

# **LINEAR ALGEBRA II**

Anita Naolekar, notes by Ramdas Singh

Second Semester

# List of Symbols

# Contents

1	PERMUTATION GROUPS	1
1.1	Even and Odd Permutations . . . . .	1
1.2	The Determinant . . . . .	2
2	EIGENVECTORS AND EIGENVALUES	7
2.1	Linear Transformers and an Introduction . . . . .	7
2.1.1	Linear Operators . . . . .	8
2.1.2	Eigenvectors and Eigenvalues . . . . .	8
2.2	Finding Eigenvalues and Eigenvectors . . . . .	9
2.2.1	Eigenspace . . . . .	10
2.3	Diagonalizability . . . . .	10
2.4	Polynomials . . . . .	11
2.4.1	Interaction with Linear Operators . . . . .	11
2.5	Triangularizability . . . . .	14
2.5.1	Determinant of Partitioned Matrices . . . . .	15
2.6	On the Characteristic and Minimal Polynomials . . . . .	15
3	INNER PRODUCT SPACES	17
3.1	An Introduction . . . . .	17
3.2	The Notion of Length and Orthogonality . . . . .	18
3.3	Orthogonality and Orthonormality . . . . .	19
3.3.1	The Orthogonal Complement . . . . .	20
3.3.2	Orthogonal and Unitary Matrices . . . . .	21
	Appendices	23
A	Appendix	25
	Index	27

## Chapter 1

# PERMUTATION GROUPS

*January 3rd.*

Let  $S_n$  denote the set of all bijections (permutations) on the set  $\{1, 2, \dots, n\}$ . If  $\sigma, \tau \in S_n$ , let us define  $\sigma\tau$  to be the bijection defined as

$$(\sigma\tau)(i) = \sigma(\tau(i)) \forall 1 \leq i \leq n. \quad (1.1)$$

This gives us a binary operation on  $S_n$  which is associative, and  $S_n$  will then contain the identity permutation  $1$  such that  $\sigma 1 = 1\sigma = \sigma$  for all  $\sigma \in S_n$ . For every such  $\sigma$ , we can also find a  $\sigma^{-1} \in S_n$  such that  $\sigma\sigma^{-1} = \sigma^{-1}\sigma = 1$ . The set  $S_n$  equipped with this binary operation, thus, forms a group. In this case, we call  $S_n$  as the *symmetric group* of degree  $n$ . We now define a cycle in regards to permutations.

**Definition 1.1.** A *cycle* is a string of positive integers, say  $(i_1, i_2, \dots, i_k)$ , which represents the permutation  $\sigma \in S_n$  (with  $k \leq n$ ) such that  $\sigma(i_j) = i_{j+1}$  for all  $1 \leq j \leq k-1$ , and  $\sigma(i_k) = i_1$ , and fixes all other integers.

We also note that  $S_3$  is the smallest Abelian group possible, upto isomorphism.  $S_3$  is one of the only two groups of order 6, and can be written as

$$S_3 = \{1, \sigma = (1, 2, 3), \sigma^2 = (1, 3, 2), \tau = (1, 2), \sigma\tau = (1, 3), \tau\sigma = (2, 3)\}. \quad (1.2)$$

Some other observations arise. We find that  $\sigma^3 = \tau^2 = 1$ , and that  $\tau\sigma = \sigma^2\tau$ . We notice another fact via this  $\sigma$ ;

**Remark 1.2.** A  $k$ -cycle  $\sigma = (i_1, i_2, \dots, i_k)$  is of order  $k$ , that is,  $\sigma^k = 1$ .

**Definition 1.3.** Two cycles in  $S_n$  are called disjoint if they have no integer in common.

We note that if  $\sigma$  and  $\tau$  are two disjoint cycles in  $S_n$  then  $\sigma$  and  $\tau$  commute, that is,  $\sigma\tau = \tau\sigma$ .

