that $c_{i,j} = 0$ for all $1 \leq i \leq k, 1 \leq j \leq n_i$. Hence,

$$\alpha_i = 0$$

for all $1 \leq i \leq k$ as required.

□

Definition 5.4. If any (and hence all) of the above conditions hold, then we say that W is an *(internal) direct sum* of the W_i, and we write

$$W = W_1 \oplus W_2 \oplus \ldots \oplus W_k$$

Example 5.5.

(i) If V is a vector space and $\mathcal{B} = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ is any basis for V, let $W_i := \text{span}\{\alpha_i\}$. Then
$$V = W_1 \oplus W_2 \oplus \ldots \oplus W_n$$

(ii) Let $V = F^{n\times n}$ be the space of $n \times n$ matrices over F. Let W_1 denote the set of all symmetric matrices (a matrix $A \in V$ is symmetric if $A = A^t$), and let W_2 denote the set of all skew-symmetric matrices (a matrix $A \in V$ is skew-symmetric if $A = -A^t$). Then, (Check!) that

$$V = W_1 \oplus W_2$$

(iii) Let $T \in L(V)$ be a linear operator and let $c_1, c_2, \ldots, c_k$ denote the distinct characteristic values of T, and let $W_i := \ker(T - c_iI)$ denote the corresponding characteristic subspaces. Then, $W_1, W_2, \ldots, W_k$ are independent by Lemma 2.12. Hence, if T is diagonalizable, then

$$V = W_1 \oplus W_2 \oplus + \ldots \oplus W_k$$

Definition 5.6. Let V be a vector space. An operator $E \in L(V)$ is called a *projection* if $E^2 = E$.

Remark 5.7. Let $E \in L(V)$ be a projection.

(i) If we set $R := E(V)$ to be the range of E and $N := \ker(E)$, then note that $\beta \in R$ if and only if $E\beta = \beta$

Proof. If $\beta = E\beta$, then $\beta \in R$. Conversely, if $\beta \in R$, then $\beta = E(\alpha)$ for some $\alpha \in V$, so that $E\beta = E^2(\alpha) = E(\alpha) = \beta$ □

(ii) Hence, $V = R \oplus N$

Proof. If $\alpha \in V$, then write $\alpha = E\alpha + (\alpha - E(\alpha))$ and observe that $E(\alpha) \in R$ and $(\alpha - E(\alpha)) \in N$. Furthermore, if $\beta \in R \cap N$, then $E(\beta) = 0$. But by part (i), we have $\beta = E(\beta) = 0$. Hence, $R \cap N = \{0\}$. Thus, R and N are independent. □

(iii) Now, suppose W_1 and W_2 are two subspaces of V such that $V = W_1 \oplus W_2$ then any vector $\alpha \in V$ can be expressed uniquely in the form $\alpha = \alpha_1 + \alpha_2$ with $\alpha_i \in W_i$ for $i = 1, 2$. Now, the map $E : V \to V$ given by $\alpha \mapsto \alpha_1$ is a linear map (Check!) and is a projection (Check!) with range W_1 and kernel W_2 (Check all of this!). This is called the *projection onto* W_1 *(along* W_2*)*.

(iv) Note that any projection is diagonalizable. If R and N as above, choose ordered bases $\mathcal{B}_1$ for R and $\mathcal{B}_2$ for N. Then, by Lemma 5.3, $\mathcal{B} = (\mathcal{B}_1, \mathcal{B}_2)$ is an ordered basis for V. Now, note that, for each $\beta \in \mathcal{B}_1$, we have $E(\beta) = \beta$ and for each $\beta \in \mathcal{B}_2$, we have $E(\beta) = 0$. Hence, the matrix

$$[E]_{\mathcal{B}} = \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix}$$

where I denotes the $k \times k$ identity matrix, where $k = \dim(R)$.

Projections may be used to describe direct-sum decompositions: Suppose

$$V = W_1 \oplus W_2 \oplus \ldots \oplus W_k$$

then any vector $\alpha \in V$ can be expressed uniquely as

$$\alpha = \alpha_1 + \alpha_2 + \ldots + \alpha_k$$

with $\alpha_i \in W_i$ for all $1 \leq i \leq k$. Then, the map $E_j : V \to V$ defined by $\alpha \mapsto \alpha_j$ is a projection onto W_j, and

$$\ker(E_j) = W_1 + W_2 + \ldots + W_{j-1} + W_{j+1} + \ldots + W_k$$

Hence, in operator theoretic terms, this means that the identity operator decomposes as a sum of projections

$$I = E_1 + E_2 + \ldots + E_k$$

This leads to the next theorem.

Theorem 5.8. *If $V = W_1 \oplus W_2 \oplus \ldots \oplus W_k$, then there exist k linear operators $E_1, E_2, \ldots, E_k$ in $L(V)$ such that*

(i) Each E_j is a projection (ie. $E_j^2 = E_j$)

(ii) $E_iE_j = 0$ if $i \neq j$ (ie. they are mutually orthogonal*)*

(iii) $I = E_1 + E_2 + \ldots + E_k$

(iv) The range of E_j is W_j for all $1 \leq j \leq k$.

Conversely, if $E_1, E_2, \ldots, E_k \in L(V)$ are linear operators satisfying the conditions (i)-(iv), then

$$V = W_1 \oplus W_2 \oplus \ldots \oplus W_k$$

where W_i is the range of E_i.

Proof. We only prove the converse direction as one direction is described above. So suppose $E_1, E_2, \ldots, E_k \in L(V)$ are operators satisfying (i)-(iv), then we wish to show that

$$V = W_1 \oplus W_2 \oplus \ldots \oplus W_k$$

where W_i is the range of E_i.

(i) Since $I = E_1 + E_2 + \ldots E_k$, for any $\alpha \in V$, we have

$$\alpha = E_1(\alpha) + E_2(\alpha) + \ldots + E_k(\alpha)$$

and $E_i(\alpha) \in W_i$ for all $1 \leq i \leq k$. Hence,

$$V = W_1 + W_2 + \ldots + W_k$$

(ii) To show that the W_i are independent, suppose $\alpha_i \in W_i$ are such that

$$\alpha_1 + \alpha_2 + \ldots + \alpha_k = 0$$

Note that, for each $1 \leq i \leq k$, we have $\alpha_i = E_i(\alpha_i)$. So fix $1 \leq j \leq k$, and consider

$$0 = E_j(\alpha_1 + \alpha_2 + \ldots + \alpha_k) = E_j(E_1(\alpha_1) + E_2(\alpha_2) + \ldots + E_k(\alpha_k))$$

But $E_iE_j = 0$ if $i \neq j$, so we have

$$E_j(E_j(\alpha_j)) = 0$$

But $E_j(E_j(\alpha_j)) = \alpha_j$ so that $\alpha_j = 0$ for all $1 \leq j \leq k$. Hence, the $W_1, W_2, \ldots, W_k$ are independent as required.

□

6. Invariant Direct Sums

Given an operator $T \in L(V)$ on a vector space, we wish to find a decomposition

$$V = W_1 \oplus W_2 \oplus \ldots \oplus W_k$$

where each W_i is invariant under T. Write $T_i := T|_{W_i} \in L(W_i)$, then, for any vector $\alpha \in V$, write α uniquely as a sum

$$\alpha = \alpha_1 + \alpha_2 + \ldots + \alpha_k$$

with $\alpha_i \in W_i$, then

$$T(\alpha) = T_1(\alpha_1) + T_2(\alpha_2) + \ldots + T_k(\alpha_k)$$

If this happens, we say that T is a *direct sum* of the T_i's.

In terms of matrices, this is what this means: Given an ordered basis $\mathcal{B}_i$ of W_i, the basis $\mathcal{B} = (\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_k)$ is an ordered basis for V. Since W_i is T invariant, we see that

$$[T]_{\mathcal{B}} = \begin{pmatrix} A_1 & 0 & 0 & \ldots & 0 \\ 0 & A_2 & 0 & \ldots & 0 \\ 0 & 0 & A_3 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & A_k \end{pmatrix}$$

where

$$A_i = [T_i]_{\mathcal{B}_i}$$

Hence, $[T]_{\mathcal{B}}$ is a block diagonal matrix, and we say that $[T]_{\mathcal{B}}$ is a direct sum of the A_i's.

What we need to begin this discussion is an understanding of such a decomposition in terms of projections.

Theorem 6.1. *Let V be a vector space and $W_1, W_2, \ldots W_k$ be subspaces such that*

$$V = W_1 \oplus W_2 \oplus \ldots \oplus W_k$$

and let E_i be the projection onto W_i (as in Theorem 5.8). Let $T \in L(V)$, then each W_i is T-invariant if and only if

$$TE_i = E_iT$$

for all $1 \leq i \leq k$.

Proof.

(i) If $TE_i = E_iT$ for each $1 \leq i \leq k$, then for any $\alpha \in W_i$, we have $E_i(\alpha) = \alpha$, so

$$T(\alpha) = TE_i(\alpha) = E_iT(\alpha) \in W_i$$

Hence, W_i is T-invariant.

(ii) Conversely, if W_i is T-invariant, then for any $\alpha \in V$, we have $E_i(\alpha) \in W_i$, so

$$TE_i(\alpha) \in W_i$$

Hence,

$$E_iTE_i(\alpha) = TE_i(\alpha) \text{ and } E_jTE_i(\alpha) = 0 \quad \text{if } j \neq i$$

Now recall that we have

$$\alpha = E_1(\alpha) + E_2(\alpha) + \ldots + E_k(\alpha)$$

Hence,

$$E_iT(\alpha) = E_i\left(TE_1(\alpha) + TE_2(\alpha) + \ldots + TE_k(\alpha)\right) = E_iTE_i(\alpha)$$

But $TE_i(\alpha) \in W_i$ so $E_iTE_i(\alpha) = TE_i(\alpha)$. Hence,

$$E_iT(\alpha) = TE_i(\alpha)$$

This is true for any $\alpha \in V$, so we are done.

□

The next theorem may be thought of as an 'operator theoretic' characterization of diagonalizability (See Theorem 2.13).

Theorem 6.2. *Let V be a vector space and $T \in L(V)$ be an operator with $c_1, c_2, \ldots, c_k$ its distinct characteristic values. Suppose T is diagonalizable, then there exist operators $E_1, E_2, \ldots, E_k \in L(V)$ such that*

(i) $T = c_1E_1 + c_2E_2 + \ldots + c_kE_k$

(ii) $I = E_1 + E_2 + \ldots + E_k$

(iii) $E_iE_j = 0$ *if* $i \neq j$.

(iv) Each E_i *is a projection.*

(v) E_i *is the projection onto* $\ker(T - c_iI)$.

Conversely, suppose there are non-zero operators $E_1, E_2, \ldots, E_k \in L(V)$ and distinct scalars $c_1, c_2, \ldots, c_k$ satisfying conditions (i),(ii) and (iii), then

- *T is diagonalizable.*
- *The constants $c_1, c_2, \ldots, c_k$ are the distinct characteristic values of T.*
- *Conditions (iv) and (v) are satisfied.*

Proof.

(i) Suppose T is diagonalizable with distinct characteristic values $c_1, c_2, \ldots, c_k$. Let $W_i := \ker(T - c_iI)$, then by Theorem 2.13,

$$V = W_1 \oplus W_2 \oplus \ldots \oplus W_k$$

Let E_i be the projection onto W_i, then by Theorem 5.8, conditions (ii), (iii), (iv) and (v) are satisfied. Furthermore, if $\alpha \in V$, then write

$$\alpha = E_1(\alpha) + E_2(\alpha) + \ldots + E_k(\alpha)$$

so that

$$T(\alpha) = TE_1(\alpha) + TE_2(\alpha) + \ldots + TE_k(\alpha)$$

But $E_1(\alpha) \in W_1 = \ker(T - c_1I)$, so $TE_1(\alpha) = c_1E_1(\alpha)$. Similarly, we get

$$T(\alpha) = c_1E_1(\alpha) + c_2E_2(\alpha) + \ldots + c_kE_k(\alpha)$$

This is true for any $\alpha \in V$, so condition (i) holds.

(ii) Now suppose there are non-zero operators $E_1, E_2, \ldots E_k \in L(V)$ and distinct scalars $c_1, c_2, \ldots, c_k$ satisfying conditions (i), (ii) and (ii), then

- Since $I = E_1 + E_2 + \ldots + E_k$ and $E_iE_j = 0$ if $i \neq j$, we have

$$E_i = E_iI = E_i(E_1 + E_2 + \ldots + E_k) = E_i^2$$

Hence, E_i is a projection.

- Once again, since $T = c_1E_1 + c_2E_2 + \ldots + c_kE_k$ and $E_iE_j = 0$ if $i \neq j$, we have
$$TE_i = (c_1E_1 + c_2E_2 + \ldots + c_kE_k)E_i = c_iE_i^2 = c_iE_i$$
Since E_i is non-zero consider W_i to be the range of E_i, then for any $\beta \in W_i$, we have $E_i(\beta) = \beta$, so
$$T(\beta) = TE_i(\beta) = c_iE_i(\beta) = c_i\beta$$
Since $W_i \neq \{0\}$, it follows that each c_i is a characteristic value of T. Furthermore, we have that
$$W_i \subset \ker(T - c_iI)$$
- To show equality, suppose $\alpha \in \ker(T - c_iI)$, then
$$0 = (T - c_i)\alpha = \sum_{j=1}^{n}(c_j - c_i)E_i(\alpha)$$
For all $j \neq i$, it follows that $E_j(\alpha) = 0$ (Why?). Hence,
$$\alpha = E_i(\alpha) \in W_i$$
Thus, $W_i = \ker(T - c_i)$
- It remains to show that T does not have any *other* characteristic values other than the c_i: For any scalar $c \in F$ with $c \neq c_i$ for all $1 \leq i \leq k$, we have
$$(T - cI) = (c_1 - c)E_1 + (c_2 - c)E_2 + \ldots + (c_k - c)E_k$$
so if $\alpha \in V$ such that $(T - cI)\alpha = 0$, then
$$(c_i - c)E_i(\alpha) = 0(\text{Why?})$$
Since $c_i \neq c$, it follows that $E_i(\alpha) = 0$. This is true for all $1 \leq i \leq k$, so
$$\alpha = E_1(\alpha) + E_2(\alpha) + \ldots + E_k(\alpha) = 0$$
Thus, $c_1, c_2, \ldots, c_k$ are *all* the characteristic values of T.
- Finally, we wish to show that T is diagonalizable, but since the c_i are all distinct, the subspaces $W_1, W_2, \ldots, W_k$ are independent (see the proof of Lemma 2.12). Since
$$I = E_1 + E_2 + \ldots + E_k$$
we have
$$V = W_1 + W_2 + \ldots + W_k$$
so by Theorem 2.13, T is diagonalizable.

□

Theorem 6.3. *Let T be a diagonalizable operator, and express T as*

$$T = c_1E_1 + c_2E_2 + \ldots + c_kE_k$$

as in Theorem 6.2. Then, for any polynomial $g \in F[x]$, we have

$$g(T) = g(c_1)E_1 + g(c_2)E_2 + \ldots + g(c_k)E_k$$

Proof. By linearity of both sides, it suffices to verify the theorem if $g = x^m$. Suppose $g = x^2$, then

$$\begin{aligned} g(T) = T^2 &= \left(\sum_{i=1}^n c_iE_i\right)\left(\sum_{j=1}^n c_jE_j\right) \\ &= \sum_{i,j=1}^n c_ic_jE_iE_j \\ &= \sum_{i=1}^n c_i^2E_i^2 \\ &= \sum_{i=1}^n c_i^2E_i \\ &= \sum_{i=1}^n g(c_i)E_i \end{aligned}$$

Now proceed by induction on m (Do it!). □

Example 6.4.

(i) Express a diagonalizable operator T as

$$T = c_1E_1 + c_2E_2 + \ldots + c_kE_k$$

as in Theorem 6.2, and for $1 \leq j \leq k$, let p_j denote the Lagrange polynomials

$$p_j = \prod_{i\neq j} \frac{(x - c_i)}{(c_j - c_i)}$$

Then, we have $p_j(c_i) = \delta_{i,j}$. Hence, by Theorem 6.3,

$$p_j(T) = E_j$$

Thus, the E_j not only commute with T, they may be expressed as polynomials in T.

(ii) As a consequence of Theorem 6.3, for any polynomial $g \in F[x]$, we have

$$g(T) = 0 \Leftrightarrow g(c_i) = 0 \quad \forall 1 \leq i \leq k$$

In particular, if p is the minimal polynomial of T

$$p = (x - c_1)(x - c_2) \ldots (x - c_k)$$

(See Remark 3.8). Then $p(T) = 0$

This observation leads us to another proof of Theorem 4.17.

Theorem 6.5. *Let $T \in L(V)$ be a linear operator whose minimal polynomial is a product of distinct linear terms. Then, T is diagonalizable.*

Proof. Write the minimal polynomial of T as

$$p = (x - c_1)(x - c_2) \ldots (x - c_k)$$

Let $p_j \in F[x]$ be the Lagrange polynomial as above

$$p_j = \prod_{i \neq j} \frac{(x - c_i)}{(c_j - c_i)}$$

Then, $p_j(c_i) = \delta_{i,j}$. If $g \in F[x]$ is any polynomial of degree $\leq (k-1)$, then

$$g = \sum_{i=1}^{n} g(c_i) p_i$$

by the Lagrange interpolation formula (Remark IV.3.2). In particular, taking $g = 1$, the scalar polynomial, and taking $g = x$, we have

$$\begin{aligned} 1 &= p_1 + p_2 + \ldots + p_k \\ x &= c_1 p_1 + c_2 p_2 + \ldots + c_k p_k \end{aligned}$$

(Note that we are implicitly assuming that $k \geq 2$ - check what happens when $k = 1$). So if we set

$$E_j = p_j(T)$$

Then, we have

$$I = E_1 + E_2 + \ldots + E_k, \text{ and}$$
$$T = c_1 E_1 + c_2 E_2 + \ldots + c_k E_k$$

Now, for any pair $i \neq j$, consider the polynomial $p_i p_j$. Note that,

$$(x - c_r) \mid p_i p_j$$

for all $1 \leq r \leq k$. Hence, the minimal polynomial divides p_ip_j,

$$p \mid p_ip_j$$

But by the earlier observation, $p(T) = 0$. Hence,

$$E_iE_j = p_i(T)p_j(T) = 0$$

Finally, we note that $E_i = p_i(T) \neq 0$ because $\deg(p_i) < \deg(p)$ and p is the minimal polynomial of T. Since the $c_1, c_2, \ldots, c_k$ are all distinct, we may apply Theorem 6.2 to conclude that T is diagonalizable. □

(End of Week 10)

7. Simultaneous Triangulation; Simultaneous Diagonalization

Let V be a finite dimensional vector space, and let $\mathcal{F} \subset L(V)$ be a collection of linear operators on V. We wish to know when we can find an ordered basis $\mathcal{B}$ of V such that $[T]_{\mathcal{B}}$ is triangular (or diagonal) for all $T \in \mathcal{F}$. We will assume that $\mathcal{F}$ is a *commuting* family of operators, ie.

$$TS = ST$$

for all $S, T \in \mathcal{F}$.

Definition 7.1. For a subspace $W \subset V$, we say that W is $\mathcal{F}$-invariant if $T(W) \subset W$ for all $T \in \mathcal{F}$.

Compare the next lemma to Lemma 4.14.

Lemma 7.2. *Let $\mathcal{F}$ be a commuting family of triangulable operators on V, and let W be a proper $\mathcal{F}$-invariant subspace of V. Then, there exists $\alpha \in V$ such that*

(i) $\alpha \notin W$

(ii) For each $T \in \mathcal{F}, T\alpha$ is in the subspace spanned by W and α.

Proof. We first assume that $\mathcal{F}$ is finite, and write it as $\mathcal{F} = \{T_1, T_2, \ldots, T_n\}$.

(i) Since $T_1 \in \mathcal{F}$ is triangulable, the minimal polynomial of T_1 is a product of linear terms (by Theorem 4.15). By Lemma 4.14, there exists $\beta_1 \in V$ and a scalar $c_1 \in F$ such that

- $\beta_1 \notin W$
- $(T_1 - c_1I)\beta_1 \in W$

So, we set

$$V_1 := \{\beta \in V : (T_1 - c_1I)\beta \in W\}$$

Then, V_1 is a subspace of V (Check!) and $W \subset V_1$, $V_1 \neq W$.

(ii) We claim that V_1 is $\mathcal{F}$-invariant: Let $S \in \mathcal{F}$, then for any $\beta \in V_1$, we have

$$(T_1 - c_1 I)\beta \in W$$

Since W is $\mathcal{F}$-invariant, we have

$$S(T_1 - c_1 I)\beta \in W$$

But S and T_1 commute, so

$$(T_1 - c_1 I)S(\beta) \in W$$

By construction, this implies $S(\beta) \in V_1$. Hence, $S(V_1) \subset V_1$ for all $S \in \mathcal{F}$, proving the claim.

(iii) Let $U_2 = T_2|V_1$ be the restriction of T_2 to V_1. Since T_2 is diagonalizable, the minimal polynomial of T_2 is also a product of linear terms by Theorem 4.15. Since U_2 is a restriction of T_2, the minimal polynomial of U_2 is a divisor of the minimal polynomial of T_2 by Lemma 4.5. Hence, the minimal polynomial of U_2 is also a product of linear terms. Applying Lemma 4.14 to U_2 (the ambient vector space is now V_1, and W is the still the proper invariant subspace), there exists $\beta_2 \in V_1$ and a scalar $c_2 \in F$ such that

- $\beta_2 \notin W$
- $(T_2 - c_2 I) \in W$

Note that, since $\beta_2 \in V_1$, we also have

- $(T_1 - c_1 I)\beta_2 \in W$

Once again, set

$$V_2 := \{\beta \in V_1 : (T_2 - c_2 I)\beta \in W\}$$

Then, V_2 is a subspace of V_1, and properly contains W. Applying the same logic again, we see that V_2 is also $\mathcal{F}$-invariant. Therefore, we may set $U_3 := T_3|_{V_2}$ and proceed as before.