**Proposition 1.4.** Every  $\sigma$  in  $S_n$  can be written uniquely as a product of disjoint cycles.

Every cycle can be written as a product of 2-cycles. 2-cycles are called *transpositions*. This can easily be seen as

$$(a_1, a_2, \dots, a_n) = (a_1, a_n)(a_1, a_{n-1}) \cdots (a_1, a_3)(a_1, a_2). \quad (1.3)$$

## 1.1 Even and Odd Permutations

Let  $x_1, x_2, \dots, x_n$  be indeterminates, and let

$$\Delta = \prod_{1 \leq i < j \leq n} (x_i - x_j). \quad (1.4)$$

Let  $\sigma \in S_n$ , and define

$$\sigma(\Delta) = \prod_{1 \leq i < j \leq n} (x_{\sigma(i)} - x_{\sigma(j)}). \quad (1.5)$$

We find that  $\sigma(\Delta) = \pm\Delta$ . Based on this, we classify permutations as odd or even.

**Definition 1.5.** A permutation  $\sigma$  is said to be an *even permutation* if  $\sigma(\Delta) = \Delta$ , and is said to be an *odd permutation* if  $\sigma(\Delta) = -\Delta$ . The sign of a permutation  $\sigma$ , denoted by  $\epsilon(\sigma)$ , is  $+1$  if  $\sigma$  is even, and is  $-1$  if  $\sigma$  is odd. So,  $\sigma(\Delta) = \epsilon(\sigma)\Delta$ .

**Proposition 1.6.** The map  $\epsilon : S_n \rightarrow \{-1, +1\}$ , where  $\epsilon(\sigma)$  is the sign of  $\sigma$ , is a homomorphism, that is,  $\epsilon(\sigma\tau) = \epsilon(\sigma)\epsilon(\tau)$  for all  $\sigma, \tau \in S_n$ .

*Proof.* Start with  $\tau(\Delta)$ ;

$$\tau(\Delta) = \prod_{1 \leq i < j \leq n} (x_{\tau(i)} - x_{\tau(j)}). \quad (1.6)$$

Let there be  $k$  factors of this polynomial where  $\tau(i) > \tau(j)$  with  $i < j$ . We find that  $\tau(\Delta) = (-1)^k \Delta$ , and so,  $\epsilon(\tau) = (-1)^k$ . Now,  $\sigma\tau(\Delta)$  has exactly  $k$  factors of the form  $x_{\sigma(j)} - x_{\sigma(i)}$ , with  $j > i$ . Bringing out a factor  $(-1)^k$ , we find that  $\sigma\tau(\Delta)$  has all factors of the form  $x_{\sigma(i)} - x_{\sigma(j)}$ , with  $j > i$ . Thus,

$$\epsilon(\sigma\tau)\Delta = \sigma\tau(\Delta) = (-1)^k \prod_{1 \leq i < j \leq n} (x_{\sigma(i)} - x_{\sigma(j)}) = (-1)^k \sigma(\Delta) = (-1)^k \epsilon(\sigma)\Delta = \epsilon(\tau)\epsilon(\sigma)\Delta. \quad (1.7)$$

Cancelling out the  $\Delta$ , we find  $\epsilon(\sigma\tau) = \epsilon(\sigma)\epsilon(\tau)$ . ■

$\epsilon$  is a homomorphism to an Abelian group, so  $\epsilon(\sigma\tau) = \epsilon(\sigma)\epsilon(\tau) = \epsilon(\tau)\epsilon(\sigma)$ .

**Proposition 1.7.** If  $\lambda = (i, j)$  is a transposition, then  $\epsilon(\lambda) = -1$ .

*Proof.* If  $\lambda = (1, 2) \in S_n$ , it is easy to show that

$$\lambda(\Delta) = (x_1 - x_2) \cdots (x_1 - x_n)(x_2 - x_3) \cdots (x_2 - x_n) \cdots = (-1)(\Delta). \quad (1.8)$$

Now, if  $\sigma = (i, j)$ , with  $(i, j) \neq (1, 2)$ , then  $(i, j) = \lambda(1, 2)\lambda$  where  $\lambda$  interchanges 1 and  $i$ , and interchanges 2 and  $j$ . Using that fact that  $\epsilon$  is a homomorphism,  $\epsilon(\sigma) = -1$ . ■

A cycle  $\sigma$  of length  $k$  is an even permutation if and only if  $k$  is odd. This is because it can be decomposed into  $k - 1$  transpositions, and we would then have  $\epsilon(\sigma) = (-1)^{k-1} = 1$  (using the fact that  $\epsilon$  is a homomorphism). Some more corollaries of the previous proposition include the fact that  $\epsilon$  is a surjective map, and that  $\epsilon(\sigma^{-1}) = \epsilon(\sigma)$ .