(iv) Proceeding this way, after finitely many steps, we arrive at a vector $\beta_n \in V$ such that

- $\beta_n \notin W$
- For each $1 \leq i \leq n$, there exist scalars $c_i \in F$ such that

$$(T_i - c_i I)\beta_n \in W$$

Thus, $\alpha := \beta_n$ works.

Now suppose $\mathcal{F}$ is not finite, choose a maximal linearly independent set $\mathcal{F}_0 \subset \mathcal{F}$ (ie. $\mathcal{F}_0$ is a basis for the subspace of $L(V)$ spanned by $\mathcal{F}$). Since $L(V)$ is finite dimensional (see Theorem III.2.4), $\mathcal{F}_0$ is finite. Therefore, by the first part of the proof, there exists $\alpha \in V$ such that

- $\alpha \notin W$
- $T\alpha \in \text{span}(W \cup \{\alpha\})$ for all $T \in \mathcal{F}_0$

Now if $T \in \mathcal{F}$, then T is a linear combination of elements from $\mathcal{F}_0$. Thus,

$$T\alpha \in \text{span}(W \cup \{\alpha\})$$

as well (since $\text{span}(W \cup \{\alpha\})$ is a subspace). Thus, α works for all of $\mathcal{F}$. □

The proof of the next theorem is identical to that of Theorem 4.15, except we replace a single operator by the set $\mathcal{F}$ and apply Lemma 7.2 instead of Lemma 4.14.

Theorem 7.3. *Let V be a finite dimensional vector space over a field F, and let $\mathcal{F}$ be a commuting family of triangulable operators on V. Then, there exists an ordered basis $\mathcal{B}$ of V such that $[T]_{\mathcal{B}}$ is upper triangular for each $T \in \mathcal{F}$.*

Corollary 7.4. *Let $\mathcal{F}$ be a commuting family of $n \times n$ matrices over a field F. Then, there exists an invertible matrix P such that $P^{-1}AP$ is triangular for all $A \in \mathcal{F}$.*

Theorem 7.5. *Let $\mathcal{F}$ be a commuting family of diagonalizable operators on a finite dimensional vector space V. Then, there exists an ordered basis $\mathcal{B}$ such that $[T]_{\mathcal{B}}$ is diagonal for all $T \in \mathcal{F}$.*

Proof. We induct on $n := \dim(V)$. If $n = 1$, there is nothing to prove, so assume that $n \geq 2$ and that the theorem is true over any vector space W with $\dim(W) < n$.

Now, if every operator in $\mathcal{F}$ is a scalar multiple of the identity, there is nothing to prove, so assume that this is not the case, and choose an operator $T \in \mathcal{F}$ that is not a scalar multiple of the identity. Let $\{c_1, c_2, \ldots, c_k\}$ be the characteristic values of T. Since T is diagonalizable and not a scalar multiple of I, it follows that $k > 1$. Set

$$W_i := \ker(T - c_iI), \qquad 1 \leq i \leq k$$

Then, $\dim(W_i) < n$ for all $1 \leq i \leq k$.

So fix $1 \leq i \leq k$, then, each W_i is $\mathcal{F}$-invariant (Check!). Let $\mathcal{F}_i := \{S_i := S|_{W_i} \colon S \in \mathcal{F}\}$, then $\mathcal{F}_i \subset L(W)$ is a commuting family of linear operators on W_i. Now, let $S \in \mathcal{F}$ be fixed, then S is diagonalizable, so its minimal polynomial is a product of distinct linear terms by Theorem 4.17. But the minimal polynomial of S_i divides the minimal polynomial of S (by Lemma 4.5). Hence, S_i is diagonalizable by Theorem 4.17 as well. Thus, the induction hypothesis applies, and there is an ordered basis $\mathcal{B}_i$ of W_i such that $[S_i]_{\mathcal{B}_i}$ is diagonal for all $S_i \in \mathcal{F}_i$.

Doing this for each $1 \leq i \leq k$, we obtain ordered bases $\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_k$ for $W_1, W_2, \ldots, W_k$ respectively. Now, by Theorem 2.13

$$\mathcal{B} = (\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_k)$$

is an ordered basis for V and each $S \in \mathcal{F}$ is diagonal in this basis. □

8. The Primary Decomposition Theorem

In the previous few sections, we have looked for a way to decompose an operator into simpler pieces. Specifically, given an operator $T \in L(V)$ on a finite dimensional vector space V, one looks for a decomposition of V into a direct sum of T-invariant subspaces

$$V = W_1 \oplus W_2 \oplus \ldots \oplus W_k$$

such that $T_i := T|_{W_i}$ is, in some sense, elementary. In the two theorems we proved (see Theorem 4.17 and Theorem 4.15), we found that we could do so provided the minimal polynomial of T factored as a product of linear terms.

However, this poses two possible problems: The minimal polynomial may not have enough roots (for instance, if the field is not algebraically closed), or if the characteristic subspaces do not span the entire vector space (this prevents diagonalizability). Now, we wish to find a general theorem that holds for *any* operator over *any* field.

This theorem is general, and therefore quite flexible, as it relies on only one fact, that holds over any field; namely, that any polynomial can be expressed uniquely as a product of irreducible polynomials (See Theorem IV.5.6). [It does, however, have the drawback of not being as powerful as Theorem 4.17 or Theorem 4.15.]

Theorem 8.1. *(Primary Decomposition Theorem) Let $T \in L(V)$ be a linear operator over a finite dimensional vector space over a field F. Let $p \in F[x]$ be the minimal polynomial of T, and express p as*

$$p = p_1^{r_1} p_2^{r_2} \ldots p_k^{r_k}$$

where the p_i are distinct irreducible polynomials in $F[x]$ and the r_i are positive integers. Set

$$W_i := \ker(p_i(T)^{r_i}), \qquad 1 \leq i \leq k$$

Then

(i) $V = W_1 \oplus W_2 \oplus \ldots \oplus W_k$

(ii) Each W_i is invariant under T.

(iii) If $T_i := T|_{W_i}$, then the minimal polynomial of T_i is $p_i^{r_i}$.

Proof.

(i) For $1 \leq i \leq k$, set

$$f_i := \prod_{j \neq i} p_j^{r_j} = \frac{p}{p_i^{r_i}}$$

Then, the polynomials $f_1, f_2, \ldots, f_k$ are relatively prime. Hence, by Example IV.4.17, there exist polynomials $g_1, g_2, \ldots, g_k \in F[x]$ such that

$$1 = f_1 g_1 + f_2 g_2 + \ldots + f_k g_k$$

Set $h_i := f_i g_i$, and set $E_i := h_i(T)$. We now use these operators E_i to construct a direct sum decomposition as in Theorem 5.8.

(ii) First note that

$$E_1 + E_2 + \ldots + E_k = I \tag{VI.3}$$

Now, if $i \neq j$, then $p_s^{r_s} \mid f_i f_j$ for all $1 \leq s \leq k$. Hence, $p \mid f_i f_j$. Thus,

$$f_i(T)f_j(T) = 0$$

Hence, $E_iE_j = 0$ if $i \neq j$.

(iii) Now, we claim that W_i is the range of E_i

- Suppose α is in the range of E_i, then $E_i(\alpha) = \alpha$. Then

$$\begin{aligned} p_i(T)^{r_i}\alpha &= p_i(T)^{r_i}h_i(T)\alpha \\ &= p_i(T)^{r_i}f_i(T)g_i(T)\alpha \\ &= p(T)g_i(T)\alpha \\ &= g_i(T)p(T)\alpha = 0 \end{aligned}$$

because $p(T) = 0$. Hence, $\alpha \in \ker(p_i(T)^{r_i}) = W_i$

- Conversely, if $\alpha \in W_i$, then $p_i(T)^{r_i}\alpha = 0$. But, for any $j \neq i$, we have

$$p_i^{r_i} \mid f_j$$

Therefore, $E_j\alpha = f_j(T)g_j(T)\alpha = g_j(T)f_j(T)\alpha = 0$. But by Equation VI.3, we have

$$\alpha = E_1(\alpha) + E_2(\alpha) + \ldots + E_k(\alpha) = E_i(\alpha)$$

and so α is in the range of E_i as required.

(iv) By construction, each W_i is invariant under T, since it is the kernel of an operator that commutes with T (See Example 4.2).

(v) We consider $T_i := T|_{W_i}$, and we wish to show that the minimal polynomial of T_i is $p_i^{r_i}$.

- Since $p_i(T)^{r_i}$ is zero on W_i, by definition, we have $p_i(T_i)^{r_i} = 0$
- Now suppose g is any polynomial such that $g(T_i) = 0$. We wish to show that

$$p_i^{r_i} \mid g$$

If α is in the range of E_i, then $g(T)\alpha = 0$. Thus, for all $\alpha \in V$,

$$g(T)E_i\alpha = 0 \Rightarrow g(T)f_i(T)g_i(T)\alpha = 0$$

This is true for all $\alpha \in V$, so $g(T)f_i(T)g_i(T) = 0$. Thus,

$$p \mid gf_ig_i \Rightarrow p_i^{r_i} \mid gf_ig_i \tag{VI.4}$$

Now consider the expression

$$f_1g_1 + f_2g_2 + \ldots + f_kg_k = 1$$

By construction,

$$p_i \mid f_j, \text{ if } i \neq j$$

So if $p_i \mid f_i g_i$, then it would follow that

$$p_i \mid 1$$

This contradicts the assumption that p_i is irreducible. Hence, $p_i \nmid f_i g_i$. Thus,

$$(p_i^{r_i}, f_i g_i) = 1$$

By Equation VI.4, and Euclid's Lemma (See the *proof* of Theorem IV.5.4), it follows that

$$p_i^{r_i} \mid g$$

Hence, the minimal polynomial of T_i must be $p_i^{r_i}$.

□

Corollary 8.2. *Let $T \in L(V)$ and let $E_1, E_2, \ldots, E_k$ be the projections occurring in the primary decomposition of T. Then,*

(i) Each E_i is a polynomial in T.

(ii) If $U \in L(V)$ is a linear operator that commutes with T, then U commutes with each E_i. Hence, each W_i is invariant under U.

Remark 8.3. Suppose that $T \in L(V)$ is such that the minimal polynomial is a product of (not necessarily distinct) linear terms

$$p = (x - c_1)^{r_1}(x - c_2)^{r_2} \ldots (x - c_k)^{r_k}$$

Let $E_1, E_2, \ldots, E_k$ be the projections occurring in the primary decomposition of T. Then the range of E_i is the null space of $(T - c_i)^{r_i}$ and is denoted by W_i. We set

$$D := c_1E_1 + c_2E_2 + \ldots c_kE_k$$

Then, by Theorem 6.2, D is diagonalizable, and is called the *diagonalizable part* of T. Note that, since each E_i is a polynomial in T, D is also a polynomial in T. Now, since $E_1 + E_2 + \ldots + E_k = I$, we have equations

$$\begin{aligned} &T = TE_1 + TE_2 + \ldots + TE_k \\ &D = c_1E_1 + c_2E_2 + \ldots + c_kE_k, \text{ and we set} \\ &N := T - D = (T - c_1)E_1 + (T - c_2)E_2 + \ldots + (T - c_k)E_k \end{aligned}$$

Then, N is also a polynomial in T. Hence,

$$ND = DN$$

Also, since the E_j are mutually orthogonal projections, we have

$$N^2 = (T - c_1)^2E_1 + (T - c_2)^2E_2 + \ldots + (T - c_k)^2E_k$$

Thus proceeding, if $r \geq \max\{r_i\}$, we have

$$N^r = (T - c_1)^rE_1 + (T - c_2)^rE_2 + \ldots + (T - c_k)^rE_k = 0$$

Definition 8.4. An operator $N \in L(V)$ is said to be *nilpotent* if there exists $r \in \mathbb{N}$ such that $N^r = 0$.

We leave the proof of the next lemma as an exercise.

Lemma 8.5. *If $T \in L(V)$ is both diagonalizable and nilpotent, then $T = 0$.*

Theorem 8.6. *Let $T \in L(V)$ be a linear operator whose minimal polynomial is a product of linear terms. Then, there exists a diagonalizable operator D and a nilpotent operator N such that*

(i) $T = D + N$

(ii) $ND = DN$

Furthermore, the operators D, N satisfying conditions (i) and (ii) are unique.

Proof. We have just proved the existence of these operators above. As for uniqueness, suppose

$$T = D' + N'$$

where D' is diagonalizable, N' is nilpotent, and $D'N' = N'D'$. Then, $D'T = TD'$ and $N'T = TN'$, and therefore, D' and N' both commute with any polynomial in T. In particular, we conclude that D' and N' commute with both D and N. Now, consider the equation

$$D + N = D' + N'$$

We conclude that

$$D - D' = N' - N$$

Now, the left hand side of this equation is the difference of two diagonalizable operators that commute with each other. Therefore, by Theorem 7.5, D and D' are simultaneously diagonalizable. This forces $(D - D')$ to be diagonalizable.

Now, consider the right hand side. N' and N are both commuting nilpotent operators. Suppose that $N^{r_1} = N'^{r_2} = 0$, consider

$$(N' - N)^r = \sum_{j=0}^{r} \binom{r}{j} N'^j N^{r-j}$$

which holds because they commute. So if we take $r := \max\{r_1, r_2\}$, then we conclude that

$$(N' - N)^r = 0$$

Hence, we have that $(D - D') = (N' - N)$ is both diagonalizable and nilpotent. By Lemma 8.5, we conclude that

$$D = D' \text{ and } N = N'$$

as required. □

Corollary 8.7. *Let V be a finite dimensional vector space over $\mathbb{C}$ and $T \in L(V)$ be an operator. Then, there exists a unique pair of operators (D, N) such that D is diagonalizable, N is nilpotent, $DN = ND$, and*

$$T = D + N$$

Furthermore, these operators are both polynomials in T.

Let us understand what the matrix equivalent of this theorem looks like. The next lemma states that an upper triangular matrix is nilpotent if all its diagonal entries are zero.

Lemma 8.8. *Let $A = (A_{i,j})$ be an $n \times n$ matrix such that $A_{i,j} = 0$ if $i \leq j$. Then, A is nilpotent.*

Proof. Consider

$$A^2(i,j) = \sum_{k=1}^{n} A_{i,k}A_{k,j} = \sum_{k=j+1}^{n} A_{i,k}A_{k,j}$$

Hence,

$$A^2(i,j) = 0 \text{ if } i \leq j+1$$

Thus proceeding, we see that

$$A^r(i,j) = 0 \text{ if } i \leq j+r-1$$

It follows that $A^n = 0$. □

Remark 8.9. Let $T \in L(V)$ be linear operator whose minimal polynomial is a product of distinct linear terms. Then, by Theorem 4.15, there is a basis $\mathcal{B}$ of V such that

$$A := [T]_{\mathcal{B}}$$

is upper triangular. Write

$$A = B + C$$

where B is diagonal, and $C = (C_{i,j})$ has the property that $C_{i,j} = 0$ for all $i \leq j$. By Lemma 8.8, C is nilpotent. Let $D, N \in L(V)$ be the unique linear operators such that

$$[D]_{\mathcal{B}} = B \text{ and } [N]_{\mathcal{B}} = C$$

Then, $T = D + N$ is the decomposition of T guaranteed by Theorem 8.6.

The next example is a continuation of Example 2.6.

Example 8.10. For instance, $T \in L(\mathbb{R}^3)$ is represented in the standard basis by the matrix

$$A = \begin{pmatrix} 3 & 1 & -1 \\ 2 & 2 & -1 \\ 2 & 2 & 0 \end{pmatrix}$$

(i) By our earlier calculations, the characteristic polynomial of A is

$$f = (x-2)^2(x-1)$$

For $c = 1$, we have $\ker(A - I) = \operatorname{span}\{\alpha_1\}$ where $\alpha_1 = (1, 0, 2)$, and for $c = 2$, we have $\ker(A - 2I) = \operatorname{span}\{\alpha_2\}$ where $\alpha_2 = (1, 1, 2)$.

(ii) Now consider

$$(A-2I)^2 = \begin{pmatrix} 1 & 1 & -1 \\ 2 & 0 & -1 \\ 2 & 2 & -2 \end{pmatrix} \begin{pmatrix} 1 & 1 & -1 \\ 2 & 0 & -1 \\ 2 & 2 & -2 \end{pmatrix} = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 0 & 0 \\ 2 & -2 & 0 \end{pmatrix}$$

This is row-equivalent to the matrix

$$C = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

which has nullity 2. Apart from α_2, if we set $\alpha_3 = (1, 1, 0)$, then

$$C\alpha_3 = 0$$

Since α_3 is not a scalar multiple of α_2, we conclude that

$$\ker((A-2I)^2) = \operatorname{span}\{\alpha_2, \alpha_3\}$$

(iii) Observe that

$$A\alpha_3 = \begin{pmatrix} 4 \\ 4 \\ 4 \end{pmatrix} = 2\alpha_2 + 2\alpha_3$$

(iv) Hence, $\mathcal{B} = \{\alpha_1, \alpha_3, \alpha_2\}$ is an ordered basis for $\mathbb{R}^3$, and

$$[T]_{\mathcal{B}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 2 \\ 0 & 0 & 2 \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}}_{\widehat{D}} + \underbrace{\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix}}_{\widehat{N}}$$

Observe that

$$\widehat{D}\widehat{N} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 4 \\ 0 & 0 & 0 \end{pmatrix} = \widehat{N}\widehat{D}$$

The first matrix on the right hand side is diagonal, and the second is nilpotent (by Lemma 8.8). Hence, if $D, N \in L(\mathbb{R}^3)$ are the unique linear operators such that

$$[D]_{\mathcal{B}} = \widehat{D}, \text{ and } [N]_{\mathcal{B}} = \widehat{N}$$

Then, D is diagonalizable, N is nilpotent, $DN = ND$, and

$$T = D + N$$

VII. The Rational and Jordan Forms

1. Cyclic Subspaces and Annihilators

Definition 1.1. Let $T \in L(V)$ be a linear operator on a finite dimensional vector space V and let $\alpha \in V$. The cyclic subspace generated by α is

$$Z(\alpha; T) := \{g(T)\alpha : g \in F[x]\} = \text{span}\{\alpha, T\alpha, T^2\alpha, \ldots\}$$

Lemma 1.2. *$Z(\alpha; T)$ is the smallest subspace of V that contains α and is invariant under T.*

Proof. Clearly, $Z(\alpha; T)$ is a subspace, it contains α and is invariant under T. So suppose M is any other subspace which contains α and is T=invariant, then $\alpha \in M$ implies that $T\alpha \in M$, and so $T^2(\alpha) \in M$ and so on. Hence,

$$T^k\alpha \in M$$

for all $k \geq 0$, whence $Z(\alpha; T) \subset M$. □

Definition 1.3. A vector $\alpha \in V$ is said to be a *cyclic* vector for T if $Z(\alpha; T) = V$.

Example 1.4.

(i) If $\alpha = 0$, then $Z(\alpha; T) = \{0\}$ for any operator T.

(ii) $Z(\alpha; T)$ is one dimensional if and only if α is a characteristic vector of T.

Proof. If α is a characteristic vector, then there exists $c \in F$ such that $T\alpha = c\alpha$. Hence, for any polynomial $g \in F[x]$, we have

$$g(T)\alpha = g(c)\alpha \in \text{span}\{\alpha\}$$

Thus, $Z(\alpha; T) = \text{span}\{\alpha\}$ and is therefore one dimensional. Conversely, if $Z(\alpha; T)$ is one-dimensional, then $\alpha \neq 0$ by part (i). Hence,

$$Z(\alpha; T) = \text{span}\{\alpha\}$$

In particular, since $T\alpha \in Z(\alpha; T)$, there exists $c \in F$ such that $T\alpha = c\alpha$ as required. □

(iii) If $T = I$ is the identity operator, then $Z(\alpha; T) = \text{span}\{\alpha\}$ for any vector $\alpha \in V$. In particular, if $\dim(V) > 1$, I does not have a cyclic vector.

(iv) Let $T \in L(\mathbb{R}^2)$ be the operator given in the standard basis by the matrix

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

- If $\alpha = \epsilon_2$, then $T(\alpha) = 0$, and so α is a characteristic vector of T. Hence,

$$Z(\alpha; T) = \text{span}\{\alpha\}$$

- If $\alpha = \epsilon_1$, then $T(\alpha) = \epsilon_2$. Hence, $Z(\alpha; T)$ contains both ϵ_1 and ϵ_2, whence

$$Z(\alpha; T) = \mathbb{R}^2$$

Definition 1.5. Given an operator $T \in L(V)$ and a vector $\alpha \in V$, the *T-annihilator of α* is the set

$$M(\alpha; T) := \{g \in F[x] : g(T)\alpha = 0\}$$

Lemma 1.6. *For any non-zero $T \in L(V)$, $M(\alpha; T)$ is a non-zero ideal of $F[x]$.*

Proof. That $M(\alpha; T)$ is an ideal is easy to check (Do it!). Also, the minimal polynomial of T is non-zero and is contained in $M(\alpha; T)$. Hence, $M(\alpha; T) \neq \{0\}$. □

The next definition is a slight abuse of notation.

Definition 1.7. The *T-annihilator of α* is the unique monic generator p_α of $M(\alpha; T)$

Remark 1.8. Since the minimal polynomial of T is contained in $M(\alpha; T)$, it follows that p_α divides the minimal polynomial. Furthermore, note that

$$\deg(p_\alpha) > 0$$

if $\alpha \neq 0$. (Why?)

Theorem 1.9. *Let $\alpha \in V$ be a non-zero vector and p_α be the T-annihilator of α.*

(i) $\deg(p_\alpha) = \dim(Z(\alpha; T))$.

(ii) If $\dim(p_\alpha) = k$, *then the set* $\{\alpha, T\alpha, T^2\alpha, \ldots, T^{k-1}\alpha\}$ *forms a basis for* $Z(\alpha; T)$.

(iii) If U is the linear operator on $Z(\alpha; T)$ induced by T, then the minimal polynomial of U is p_α.

Proof.

(i) We prove (i) and (ii) simultaneously.

- Let $S := \{\alpha, T\alpha, \ldots, T^{k-1}\alpha\} \subset Z(\alpha; T)$. If S is linearly dependent, then there would be a non-zero polynomial $q \in F[x]$ such that

$$q(T)\alpha = 0$$

and $\deg(q) < k$. This is impossible because p_α is the polynomial of least degree with this property. Hence, S is linearly independent.

- We claim that S spans $Z(\alpha;T)$: Suppose $g \in F[x]$, then by Euclidean division Theorem IV.4.2, there exists polynomials $q, r \in F[x]$ such that

  $$g = qp_\alpha + r$$

  and either $r = 0$ or $\deg(r) < k$. Then, since $p_\alpha(T)\alpha = 0$, it follows that

  $$g(T)\alpha = r(T)\alpha$$

  But, $r(T)\alpha \in \text{span}(S)$. Thus, S spans $Z(\alpha;T)$.

(ii) Now consider part (iii).

  (i) Since $p_\alpha(T)\alpha = 0$, it follows that

  $$p_\alpha(U)g(T)\alpha = p_\alpha(T)g(T)\alpha = g(T)p_\alpha(T)\alpha = 0$$

  This is true for any $g \in F[x]$. Hence,

  $$p_\alpha(U) = 0$$

  (ii) Now suppose $h \in F[x]$ is a non-zero polynomial of degree $< k$ such that $h(U) = 0$, then

  $$0 = h(U)\alpha = h(T)\alpha$$

  This is impossible because p_α is the polynomial of least degree with this property. Hence, $h(U) \neq 0$.

  Thus, p_α is the minimal polynomial of U.

□

Corollary 1.10. *If $\alpha \in V$ is a cyclic vector for T, then the minimal polynomial of T, the characteristic polynomial of T, and p_α all coincide.*

Proof. Let f and p denote the characteristic and minimal polynomials of T respectively. Then, by Cayley-Hamilton,

$$p \mid f$$

Furthermore, $p(T) = 0$ implies that $p(T)\alpha = 0$. Hence,

$$p_\alpha \mid p$$

Since all three polynomials are monic, it now suffices to prove that

$$p_\alpha = f$$

But $p_\alpha \mid f$ and by Theorem 1.9,

$$\deg(p_\alpha) = \dim(Z(\alpha;T)) = \dim(V) = \deg(f)$$

Hence, $p_\alpha = f$ and we are done. □

Remark 1.11. Let $U \in L(W)$ is a linear operator on a finite dimensional vector space W with cyclic vector α. Then, write the minimal polynomial of U as

$$p = p_\alpha = c_0 + c_1 x + \ldots + c_k x^k$$

Then, by Theorem 1.9, the set

$$\mathcal{B} = \{\alpha, U\alpha, U^2\alpha, \ldots, U^{k-1}\alpha\} =: \{\alpha_1, \alpha_2, \ldots, \alpha_k\}$$

forms an ordered basis for W. Now, if $0 \leq i \leq k-1$, then

$$U(\alpha_i) = \alpha_{i+1}$$

And, since $p_\alpha(U) = 0$, we have

$$c_0\alpha + c_1 U(\alpha) + c_2 U^2(\alpha) + \ldots + c_{k-1}U^{k-1}(\alpha) + U^k(\alpha) = 0$$

Hence,

$$U(\alpha_k) = -c_0\alpha_1 - c_1\alpha_2 - \ldots - c_{k-1}\alpha_k$$

Thus, the matrix of U in this basis is given by

$$[U]_\mathcal{B} = \begin{pmatrix} 0 & 0 & 0 & \ldots & 0 & -c_0 \\ 1 & 0 & 0 & \ldots & 0 & -c_1 \\ 0 & 1 & 0 & \ldots & 0 & -c_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 1 & -c_{k-1} \end{pmatrix}$$

Definition 1.12. Let $f \in F[x]$ be a monic polynomial written as

$$f = c_0 + c_1 x + c_2 x^2 + \ldots + c_k x^k$$

Then, the *companion matrix* of f is the matrix

$$A = \begin{pmatrix} 0 & 0 & 0 & \ldots & 0 & -c_0 \\ 1 & 0 & 0 & \ldots & 0 & -c_1 \\ 0 & 1 & 0 & \ldots & 0 & -c_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 1 & -c_{k-1} \end{pmatrix}$$

Theorem 1.13. *Let U be an operator on a finite dimensional vector space W. Then, U has a cyclic vector if and only if there is some ordered basis $\mathcal{B}$ of W such that the matrix $[U]_\mathcal{B}$ is the companion matrix of the minimal polynomial of U.*

Proof. We have just proved one direction: If U has a cyclic vector, then there is an ordered basis $\mathcal{B}$ such that $[U]_\mathcal{B}$ is the companion matrix to the minimal polynomial.

Conversely, suppose there is a basis $\mathcal{B}$ such that $[U]_{\mathcal{B}}$ is the companion matrix to the minimal polynomial p, then write $\mathcal{B} = \{\alpha_1, \alpha_2, \ldots, \alpha_k\}$. Observe that, by construction,

$$\alpha_i = U^{i-1}(\alpha_1)$$

for all $i \geq 1$. Thus, $Z(\alpha_1; U) = V$ because it contains $\mathcal{B}$. Thus, α_1 is a cyclic vector for U. □

Corollary 1.14. *If A is the companion matrix of a monic polynomial $p \in F[x]$, then p is both the minimal polynomial and characteristic polynomial of A.*

Proof. Let $T \in L(F^n)$ be the linear operator whose matrix in the standard basis is A. Then, T has a cyclic vector by Theorem 1.13. By Corollary 1.10, the minimal and characteristic polynomials of T coincide. Therefore, it suffices to show that the characteristic polynomial of A is p. We prove this by induction on $k := \deg(p)$.

(i) If $\deg(p) = 1$, then write $p = c_0 + x$, then

$$A = \begin{pmatrix} -c_0 \end{pmatrix}$$

so the characteristic polynomial of A is clearly $f = x + c_0 = p$.

(ii) Suppose that the theorem is true for any polynomial q with $\deg(q) < k$, and write

$$A = \begin{pmatrix} 0 & 0 & 0 & \ldots & 0 & -c_0 \\ 1 & 0 & 0 & \ldots & 0 & -c_1 \\ 0 & 1 & 0 & \ldots & 0 & -c_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 1 & -c_{k-1} \end{pmatrix}$$

Hence, the characteristic polynomial of A is given by

$$\begin{aligned} f &= \det \begin{pmatrix} x & 0 & 0 & \ldots & 0 & c_0 \\ -1 & x & 0 & \ldots & 0 & c_1 \\ 0 & -1 & x & \ldots & 0 & c_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & -1 & x + c_{k-1} \end{pmatrix} \\ &= x \det \begin{pmatrix} x & 0 & \ldots & 0 & c_1 \\ -1 & x & \ldots & 0 & c_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & -1 & x + c_{k-1} \end{pmatrix} + \det \begin{pmatrix} 0 & 0 & \ldots & 0 & c_0 \\ -1 & x & \ldots & 0 & c_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & -1 & x + c_{k-1} \end{pmatrix} \end{aligned}$$

Now, the first matrix that appears on the right hand side is the companion matrix to the polynomial

$$q = c_1 + c_2 x + \ldots + c_{k-1} x^{k-2} + x^{k-1}$$

So by induction, the first term is

$$x(c_1 + c_2x + \ldots + c_{k-1}x^{k-2} + x^{k-1})$$

Now by expanding along the first row of the second term, we get

$$(-1)^{k-1}c_0 \det \begin{pmatrix} -1 & x & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & -1 \end{pmatrix}$$

This matrix is an upper triangular matrix with -1's along the diagonal. Hence, this term becomes

$$(-1)^{k-1}c_0(-1)^{k-1} = c_0$$

Thus, we get

$$f = x(c_1 + c_2x + \ldots + c_{k-1}x^{k-2} + x^{k-1}) + c_0 = p$$

as required.

□

(End of Week 11)

2. Cyclic Decompositions and the Rational Form

The goal of this section is to show that, for a given operator $T \in L(V)$, there exist vectors $\alpha_1, \alpha_2, \ldots, \alpha_r \in V$ such that

$$V = Z(\alpha_1; T) \oplus Z(\alpha_2; T) \oplus \ldots \oplus Z(\alpha_r; T)$$

In other words, V is the direct sum of T-invariant subspaces such that each subspace has a T-cyclic vector. This is a hard theorem, and we break it into many smaller pieces, so it is easier to digest.

Definition 2.1. Let $W \subset V$ be a subspace of V. A subspace $W' \subset V$ is said to be *complementary* to W if $V = W \oplus W'$.

Remark 2.2. Let $T \in L(V)$ and $W \subset V$ be a T-invariant subspace, and suppose there is a complementary subspace $W' \subset V$ which is also T-invariant. Then, for any $\beta \in V$, there exist $\gamma \in W$ and $\gamma' \in W'$ such that

$$\beta = \gamma + \gamma'$$

Since W and W' are both T-invariant, for any polynomial $f \in F[x]$, we have

$$f(T)\beta = f(T)\gamma + f(T)\gamma'$$

where $f(T)\gamma \in W$ and $f(T)\gamma' \in W'$. Hence, if $f(T)\beta \in W$, then it must happen that $f(T)\gamma' = 0$, and in that case,

$$f(T)\beta = f(T)\gamma$$

This leads to the next definition.

Definition 2.3. Let $T \in L(V)$ and $W \subset V$ a subspace. We say that W is T-*admissible* if

(i) W is T-invariant.

(ii) For any $\beta \in V$ and $f \in F[x]$, if $f(T)\beta \in W$, then there exists $\gamma \in W$ such that
$$f(T)\beta = f(T)\gamma$$

We now make some comments about T-conductors.

Remark 2.4. Fix $T \in L(V)$, and a T-invariant subspace $W \subset V$.

(i) Given a vector $\alpha \in V$, we had defined (See Definition VI.4.9) the T-conductor of α into W as the ideal
$$S(\alpha; W) := \{f \in F[x] : f(T)\alpha \in W\}$$
We saw in Lemma VI.4.11 that $S(\alpha; W)$ is an ideal of $F[x]$, and its unique monic generator is also termed the T-conductor of α into W. We denote this polynomial by
$$s(\alpha; W)$$

(ii) Now, different vectors have different T-conductors. Furthermore, for any $\alpha \in W$, we have
$$\deg(s(\alpha; W)) \leq \dim(V)$$
because the minimal polynomial has degree $\leq \dim(W)$ (by Cayley-Hamilton), and every T-conductor must divide the minimal polynomial. We say that a vector $\beta \in V$ is *maximal with respect to* W if
$$\deg(s(\beta; W)) = \max_{\alpha \in V} \deg(s(\alpha; W))$$
In other words, the degree of the T-conductor of β is the highest among all T-conductors into W. Note that such a vector always exists.

The next few lemmas are all part of a larger theorem - the cyclic decomposition theorem. The textbook states it as one theorem, but we have broken into smaller parts for ease of reading.

Lemma 2.5. *Let $T \in L(V)$ be a linear operator and $W_0 \subsetneq V$ be a proper T-invariant subspace. Then, there exist non-zero vectors $\beta_1, \beta_2, \ldots, \beta_r$ in V such that*

(i) $V = W_0 + Z(\beta_1; T) + Z(\beta_2; T) + \ldots + Z(\beta_r; T)$

(ii) For $1 \leq k \leq r$, if we set
$$W_k := W_0 + Z(\beta_1; T) + Z(\beta_2; T) + \ldots + Z(\beta_k; T)$$
Then each β_k is maximal with respect to W_{k-1}.

Proof. Since W_0 is T-invariant, there exists $\beta_1 \in V$ which is maximal with respect to W_0. Note that, by Example VI.4.10, for any $\alpha \in W_0$, $s(\alpha; W) = 1$ if and only if $\alpha \in W_0$. Since $W \neq V$, there exists some $\beta \in V$ such that $\beta \notin W$. Hence,

$$\deg(s(\beta; W)) > 0$$

Therefore, $\beta_1 \notin W$ and

$$\deg(s(\beta_1; W)) > 0$$

Hence, if we set

$$W_1 := W_0 + Z(\beta_1; T)$$

Then $W_0 \subsetneq W_1$. If $W_1 = V$, then there is nothing more to do. If not, we observe that W_1 is also T-invariant, and repeat the process. Each time, we increase the dimension by at least one, so since V is finite dimensional, this process must terminate after finitely many steps. □

Remark 2.6. In the decomposition above, note that

$$W_k = W_0 + Z(\beta_1; T) + Z(\beta_2; T) + \ldots + Z(\beta_k; T)$$

Hence, if $\alpha \in W_k$, then one can express α as a sum

$$\alpha = \beta_0 + g_1(T)\beta_1 + g_2(T)\beta_2 + \ldots + g_k(T)\beta_k$$

for some polynomials $g_i \in F[x]$, and $\beta_0 \in W_0$. Note, however, that this expression may not be unique.

Lemma 2.7. *Let $T \in L(V)$ and $W_0 \subsetneq V$ be a T-admissible subspace. Let $\beta_1, \beta_2, \ldots, \beta_r \in V$ be vectors satisfying the conditions of Lemma 2.5. Fix a vector $\beta \in V$ and $1 \leq k \leq r$, and set*

$$f := s(\beta; W_{k-1})$$

to be the T-conductor of β into W_{k-1}. Suppose further that $f(T)\beta \in W_{k-1}$ has the form

$$f(T)\beta = \beta_0 + \sum_{i=1}^{k-1} g_i(T)\beta_i \tag{VII.1}$$

where $\beta_0 \in W_0$. Then

(i) $f \mid g_i$ for all $1 \leq i \leq k-1$

(ii) $\beta_0 = f(T)\gamma_0$ for some $\gamma_0 \in W_0$.

Proof. If $k = 1$, then $f = s(\beta; W_0)$ so that

$$f(T)\beta = \beta_0 \in W_0$$

Hence, part (i) of the conclusion is vacuously true, and part (ii) is precisely the requirement that W_0 is T-admissible.

Now suppose $k \geq 2$.

(i) For each $1 \leq i \leq k-1$, apply Euclidean division (Theorem IV.4.2) to write

$$g_i = fh_i + r_i, \text{ such that } r_i = 0 \text{ or } \deg(r_i) < \deg(f)$$

We wish to show that $r_i = 0$ for all $1 \leq i \leq k-1$. Suppose not, then there is a largest value $1 \leq j \leq k-1$ such that $r_j \neq 0$.

(ii) To that end, set

$$\gamma := \beta - \sum_{i=1}^{k-1} h_i(T)\beta_i \in W_{k-1}$$

Then, $\gamma - \beta \in W_{k-1}$, so (Why?)

$$s(\gamma; W_{k-1}) = s(\beta; W_{k-1}) = f$$

Furthermore,

$$f(T)\gamma = \beta_0 + \sum_{i=1}^{k-1} r_i(T)\beta_i$$

With j as above, we get

$$f(T)\gamma = \beta_0 + \sum_{i=1}^{j} r_i(T)\beta_i \text{ and } r_j \neq 0, \deg(r_j) < \deg(f) \tag{VII.2}$$

(iii) Now set $p := s(\gamma, W_{j-1})$. Since $W_{j-1} \subset W_{k-1}$, we have

$$f = s(\gamma; W_{k-1}) \mid p$$

So choose $g \in F[x]$ such that $p = fg$. Applying $g(T)$ to Equation VII.2, we get

$$p(T)\gamma = g(T)\beta_0 + \sum_{i=1}^{j-1} g(T)r_i(T)\beta_i + g(T)r_j(T)\beta_j$$

By definition, $p(T)\gamma \in W_{j-1}$, and the first two terms on the right-hand-side are also in W_{j-1}. Hence,

$$g(T)r_j(T)\beta_j \in W_{j-1}$$

Hence, $s(\beta_j; W_{j-1}) \mid gr_j$. By condition (ii) of Lemma 2.5, we have

$$\begin{aligned}\deg(gr_j) \geq \deg(s(\beta_j; W_{j-1}))& \\ &\geq \deg(s(\gamma; W_{j-1})) \\ &= \deg(p) = \deg(fg)\end{aligned}$$

Hence, it follows that

$$\deg(r_j) \geq \deg(f)$$

which is a contradiction. Thus, it follows that $r_i = 0$ for all $1 \leq i \leq k-1$. Hence,

$$f \mid g_i$$

for all $1 \leq i \leq k-1$.

(iv) Finally, back in Equation VII.1, we have

$$f(T)\beta = \beta_0 + \sum_{i=1}^{k-1} f(T)h_i(T)\beta_i$$

Hence, we conclude that

$$\beta_0 = f(T)\gamma$$

with γ defined as above. This completes the proof.

□

Recall that (Definition VI.4.9), for an operator $T \in L(V)$ and a vector $\alpha \in V$, the T-annihilator of α is the unique monic generator of the ideal

$$S(\alpha; \{0\}) := \{f \in F[x] : f(T)\alpha = 0\}$$

Theorem 2.8 (Cyclic Decomposition Theorem - Existence)**.** *Let $T \in L(V)$ be a linear operator on a finite dimensional vector space V, and let W_0 be a proper T-admissible subspace of V. Then, there exist non-zero vectors $\alpha_1, \alpha_2, \ldots, \alpha_r$ in V with respective T-annihilators $p_1, p_2, \ldots, p_r$ such that*

(i) $V = W_0 \oplus Z(\alpha_1; T) \oplus Z(\alpha_2; T) \oplus \ldots \oplus Z(\alpha_r; T)$

(ii) $p_k \mid p_{k-1}$ *for all* $k = 2, 3, \ldots, r$.

Note that the above theorem is typically applied with $W_0 = \{0\}$.

Proof.

(i) Start with vectors $\beta_1, \beta_2, \ldots, \beta_r \in V$ satisfying the conditions of Lemma 2.5. To each vector $\beta = \beta_k$ and the T-conductor $f = p_k$, we apply Lemma 2.7 to obtain

$$p_k(T)\beta_k = p_k(T)\gamma_0 + \sum_{i=1}^{k-1} p_k(T)h_i(T)\beta_i$$

where $\gamma_0 \in W_0$ and $h_1, h_2, \ldots, h_r$ are polynomials. Now set

$$\alpha_k := \beta_k - \gamma_0 - \sum_{i=1}^{k-1} h_i(T)\beta_i \tag{VII.3}$$

Then, $\beta_k - \alpha_k \in W_{k-1}$, so (Why?)

$$s(\alpha_k; W_{k-1}) = s(\beta_k; W_{k-1}) = p_k$$

(ii) Note that

$$p_k(T)\alpha_k = 0$$

But, $p_k = s(\alpha_k; W_{k-1})$. Hence, it follows that, for any polynomial $f \in F[x]$, if $f(T)\alpha \in W_{k-1}$, then $p_k \mid f$. But this implies that $f(T)\alpha = 0$. Hence,

$$W_{k-1} \cap Z(\alpha_k; T) = \{0\}$$

(iii) Now if $\alpha \in W_k$, then since $W_k = W_{k-1} + Z(\beta_k; T)$, we write

$$\alpha = \beta + g(T)\beta_k$$

for some $g \in F[x]$ and $\beta \in W_{k-1}$. Using Equation VII.3, we have

$$g(T)\beta_k = g(T)\alpha_k + g(T)\gamma_0 + \sum_{i=1}^{k-1} g(T)h_i(T)\beta_i$$

Hence, if $\beta' := \beta + g(T)\gamma_0 + \sum_{i=1}^{k-1} g(T)h_i(T)\beta_i$, we have $\beta' \in W_{k-1}$ and

$$\alpha = \beta' + g(T)\alpha_k$$

Thus,

$$W_k = W_{k-1} + Z(\alpha_k; T)$$

(iv) By the previous step, we conclude that

$$W_k = W_{k-1} \oplus Z(\alpha_k; T)$$

By induction, we have

$$W_k = W_0 \oplus Z(\alpha_1; T) \oplus Z(\alpha_2; T) \oplus \ldots \oplus Z(\alpha_k; T)$$

In particular, we have

$$V = W_0 \oplus Z(\alpha_1; T) \oplus Z(\alpha_2; T) \oplus \ldots \oplus Z(\alpha_r; T)$$

(v) To verify condition (ii), observe that $p_i(T)\alpha_i = 0$ for all $1 \leq i \leq r$, so

$$p_k(T)\alpha_k = 0 = 0 + p_1(T)\alpha_1 + p_2(T)\alpha_2 + \ldots + p_{k-1}(T)\alpha_{k-1}$$

Taking $\beta = \alpha_k$ and $f = p_k$ in Lemma 2.7, we conclude that

$$p_k \mid p_i$$

for all $1 \leq i \leq k-1$.

□

Definition 2.9. Let $T \in L(V)$ be a fixed operator, $f \in F[x]$ a polynomial, and $W \subset V$ a subspace. We write

$$fW := \{f(T)\alpha : \alpha \in W\}$$

Note that this is also a subspace of V.