If, for  $\sigma \in S_n$ ,  $\sigma$  can be decomposed as  $\sigma_1\sigma_2 \cdots \sigma_k$ , where  $\sigma_i$  is a  $m_i$ -cycle, then  $\epsilon(\sigma_i) = (-1)^{m_i-1}$ , and  $\epsilon(\sigma) = (-1)^{(\sum m_i) - k}$ .

**Proposition 1.8.**  $\sigma$  is an odd permutation if and only if the number of cycles of even length in its cycle decomposition is odd.

## 1.2 The Determinant

**Definition 1.9.** If  $A = (a_{ij})$  is a square matrix of order  $n$ , then the *determinant* of  $A$  is defined as

$$\det A = \sum_{\sigma \in S_n} \epsilon(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)}. \quad (1.9)$$

Using this definition of the determinant of a square matrix, one may derive the usual determinant properties with ease.

January 7th.

**Remark 1.10.** The following properties may be inferred:

- If  $A$  contains a row of zeroes, or a column of zeroes, then  $\det A = 0$ .
- $\det I_n = 1$ .
- The determinant of a diagonal matrix is the product of the diagonal elements. This is because if  $\sigma \in S_N$  is not the identity permutation, then there exists at least one element in the corresponding term where  $i \neq \sigma(i)$ , and  $a_{i\sigma(i)}$  makes the term zero. For the identity transformation, it contains only those elements of the form  $a_{ii}$ .

Other non-trivial properties may also be shown with ease.

**Corollary 1.11.** *If  $A$  is an upper triangular matrix, then  $\det A$  is the product of the diagonal entries.*

*Proof.* If  $a_{1\sigma(1)} \cdots a_{n\sigma(n)} \neq 0$ , then  $a_{n\sigma(n)} \neq 0$ , that is,  $\sigma(n) = n$ , as  $a_{ni} = 0 \ \forall \ i < n$ . Again,  $a_{(n-1)\sigma(n-1)} \neq 0$  leads us to conclude that  $\sigma(n-1) = n-1$  as  $\sigma$  is a bijection and has to lead to a non-zero element. By similar logic,  $\sigma(i) = i$  for all valid  $i$ . So,  $\sigma$  is the identity permutation. ■

**Corollary 1.12.** *If  $A$  is a lower triangular matrix, then  $\det A$  is the product of the diagonal entries.*

*Proof.* The proof of this is similar to the previous proof if we consider that the determinant of the transpose of a matrix is equal to the determinant of said matrix. ■

**Theorem 1.13.** *The determinant of a matrix is equal to the determinant of its transpose, that is,  $\det A = \det A^t$  for a square matrix  $A$ .*

*Proof.* The proof is left as an exercise to the reader. ■

**Proposition 1.14.** *Let  $B$  be obtained from  $A$  by multiplying a row (or column) of  $A$  by a non-zero scalar,  $\alpha$ . Then,  $\det B = \alpha \det A$ .*

*Proof.* The proof is left as an exercise to the reader. ■

**Proposition 1.15.** *If  $B$  is obtained from  $A$  by interchanging any two rows (or columns) of  $A$ , then  $\det B = -\det A$ .*

*Proof.* Let  $B$  be obtained from  $A$  by interchanging the rows  $k$  and  $l$ , with  $k < l$ . We then have

$$\begin{aligned} \det B &= \sum_{\sigma \in S_n} \epsilon(\sigma) b_{1\sigma(1)} b_{2\sigma(2)} \cdots b_{n\sigma(n)} \\ &= \sum_{\sigma \in S_n} \epsilon(\sigma) a_{1\sigma(1)} \cdots a_{(k-1)\sigma(k-1)} a_{l\sigma(k)} a_{(k+1)\sigma(k+1)} \cdots a_{k\sigma(l)} \cdots a_{n\sigma(n)}. \end{aligned} \quad (1.10)$$