Lemma 2.10. *Let $T \in L(V)$ and $f \in F[x]$.*

(i) If $\alpha \in V$, then $fZ(\alpha; T) = Z(f(T)\alpha; T)$.

(ii) If $V = V_1 \oplus V_2 \oplus \ldots \oplus V_k$ where each V_i is T-invariant, then

$$fV = fV_1 \oplus fV_2 \oplus \ldots \oplus fV_k$$

(iii) If $\alpha, \gamma \in V$ both have the same T-annihilator, then $f(T)\alpha$ and $f(T)\gamma$ have the same T-annihilator. Therefore,

$$\dim(Z(f(T)\alpha; T)) = \dim(Z(f(T)\gamma; T))$$

Proof.

(i) We prove containment both ways: If $\beta \in Z(\alpha; T)$, then write $\beta = g(T)\alpha$, so that

$$f(T)\beta = f(T)g(T)\alpha = g(T)f(T)\alpha \in Z(f(T)\alpha; T)$$

Hence, $fZ(\alpha; T) \subset Z(f(T)\alpha; T)$. The reverse containment is similar.

(ii) If $\beta \in V$, then write

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

where $\beta_i \in V_i$. Then, since each V_i is T-invariant, we have

$$f(T)\beta = f(T)\beta_1 + f(T)\beta_2 + \ldots + f(T)\beta_k \in fV_1 + fV_2 + \ldots + fV_k$$

Now, suppose $\gamma \in fV_j \cap (fV_1 + fV_2 + \ldots + fV_{j-1})$, then write

$$\gamma = f(T)\beta_j = f(T)\beta_1 + f(T)\beta_2 + \ldots + f(T)\beta_{j-1}$$

where $\beta_i \in V_i$ for all $1 \leq i \leq j$. But each V_i is T-invariant, so

$$f(T)\beta_i \in V_i$$

for all $1 \leq i \leq j$. By Lemma VI.5.3,

$$V_j \cap (V_1 + V_2 + \ldots + V_{j-1}) = \{0\}$$

Hence, $\gamma = f(T)\beta_j = 0$. Thus,

$$fV_j \cap (fV_1 + fV_2 + \ldots + fV_{j-1}) = \{0\}$$

So by Lemma VI.5.3, we get part (ii).

(iii) Suppose p denotes the common annihilator of α and γ, and let g and h denote the annihilators of $f(T)\alpha$ and $f(T)\gamma$ respectively. Then

$$g(T)f(T)\alpha = 0$$

Hence, $p \mid gf$, so that

$$g(T)f(T)\gamma = 0$$

Hence, $h \mid g$. Similarly, $g \mid h$. Since both are monic, we conclude that $g = h$. Finally, the fact that

$$\dim(Z(f(T)\alpha; T)) = \dim(Z(f(T)\gamma; T))$$

follows from Theorem 1.9.

□

Definition 2.11. Let $T \in L(V)$ be a linear operator and $W \subset V$ be a T-invariant subspace. Define

$$S(V;W) := \{f \in F[x] : f(T)\alpha \in W \quad \forall \alpha \in V\}$$

Then, it is clear (as in Lemma VI.4.11) that $S(V;W)$ is a non-zero ideal of $F[x]$. We denote its unique monic generator by p_W.

Theorem 2.12 (Cyclic Decomposition Theorem - Uniqueness). *Let $T \in L(V)$ be a linear operator on a finite dimensional vector space V and let W_0 be a proper T-admissible subspace of V. Then, suppose we are given non-zero vectors $\alpha_1, \alpha_2, \ldots, \alpha_r$ in V with respective T-annihilators $p_1, p_2, \ldots, p_r$ such that*

(i) $V = W_0 \oplus Z(\alpha_1;T) \oplus Z(\alpha_2;T) \oplus \ldots \oplus Z(\alpha_r;T)$

(ii) $p_k \mid p_{k-1}$ *for all* $k = 2, 3, \ldots, r$.

And suppose we are given non-zero vectors $\gamma_1, \gamma_2, \ldots, \gamma_s \in V$ with T-annihilators $g_1, g_2, \ldots, g_s$ such that

(i) $V = W_0 \oplus Z(\gamma_1;T) \oplus Z(\gamma_2;T) \oplus \ldots \oplus Z(\gamma_s;T)$

(ii) $g_k \mid g_{k-1}$ *for all* $k = 2, 3, \ldots, s$.

Then, $r = s$ and $\gamma_i = \alpha_i$ for all $1 \leq i \leq r$.

Proof. (i) We begin by showing that $p_1 = g_1$. In fact, we show that

$$p_1 = g_1 = p_{W_0}$$

For $\beta \in V$, we write

$$\beta = \beta_0 + f_1(T)\gamma_1 + f_2(T)\gamma_2 + \ldots + f_s(T)\gamma_s$$

for some $\beta_0 \in W_0$ and polynomials $f_i \in F[x]$. Then

$$g_1(T)\beta = g_1(T)\beta_0 + \sum_{i=1}^{s} g_1(T)f_i(T)\gamma_i$$

By hypothesis, $g_i \mid g_1$ for all $i \geq 1$. Since $g_i(T)\gamma_i = 0$, we conclude that

$$g_1(T)\beta = g_1(T)\beta_0 \in W_0$$

Hence, $g_1 \in S(V;W_0)$.

(ii) Now suppose $f \in S(V; W_0)$, then, in particular, $f(T)\gamma_1 \in W_0$. But, $f(T)\gamma_1 \in Z(\gamma_1; T)$ as well. Since this is a direct sum decomposition,

$$f(T)\gamma_1 = 0$$

But g_1 is the T-annihilator of γ_1, so it follows that $g_1 \mid f$. Hence, we conclude that

$$g_1 = p_{W_0}$$

By symmetry, $p_1 = p_{W_0}$ as well. Hence, $g_1 = p_1$.

(iii) Now suppose $r \geq 2$, then

$$\dim(W_0) + \dim(Z(\alpha_1; T)) < \dim(V)$$

Since $p_1 = g_1$, by Lemma 2.10, we have

$$\dim(Z(\alpha_1; T)) = \dim(Z(\gamma_1; T))$$

Hence,

$$\dim(W_0) + \dim(Z(\gamma_1; T)) < \dim(V)$$

Thus, we must have that $s \geq 2$ as well.

(iv) We now show that $p_2 = g_2$. Observe that, by Lemma 2.10, we have

$$p_2 V = p_2 W_0 \oplus Z(p_2(T)\alpha_1; T) \oplus Z(p_2(T)\alpha_2; T) \oplus \ldots \oplus Z(p_2(T)\alpha_r; T)$$

Similarly, we have

$$p_2 V = p_2 W_0 \oplus Z(p_2(T)\gamma_1; T) \oplus Z(p_2(T)\gamma_2; T) \oplus \ldots \oplus Z(p_2(T)\gamma_s; T) \qquad \text{(VII.4)}$$

However, since $p_i \mid p_2$ for all $i \geq 2$, we have $p_2(T)\alpha_i = 0$ for all $i \geq 2$. Thus, the first sum reduces to

$$p_2 V = p_2 W_0 \oplus Z(p_2(T)\alpha_1; T) \qquad \text{(VII.5)}$$

Now note that $p_1 = g_1$. Therefore, by Lemma 2.10, we conclude that

$$\dim(Z(p_2(T)\alpha_1; T)) = \dim(Z(p_2(T)\gamma_1; T))$$

By comparing Equation VII.4 and Equation VII.5, we conclude that

$$\dim(Z(p_2(T)\gamma_i; T)) = 0$$

for all $i \geq 2$. Hence,

$$p_2(T)\gamma_i = 0$$

for all $i \geq 2$. In particular, we must have $g_2 \mid p_2$.

(v) By symmetry, we have $p_2 \mid g_2$ as well. Therefore, $p_2 = g_2$.

(vi) By proceeding in this way, we conclude that $r = s$ and $p_i = g_i$ for all $1 \leq i \leq r$. □

Remark 2.13. Let $T \in L(V)$ be a linear operator. Applying Theorem 2.8 with $W_0 = \{0\}$ gives a decomposition of V into a direct sum of cyclic subspaces

$$V = Z(\alpha_1; T) \oplus Z(\alpha_2; T) \oplus \ldots \oplus Z(\alpha_r; T)$$

and let $p_1, p_2, \ldots, p_r$ be the T-annihilators of α_i so that $p_k \mid p_{k-1}$ for all $k \geq 2$.

Now, consider the restriction T_i of T to $Z(\alpha_i; T)$, and let $\mathcal{B}_i$ be the ordered basis of $Z(\alpha_i; T)$ from Theorem 1.9

$$\mathcal{B}_i = \{\alpha_i, T(\alpha_i), T^2(\alpha_i), \ldots, T^{k_i - 1}(\alpha_i)\}$$

where $k_i = \deg(p_i)$. Then the matrix

$$A_i = [T_i]_{\mathcal{B}_i}$$

is the companion matrix of p_i (See Remark 1.11). Furthermore, if we take $\mathcal{B} := (\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_r)$, then $\mathcal{B}$ is an ordered basis of V (by Lemma VI.5.3), and the matrix of T in this basis has the form

$$[T]_{\mathcal{B}} = \begin{pmatrix} A_1 & 0 & 0 & \ldots & 0 \\ 0 & A_2 & 0 & \ldots & 0 \\ 0 & 0 & A_3 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & A_r \end{pmatrix}$$

Furthermore, by Theorem 2.12, the polynomials occurring in this decomposition are unique, and therefore, the companion matrices are also unique.

Definition 2.14. An $n \times n$ matrix over a field F is said to be in *rational form* if A can be expressed as a direct sum of matrices $A_1, A_2, \ldots, A_r$, where each A_i is the companion matrix of a monic polynomial $p_i \in F[x]$, and, furthermore, $p_i \mid p_{i-1}$ for all $i = 2, 3, \ldots, r$.

Therefore, a consequence of the existence and uniqueness of the cyclic decomposition gives us the following result.

Corollary 2.15. *Let B denote an $n \times n$ matrix over a field F. Then, B is similar to one and only one matrix A over F which is in rational form.*

Proof. Let $T \in L(F^n)$ be the linear operator given by $T(X) = BX$. Then, by Remark 2.13, there is an ordered basis $\mathcal{B}$ of F^n such that

$$A := [T]_{\mathcal{B}}$$

is in rational form. Now, B is similar to A by Theorem III.4.8. Let $p_i, 1 \leq i \leq r$ denote the polynomials associated to this matrix A.

Now suppose C is another matrix in rational form, expressed as a direct sum of matrices $C_i, 1 \leq i \leq s$, where each C_i is the companion matrix of a monic polynomial $g_i \in F[x]$ satisfying $g_i \mid g_{i-1}$ for all $i \geq 2$. Let $\mathcal{B}_i \subset \mathcal{B}$ be the basis for the subspace W_i of F^n corresponding to the summand C_i. Then,

$$[C_i]_{\mathcal{B}_i}$$

is the companion matrix of g_i. By Corollary 1.14, the minimal and characteristic polynomials of C_i are both g_i. By Theorem 1.13, C_i has a cyclic vector β_i. Thus,

$$W_i = Z(\beta_i; T)$$

so that

$$F^n = Z(\beta_1; T) \oplus Z(\beta_2; T) \oplus \ldots \oplus Z(\beta_s; T)$$

By the uniqueness in Theorem 2.12, it follows that $r = s$ and $g_i = p_i$ for all $1 \leq i \leq r$. Hence,

$$C = A$$

as required. □

Definition 2.16. Let $T \in L(V)$, and consider the (unique) polynomials $p_1, p_2, \ldots, p_r$ occurring in the cyclic decomposition of T. These are called the *invariant factors* of T.

These invariant factors are uniquely determined by T. Furthermore, we have the following fact.

Lemma 2.17. *Let $T \in L(V)$ be a linear operator with invariant factors $p_1, p_2, \ldots, p_r$ satisfying $p_i \mid p_{i-1}$ for all $i = 2, 3, \ldots, r$. Then,*

(i) The minimal polynomial of T is p_1

(ii) The characteristic polynomial of T is $p_1 p_2 \ldots p_r$.

Proof.

(i) Assume without loss of generality that $V \neq \{0\}$. Apply Theorem 2.8 and write

$$V = Z(\alpha_1; T) \oplus Z(\alpha_2; T) \oplus \ldots \oplus Z(\alpha_r; T)$$

where the T-annihilators $p_1, p_2, \ldots, p_r$ of $\alpha_1, \alpha_2, \ldots, \alpha_r$ are such that $p_k \mid p_{k-1}$ for all $k = 2, 3, \ldots, r$. If $\alpha \in V$, then write

$$\alpha = f_1(T)\alpha_1 + f_2(T)\alpha_2 + \ldots + f_r(T)\alpha_r$$

for some polynomials $f_i \in F[x]$. For each $1 \leq i \leq r$, $p_i \mid p_1$, so it follows that

$$p_1(T)\alpha = \sum_{i=1}^{r} p_1(T) f_i(T)\alpha_i = \sum_{i=1}^{r} f_i(T) p_1(T)\alpha_i = 0$$

Hence, $p_1(T) = 0$. Furthermore, since p_1 is the T-annihilator of α_1, it follows that, for any polynomial $q \in F[x]$, if $\deg(q) < \deg(p_1)$, then $q(T)\alpha_1 \neq 0$. Hence, p_1 is the minimal polynomial of T.

(ii) As for the characteristic polynomial of T, consider the basis $\mathcal{B}$ in Remark 2.13 such that

$$A := [T]_{\mathcal{B}}$$

is in rational form. Write

$$A = \begin{pmatrix} A_1 & 0 & 0 & \dots & 0 \\ 0 & A_2 & 0 & \dots & 0 \\ 0 & 0 & A_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & A_r \end{pmatrix}$$

where each A_i is the companion matrix of p_i. Then, the characteristic polynomial of T is the characteristic polynomial of A. However, A is a block-diagonal matrix, so the characteristic polynomial of A is given by the product of all the characteristic polynomials of the A_i (This follows from Theorem V.4.3). But, by Corollary 1.14, the characteristic polynomial of A_i is p_i. Hence, the characteristic polynomial of T is $p_1 p_2 \dots p_r$.

□

Recall that (Corollary 1.10) if $T \in L(V)$ has a cyclic vector, then its characteristic and minimal polynomials both coincide. The next corollary is a converse of this fact.

Corollary 2.18. *Let $T \in L(V)$ be a linear operator on a finite dimensional vector space.*

(i) There exists $\alpha \in V$ such that the T-annihilator of α is the minimal polynomial of T.

(ii) T has a cyclic vector if and only if the characteristic and minimal polynomials of T coincide.

Proof.

(i) Take $\alpha = \alpha_1$, then p_1 is the T-annihilator of α, which is also the minimal polynomial of T by Lemma 2.17.

(ii) If T has a cyclic vector, then the characteristic and minimal polynomial coincide by Corollary 1.10. So suppose the characteristic and minimal polynomial coincide, and label it $p \in F[x]$. Then, $\deg(p) = \dim(V)$. So if we choose $\alpha \in V$ from part (i), then, by Theorem 1.9,

$$\dim(Z(\alpha; T)) = \deg(p) = \dim(V)$$

So $V = Z(\alpha; T)$, and thus, T has a cyclic vector.

□

Example 2.19.

(i) Suppose $T \in L(V)$ where $\dim(V) = 2$, and consider two possible cases:

   (i) The minimal polynomial of T has degree 2: Then the minimal polynomial and characteristic polynomial must coincided, so T has a cyclic vector by Corollary 2.18. Hence, there is a basis $\mathcal{B}$ of V such that the matrix of T is
$$[T]_{\mathcal{B}} = \begin{pmatrix} 0 & -c_0 \\ 1 & -c_1 \end{pmatrix}$$
   is the companion matrix of its minimal polynomial.

   (ii) The minimal polynomial of T has degree 1, then T is a scalar multiple of the identity. Thus, there is a scalar $c \in F$ such that, for any basis $\mathcal{B}$ of V, one has
$$[T]_{\mathcal{B}} = \begin{pmatrix} c & 0 \\ 0 & c \end{pmatrix}$$

(ii) LeT $T \in L(\mathbb{R}^3)$ be the linear operator represented in the standard ordered basis by
$$A = \begin{pmatrix} 5 & -6 & -6 \\ -1 & 4 & 2 \\ 3 & -6 & -4 \end{pmatrix}$$
In Example VI.2.15, we calculated the characteristic polynomial of T to be
$$f = (x-1)(x-2)^2$$
Furthermore, we showed that T is diagonalizable. Since the minimal and characteristic polynomials must share the same roots (by Theorem VI.3.7) and the minimal polynomial must be a product of distinct linear factors (by Theorem VI.4.17), it follows that the minimal polynomial of T is
$$p = (x-1)(x-2)$$
So consider the cyclic decomposition of T, given by
$$\mathbb{R}^3 = Z(\alpha_1; T) \oplus Z(\alpha_2; T) \oplus \ldots \oplus Z(\alpha_r; T)$$
By Theorem 1.9, $\dim(Z(\alpha_1; T))$ is the degree of its T-annihilator. However, by Lemma 2.17, this T-annihilator is the minimal polynomial of T. Thus,
$$\dim(Z(\alpha_1; T)) = \deg(p) = 2$$
Since $\dim(\mathbb{R}^3) = 3$, there can be atmost one more summand, so $r = 2$. Hence
$$\mathbb{R}^3 = Z(\alpha_1; T) \oplus Z(\alpha_2; T)$$
And furthermore, $\dim(Z(\alpha_2; T)) = 1$. Hence, α_2 must be a characteristic vector of T by Example 1.4. Furthermore, the T-annihilator of α_2, denoted by p_2 must satisfy
$$pp_2 = f$$

by Lemma 2.17. Hence,

$$p_2 = (x-2)$$

so the characteristic value associated to α_2 is 2. Thus, the rational form of T is

$$B = \begin{pmatrix} 0 & -2 & 0 \\ 1 & 3 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

where the upper left-hand block is the companion matrix of the polynomial

$$p = (x-1)(x-2) = x^2 - 3x + 2$$

We had seen in Remark 2.2 that if $W_0 \subset V$ is a T-invariant subspace which has a complementary subspace that is also T-invariant, then W_0 is T-admissible. The next corollary is the converse of this fact.

Corollary 2.20. *Let $T \in L(V)$ be a linear operator on a finite dimensional vector space and W_0 be T-admissible subspace of V. Then there is a subspace W_0' that is complementary to W_0 that is also T-invariant.*

Proof. Let W_0 be a T-admissible subspace. If $W_0 = V$, then take $W_0' = \{0\}$. If not, then apply Theorem 2.8 to write

$$V = W_0 \oplus Z(\alpha_1; T) \oplus Z(\alpha_2; T) \oplus \ldots \oplus Z(\alpha_r; T)$$

and take

$$W_0' := Z(\alpha_1; T) \oplus Z(\alpha_2; T) \oplus \ldots \oplus Z(\alpha_r; T)$$

□

Theorem 2.21 (Generalized Cayley-Hamilton Theorem). *Let $T \in L(V)$ be a linear operator on a finite dimensional vector space with minimal polynomial p and characteristic polynomial f. Then,*

(i) $p \mid f$

(ii) *p and f have the same prime factors, except for multiplicities.*

(iii) *If the prime factorization of p is given by*

$$p = f_1^{r_1} f_2^{r_2} \ldots f_k^{r_k}$$

then the prime factorization of f is

$$f = f_1^{d_1} f_2^{d_2} \ldots f_k^{d_k}$$

where

$$d_i = \frac{\dim(\ker(f_i(T)^{r_i}))}{\deg(f_i)}$$

Proof. Consider invariant factors $p_1, p_2, \ldots, p_r$ of T such that $p_i \mid p_{i-1}$ for all $i \geq 2$. Then, by Lemma 2.17, we have

$$p_1 = p \text{ and } f = p_1 p_2 \ldots p_r$$

Therefore, part (i) follows. Furthermore, if $q \in F[x]$ is an irreducible polynomial such that

$$q \mid f$$

then by Theorem IV.5.4, there exists $1 \leq i \leq r$ such that $q \mid p_i$. However, $p_i \mid p_1 = p$, so

$$q \mid p$$

Thusm part (ii) follows as well.

Finally, consider the primary decomposition of T Theorem VI.8.1. Here, we get

$$V = W_1 \oplus W_2 \oplus \ldots \oplus W_k$$

where each $W_i = \ker(f_i(T)^{r_i})$ and the minimal polynomial of $T_i = T|_{W_i}$ is $f_i^{r_i}$. Now, apply part (ii) of this result to the operator T_i. The minimal polynomial of T_i is $f_i^{r_i}$, so the characteristic polynomial of T_i must be of the form

$$f_i^{d_i}$$

for some $d_i \geq r_i$. Furthermore, it is clear that $d_i \deg(f_i)$ is the degree of this characteristic polynomial, so

$$d_i \deg(f_i) = \dim(W_i)$$

Hence,

$$d_i = \frac{\dim(W_i)}{\deg(f_i)} = \frac{\dim(\ker(f_i(T)^{r_i})}{\deg(f_i)}$$

But, as discussed in Lemma 2.17, the characteristic polynomial of T is the product of the characteristic polynomials of the T_i. Hence,

$$f = f_1^{d_1} f_2^{d_2} \ldots f_k^{d_k}$$

as required. □

(End of Week 12)

3. The Jordan Form

In this section, we wish to give another description of a linear operator $T \in L(V)$ in terms of a 'simple' matrix. This time, we start with the observation from Theorem VI.8.6, that T can be expressed as a sum

$$T = D + N$$

where D is diagonal, N is nilpotent, and $DN = ND$. We begin with the following observation about the cyclic decomposition of a nilpotent operator.