As  $\sigma$  runs through all elements in  $S_n$ ,  $\tau = \sigma(k, l)$  also runs through all  $S_n$ . Hence, via  $\epsilon(\tau) = -\epsilon(\sigma)$ , the equation now looks like

$$\det B = - \sum_{\tau \in S_n} \epsilon(\tau) a_{1\tau(1)} \cdots a_{l\tau(l)} \cdots a_{k\tau(k)} \cdots a_{n\tau(n)} = -\det A. \quad (1.11)$$

■

**Proposition 1.16.** *If two rows (or columns) of  $A$  are equal, then  $\det A = 0$ .*

*Proof.* Suppose that the rows  $k$  and  $l$  of  $A$  are equal. Interchanging will alter the determinant by  $-1$ , so  $\det A = -\det A \implies 2\det A = 0 \implies \det A = 0$  if  $2 \neq 0$  in the field  $F$  from where the elements of  $A$  arrive.

If  $2 = 0$  in  $F$ , that is,  $F$  is of characteristic 2, we pair the  $\sigma$  term in the expression of  $\det A$  with the term  $\tau$  where  $\tau = \sigma(k, l)$ . The terms corresponding to  $\sigma$  and  $\tau$  in the expressions are the same, differing in only the sign. Hence,  $\det A = 0$ . ■

**Theorem 1.17.** *For a fixed  $k$ , let the row  $k$  of  $A$  be the sum of the two row vectors  $X^t$  and  $Y^t$ , that is,  $a_{kj} = x_j + y_j$  for all  $1 \leq j \leq n$ . Then  $\det A = \det B + \det C$  where  $B$  is obtained from  $A$  by replacing the row  $k$  of  $A$  by the row vector  $X^t$ , and  $C$  is obtained from  $A$  by replacing the row  $k$  of  $A$  by the row vector  $Y^t$ .*

*Proof.* We utilize the fact that  $a_{kj} = x_j + y_j$ . We have

$$\begin{aligned} \det A &= \sum_{\sigma \in S_n} \epsilon(\sigma) a_{1\sigma(1)} \cdots a_{k\sigma(k)} \cdots a_{n\sigma(n)} \\ &= \left( \sum_{\sigma \in S_n} \epsilon(\sigma) a_{1\sigma(1)} \cdots x_{\sigma(k)} \cdots a_{n\sigma(n)} \right) + \left( \sum_{\sigma \in S_n} \epsilon(\sigma) a_{1\sigma(1)} \cdots y_{\sigma(k)} \cdots a_{n\sigma(n)} \right) \\ &= \det B + \det C. \end{aligned}$$

■

**Proposition 1.18.** *If a scalar multiple of a row (or column) is added to a row (or column) of a matrix, the determinant remains unchanged.*

*Proof.* The proof follows immediately from the previously proved properties. ■

January 10th.

**Definition 1.19.** For  $a_{ij} \in A$ , the *cofactor* of  $a_{ij}$  is  $A_{ij} = (-1)^{i+j} \det M_{ij}$ , where  $M_{ij}$  is the  $(n-1) \times (n-1)$  matrix obtained from  $A$  by deleting the  $i^{\text{th}}$  row and  $j^{\text{th}}$  column of  $A$ .

**Lemma 1.20.** *Fix  $k, j$ . If  $a_{kl} = 0$  for all  $l \neq j$ , then  $\det A = a_{kj} A_{kj}$ .*

*Proof.* Take  $A$  to be a  $n \times n$  matrix. We deal in cases.