Definition 3.1. A $k \times k$ *elementary nilpotent matrix* is a matrix of the form

$$A = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 \end{pmatrix}$$

Note that such a matrix A (being a lower triangular matrix with zeroes along the diagonal), is a nilpotent matrix by an analogue of Lemma VI.8.8.

Lemma 3.2. *Let N be a nilpotent operator on a finite dimensional vector space V. Then, there is an ordered basis $\mathcal{B}$ of V such that*

$$A := [N]_{\mathcal{B}} = \begin{pmatrix} A_1 & 0 & 0 & \dots & 0 \\ 0 & A_2 & 0 & \dots & 0 \\ 0 & 0 & A_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & A_r \end{pmatrix}$$

where each A_i is a $k_i \times k_i$ elementary nilpotent matrix. Here, $k_1, k_2, \dots, k_r$ are positive integers such that

$$k_1 + k_2 + \ldots + k_r = n \text{ and } r = \text{nullity}(N)$$

Furthermore, we may arrange that

$$k_1 \geq k_2 \geq \ldots \geq k_r$$

Proof.

(i) Consider the cyclic decomposition of N obtained from Theorem 2.8

$$V = Z(\alpha_1; N) \oplus Z(\alpha_2; N) \oplus \ldots Z(\alpha_r; N) \tag{VII.6}$$

where $\alpha_1, \alpha_2, \dots, \alpha_r \in V$ are non-zero vectors with T-annihilators $p_1, p_2, \dots, p_r$ such that $p_{i+1} \mid p_i$ for all $i = 1, 2, \dots, r-1$. Since N is nilpotent, the minimal polynomial of N is

$$p = x^k$$

for some $k \leq n$. Since each T-annihilator divides the minimal polynomial (See Remark 1.8), each p_i is of the form x^{k_i}, and the divisibility condition implies that

$$k_1 \geq k_2 \geq \ldots \geq k_r$$

Furthermore, by Lemma 2.17, we know that $p_1 = p$, so that $k_1 = k$.

(ii) Now, the companion matrix for x^{k_i} is precisely the $k_i \times k_i$ elementary nilpotent matrix. Thus, one obtains an ordered basis $\mathcal{B}$ such that

$$A := [T]_{\mathcal{B}}$$

has the required form.

(iii) We now verify that $r = \text{nullity}(N)$. To do this, we show that the set

$$S := \{N^{k_1-1}\alpha_1, N^{k_2-1}\alpha_2, \ldots, N^{k_r-1}\alpha_r\}$$

forms a basis for $\ker(N)$.

- Note that $N^{k_i-1}\alpha_i \in Z(\alpha_i; N)$. Since these subspaces as independent, it follows that S is linearly independent.
- To show that S spans $\ker(N)$, fix $\alpha \in \ker(N)$. By the decomposition of Equation VII.6, we write

  $$\alpha = f_1(N)\alpha_1 + f_2(N)\alpha_2 + \ldots + f_r(N)\alpha_r$$

  for some polynomials $f_i \in F[x]$. Furthermore, by Theorem 1.9, we may assume that

  $$\deg(f_i) < \deg(p_i) = k_i$$

  for all $1 \leq i \leq r$. Now observe that

  $$0 = N(\alpha) = \sum_{i=1}^{r} Nf_i(N)\alpha_i$$

  Once again, since the decomposition of Equation VII.6 is a direct sum decomposition, it follows that

  $$Nf_i(N)\alpha_i = 0$$

  Hence, $p_i \mid xf_i$. But since $\deg(f_i) < k_i$. it follows that

  $$f_i = c_i x^{k_i-1}$$

  for some constant $c_i \in F$. Thus,

  $$\alpha = \sum_{i=1}^{r} c_i N^{k_i-1}\alpha_i$$

  Thus, S spans $\ker(N)$ as required.

□

Definition 3.3. A $k \times k$ *elementary Jordan matrix with characteristic value* c is a matrix of the form

$$A = \begin{pmatrix} c & 0 & 0 & \dots & 0 & 0 \\ 1 & c & 0 & \dots & 0 & 0 \\ 0 & 1 & c & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & c \end{pmatrix}$$

Theorem 3.4. *Let $T \in L(V)$ be a linear operator over a finite dimensional vector space V whose characteristic polynomial factors as a product of linear terms. Then, there is an ordered basis $\mathcal{B}$ of V such that the matrix*

$$A := [T]_{\mathcal{B}} = \begin{pmatrix} A_1 & 0 & 0 \dots & 0 & \\ 0 & A_2 & 0 & \dots & 0 \\ 0 & 0 & A_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & A_k \end{pmatrix}$$

where, for each $1 \leq i \leq k$, there are distinct scalars c_i such that A_i is of the form

$$A_i = \begin{pmatrix} J_1^{(i)} & 0 & 0 & \dots & 0 \\ 0 & J_2^{(i)} & 0 & \dots & 0 \\ 0 & 0 & J_3^{(i)} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & J_{n_i}^{(i)} \end{pmatrix}$$

where each $J_j^{(i)}$ is an elementary Jordan matrix with characteristic value c_i. Furthermore, the sizes of matrices $J_j^{(i)}$ decreases as j increases. Furtheremore, this form is uniquely associated to T.

Proof.

(i) Existence: We may assume that the characteristic polynomial of T is of the form

$$f = (x - c_1)^{d_1}(x - c_2)^{d_2} \dots (x - c_k)^{d_k}$$

where $c_1, c_2, \dots, c_k$ are distinct scalars and $d_i \geq 1$. By Theorem 2.21, the minimal polynomial is of the form

$$p = (x - c_1)^{r_1}(x - c_2)^{r_2} \dots (x - c_k)^{r_k}$$

for some integers $0 < r_k \leq d_k$. If $W_i := \ker(T - c_i I)^{r_i}$, then the primary decomposition theorem (Theorem VI.8.1) says that

$$V = W_1 \oplus W_2 \oplus \dots \oplus W_k$$

Let T_i denote the operator on W_i induced by T. Then, the minimal polynomial of T_i is

$$p_i = (x - c_i)^{r_i}$$

Let $N_i := (T_i - c_i I) \in L(W_i)$, then N_i is nilpotent, and has minimal polynomial

$$q_i := x^{r_i}$$

Furthermore,

$$T_i = N_i + c_i I$$

Now, choose an ordered basis $\mathcal{B}_i$ of W_i from Lemma 3.2 so that

$$B_i := [N_i]_{\mathcal{B}_i}$$

is a direct sum of elementary nilpotent matrices. Then,

$$A_i := [T_i]_{\mathcal{B}_i}$$

is a direct sum of elementary Jordan matrices with characteristic value c_i. Furthermore, by the constrution of Lemma 3.2, the Jordan matrices appearing in each A_i increase in size as we go down the diagonal.

(ii) Uniqueness:

- If A_i is a $d_i \times d_i$ matrix, then the characteristic polynomial of A_i is

  $$(x - c_i)^{d_i}$$

  Hence, by Theorem V.4.3, the characteristic polynomial of A (and hence of T) is

  $$f = (x - c_1)^{d_1}(x - c_2)^{d_2} \ldots (x - c_k)^{d_k}$$

  Thus, it follows that, upto ordering, $c_1, c_2, \ldots, c_k$ and $d_1, d_2, \ldots, d_k$ are uniquely determined.

- Now, the direct sum decomposition of A into the A_i gives a direct sum decomposition

  $$V = W_1 \oplus W_2 \oplus \ldots \oplus W_k$$

  We claim that $W_i = \ker((T - c_i I)^n)$ where $n = \dim(V)$. Clearly,

  $$(A - c_i I)^n \equiv 0 \text{ on } W_i$$

  Furthermore, $\det(A_j - c_i I) \neq 0$, so

  $$W_i = \ker(T - c_i)^n$$

  Hence, the subspaces W_i are uniquely determined.

- Finally, if T_i denotes the restriction of T to W_i, then the matrix A_i is the rational form of T_i. Hence, A_i is uniquely determined by the uniqueness of the rational form (Theorem 2.12).

□

Definition 3.5. An $n \times n$ matrix A that is in the form described in Theorem 3.4 is called a *Jordan matrix*, and is called the *Jordan form* of the associated linear operator.

Remark 3.6. We make some observations about a Jordan matrix A.

(i) Every entry of A not on or immediately below the principal diagonal is zero.

(ii) One the diagonal of A occur the k distinct characteristic values of T. Also, each characteristic value c_i is repeated d_i times, where d_i is the multiplicity of the c_i as a root of the characteristic polynomial.

(iii) For each i, the matrix A_i is the direct sum of n_i elementary Jordan matrices $J_j^{(i)}$ with characteristic value c_i. Furthermore,

$$n_i = \dim \ker(T - c_i I)$$

Hence, T is diagonalizable if and only if $n_i = d_i$ for all $1 \leq i \leq k$.

(iv) For each $1 \leq i \leq k$, the first block $J_1^{(i)}$ in the matrix A_i is an $r_i \times r_i$ matrix, where r_i is the multiplicity of c_i as a root of the minimal polynomial of T. This is because the minimal polynomial of the nilpotent operator $(T_i - c_i I)$ is x^{r_i}.

Corollary 3.7. *If B is an $n \times n$ matrix over a field F and if the characteristic polynomial of B factors completely over F, then B is similar over F to an $n \times n$ matrix A in Jordan form, and A is unique upto rearrangement of the order of its characteristic values.*

This matrix A is called the *Jordan form* of B. Note that the above corollary automatically applies to matrices over algebraically closed fields such as $\mathbb{C}$.

Example 3.8.

(i) Suppose $T \in L(V)$ with $\dim(V) = 2$ where V is a complex vector space. Then, the characteristic polynomial of T is either of the form

$$f = (x - c_1)(x - c_2) \text{ for } c_1 \neq c_2 \text{ or } (x - c)^2$$

In the first case, T is diagonalizable and is Jordan form is

$$A = \begin{pmatrix} c_1 & 0 \\ 0 & c_2 \end{pmatrix}$$

In the second case, the minimal polynomial of T may be either $(x - c)$ or $(x - c)^2$. If the minimal polynomial is $(x - c)$, then

$$T = cI$$

If the minimal polynomial is $(x - c)^2$, then the Jordan form of T is

$$A = \begin{pmatrix} c & 0 \\ 1 & c \end{pmatrix}$$

(ii) Let A be the 3×3 complex matrix given by

$$A = \begin{pmatrix} 2 & 0 & 0 \\ a & 2 & 0 \\ b & c & -1 \end{pmatrix}$$

The characteristic polynomial of T is

$$f = (x-2)^2(x+1)$$

- If the minimal polynomial of T is f, then A is similar to the matrix

$$B = \begin{pmatrix} 2 & 0 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

- If the minimal polynomial of T is $(x-2)(x+1)$, then A is similar to the matrix

$$B = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

Now,

$$(A-2I)(A+I) = \begin{pmatrix} 0 & 0 & 0 \\ 3a & 0 & 0 \\ ac & 0 & 0 \end{pmatrix}$$

Thus, A is similar to a diagonal matrix if and only if $a = 0$.

VIII. Inner Product Spaces

1. Inner Products

Given two vector $\alpha = (x_1, x_2, x_3), \beta = (y_1, y_2, y_3) \in \mathbb{R}^3$, the 'dot product' of these vectors is given by

$$(\alpha|\beta) = x_1y_1 + x_2y_2 + x_3y_3$$

The dot product simultaneously allows us to define two geometric concepts: The length of a vector is defined as

$$\|\alpha\| := (\alpha|\alpha)^{1/2}$$

and the angle between two vectors can be measured by

$$\theta := \cos^{-1}\left(\frac{(\alpha|\beta)}{\|a\|\|\beta\|}\right)$$

While we will not usually care about the 'angle' between two vectors in an arbitary vector space, we will care about when two vectors are *orthogonal*, ie. $(\alpha|\beta) = 0$. The abstract notion of an inner product allows us to introduce this kind of geometry into the study of vector spaces.

Note that, throughout the rest of this course, all fields will necessarily have to be either $\mathbb{R}$ or $\mathbb{C}$.

Definition 1.1. An *inner product* on a vector space V over a field F ($= \mathbb{R}$ or $\mathbb{C}$) is a function

$$V \times V \to F \text{ given by } (\alpha, \beta) \mapsto (\alpha|\beta)$$

such that, for all vectors $\alpha, \beta, \gamma \in V$ and $c \in F$, we have

(i) $(\alpha + \beta|\gamma) = (\alpha|\gamma) + (\beta|\gamma)$

(ii) $(c\alpha|\beta) = c(\alpha|\beta)$

(iii) $(\beta|\alpha) = \overline{(\alpha|\beta)}$, where $\bar{\cdot}$ denotes complex conjugation.

(iv) $(\alpha|\alpha) \geq 0$ and $(\alpha|\alpha) = 0$ if and only if $\alpha = 0$.

Note that, from (ii) and (iii), it follows that

$$(\alpha|c\beta) = \overline{c}(\alpha|\beta)$$

Example 1.2.

(i) Let $V = F^n$ with the *standard inner product*: If $\alpha = (x_1, x_2, \dots, x_n), \beta = (y_1, y_2, \dots, y_n) \in V$, we define

$$(\alpha|\beta) = \sum_{i=1}^{n} x_i\overline{y_i}$$

If $F = \mathbb{R}$, this is

$$(\alpha|\beta) = \sum_{i=1}^{n} x_i y_i$$

(ii) If $F = \mathbb{R}$ and $V = \mathbb{R}^2$, we may define another inner product by

$$(\alpha|\beta) = x_1y_1 - x_2y_1 - x_1y_2 + 4x_2y_2$$

where $\alpha = (x_1, x_2)$ and $\beta = (y_1, y_2)$. Note that

$$(\alpha|\alpha) = (x_1 - x_2)^2 + 3x_2^2$$

so it safisfies condition (iv) of Definition 1.1. The other axioms can also be verified (Check!).

(iii) Let $V = F^{n\times n}$, then the standard inner product on F^{n^2} may be borrowed to V to give

$$(A|B) = \sum_{i,j} A_{i,j}\overline{B_{i,j}}$$

We define the *conjugate transpose* of a matrix B (denoted by B^*) by

$$(B^*)_{i,j} = \overline{B_{j,i}}$$

Then, it follows that

$$(A|B) = \sum_{i,j} A_{i,j}B^*_{j,i} = \text{trace}(AB^*) = \text{trace}(B^*A)$$

(iv) Let $V = F^{n\times 1}$ be the space of $n \times 1$ (column) matrices over F and let Q be a fixed $n \times n$ invertible matrix. For $X, Y \in V$, define

$$(X|Y) := Y^*Q^*QX$$

This is an inner product on V. When $Q = I$, then this can be identified with the standard inner product from Example (i).

(v) Let $V = C[0, 1]$ be the space of all continuous, complex-valued functions defined on the interval $[0, 1]$. For $f, g \in V$, we define

$$(f|g) := \int_0^1 f(t)\overline{g(t)}dt$$

Then, this is an inner product on $C[0, 1]$.

(vi) Let W be a vector space and $(\cdot|\cdot)$ be a fixed inner product on W. We may construct new inner products as follows. Let $T : V \to W$ be a fixed injective (non-singular) linear transformation, and define $p_T : V \times V \to F$ by

$$p_T(\alpha, \beta) := (T\alpha|T\beta)$$

Then this defines an inner product p_T on V. We give some special cases of this example.

- Let V be a finite dimensional vector space with a fixed ordered basis $\mathcal{B} = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$. Let $W = F^n$ with the standard ordered basis $\{\epsilon_1, \epsilon_2, \ldots, \epsilon_n\}$ and let $T : V \to W$ be an isomorphism such that

$$T(\alpha_j) = \epsilon_j$$

for all $1 \leq j \leq n$ (See Theorem III.3.2). Then, we may use T to inherit the standard inner product from Example (i) by

$$p_T(\sum_{j=1}^n x_j\alpha_j, \sum_{i=1}^n y_i\alpha_i) = \sum_{k=1}^n x_k\overline{y_k}$$

In particular, for any basis $\mathcal{B} = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ of V, there is a inner product $(\cdot|\cdot)$ on V such that

$$(\alpha_i, \alpha_j) = \delta_{i,j}$$

for all $1 \leq i, j \leq n$.

- Now take $V = W = C[0, 1]$ and $T : V \to W$ be the operator

$$T(f)(t) := tf(t)$$

Then, T is non-singular (Check!), and the new inner product on V inherited from the inner product from Example (v) is

$$p_T(f, g) := \int_0^1 f(t)\overline{g(t)}t^2 dt$$

Remark 1.3. Let $(\cdot|\cdot)$ be a fixed inner product on a vector space V. Then,

$$(\alpha|\beta) = \text{Re}(\alpha|\beta) + i\text{Im}(\alpha|\beta)$$

But, for any $z \in \mathbb{C}$, one has $\text{Im}(z) = \text{Re}(-iz)$, so

$$\text{Im}(\alpha|\beta) = \text{Re}[-i(\alpha|\beta)] = \text{Re}(\alpha|i\beta)$$

Hence,

$$(\alpha|\beta) = \text{Re}(\alpha|\beta) + \text{Re}(\alpha|i\beta)$$

Thus, the inner product is completely determined by its ‘real part’.

Definition 1.4. Let V be an inner product space. For a vector $\alpha \in V$, the *norm* of α is the scalar

$$\|\alpha\| := (\alpha|\alpha)^{1/2}$$

Note that this is well-defined because $(\alpha|\alpha) \geq 0$ for all $\alpha \in V$. Furthermore, axiom (iv) implies that $\|\alpha\| = 0$ if and only if $\alpha = 0$. Thus, the norm of a vector may profitably be thought of as the 'length' of the vector.

Remark 1.5. The norm and inner product are intimately related to each other. For instance, one has (Check!)

$$\|\alpha \pm \beta\|^2 = \|a\|^2 \pm 2\text{Re}(\alpha|\beta) + \|b\|^2$$

for all $\alpha, \beta \in V$. Hence, if $F = \mathbb{R}$, one has

$$(\alpha|\beta) = \frac{1}{4}\|\alpha + \beta\|^2 - \frac{1}{4}\|\alpha - \beta\|^2$$

and if $F = \mathbb{C}$, one has

$$(\alpha|\beta) = \frac{1}{4}\|\alpha + \beta\|^2 - \frac{1}{4}\|\alpha - \beta\|^2 + \frac{i}{4}\|\alpha + i\beta\|^2 - \frac{i}{4}\|\alpha - i\beta\|^2$$

(Please verify this statement!). These equations show that the inner product may be recovered from the norm, and are called the *polarization identities.* They may be written (in the complex case) as

$$(\alpha|\beta) = \frac{1}{4}\sum_{n=1}^{4} i^n\|\alpha + i^n\beta\|^2$$

Note that this identity holds regardless of whether V is finite dimensional or not.

Definition 1.6. Let V be a finite dimensional inner product space and $\mathcal{B} = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ be a fixed ordered basis of V. Define $G \in F^{n\times n}$ by

$$G_{i,j} = (\alpha_i, \alpha_j)$$

This matrix G is called the *matrix of the inner product in the ordered basis $\mathcal{B}$.*

Note that, for any $\alpha, \beta \in V$, we may write

$$\alpha = \sum_{i=1}^{n} x_i\alpha_i \text{ and } \beta = \sum_{i=1}^{n} y_i\alpha_i$$

Then

$$\begin{aligned}(\alpha|\beta) = (\sum_{i=1}^{n} \alpha_i|\beta) \\ &= \sum_{i=1}^{n} x_i(\alpha_i|\beta) \\ &= \sum_{i=1}^{n} x_i(\alpha_i|\sum_{j=1}^{n} y_j\alpha_j) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} x_i\overline{y_i}(\alpha_i|\alpha_j) \\ &= Y^*GX\end{aligned}$$

where X and Y are the coordinate matrices of α and β in the ordered basis $\mathcal{B}$.

Remark 1.7. Let G be a matrix of the inner product in a fixed ordered basis $\mathcal{B}$, then

(i) G is hermitian: $G = G^*$ because

$$G^*_{i,j} = \overline{G_{i,j}} = \overline{(\alpha_j|\alpha_i)} = (\alpha_i|\alpha_j) = G_{i,j}$$

(ii) Furthermore, for any $X \in F^{n\times 1}$ we have

$$X^*GX > 0$$

if $X \neq 0$. Hence, G is invertible (because if $GX = 0$, then $X \neq 0$ would violate this condition). Furthermore, for any scalars $x_1, x_2, \ldots, x_n \in F$, we have

$$\sum_{i,j} x_i G_{i,j} x_j > 0$$

This implies, in particular, that $G_{i,i} > 0$ for all $1 \leq i \leq n$. However, this condition alone is not suffices to ensure that the matrix G is a matrix of an inner product.

(iii) However, if G is an $n \times n$ matrix such that

$$\sum_{i,j} x_i G_{i,j} x_j > 0$$

for any scalars $x_1, x_2, \ldots, x_n \in F$ not all zero, then G defines an inner product on V by

$$(\alpha|\beta) = Y^*GX$$

where X and Y are the coordinate matrices of α and β in the ordered basis $\mathcal{B}$.

2. Inner Product Spaces

Definition 2.1. An *inner product space* is a (real or complex) vector space V together with a fixed inner product on it.

Recall that, for a vector $\alpha \in V$, we write $\|\alpha\| := (\alpha|\alpha)^{1/2}$

Theorem 2.2. *Le V be an inner product space. Then, for any $\alpha, \beta \in V$ and scalar $c \in F$, we have*

(i) $\|c\alpha\| = |c|\|\alpha\|$

(ii) $\|a\| \geq 0$ *and* $\|\alpha\| = 0$ *if and only if* $\alpha = 0$

(iii) $|(\alpha|\beta)| \leq \|a\|\|\beta\|$

(iv) $\|\alpha + \beta\| \leq \|\alpha\| + \|\beta\|$

Proof.