- Case I:  $k = j = n$ . In the expansion of the determinant,

$$\det A = \sum_{\sigma \in S_n} \epsilon(\sigma) a_{1\sigma(1)} \cdots a_{n\sigma(n)},$$

only those  $\sigma$ 's survive where  $\sigma(n) = n$ . These  $\sigma$ 's can be thought of as permutations of  $S_{n-1}$  instead. The sign of  $\sigma \in S_n$  and  $\sigma \in S_{n-1}$  is the same as  $n$  is fixed. Thus, we get

$$a_{nn} \sum_{\sigma \in S_{n-1}} \epsilon(\sigma) a_{1\sigma(1)} \cdots a_{(n-1)\sigma(n-1)} = a_{nn} \det M_{nn} = (-1)^{n+n} a_{nn} A_{nn} = a_{nn} A_{nn}. \quad (1.12)$$

- Case II:  $(k, j) \neq (n, n)$ . We construct a matrix  $B$  by interchanging  $n-k$  rows and  $n-j$  columns to bring  $a_{ij}$  to the position  $(n, n)$ . Thus, we have  $\det B = (-1)^{n-k+n-j} \det A = (-1)^{k+j} \det A$ . But  $B = a_{kj} \det M_{kj}$ , so

$$\det A = (-1)^{k+j} a_{kj} \det M_{kj} = a_{kj} A_{kj}. \quad (1.13)$$

■

**Theorem 1.21.** *Let  $A$  be a  $n \times n$  matrix, and let  $1 \leq k \leq n$ . Then,  $\det A = \sum_{j=1}^n a_{kj} A_{kj}$ , expansion by the  $k^{\text{th}}$  row.*

*Proof.* Write out the  $k^{\text{th}}$  row of  $A$  as  $x_1^t + \cdots + x_n^t$ , where  $x_i = (0, \dots, 0, a_{ki}, 0, \dots, 0)^t$ , and all the other rows remaining are the same. Writing the matrix  $A$  as the sum of  $n$  matrices where each matrix is the same as  $A$  but with a row that looks like  $x_i^t$ , we can easily show that  $\det A = \sum_{j=1}^n a_{kj} A_{kj}$ . ■

**Example 1.22.** Let  $n \geq 1$ , and let  $A_n = \begin{pmatrix} a_1^{n-1} & a_1^{n-2} & \cdots & a_1 & 1 \\ a_2^{n-1} & a_2^{n-2} & \cdots & a_2 & 1 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_n^{n-1} & a_n^{n-2} & \cdots & a_n & 1 \end{pmatrix}$ . Then,  $\det A_n = \prod_{1 \leq i < j \leq n} (a_i - a_j)$ .

*Proof.* If  $a_i = a_j$  for some  $i \neq j$ , then  $\det A_n = 0$  as two rows are then identical. Hence, assume that the  $a_i$ 's are distinct. Now construct

$$B_n = \begin{pmatrix} x_1^{n-1} & x_1^{n-2} & \cdots & x_1 & 1 \\ a_2^{n-1} & a_2^{n-2} & \cdots & a_2 & 1 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_n^{n-1} & a_n^{n-2} & \cdots & a_n & 1 \end{pmatrix}. \quad (1.14)$$

Notice that  $\det B_n \in F[x]$ , where  $F$  is the field, and  $x$  is an indeterminate.  $\det B$  is also of degree  $(n-1)$ ; let us call this polynomial  $f(x)$ . Each of  $a_2, \dots, a_n$  are roots of  $f(x)$ , so  $f(x)$  must be of the form  $f(x) = C(x - a_2) \cdots (x - a_n)$ . Equating coefficients of  $x^{n-1}$ , we get

$$C = \prod_{2 \leq i < j \leq n} (a_i - a_j) = \det \begin{pmatrix} a_2^{n-2} & \cdots & a_2 & 1 \\ \cdots & \cdots & \cdots & \cdots \\ a_n^{n-2} & \cdots & a_n & a_1 \end{pmatrix}. \quad (1.15)$$

Thus, we must have

$$f(x) = \left( \prod_{2 \leq i < j \leq n} (a_i - a_j) \right) (x - a_2) \cdots (x - a_n) \quad (1.16)$$

$$\implies \det A_n = f(1) = \prod_{1 \leq i < j \leq n} (a_i - a_j). \quad (1.17)$$

■

**Example 1.23.** Show that there exists a unique polynomial of degree  $n$  that takes arbitrary prescribed values at the  $(n+1)$  points  $x_0, x_1, \dots, x_n$ .