(i) We have $\|c\alpha\|^2 = (c\alpha|c\alpha) = c(\alpha|c\alpha) = c\overline{c}(\alpha|\alpha) = |c|^2\|\alpha\|^2$

(ii) This is also obvious from the axioms.

(iii) If $\alpha = 0$, there is nothing to prove since both sides are zero, so assume $\alpha \neq 0$. Then set

$$\gamma := \beta - \frac{(\beta|\alpha)}{\|\alpha\|^2}\alpha$$

Then, $(\gamma|\alpha) = 0$. Furthermore

$$\begin{aligned} 0 \leq \|\gamma\|^2 &= (\gamma|\gamma) \\ &= \left(\beta - \frac{(\beta|\alpha)}{\|\alpha\|^2}\alpha|\beta - \frac{(\beta|\alpha)}{\|\alpha\|^2}\alpha\right) \\ &= (\beta|\beta) - \frac{(\beta|\alpha)(\alpha|\beta)}{\|\alpha\|^2} \\ &= \|\beta\|^2 - \frac{|(\alpha|\beta)|}{\|\alpha\|^2} \end{aligned}$$

Hence,

$$|(\alpha|\beta)| \leq \|\alpha\|\|\beta\|$$

(iv) Now observe that

$$\|\alpha + \beta\|^2 = (\alpha + \beta|\alpha + \beta) = (\alpha|\alpha) + (\beta|\beta) + 2\text{Re}(\alpha|\beta)$$

Using part (iii), we have

$$2\text{Re}(\alpha|\beta) \leq 2|(\alpha|\beta)| \leq 2\|\alpha\|\|\beta\|$$

Hence,

$$\|\alpha + \beta\|^2 \leq (\|\alpha\| + \|\beta\|)^2$$

and so $\|\alpha + \beta\| \leq \|\alpha\| + \|\beta\|$.

□

Remark 2.3. The inequality in part (iii) is an important fact, called the *Cauchy-Schwartz inequality*. In fact, the proof shows more: If $\alpha, \beta \in V$ are two vectors such that equality holds in the Cauchy-Schwartz inequality, then

$$\gamma := \beta - \frac{(\beta|\alpha)}{\|\alpha\|^2}\alpha = 0$$

Hence, it follows that

$$\beta = c\alpha$$

for some scalar $c \in F$. Conversely, if $\beta = c\alpha$, then equality holds as well (Check!). Therefore, if $\{\alpha, \beta\}$ is a linearly independent set, then the inequality is *strict*.

Example 2.4.

(i) Applying the Cauchy-Schwartz inequality to the standard inner product on $\mathbb{C}^n$ gives

$$|\sum_{i=1}^n x_i\overline{y_i}| \leq \left(\sum_{k=1}^n |x_k|^2\right)^{1/2} \left(\sum_{k=1}^n |y_k|^2\right)^{1/2}$$

(ii) For matrices $A, B \in F^{n\times n}$, one has

$$|\text{trace}(B^*A)| \leq \text{trace}(A^*A)^{1/2}\text{trace}(B^*B)^{1/2}$$

(iii) For continuous functions $f, g \in C[0,1]$, one has

$$\left|\int_0^1 f(t)\overline{g(t)}dt\right| \leq \left(\int_0^1 |f(t)|^2 dt\right)^{1/2} \left(\int_0^1 |g(t)|^2 dt\right)^{1/2}$$

Definition 2.5. Let V be an inner product space.

(i) Two vectors $\alpha, \beta \in V$ are said to be *orthogonal* if $(\alpha|\beta) = 0$.

(ii) A set $S \subset V$ is said to be *orthogonal* if any two distinct vectors in S are orthogonal.

(iii) A set $S \subset V$ is said to be *orthonormal* if it is orthogonal, and $\|\alpha\| = 1$ for all $\alpha \in S$.

Example 2.6.

(i) The zero vector is orthogonal to any other vector.

(ii) If F^n is endowed with the standard inner product, then the standard basis $\{\epsilon_1, \epsilon_2, \ldots, \epsilon_n\}$ is an orthonormal set.

(iii) The vectors (x, y) and $(x, -y) \in \mathbb{R}^2$ are orthogonal ($\mathbb{R}^2$ is equipped with the standard inner product).

(iv) Let $V = \mathbb{C}^{n\times n}$, the space of complex $n \times n$ matrices, and let $E^{p,q}$ be the matrix

$$E^{p,q}_{i,j} = \delta_{p,i}\delta_{q,j}$$

If V is given the inner product of Example 1.2 (iii), then

$$(E^{p,q}|E^{r,s}) = \text{trace}(E^{p,q}E^{s,r}) = \delta_{q,s}\delta_{p,r}$$

Thus, $S = \{E^{p,q} : 1 \leq p, q \leq n\}$ is an orthonormal set.

(v) Let $V = C[0, 1]$. For $n \in \mathbb{N}$, set

$$f_n(x) := \cos(2\pi nx), \text{ and } g_n(x) = \sin(2\pi nx)$$

Then, the set $S = \{1, f_1, g_1, f_2, g_2, \ldots\}$ is an infinite orthogonal set. If we consider complex-valued functions, we may take

$$h_n(x) := f_n(x) + ig_n(x) = e^{2\pi inx}$$

Then, the set $\{h_n : n = \pm1, \pm2, \ldots\}$ is an infinite orthogonal set.

Theorem 2.7. *An orthogonal set of non-zero vectors is linearly independent.*

Proof. Let $S \subset V$ be an orthogonal set, and let $\alpha_1, \alpha_2, \ldots, \alpha_n \in S$ and $c_i \in F$ be scalars such that

$$c_1\alpha_1 + c_2\alpha_2 + \ldots + c_n\alpha_n = 0$$

For $1 \leq j \leq n$, we take an inner product with α_j to get

$$0 = c_j(\alpha_j|\alpha_j)$$

Since $\alpha_j \neq 0$, this forces $c_j = 0$ for all $1 \leq j \leq n$. Thus, every finite subset of S is linearly independent. So S is linearly independent (See Remark II.3.2 (vi)). □

Corollary 2.8. *Let $S = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ is an orthogonal set of non-zero vectors, and $\beta \in span(S)$, then*

$$\beta = \sum_{i=1}^n \frac{(\beta|\alpha_i)}{\|\alpha_i\|^2}\alpha_i$$

Proof. Write

$$\beta = \sum_{i=1}^n c_i\alpha_i$$

and take an inner product with α_j to see that

$$(\beta|\alpha_j) = c_j(\alpha_j|\alpha_j) = c_j\|\alpha_j\|^2$$

Solving for c_j gives the required expression. □

Theorem 2.9 (Gram-Schmidt Orthogonalization)**.** *Let V be an inner product space and $\{\beta_1, \beta_2, \ldots, \beta_n\} \subset V$ be a set of linearly independent vectors. Then, there exists orthogonal vectors $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ such that*

$$\textit{span}\{\alpha_1, \alpha_2, \ldots, \alpha_n\} = \textit{span}\{\beta_1, \beta_2, \ldots, \beta_n\}$$

Proof. We proceed by induction on n. If $n = 1$, we set $\alpha_1 = \beta_1$.

If $n > 1$: Assume, by induction, that we have constructed orthogonal vectors $\{\alpha_1, \alpha_2, \ldots, \alpha_{n-1}\}$ such that

$$\text{span}\{\alpha_1, \alpha_2, \ldots, \alpha_{n-1}\} = \text{span}\{\beta_1, \beta_2, \ldots, \beta_{n-1}\}$$

We now define

$$\alpha_n := \beta_n - \sum_{k=1}^{n-1} \frac{(\beta_n|\alpha_k)}{\|\alpha_k\|^2}\alpha_k$$

If $\alpha_n = 0$, then $\beta_n \in \text{span}\{\alpha_1, \alpha_2, \ldots, \alpha_{n-1}\} = \text{span}\{\beta_1, \beta_2, \ldots, \beta_{n-1}\}$. This contradicts the assumption that the set $\{\beta_1, \beta_2, \ldots, \beta_n\}$ is linearly independent. Hence,

$$\alpha_n \neq 0$$

Furthermore, if $1 \leq j \leq n - 1$,

$$\begin{aligned}(\alpha_n|\alpha_j) &= (\beta_n|\alpha_j) - \sum_{k=1}^{n-1} \frac{(\beta_n|\alpha_k)}{\|\alpha_k\|^2}(\alpha_k|\alpha_j) \\ &= (\beta_n|\alpha_j) - \frac{(\beta_n|\alpha_j)}{\|\alpha_j\|^2}(\alpha_j|\alpha_j) \\ &= 0\end{aligned}$$

Since $\{\alpha_1, \alpha_2, \ldots, \alpha_{n-1}\}$ is orthogonal, this shows that the set $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ is orthogonal. Now, it is clear that

$$\alpha_n \in \text{span}\{\beta_n, \alpha_1, \alpha_2, \ldots, \alpha_{n-1}\} = \text{span}\{\beta_n, \beta_1, \beta_2, \ldots, \beta_{n-1}\}$$

and so

$$\text{span}\{\alpha_1, \alpha_2, \ldots, \alpha_n\} \subset \text{span}\{\beta_1, \beta_2, \ldots, \beta_n\}$$

However, $\text{span}\{\beta_1, \beta_2, \ldots, \beta_n\}$ has dimension n, and $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ has n elements and is linearly independent by Theorem 2.7. Thus,

$$\text{span}\{\alpha_1, \alpha_2, \ldots, \alpha_n\} = \text{span}\{\beta_1, \beta_2, \ldots, \beta_n\}$$

This completes the proof. □

Corollary 2.10. *Every finite dimensional inner product space has an orthonormal basis.*

Proof. We start with any basis $\mathcal{B} = \{\beta_1, \beta_2, \ldots, \beta_n\}$ of V. By Gram-Schmidt orthogonalization (Theorem 2.9), there is an orthogonal set $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ such that

$$\operatorname{span}\{\alpha_1, \alpha_2, \ldots, \alpha_n\} = V$$

Now simply take $\{\alpha_1', \alpha_2', \ldots, \alpha_n'\}$ where

$$\alpha_j' := \frac{\alpha_j}{\|\alpha_j\|^2}$$

□

Example 2.11.

(i) Consider the vectors

$$\begin{aligned}\beta_1 &:= (3,0,4)\\ \beta_2 &:= (-1,0,7)\\ \beta_3 &:= (2,9,11)\end{aligned}$$

in $\mathbb{R}^3$ equipped with the standard inner product. Applying the Gram-Schmidt process, we obtain the following vectors

$$\begin{aligned}\alpha_1 &= (3,0,4)\\ \alpha_2 &= (-1,0,7) - \frac{((-1,0,7)|(3,0,4))}{\|(3,0,4)\|}(3,0,4)\\ &= (-1,0,7) - \frac{25}{25}(3,0,4)\\ &= (-1,0,7) - (3,0,4)\\ &= (-4,0,3)\\ \alpha_3 &= (2,9,11) - \frac{((2,9,11)|(3,0,4))}{\|(3,0,4)\|}(3,0,4) - \frac{((2,9,11)|(-4,0,3))}{\|(-4,0,3)\|}(-4,0,3)\\ &= (2,9,11) - 2(3,0,4) - (-4,0,3)\\ &= (0,9,0)\end{aligned}$$

The vectors $\{\alpha_1, \alpha_2, \alpha_3\}$ are mutually orthogonal and non-zero, so they form a basis for $\mathbb{R}^3$. To express a vector $\beta \in \mathbb{R}^3$ as a linear combination of these vectors, we may use Corollary 2.8 and write

$$\beta = \sum_{i=1}^{n} \frac{(\beta|\alpha_i)}{\|\alpha_i\|^2}\alpha_i$$

If $\beta = (x_1, x_2, x_3)$, this reduces to

$$(x_1, x_2, x_3) = \frac{3x_1 + 4x_3}{25}\alpha_1 + \frac{-4x_1 + 3x_3}{25}\alpha_2 + \frac{x_2}{9}\alpha_3$$

For instance,

$$(1,2,3) = \frac{3}{5}(3,0,4) + \frac{1}{5}(-4,0,3) + \frac{2}{9}(0,9,0)$$

Equivalently, the dual basis $\{f_1, f_2, f_3\}$ to the basis $\{\alpha_1, \alpha_2, \alpha_3\}$ is given by

$$\begin{aligned} f_1(x_1,x_2,x_3) &= \frac{3x_1 + 4x_3}{25} \\ f_2(x_1,x_2,x_3) &= \frac{-4x_1 + 3x_3}{25} \\ f_3(x_2,x_2,x_3) &= \frac{x_2}{9} \end{aligned}$$

Finally, observe that the orthonormal basis one obtains from this process is

$$\begin{aligned} \alpha_1' &= \frac{1}{5}(3,0,4) \\ \alpha_2' &= \frac{1}{5}(-4,0,3) \\ \alpha_3' &= (0,1,0) \end{aligned}$$

The Gram-Schmidt process is itself obtained as a special case of an interesting geometrical notion; that of a projection onto a subspace. Given a subspace W of an inner product space V and a vector $\alpha \in W$, one is often interested in a vector $\alpha \in W$ that is closest to β. If $\beta \in W$, this vector would be β of course, but in general, we would like a way of computing this vector α from β.

Definition 2.12. Given a subspace W of an inner product space V and a vector $\beta \in V$, a *best approximation* to β by vectors in W is a vector $\alpha \in W$ such that

$$\|\beta - \alpha\| \leq \|\beta - \gamma\|$$

for all $\gamma \in W$.

Note that we do not know, as yet, if such a vector exists. However, if one thinks about the problem geometrically in $\mathbb{R}^2$ or $\mathbb{R}^3$, one observes that is one is looking for a vector $\alpha \in W$ such that $(\beta - \alpha)$ is *perpendicular* to W.

Theorem 2.13. *Let W be a subspace of an inner product space V and let $\beta \in V$.*

(i) The vector $\alpha \in W$ is a best approximation to β by vectors in W if and only if

$$(\beta - \alpha|\gamma)$$

for all $\gamma \in W$.

(ii) If a best approximation to β by vectors in W exists, then it is unique.

(iii) If W is finite dimensional and $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ is any orthonormal basis for W, then the vector

$$\alpha := \sum_{k=1}^{n} (\beta|\alpha_k)\alpha_k$$

is the (unique) best approximation to β by vectors in W.

Proof.

(i)

- Suppose $\alpha \in W$ is a best approximation to β by vectors in W and $\gamma \in W$, then

$$\begin{aligned}\|\beta - \gamma\|^2 &= \|(\beta - \alpha) + (\alpha - \gamma)\|^2 \\ &= \|\beta - \alpha\|^2 + 2\text{Re}(\beta - \alpha|\alpha - \gamma) + \|\alpha - \gamma\|^2 \\ &\geq \|\beta - \alpha\|^2\end{aligned}$$

Hence,

$$2\text{Re}(\beta - \alpha|\alpha - \gamma) + \|\alpha - \gamma\|^2 \geq 0$$

for all $\gamma \in W$. Replacing γ by $\tau := \alpha + \gamma \in W$, we conclude that

$$2\text{Re}(\beta - \alpha|\tau) + \|\tau\|^2 \geq 0$$

for all $\tau \in W$. In particular, if $\gamma \in W$ is such that $\gamma \neq \alpha$, then we may set

$$\tau := -\frac{(\beta - \alpha|\alpha - \gamma)}{\|\alpha - \gamma\|^2}(\alpha - \gamma)$$

Then the inequality reduces to the statement

$$-2\frac{|(\beta - \alpha|\alpha - \gamma)|^2}{\|\alpha - \gamma\|^2} + \frac{|(\beta - \alpha|\alpha - \gamma)|^2}{\alpha - \gamma\|^2} \geq 0$$

But this last inequality holds if and only if

$$(\beta - \alpha|\alpha - \gamma) = 0$$

This must hold for all $\gamma \in W$ with $\gamma \neq \alpha$, so we conclude t hat

$$(\beta - \alpha|\gamma) = 0$$

for all $\gamma \in W$.

- Conversely, suppose that $(\beta - \alpha|\gamma) = 0$ for all $\gamma \in W$, then we have (as above)

$$\begin{aligned}\|\beta - \gamma\|^2 &= \|(\beta - \alpha) + (\alpha - \gamma)\|^2 \\ &= \|\beta - \alpha\|^2 + 2\text{Re}(\beta - \alpha|\alpha - \gamma) + \|\alpha - \gamma\|^2\end{aligned}$$

However, $\alpha \in W$ so $\alpha - \gamma \in W$, so that

$$(\beta - \alpha | \alpha - \gamma) = 0$$

Thus,

$$\|\beta - \gamma\|^2 = \|\beta - \alpha\|^2 + \|\alpha - \gamma\|^2 \geq \|\beta - \alpha\|^2$$

This is true for any $\gamma \in W$, so α is a best approximation to β by vectors in W.

(ii) Now we show uniqueness: If α and α' are two best approximations to β by vectors in W, then $\alpha, \alpha' \in W$ and by part (i), we have

$$(\beta - \alpha | \gamma) = (\beta - \alpha' | \gamma) = 0$$

for all $\alpha \in W$. In particular,

$$\|\alpha - \alpha'\|^2 = (\alpha - \alpha' | \alpha - \alpha') = (\alpha - \beta | \alpha - \alpha') + (\beta - \alpha' | \alpha - \alpha') = 0 + 0 = 0$$

Hence, $\alpha = \alpha'$.

(iii) Now suppose W is finite dimensional and $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ is an orthonormal basis for W. Then, for any $\gamma \in W$, one has

$$\gamma = \sum_{k=1}^n (\gamma | \alpha_k) \alpha_k$$

by Corollary 2.8. If $\alpha \in W$ is such that $(\beta - \alpha | \gamma) = 0$ for all $\gamma \in W$, then one has

$$(\beta | \alpha_k) - (\alpha | \alpha_k) = (\beta - \alpha | \alpha_k) = 0$$

Hence,

$$(\alpha | \alpha_k) = (\beta | \alpha_k)$$

for all $1 \leq k \leq n$. Hence,

$$\alpha = \sum_{k=1}^n (\alpha | \alpha_k) \alpha_k = \sum_{k=1}^n (\beta | \alpha_k) \alpha_k$$

□

Definition 2.14. Let S be a subset of an inner product space V. The *orthogonal complement* of S is the set

$$S^\perp := \{\beta \in V : (\beta | \alpha) = 0 \quad \forall \alpha \in S\}$$

Note that (Check!) $S^\perp$ is a subspace of V regardless of whether S is a subspace or not. Furthermore, if $0 \in S$, then

$$S \cap S^\perp = \{0\}$$

Definition 2.15. Let W be a subspace of an inner product space V.

(i) If $\beta \in V$, the best approximation $\alpha \in W$ to β by vectors in W is called the *orthogonal projection of* β *on* W.

(ii) If W is finite dimensional, define a map $E : V \to W$ which sends β to α. This is exists and is well-defined by Theorem 2.13.

Corollary 2.16. *Let* W *be a finite dimensional subspace of an inner product space* V, *and let* E *denote the orthogonal projection of* V *on* W. *Then, the map*

$$\beta \mapsto \beta - E(\beta)$$

is the orthogonal projection of V *onto* $W^{\perp}$.

Proof.

(i) Fix $\beta \in V$ and set $\gamma := \beta - E(\beta)$. If $\eta \in W$, then

$$(\beta - E(\beta)|\eta) = 0$$

by Theorem 2.13. Hence,

$$\gamma \in W^{\perp}$$

(ii) Furthermore, if $\tau \in W^{\perp}$, then

$$\|\beta-\tau\|^2 = \|E(\beta)+\beta-E(\beta)-\tau\|^2 = \|E(\beta)\|^2+\|\beta-E(\beta)-\tau\|^2+2\text{Re}(E(\beta)|\beta-E(\beta)-\tau)$$

However, $E(\beta) \in W$, and $\beta - E(\beta) \in W^{\perp}$ so

$$\beta - E(\beta) - \tau \in W^{\perp}$$

whence $(E(\beta)|\beta - E(\beta) - \tau) = 0$. Thus,

$$\|\beta - \tau\|^2 = \|E(\beta)\|^2 + \|\beta - E(\beta) - \tau\|^2 \geq \|E(\beta)\|^2 = \|\beta - (\beta - E(\beta))\|^2$$

Hence, $\beta - E(\beta)$ is a best approximation to β by vectors in $W^{\perp}$.

□

Theorem 2.17. *Let* W *be a finite dimensional subspace of an inner product space* V, *and let* $E : V \to W$ *denote the orthogonal projection of* V *on* W. *Then,*

(i) E *is a linear transformtion*

(ii) E *is idempotent (ie.* $E^2 = E$*)*

(iii) $\ker(E) = W^{\perp}$

(iv) Furthermore,

$$V = W \oplus W^{\perp}$$

Proof.

(i) If $\alpha, \beta \in V$ and $c \in F$, then set

$$\gamma := cE(\alpha) + E(\beta)$$

Now, since $E(\alpha), E(\beta) \in W$, we have that $\gamma \in W$. Furthermore, we know that

$$\alpha - E(\alpha) \in W^{\perp} \text{ and } \beta - E(\beta) \in W^{\perp}$$

by Theorem 2.13. Hence,

$$(c\alpha + \beta) - \gamma \in W^{\perp}$$

since $W^{\perp}$ is a subspace of V. Therefore, it follows from Theorem 2.13 that

$$E(c\alpha + \beta) = \gamma$$

Hence, E is linear.

(ii) If $\gamma \in W$, then clear $E(\gamma) = \gamma$ since γ is the best approximation to itself. Hence, if $\beta \in V$, one has

$$E(E(\beta)) = E(\beta)$$

Thus, $E^2 = E$.