## Chapter 2

# EIGENVECTORS AND EIGENVALUES

### 2.1 Linear Transformers and an Introduction

Let  $\mathcal{B} = (v_1, \dots, v_n)$  be a basis of vector space  $V$  and  $\mathcal{C} = (w_1, \dots, w_m)$  be a basis of a vector space  $W$ . As these are bases, given a  $v \in V$ , there exists a unique  $X \in F^n$  such that  $v = \mathcal{B}X$ , called the *coordinate vector* of  $v$  with respect to the basis  $\mathcal{B}$ . We note that since the mapping from a  $v \in V$  to a  $X \in F^n$  is linear in nature and is bijection, the vector spaces  $V$  and  $F^n$  are isomorphic to each other. Similarly, a mapping that takes  $w \in W$  to  $Y \in F^m$  shows that  $W$  and  $F^m$  are isomorphic to each other.

Now suppose that there exists a linear map that takes  $v \mapsto Tv$  with  $v \in V$  and  $Tv \in W$ . This transformer  $T$  is with respect to the bases  $\mathcal{B}$  and  $\mathcal{C}$  of  $V$  and  $W$ , respectively. We construct the  $m \times n$  matrix  $A$  so that the  $j^{\text{th}}$  column of  $A$  is the coordinate vector of  $Tv_j$  with respect to the basis  $\mathcal{C}$ . We will then have  $T(\mathcal{B}) = \mathcal{C}A$ . For any vector  $v \in V$ , we have

$$\begin{aligned} v &= \mathcal{B}X = v_1x_1 + \dots + v_nx_n \\ \implies T(v) &= T(v_1)x_1 + \dots + T(v_n)x_n = (T(v_1), \dots, T(v_n)) \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix} = T(\mathcal{B})X = (\mathcal{C}A)X \end{aligned} \quad (2.1)$$

$$= (w_1, \dots, w_m)AX; \quad (2.2)$$

the coordinate vector of  $Tv$  with respect to the basis  $AX$ . In fact, if we denote the isomorphism from  $V$  to  $F^n$  by  $\phi_{\mathcal{B}}$  and the isomorphism from  $W$  to  $F^m$  by  $\phi_{\mathcal{C}}$ , we get  $\phi_{\mathcal{C}} \circ T = (\text{mult. by } A) \circ \phi_{\mathcal{B}}$ .

The next theorem will be divided into two parts.

**Theorem 2.1.** 1. *The vector space form. Let  $T : V \rightarrow W$  be a linear mapping between finite dimensional vector spaces  $V$  and  $W$ , of dimensions  $n$  and  $m$  respectively. There are bases  $\mathcal{B}$  and  $\mathcal{C}$  of  $V$  and  $W$  respectively such that the matrix of  $T$  with respect to the bases  $\mathcal{B}$  and  $\mathcal{C}$  looks like*

$$\begin{pmatrix} I_r & O_{r \times (n-r)} \\ O_{(m-r) \times r} & O_{(m-r) \times (n-r)} \end{pmatrix}_{m \times n}.$$

2. *The matrix form. If  $A$  is a  $m \times n$  matrix, then there exists an invertible matrix  $Q_{m \times m}$  and an invertible matrix  $P_{n \times n}$  such that  $Q^{-1}AP$  is of the form  $\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$ , where  $r$  is the rank of  $A$ .*

3. *In fact, both these forms of the theorem are equivalent.*

*Proof.* 1. Let  $(u_1, \dots, u_{n-r})$  be a basis of  $\ker T$ . We can extend this to a basis  $\mathcal{B}$  by appending independent vectors that do not belong to the kernel of  $T$ , that is,  $(v_1, \dots, v_r, u_1, \dots, u_{n-r})$ . Let  $(Tv_1, \dots, Tv_r)$  be a basis of  $\text{Im} T$ . We can extend this to a basis of  $W$ , say  $\mathcal{C} = (w_1, \dots, w_r, w_{r+1}, \dots, w_m)$ , where  $w_i = Tv_i$  for  $1 \leq i \leq r$ . These bases are the desired ones.

2.  $P$  is a sequence of column operations, multiplied to form a matrix, and  $Q^{-1}$  is a sequence of row operations, multiplied to form a matrix, that get the matrix  $A$  into the desired form. These are our desired  $P$  and  $Q$ .
3. Suppose the vector space form holds. Let  $A$  be a  $m \times n$  matrix over  $F$ , with  $A : F^n \rightarrow F^m$  defined as  $X \mapsto AX$ . There then exists a basis  $\mathcal{B}$  of  $F^n$  and a basis  $\mathcal{C}$  of  $F^m$  such that the linear map  $A$  with respect to the bases  $\mathcal{B}$  and  $\mathcal{C}$  has the desired matrix. We then have  $\mathcal{B} = I_n P_{n \times n}$  and  $\mathcal{C} = I_m Q_{m \times m}$ , with both  $P$  and  $Q$  invertible. We claim that the matrix of the linear mapping  $A$  with respect to the bases  $\mathcal{B}$  and  $\mathcal{C}$  is  $Q^{-1}AP$ . ■

January 16th.

**Proposition 2.2.** 1. Let  $T : V \rightarrow W$  be a linear map, and  $A$  the matrix of  $T$  with respect to the bases  $\mathcal{C}$  and  $\mathcal{C}$  of  $V$  and  $W$  respectively. Let  $\mathcal{B}'$  and  $\mathcal{C}'$  be new bases of  $V$  and  $W$  respectively, and let the change of basis matrices be given by  $\mathcal{B}' = \mathcal{B}P$  and  $\mathcal{C}' = \mathcal{C}Q$ . Then the matrix of  $T$  with respect to  $\mathcal{B}'$  and  $\mathcal{C}'$  is  $Q^{-1}AP$ .

2. If  $A' = Q_1^{-1}AP_1$ , where  $P_1$  and  $Q_1$  are  $n \times n$  and  $m \times m$  invertible matrices, respectively, then  $A'$  is the matrix of  $T$  with respect to the bases  $\mathcal{B}P_1$  and  $\mathcal{C}Q_1$ .

*Proof.* Let the coordinate vector of  $v$  with respect to the basis  $\mathcal{B}'$  be  $X'$ . We claim that the coordinate vector of  $Tv$  with respect to the basis  $\mathcal{C}'$  is  $Y'$ , where  $Y' = (Q^{-1}AP)X'$ . We assume that  $\mathcal{B}' = \mathcal{B}P_{n \times n}$ ,  $\mathcal{C}' = \mathcal{C}Q_{m \times m}$ , and  $T(\mathcal{B}) = \mathcal{C}A_{m \times n}$ . If  $v = \mathcal{B}X$ , then  $T(v) = \mathcal{C}(AX)$ . If we let  $v = \mathcal{B}'X' = v'_1x'_1 + \dots + v'_nx'_n$ , then

$$T(v) = \mathcal{C}'Y' = (\mathcal{C}Q)' = \mathcal{C}(QY') = \mathcal{C}(APX') \implies QY' = APX' \implies Y' = (Q^{-1}AP)X' \quad (2.3)$$

To prove the second part, we will show that the first part implies it. Let  $A_{m \times n}$  be a matrix. Let  $T_A$  be the linear map from  $\mathbb{R}^n \rightarrow \mathbb{R}^m$  given by multiplication by  $A$ , that is  $T_A : \mathbb{R}^n \rightarrow \mathbb{R}^m$  given by  $X \mapsto AX$ . By the first part, there exist bases  $P_{n \times n}$  and  $Q_{m \times m}$ , both invertible, such that with respect  $P$  and  $Q$ , the matrix of  $T_A$  looks like  $\begin{pmatrix} I & O \\ O & O \end{pmatrix}$ , that is,  $Q^{-1}AP = \begin{pmatrix} I & O \\ O & O \end{pmatrix}$ . ■

### 2.1.1 Linear Operators

Let  $T : V_{\mathcal{B}} \rightarrow V_{\mathcal{B}}$ . Let  $A$  be the matrix of  $T$  with respect to the basis  $\mathcal{B}$ . The other matrices of  $T$  with respect to new bases are  $P^{-1}AP$ , where  $P_{n \times n}$  is invertible. Also, the fact that  $T$  is bijective, one-one, or onto are all equivalent for a finite dimensional vector space  $V$ .