(iii) Let $\beta \in V$, then $E(\beta) \in W$ is the unique vector such that $\beta - E(\beta) \in W^{\perp}$. Hence, if $\beta \in W^{\perp}$, then

$$E(\beta) = 0$$

Conversely, if $\beta \in V$ is such that $E(\beta) = 0$, then $\beta = \beta - E(\beta) \in W^{\perp}$. Thus,

$$\ker(E) = W^{\perp}$$

(iv) Finally, if $\beta \in V$, then we may write

$$\beta = E(\beta) + (\beta - E(\beta))$$

By Corollary 2.16, we have

$$E(\beta) \in W \text{ and } (\beta - E(\beta)) \in W^{\perp}$$

Thus, $V = W + W^{\perp}$. Since $W \cap W^{\perp} = \{0\}$, it follows that this is a direct sum decomposition.

□

Corollary 2.18. *Let W be a finite dimensional space of an inner product space V and let E denote the orthogonal projection of V on W. Then, $(I - E)$ is the orthogonal projection of V on $W^{\perp}$. It is an idempotent linear transformtion with range $W^{\perp}$ and kernel W.*

Proof. We know that $(I - E)$ maps V to $W^\perp$ by Corollary 2.16. Since E is a linear transformation by Theorem 2.17, it follows that $(I - E)$ is also linear. Furthermore,

$$(I-E)^2 = (I-E)(I-E) = I + E^2 - E - E = I + E - E - E = I - E$$

Finally, observe that, for any $\beta \in V$, one has $(I - E)\beta = 0$ if and only if $\beta = E(\beta)$. This happens (Check!) if and only if $\beta \in W$. □

Remark 2.19. The Gram-Schmidt orthogonalization process (Theorem 2.9) may now be described geometrically as follows: Given a linearly independent set $\{\beta_1, \beta_2, \ldots, \beta_n\}$ in an inner product space V, define operators $P_1, P_2, \ldots, P_n$ as follows:

$$P_1 = I$$

and, for $k > 1$, set P_k to be the orthogonal projection of V on the orthogonal complement of

$$W_k := \text{span}\{\beta_1, \beta_2, \ldots, \beta_{k-1}\}$$

Such a map exists by Corollary 2.18. The Gram-Schmidt orthogonalization now yields vectors

$$\alpha_k := P_k(\beta_k), 1 \leq k \leq n$$

Corollary 2.20 (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 $\beta \in V$, then*

$$\sum_{k=1}^n \frac{|(\beta|\alpha_k)|^2}{\|\alpha_k\|^2} \leq \|\beta\|^2$$

and equality holds if and only if

$$\beta \in \textit{span}\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$$

Proof. Set

$$\gamma := \sum_{k=1}^n \frac{(\beta|\alpha_k)}{\|\alpha_k\|^2}\alpha_k$$

and

$$\delta := \beta - \gamma$$

Then, $(\gamma|\delta) = 0$. Hence,

$$\|\beta\|^2 = \|\gamma\|^2 + \|\delta\|^2 \geq \|\gamma\|^2$$

Finally, observe that

$$\|\gamma\|^2 = \left(\sum_{k=1}^n \frac{(\beta|\alpha_k)}{\|\alpha_k\|^2}\alpha_k | \sum_{k=1}^n \frac{(\beta|\alpha_k)}{\|\alpha_k\|^2}\alpha_k\right)$$

Since $(\alpha_i|\alpha_j) = 0$ if $i \neq j$, we conclude that

$$\|\gamma\|^2 = \sum_{k=1}^n \frac{|(\beta|\alpha_k)|^2}{\|\alpha_k\|^2}$$

This proves the inequality.

Now, equality holds if and only if $\delta = 0$; or, equivalently,

$$\beta = \gamma = \sum_{k=1}^n \frac{(\beta|\alpha_k)}{\|\alpha_k\|^2}\alpha_k$$

This clearly implies that $\beta \in \text{span}\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$. Conversely, if $\beta \in \text{span}\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$, then $\beta = \gamma$ by Corollary 2.8. □

Example 2.21. Let $V = C[0,1]$, the space of continuous, complex-valued functions on $[0,1]$. Then, for any $f \in C[0,1]$, one has

$$\sum_{k=-n}^n \left|\int_0^1 f(t)e^{-2\pi ikt}dt\right|^2 \leq \int_0^1 |f(t)|^2 dt$$

(End of Week 13)

3. Linear Functionals and Adjoints

Remark 3.1.

(i) Let V be a vector space over a field F. Recall (Definition III.5.1) that a linear functional is a linear transformation $L : V \to F$. Furthermore, we write $V^* := L(V,F)$ for the set of all linear functionals on V (See Definition III.5.3).

(ii) Consider the case $V = F^n$. For a fixed n-tuple $\beta := (a_1, a_2, \ldots, a_n) \in V$, there is an associated linear functional $L_\beta : V \to F$ given by

$$L_\beta(x_1, x_2, \ldots, x_n) := \sum_{i=1}^n a_i x_i$$

Furthermore, every linear functional on F^n is of this form (See Example III.5.2).

(iii) Now suppose V is an arbitrary inner product space with inner product $(\cdot|\cdot)$. For a fixed vector $\beta \in V$, there is an associated linear functional $L_\beta : V \to F$ given by

$$L_\beta(\alpha) := (\alpha|\beta)$$

Note that this map is linear by the axioms of the inner product (Definition 1.1).

We now show, just as in the case of F^n, that every linear functional on V is of this form, provided V is finite dimensional.

Theorem 3.2. *Let V be a finite dimensional inner product space and $L : V \to F$ be a linear functional on V. Then, there exists a unique vector $\beta \in V$ such that*

$$L(\alpha) = (\alpha|\beta)$$

for all $\alpha \in V$.

Proof.

(i) Existence: Fix an orthonormal basis $\{\alpha_1, \alpha_2, \dots, \alpha_n\}$ of V (guaranteed by Corollary 2.10). Set

$$\beta := \sum_{j=1}^{n} \overline{L(\alpha_j)}\alpha_j$$

Then, for each $1 \leq i \leq n$, we have

$$L_\beta(\alpha_i) = (\alpha_i|\beta) = \sum_{j=1}^{n} L(\alpha_j)(\alpha_i|\alpha_j) = L(\alpha_i)$$

Since L_β and L are both linear functionals that agree on a basis, it follows by Theorem III.1.4 that

$$L = L_\beta$$

as required.

(ii) Uniqueness: Suppose $\beta, \beta' \in V$ are such that $L_\beta = L_{\beta'}$, then

$$(\alpha|\beta) = (\alpha|\beta')$$

for all $\alpha \in V$. In particular, for $\alpha = \beta - \beta'$, we have

$$0 = (\alpha|\beta - \beta') = \|\beta - \beta'\|^2$$

Hence, $\beta = \beta'$.

□

Theorem 3.3. *Let V be a finite dimensional inner product space and $T \in L(V)$ be a linear operator. Then, there exists a unique linear operator $S \in L(V)$ such that*

$$(T\alpha|\beta) = (\alpha|S\beta)$$

for all $\alpha, \beta \in V$.

This operator is called the *adjoint* of T and is denoted by T^*.

Proof.

(i) Existence: Fix $\beta \in V$ and consider $L : V \to F$ by

$$L(\alpha) := (T\alpha|\beta)$$

Then L is a linear functional on V. Hence, by Theorem 3.2, there exists a unique vector $\beta' \in V$ such that

$$(\alpha|\beta') = (T\alpha|\beta)$$

for all $\alpha \in V$. Define $S : V \to V$ by

$$S(\beta) = \beta'$$

so that

$$(\alpha|S\beta) = (T\alpha|\beta)$$

for all $\alpha \in V$.

(i) S is well-defined: If $\beta \in V$, then $\beta' \in V$ is uniquely determined by the equation

$$(\alpha|\beta') = (T\alpha|\beta)$$

by Theorem 3.2. Hence, S is well-defined.

(ii) S is additive: Suppose $\beta_1, \beta_2 \in V$ are chosen and $\beta_1', \beta_2' \in V$ are such that

$$(\alpha|\beta_1') = (T\alpha|\beta_1) \text{ and } (\alpha|\beta_2') = (T\alpha|\beta_2)$$

for all $\alpha \in V$. Then, let $\beta := \beta_1 + \beta_2$, then for any $\alpha \in V$ we have

$$(T\alpha|\beta) = (T\alpha|\beta_1) + (T\alpha|\beta_2) = (\alpha|\beta_1') + (\alpha|\beta_2') = (\alpha|\beta_1' + \beta_2')$$

Hence, by definition

$$S(\beta) = \beta_1' + \beta_2'$$

as desired.

(iii) S respects scalar multiplication: Exercise (Similar to part (ii)).

Hence, we have constructed $S \in L(V)$ such that

$$(\alpha|S(\beta)) = (T\alpha|\beta)$$

for all $\alpha, \beta \in V$.

(ii) Uniqueness: Suppose $S_1, S_2 \in L(V)$ such that

$$(\alpha|S_1(\beta)) = (T\alpha|\beta) = (\alpha|S_2\beta)$$

for all $\alpha, \beta \in V$. Then, for any $\beta \in V$ fixed,

$$(\alpha|S_1(\beta)) = (\alpha|S_2(\beta))$$

for all $\alpha \in V$. As in the uniqueness of Theorem 3.2, we conclude that

$$S_1\beta = S_2\beta$$

This is true for all $\beta \in V$. Hence, $S_1 = S_2$.

□

Theorem 3.4. *Let V be a finite dimensional inner product space and let $\mathcal{B} := \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ be an ordered orthonormal basis for V. Let $T \in L(V)$ be a linear operator and*

$$A = [T]_{\mathcal{B}}$$

Then, $A_{k,j} = (T\alpha_j|\alpha_k)$.

Proof. Since $\mathcal{B}$ is an orthonormal basis, we have

$$\alpha = \sum_{k=1}^{n} (\alpha|\alpha_k)\alpha_k$$

for any fixed $\alpha \in V$ (See Corollary 2.8). By definition of A, we have

$$T\alpha_j = \sum_{k=1}^{n} A_{k,j}\alpha_k$$

Hence,

$$A_{k,j} = (T\alpha_j|\alpha_k)$$

as required. □

The next corollary identifies the adjoint of an operator in terms of the matrix of the operator in a fixed ordered orthonormal basis.

Corollary 3.5. *Let V be a finite dimensional inner product space and $T \in L(V)$ be a linear operator. If $\mathcal{B}$ is an ordered orthonormal basis for V, then the matrix*

$$[T^*]_{\mathcal{B}}$$

is the conjugate transpose of the matrix $[T]_{\mathcal{B}}$.

Proof. Set

$$A := [T]_{\mathcal{B}} \text{ and } B := [T^*]_{\mathcal{B}}$$

Then, by Theorem 3.4, we have

$$A_{k,j} = (T\alpha_j|\alpha_k) \text{ and } B_{k,j} = (T^*\alpha_j|\alpha_k)$$

But,

$$B_{k,j} = (T^*\alpha_j|\alpha_k) = (\alpha_j|T\alpha_k) = \overline{(T\alpha_k|\alpha_j)} = \overline{A_{j,k}}$$

as required. □

Remark 3.6.

(i) The adjoint of an operator T depends on the inner product.

(ii) In an arbitrary ordered basis $\mathcal{B}$ that is not necessarily orthonormal, the relationship between $[T]_{\mathcal{B}}$ and $[T^*]_{\mathcal{B}}$ is more complicated than the one in Corollary 3.5.

Example 3.7.

(i) Let $V = \mathbb{C}^{n\times 1}$, the space of complex $n \times 1$ matrix with the inner product

$$(X|Y) = Y^*X$$

Let $A \in \mathbb{C}^{n\times n}$ be an $n \times n$ matrix and define $T : V \to V$ by

$$TX := AX$$

Then, for any $Y \in V$, we have

$$(TX|Y) = Y^*AX = (A^*Y)^*X = (X, A^*Y)$$

Hence, T^* is the linear operator $Y \mapsto A^*Y$. This is, of course, just a special case of Corollary 3.5.

(ii) Let $V = \mathbb{C}^{n\times n}$ with the inner product

$$(A|B) = \text{trace}(B^*A)$$

Let M be a fixed $n \times n$ matrix and $T : V \to V$ be the map

$$T(A) := MA$$

Then, for any $B \in V$, one has

$$\begin{aligned}(TA|B) &= \text{trace}(B^*MA)\\ &= \text{trace}(MAB^*)\\ &= \text{trace}(AB^*M)\\ &= \text{trace}(A(M^*B))^*)\\ &= (A|M^*B)\end{aligned}$$

Hence, T^* is the linear operator $B \mapsto M^*B$.

(iii) If E is an orthogonal projection on a subspace W of an inner product space V, then, for any vectors $\alpha, \beta \in V$, we have

$$\begin{aligned}(E\alpha|\beta) &= (E\alpha|E\beta + (I - E)\beta)\\ &= (E\alpha|E\beta) + (E\alpha|(I - E)\beta))\end{aligned}$$

But $E\alpha \in W$ and $(I - E)\beta \in W^\perp$ by Corollary 2.16. Hence,

$$(E\alpha|\beta) = (E\alpha|E\beta)$$

Similarly,

$$(\alpha|E\beta) = (E\alpha|E\beta)$$

Hence,

$$(\alpha|E\beta) = (E\alpha|\beta)$$

By uniqueness of the adjoint, it follows that $E = E^*$.

In what follows, we will frequently use the following fact: If $S_1, S_2 \in L(V)$ are two linear operators on a finite dimensional inner product space V and

$$(S_1(\alpha)|\beta) = (S_2(\alpha)|\beta)$$

for all $\alpha, \beta \in V$. Then $S_1 = S_2$. The same thing holds if

$$(\alpha|S_1\beta) = (\alpha|S_2\beta)$$

for all $\alpha, \beta \in V$. Now we prove some algebraic properties of the adjoint.

Theorem 3.8. *Let V be a finite dimensional inner product space. Let $T, U \in L(V)$ and $c \in F$. Then*

(i) $(T+U)^* = T^* + U^*$

(ii) $(cT)^* = \overline{c}T^*$

(iii) $(TU)^* = U^*T^*$

(iv) $(T^*)^* = T$

Proof.

(i) For $\alpha, \beta \in V$, we have

$$\begin{aligned} ((T+U)\alpha|\beta) &= (T\alpha + U\alpha|\beta) \\ &= (T\alpha|\beta) + (U\alpha|\beta) \\ &= (\alpha|T^*\beta) + (\alpha|U^*\beta) \\ \Rightarrow (\alpha|(T+U)^*\beta) &= (\alpha|(T^* + U^*)\beta) \end{aligned}$$

This is true for all $\alpha, \beta \in V$, so (by the uniqueness of the adjoint), we have

$$(T+U)^* = T^* + U^*$$

(ii) Exercise.

(iii) For $\alpha, \beta \in V$ fixed, we have

$$\begin{aligned} ((TU)\alpha|\beta) &= (T(U(\alpha))|\beta) \\ &= (U(\alpha)|T^*\beta) \\ &= (\alpha|U^*(T^*(\beta)) \\ \Rightarrow (\alpha|(TU)^*\beta) &= (\alpha|(U^*T^*)\beta) \end{aligned}$$

Hence,

$$(TU)^* = U^*T^*$$

(iv) For $\alpha, \beta \in V$, we have

$$(\alpha|(T^*)^*\beta) = (T^*\alpha, \beta) = (\alpha|T\beta)$$

Hence, $T = (T^*)^*$.

□

Definition 3.9. A linear operator $T \in L(V)$ is said to be *self-adjoint* or *hermitian* if $T = T^*$.

Note that T is hermitian if and only if there is an ordered orthonormal basis $\mathcal{B}$ of V such that

$$[T]_{\mathcal{B}}$$

is a self-adjoint matrix.

Definition 3.10. Let $T \in L(V)$ be a linear operator on a finite dimensional inner product space V. Define

$$U_1 := \frac{1}{2}(T + T^*) \text{ and } U_2 := \frac{1}{2i}(T - T^*)$$

Then, U_1 and U_2 are called the *real* and *imaginary* parts of T respectively.

Note that, if $T \in L(V)$ and U_1 and U_2 are as in Definition 3.10, then U_1 and U_2 are both self-adjoint and

$$T = U_1 + iU_2 \tag{VIII.1}$$

Furthermore, suppose S_1, S_2 are two self-adjoint operators such that

$$T = S_1 + iS_2$$

Then, we have (by Theorem 3.8) that

$$T^* = S_1 - iS_2$$

Hence,

$$S_1 = \frac{T + T^*}{2} = U_1 \text{ and } S_2 = U_2$$

Hence, the expression in Equation VIII.1 is unique.

4. Unitary Operators

Recall that an isomorphism between vector spaces is a bijective linear map.

Definition 4.1. Let V and W be inner product spaecs over the same field F, and let $T : V \to W$ be a linear transformation. We say that T

(i) *preserves inner products* if $(T\alpha|T\beta) = (\alpha|\beta)$ for all $\alpha, \beta \in V$.

(ii) is an *isomorphism* if T is bijective and preserves inner products.

Note that, if T is a linear transformation of inner product spaces that preserves inner products, then for any $\alpha \in V$, one has

$$\|T\alpha\| = \|\alpha\|$$

Hence, T is necessarily injective. Furthermore, if T preserves inner products and is bijective, then T^{-1} is not only a linear map, but also preserves inner products. Hence, this notion of isomorphism of inner product spaces is an equivalence relation (Compare this with section 3). Hence, if such an isomorphism exists, we say that V and W are *isomorphic*. Compare the next theorem to Theorem III.2.15.

Theorem 4.2. *Let V and W be finite dimensional inner product spaces over the same field F, having the same dimension, and let $T : V \to W$ be a linear transformation. Then, TFAE:*

(i) T preserves inner products.

(ii) T is an isomorphism of inner product spaces.

(iii) T carries every orthonormal basis for V onto an orthonormal basis for W

(iv) T carries some orthonormal basis for V to an orthonormal basis for W.

Proof.

$(i) \Rightarrow (ii)$: If T preserves inner products, then T is injective (as mentioned above). Since $\dim(V) = \dim(W)$, it must happen that T is surjective (by Theorem III.2.15). Thus, T is a vector space isomorphism.

$(ii) \Rightarrow (iii)$: Suppose T is an isomorphism and $\mathcal{B} = \{\alpha_1, \alpha_2, \dots, \alpha_n\}$ is an orthonormal basis. Then

$$(\alpha_i, \alpha_j) = \delta_{i,j}$$

Since T preserves inner products, it follows that

$$(T\alpha_i, T\alpha_j) = \delta_{i,j}$$

Hence, the set $\{T(\alpha_1), T(\alpha_2), \dots, T(\alpha_n)\}$ is an orthonormal subset of W. Since $\dim(W) = \dim(V)$, it must form an orthonormal basis for W as well.

$(iii) \Rightarrow (iv)$: Obvious.

$(iv) \Rightarrow (i)$: Let $\mathcal{B} = \{\alpha_1, \alpha_2, \dots, \alpha_n\}$ be an orthonormal basis for V such that $\mathcal{B}' = \{T(\alpha_1), T(\alpha_2), \dots, T(\alpha_n)\}$ is an orthonormal basis for W. Then,

$$(T\alpha_k, T\alpha_j) = \delta_{k,j}$$

Now, for $\alpha, \beta \in V$ fixed, we may express them as

$$\alpha = \sum_{k=1}^n (\alpha|\alpha_k)\alpha_k \text{ and } \beta = \sum_{k=1}^n (\beta|\alpha_k)\alpha_k$$

by Corollary 2.8. Hence

$$\begin{aligned}(T\alpha|T\beta) &= \left(\sum_{k=1}^{n}(\alpha|\alpha_k)T\alpha_k|T\beta\right) \\ &= \sum_{k=1}^{n}(\alpha|\alpha_k)(T\alpha_k|T\beta) \\ &= \sum_{k=1}^{n}(\alpha|\alpha_k)\left(T\alpha_k|\sum_{j=1}^{n}(\beta|\alpha_j)T\alpha_j\right) \\ &= \sum_{k=1}^{n}\sum_{j=1}^{n}(\alpha|\alpha_k)\overline{(\beta|\alpha_j)}(T\alpha_k,T\alpha_j) \\ &= \sum_{k=1}^{n}(\alpha|\alpha_k)\overline{(\beta|\alpha_k)} \\ &= \sum_{k=1}^{n}\sum_{j=1}^{n}(\alpha|\alpha_k)\overline{(\beta|\alpha_j)}(\alpha_k|\alpha_j) \\ &= (\alpha|\beta)\end{aligned}$$

Hence, T preserves inner products.

□

Corollary 4.3. *Let V and W be two finite dimensional inner product spaces over the same field F. Then, V and W are isomorphic (as inner product spaces) if and only if $\dim(V) = \dim(W)$.*

Proof. Clearly, if $V \cong W$, then $\dim(V) = \dim(W)$. Conversely, if $\dim(V) = \dim(W)$, then one may fix orthonormal bases $\mathcal{B} = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ and $\mathcal{B}' = \{\beta_1, \beta_2, \ldots, \beta_n\}$ of V and W respectively (which exist by Corollary 2.10). Then, by Theorem III.1.4, there is a linear map $T : V \to W$ such that

$$T\alpha_j = \beta_j$$

for all $1 \leq j \leq n$. This map is an isomorphism by Theorem 4.2. □

Example 4.4.

(i) Let V be an n-dimensional inner product space. Then, for any orthonormal basis $\mathcal{B} = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ of V, one can define an isomorphism

$$T : V \to F^n$$

given by

$$T\alpha_j = \epsilon_j$$

where $\{\epsilon_1, \epsilon_2, \ldots, \epsilon_n\}$ is the standard orthonormal basis for F^n.

(ii) Let $V = \mathbb{C}^{n\times 1}$, the vector space of all complex $n \times 1$ matrices, and let $P \in \mathbb{C}^{n\times n}$ be a fixed invertible matrix. Then, if $G = P^*P$, then one can define two inner products on V by

$$(X|Y) := Y^*X \text{ and } [X|Y] := Y^*GX$$

(See Example 1.2 (iii)). We write W for the vector space with the second inner product, and define $T : W \to V$ by

$$TX := PX$$

Then, T is clearly bijective. Furthermore,

$$(TX|TY) = (PY)^*PX = Y^*P^*PX = Y^*GX = [X|Y]$$

Hence, T preserves inner products, and is thus an isomorphism of inner product spaces.

Lemma 4.5. *Let V and W be inner product spaces over the same field F and let $T : V \to W$ be a linear transformation. Then, T preserves inner products if and only if*

$$\|T\alpha\| = \|\alpha\|$$

for all $\alpha \in V$.

Proof. Clearly, if T preserves inner products, then $\|T\alpha\| = \|\alpha\|$ for all $\alpha \in V$ must hold.

Conversely, suppose T satisfies this condition, then we use the polarization identities (See Remark 1.5). For instance, if $F = \mathbb{C}$, then this takes the form

$$(\alpha|\beta) = \frac{1}{4}\sum_{n=1}^{4} i^n\|\alpha + i^n\beta\|^2$$

for all $\alpha, \beta \in V$. Hence, it follows that

$$(T\alpha|T\beta) = \frac{1}{4}\sum_{n=1}^{4} i^n\|T\alpha+i^nT\beta\|^2 = \frac{1}{4}\sum_{n=1}^{4} i^n\|T(\alpha+i^n\beta)\|^2 = \frac{1}{4}\sum_{n=1}^{4} i^n\|\alpha+i^n\beta\|^2 = (\alpha|\beta)$$

Thus, T preserves inner products (The case when $F = \mathbb{R}$ is entirely similar). □

Definition 4.6. A *unitary operator* is an operator on an inner product space that is an isomorphism onto itself.

Equivalently, it is an operator $U : V \to V$ that is surjective and preserves inner products, or equivalently, satisfies

$$\|U\alpha\| = \|\alpha\|$$

for all $\alpha \in V$. Note that, if U_1 and U_2 are both unitaries, then U_1U_2 is a unitary, and so is U_1^{-1}. Hence, the set of all unitary operators in $L(V)$ is a group.

Theorem 4.7. *Let $U \in L(V)$ be a linear operator on a finite dimensional inner product space. Then, U is a unitary operator if and only if*

$$UU^* = U^*U = I$$

Proof. Suppose U is a unitary operator, then for any $\alpha, \beta \in V$, one has

$$(U^*U\alpha|\beta) = (U\alpha|(U^*)^*\beta) = (U\alpha|U\beta) = (\alpha|\beta)$$

Hence, $U^*U = I$. Similarly, $UU^* = I$ holds as well.

Conversely, suppose $U^*U = UU^* = I$, then for any $\alpha, \beta \in V$, one has

$$(U\alpha|U\beta) = (U^*U\alpha|\beta) = (\alpha|\beta)$$

Hence, U preserves the inner product. Furthermore, if $U\alpha = 0$, then

$$\alpha = I\alpha = U^*U\alpha = 0$$

Thus, U is injective. By Theorem III.2.15, it follows that U is bijective, and thus a unitary. □

Definition 4.8. An $n \times n$ matrix A is said to be a *unitary* if $A^* = AA^* = I$.

Note that, if $A^*A = I$, then $AA^* = I$ holds automatically by Corollary I.4.9. Hence, A is a unitary matrix if and only if, for all $1 \leq i, j \leq n$, one has

$$\sum_{r=1}^{n} \overline{A_{r,j}} A_{r,i} = \delta_{i,j}$$

Thus, A is a unitary matrix if and only if the rows of A form an orthonormal collection of vectors in F^n (with the standard inner product). Similarly, using the fact that $AA^* = I$, one sees that the columns of A must also form an orthonormal collection of vectors in F^n (and hence an orthonormal basis).

Thus, a matrix A is unitary if and only if its rows (or its columns) form an orthonormal basis for F^n with the standard inner product.

Now, if $\mathcal{B}$ is an orthonormal basis for V and T is any linear operator, then, if $A := [T]_{\mathcal{B}}$, one has

$$[T^*T]_{\mathcal{B}} = [T^*]_{\mathcal{B}}[T]_{\mathcal{B}} = A^*A$$

by Corollary 3.5. The next theorem is a simple corollary of this fact.

Theorem 4.9. *Let $U \in L(V)$ be a linear operator on a finite dimensional inner product space V. Then, U is a unitary operator if and only if there is an orthonormal basis $\mathcal{B}$ of V such that $A := [U]_{\mathcal{B}}$ is a unitary matrix.*

Definition 4.10. A real or complex $n \times n$ matrix A is said to be *orthogonal* if $A^tA = AA^t = I$.

Once again, if $A^tA = I$, then one concludes that $AA^t = I$ automatically.

Now, observe that a real orthogonal matrix (ie. an orthogonal matrix with real entries) is automatically unitary. Furthermore, a unitary matrix is orthogonal if and only if its entries are all real.

Example 4.11.

(i) If $A = [c]$ is a 1×1 matrix, then A is orthogonal if and only if $c = \pm 1$ and A is unitary if and only if $\overline{c}c = 1$ (equivalently, $c = e^{i\theta}$ for some $\theta \in \mathbb{R}$).

(ii) Let

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

be a 2×2 matrix , then A is orthogonal if and only if

$$A^t = A^{-1} = \frac{1}{ad - bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

Since A is orthogonal,

$$1 = \det(A^tA) = \det(A)^2$$

so $\det(A) = \pm 1$. Hence, A is orthogonal if and only if

$$A = \begin{pmatrix} a & b \\ -b & a \end{pmatrix} \text{ or } A = \begin{pmatrix} a & b \\ b & -a \end{pmatrix}$$

where $a, b \in \mathbb{R}$ are such that $a^2 + b^2 = 1$.

(iii) For instance, if $\theta \in \mathbb{R}$, then the matrix

$$A_\theta = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$$

is an orthogonal matrix. As an operator on $\mathbb{R}^2$, this represents a rotation by θ degrees.

(iv) Let

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

Then A is unitary if and only if

$$\begin{pmatrix} \overline{a} & \overline{c} \\ \overline{b} & \overline{d} \end{pmatrix} = \frac{1}{ad - bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

If A is unitary, then

$$1 = \det(A^*A) = \det(A^*)\det(A) = \overline{\det(A)}\det(A)$$

Hence, $|\det(A)| = 1$, so $\det(A) = e^{i\theta}$ for some $\theta \in \mathbb{R}$. Hence, A is unitary if and only if

$$A = \begin{pmatrix} a & b \\ -e^{i\theta}\overline{b} & e^{i\theta}\overline{a} \end{pmatrix}$$

where $\theta \in \mathbb{R}$ and $a, b \in \mathbb{C}$ are such that $|a|^2 + |b|^2 = 1$.

Recall that the set $U(n)$ of $n \times n$ unitary matrices forms a group under multiplication. Set $T^+(n)$ to be the set of all lower-triangular matrices whose entries on the principal diagonal are all positive. Note that every such matrix is necessarily invertible (since its determinant would be non-zero). The next lemma is a short exercise. It can be proved 'by hand'; or by a proof described in the textbook (See [Hoffman-Kunze, Page 306])

Lemma 4.12. *$T^+(n)$ is a group under matrix multiplication.*

Theorem 4.13. *Let $B \in \mathbb{C}^{n\times n}$ be an $n \times n$ invertible matrix. Then, there exists a unique lower-triangular matrix M with positive entries on the principal diagonal such that $U := MB$ is unitary.*

Proof.

(i) Existence: The rows $\beta_1, \beta_2, \ldots, \beta_n$ form a basis for $\mathbb{C}^n$. Let $\alpha_1, \alpha_2, \ldots, \alpha_n$ be the vectors obtained by the Gram-Schmidt process (Theorem 2.9). For each $1 \leq i \leq k$, the set $\{\alpha_1, \alpha_2, \ldots, \alpha_k\}$ is an orthogonal basis for $\text{span}\{\beta_1, \beta_2, \ldots, \beta_k\}$, and

$$\alpha_k = \beta_k - \sum_{i=1}^{k-1} \frac{(\beta_k|\alpha_i)}{\|\alpha_i\|^2}\alpha_i$$

Set

$$C_{k,j} := \frac{(\beta_k|\alpha_i)}{\|\alpha_i\|^2}$$

Let U be the unitary matrix whose rows are

$$\frac{\alpha_1}{\|\alpha_1\|}, \frac{\alpha_2}{\|\alpha_2\|}, \ldots, \frac{\alpha_n}{\|\alpha_n\|}$$

and M be the matrix defined by

$$M_{k,j} = \begin{cases} -\frac{C_{k,j}}{\|\alpha_k\|} & : \text{ if } j < k \\ \frac{1}{\|\alpha_k\|} & : j = k \\ 0 & : j > k \end{cases}$$

Then, M is lower-triangular and the entries on its principal diagonal are all positive. Furthermore, by construction, we have

$$\frac{\alpha_k}{\|\alpha_k\|} = \sum_{j=1}^{n} M_{k,j}\beta_j$$

This implies that $U = MB$ as required.

(ii) Uniqueness: Suppose $M_1, M_2 \in T^+(n)$ are such that M_1B and M_2B are both in $U(n)$. Since $U(n)$ is a group, it follows that

$$M_1M_2^{-1} = (M_1B)(M_2B)^{-1} \in U(n)$$

But, by Lemma 4.12,

$$M_1M_2^{-1} \in T^+(n)$$

But, for any matrix $U \in U(n)$, one has

$$U^* = U^{-1}$$

Thus,

$$(M_1M_2^{-1})^* = (M_1M_2^{-1})^{-1} \in T^+(n)$$

But $(M_1M_2^{-1})^*$ is the conjugate-transpose of a lower-triangular matrix, and is thus upper-triangular. Thus, $(M_1M_2^{-1})^*$ is both lower and upper-triangular, and is thus a diagonal matrix.

However, if a diagonal matrix is unitary, then each of its diagonal entries must have modulus 1. Since the diagonal entries of $M_1M_2^{-1}$ are all positive real numbers, we thus conclude that

$$M_1M_2^{-1} = I$$

whence $M_1 = M_2$, as required.

□

We set $GL(n)$ to be the set of all $n \times n$ invertible matrices. Observe that $GL(n)$ is also a group under matrix multiplication. We conclude that

Corollary 4.14. *For any $B \in GL(n)$, there exist unique matrices $M \in T^+(n)$ and $U \in U(n)$ such that*

$$B = MU$$

Recall that two matrices $A, B \in F^{n\times n}$ are similar if there exists an invertible matrix P such that $B = P^{-1}AP$.

Definition 4.15. Let $A, B \in F^{n\times n}$ be two matrices. We say that they are

(i) *unitarily equivalent* if there exists a unitary matrix $U \in U(n)$ such that $B = U^{-1}AU$

(ii) *orthogonally equivalent* if there exists an orthogonal matrix P such that $B = P^{-1}AP$.

5. Normal Operators

The goal of this section is to answer the following question: Given a linear operator T on a finite dimensional inner product space, under what conditions does V have an orthonormal basis consisting of characteristic vectors of T? In other words, does there exist an *orthonormal* basis $\mathcal{B}$ of V such that $[T]_{\mathcal{B}}$ is diagonal?

Clearly, T must be diagonalizable in the sense of Definition VI.2.7. To see if we need something more, we begin with a necessary condition. Suppose $\mathcal{B} = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ is an orthonormal basis with the property that

$$T\alpha_j = c_j\alpha_j, \quad j = 1, 2, \ldots, n$$

Then, $[T]_{\mathcal{B}}$ is a diagonal matrix, so by Theorem V.4.6, the matrix $[T^*]_{\mathcal{B}}$ is also diagonal, with diagonal entries $\overline{c_j}$. In other words,

$$T^*\alpha_k = \overline{c_k}\alpha_k, \quad k = 1, 2, \ldots, n$$

If V is a real inner product space, then $\overline{c_k} = c_k$, so it must happen that $T = T^*$.

If V is a complex inner product space, then it must happen that

$$TT^* = T^*T$$

because any two diagonal matrices commute with each other. It turns out, this condition is enough to ensure that such a basis exists.

Definition 5.1. We say that an operator $T \in L(V)$ defined on an inner product space is *normal* if

$$TT^* = T^*T$$

Clearly, every self-adjoint operator is normal, every unitary operator is normal; however sums and products of normal operators need not be normal. We begin our study with self-adjoint operators.

Theorem 5.2. *Let V be an inner product space and $T \in L(V)$ be self-adjoint. Then,*

(i) Each characteristic value of T is real.

(ii) Characteristic vectors associated to different characteristic values are orthogonal.

Proof.

(i) Suppose $c \in F$ is a characteristic value of T with characteristic vector α, then $\alpha \neq 0$, and

$$\begin{aligned} c(\alpha|\alpha) &= (c\alpha|\alpha) = (T\alpha|\alpha) \\ &= (\alpha|T^*\alpha) = (\alpha|T\alpha) \\ &= (\alpha|c\alpha) = \overline{c}(\alpha|\alpha) \end{aligned}$$

Since $(\alpha|\alpha) \neq 0$, it follows that $c = \overline{c}$, so that $c \in \mathbb{R}$.

(ii) Suppose $T\beta = d\beta$ and $d \neq c$ and $\beta \neq 0$, then

$$\begin{aligned} c(\alpha|\beta) &= (c\alpha|\beta) = (T\alpha|\beta) \\ &= (\alpha|T^*\beta) = (\alpha|T\beta) \\ &= (\alpha|d\beta) = \overline{d}(\alpha|\beta) \\ &= d(\alpha|\beta) \end{aligned}$$

Where the last equality follows from part (i). Since $c \neq d$, it follows that $(\alpha|\beta) = 0.0$

□

Theorem 5.3. *Let $V \neq \{0\}$ be a finite dimensional inner product space and $0 \neq T \in L(V)$ be self-adjoint. Then, T has a non-zero characteristic value.*

Proof. Let $n := \dim(V) > 0$ and $\mathcal{B}$ be an orthonormal basis for V, and let

$$A := [T]_{\mathcal{B}}$$

Then, $A = A^*$. Let W be the space of all $n \times 1$ matrices over $\mathbb{C}$ with inner product

$$(X|Y) := Y^*X$$

Define $U : W \to W$ be given by $UX := AX$, then U is a self-adjoint linear operator on W, and the characteristic polynomial of U is

$$f = \det(xI - A)$$

By the fundamental theorem of algebra, f has a root $c \in \mathbb{C}$. Thus, there exists $X \in W$ non-zero such that

$$AX = cX$$

Since U is self-adjoint, it follows by Theorem 5.2 that $c \in \mathbb{R}$ is real. Now consider two cases:

(i) If V is a complex inner product space, then we immediately obtain $\alpha \in V$ such that $T\alpha = c\alpha$.

(ii) If V is a real inner product space, then A has real entries. Since $(A - cI)$ has real entries, it follows that we may choose X to have real entries. Thus, there exists $\alpha \in V$ such that

$$T\alpha = cA$$

□

Theorem 5.4. *Let V be a finite dimensional inner product space and $T \in L(V)$. If $W \subset V$ is a T-invariant subspace, then $W^\perp$ is T^*-invariant.*

Proof. Suppose W is T-invariant and $\alpha \in W^\perp$. We wish to show that $T^*(\alpha) \in W^\perp$. For this, fix $\beta \in W$, and note that

$$(T^*(\alpha)|\beta) = (\alpha|T\beta) = 0$$

because $T\beta \in W$. This is true for all $\beta \in W$, so $T^*\alpha \in W^\perp$ as required. □

Theorem 5.5. *Let V be a finite dimensional inner product space and $T \in L(V)$ be self-adjoint. Then, there is an orthonormal basis of V, each vector of which is a characteristic vector of T.*

Proof. Assume $\dim(V) > 0$. By Theorem 5.3, there exists $c \in F$ and $\alpha \in V$ such that $\alpha \neq 0$ and

$$T\alpha = c\alpha$$

Set $\alpha_1 := \alpha/\|\alpha\|$, then $\{\alpha_1\}$ is orthonormal. Hence, if $\dim(V) = 1$, then we are done.

Now suppose $\dim(V) > 1$ and assume that the theorem is true for any self-adjoint operator $S \in L(W)$ on an inner product space V' with $\dim(V') < \dim(V)$. If α_1 as above, set

$$W := \text{span}(\{\alpha_1\})$$

Then, W is T-invariant. So, by Theorem 5.4, $V' := W^\perp$ is T^*-invariant. But $T = T^*$, so we have a direct sum decomposition

$$V = W \oplus V'$$

of V into T-invariant subspaces. Now consider $S := T|_{V'} \in L(V')$. By induction hypothesis, V' has an orthonormal basis $\mathcal{B}' = \{\alpha_2, \alpha_3, \ldots, \alpha_n\}$ each vector of which is a characteristic vector of S. Thus,

$$\mathcal{B} := \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$$

is an orthonormal basis for V, each vector of which is a characteristic vector of T. □

The next corollary follows from Theorem 5.5 by applying it to the vector space $V = \mathbb{C}^{n\times 1}$. Check the details!

Corollary 5.6. *Let A be an $n \times n$ Hermitian (self-adjoint) matrix. Then, there is a unitary matrix P such that $P^{-1}AP$ is diagonal. If A is a real symmetric matrix, then there is a real orthogonal matrix P such that $P^{-1}AP$ is diagonal.*

We now look to understand normal operators.

Theorem 5.7. *Let V be a finite dimensional inner product space and $T \in L(V)$ be normal. If $c \in F$ is a characteristic value of T with characteristic vector α, then $\overline{c}$ is a characteristic value of T^* with characteristic vector α.*

Proof. Suppose S is normal and $\beta \in V$, then

$$\|S\beta\|^2 = (S\beta|S\beta) = (\beta|S^*S\beta) = (\beta|SS^*\beta) = (S^*\beta|S^*\beta) = \|S^*\beta\|^2$$

Hence, taking $S = (T - cI)$ (which is normal) and $\beta = \alpha$, we see that

$$0 = \|(T - cI)\alpha\| = \|(T^* - \bar{c}I)\alpha\|$$

Thus, $T^*\alpha = \bar{c}\alpha$ as required. □

Theorem 5.8. *Let V be a finite dimensional complex inner product space and $T \in L(V)$ be a normal operator. Then, there is an orthonormal basis of V, each vector of which is a characteristic vector of T.*

Proof. Once again, we proceed by induction on $\dim(V)$. Note that, since V is assumed to be a complex inner product space, every operator on V has at least one characteristic value (since the characteristic polynomial has a root by the fundamental theorem of algebra). Therefore, if $\dim(V) = 1$, there is nothing to prove.

Now suppose $\dim(V) > 1$ and that the theorem is true for any complex inner product space V' with $\dim(V') < \dim(V)$. Then, let $\alpha \in V$ be a characteristic vector of T associated to any fixed characteristic value $c \in \mathbb{C}$. Furthermore, taking $\alpha_1 := \alpha/\|\alpha\|$, we set

$$W := \text{span}(\{\alpha_1\})$$

Then, W is T-invaraint. Hence, $W^\perp$ is invaraint under T^* by Theorem 5.4.

However, by Theorem 5.7, W is also invariant under T^*. Hence, $W^\perp$ is invariant under $T = (T^*)^*$ as well. Thus, if

$$V' := W^\perp$$

Then V' is invariant under T and T^*. Thus, if $S := T|_{V'}$, then S is normal because $S^* = T^*|_{V'}$ and these two operators must commute. Hence, by induction hypothesis, V' has an orthonormal basis $\mathcal{B}' = \{\alpha_2, \alpha_3, \ldots, \alpha_n\}$ consisting of characteristic vectors of S. Hence,

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

is an orthonormal basis of V consisting of characteristic vectors of T. □

Note that an $n \times n$ matrix A is said to be *normal* if $AA^* = A^*A$. The next corollary follows from Theorem 5.8 as before.

Corollary 5.9. *For every normal matrix A, there is a unitary matrix U such that $U^{-1}AU$ is diagonal.*

Remark 5.10.

(i) Theorem 5.8 is an important theorem, called the *Spectral Theorem*. Its generalization to the case of infinite dimensional inner product spaces is a deep result that you may learn in your fifth year.

(ii) The Spectral theorem does not hold for *real* inner product spaces. For instance, a normal operator on such an inner product space may not even have one characteristic value. For instance, we may consider the linear operator $T \in L(\mathbb{R}^2)$ given in the standard basis by the matrix

$$A = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$$

You may check that for most values of $\theta \in \mathbb{R}$ such an operator has no (real) characteristic values. However, such an opertor is always normal.

IX. Instructor Notes

(i) Given that the semester was entirely online, my expectations were low, but the course was truly abysmal. I was simply recording videos and uploading them every week with virtually no feedback from the students.

(ii) The assessment, hampered by poor administrative guidelines, was meaningless as the students copied everything. Therefore, from my perspective, the entire semester was a wash-out.

(iii) The material though, is fine, and can be used for future courses as is.

Bibliography

[Hoffman-Kunze] K. Hoffman, R. Kunze, *Linear Algebra (2nd Edition)*, Prentice-Hall (1971)

[Conrad] K. Conrad, https://kconrad.math.uconn.edu/blurbs/grouptheory/sign.pdf