### 2.1.2 Eigenvectors and Eigenvalues

**Definition 2.3.** A non-zero vector  $v \in V$  is said to be an *eigenvector* of  $T$  if  $T(v) = \lambda v$  for some  $\lambda \in \mathbb{F}$ . If  $A$  is a  $n \times n$  matrix, a non-zero column vector  $X$  is said to be an eigenvector of  $A$  if  $AX = \lambda X$  for some  $\lambda \in \mathbb{F}$ .  $\lambda$ , in both these cases, is called the *eigenvalue* of  $v$  and  $X$  respectively.

Usually, we always disregard the zero vector being an eigenvector. If  $v$  is an eigenvector of  $T : V \rightarrow V$ , and  $v = \mathcal{B}X$  with respect to some basis  $\mathcal{B}$  of  $V$ , then  $X$  is an eigenvector of the matrix of  $T$  with respect to the basis  $\mathcal{B}$ . In fact,

$$\mathcal{B}(AX) = (\mathcal{B}A)X = T(\mathcal{B})X = T(\mathcal{B}X) = Tv = \lambda v = \lambda \mathcal{B}X = \mathcal{B}(\lambda X) \implies AX = \lambda X. \quad (2.4)$$

The converse is also true; if  $X$  is an eigenvector of  $A_{n \times n}$ , then  $X$  is also an eigenvector of  $T_A : \mathbb{R}^n \rightarrow \mathbb{R}^n$ .

**Proposition 2.4.** 0 is an eigenvalue of  $A_{n \times n}$  ( $T : V \rightarrow V$ ) if and only if  $A$  ( $T$ ) is non-invertible (not an isomorphism).

Suppose  $v$  is an eigenvector of  $T : V \rightarrow V$  with eigenvalue  $\lambda$ . Let  $W$  be the subspace spanned by  $v$ . Then every vector  $w \in W$  is an eigenvector of  $T$  with eigenvalue  $\lambda$ . The proof of this is left as an exercise.

**Definition 2.5.** Two matrices  $A'_{n \times n}$  and  $A_{n \times n}$  are called *similar matrices* if there exists an invertible matrix  $P_{n \times n}$  such that  $P^{-1}AP = A'$ .

Again let  $T : V \rightarrow V$  be a linear operator, and let  $\mathcal{B} = (v_1, \dots, v_n)$ . Suppose, with respect to the basis  $\mathcal{B}$ , the matrix of  $T$  is  $\begin{pmatrix} \lambda_1 & \cdots & \cdots & \cdots \\ 0 & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & \cdots & \cdots \end{pmatrix}$ . Then  $v_1$  is an eigenvector with eigenvalue  $\lambda_1$ .

## 2.2 Finding Eigenvalues and Eigenvectors

*January 21st.*

Let  $T : V \rightarrow V$  and let  $\mathcal{B} = (v_1, \dots, v_n)$  be a basis of  $V$ . Then the matrix of  $T$  with respect to the basis  $\mathcal{B}$  is a diagonal matrix if and only if each of the basis elements is an eigenvector. An equivalent statement for matrices is that an  $n \times n$  matrix  $A$  is similar to a diagonal matrix if and only if  $\mathbb{F}^n$  admits a basis consisting of eigenvectors of  $A$ . The proof of this is left as an exercise to the reader.

We can now discuss the computation. For a linear operator  $T : V \rightarrow V$ ,  $\lambda$  is an eigenvalue of  $T$  if and only if there exists a non-zero vector  $v$  such that  $Tv = \lambda v$ . This can be rearranged to give

$$(\lambda I_v - T)v = 0. \quad (2.5)$$

We can now consider  $\lambda I_v - T : V \rightarrow V$  to be a linear operator which maps  $v \mapsto \lambda v - Tv$ . If eigenvalues exist, this operator is a singular operator, that is, it contains a non-trivial kernel. The matrix of the operator  $\lambda I_v - T$  comes out to be  $\lambda I_n - A$ , where  $A$  is the matrix of  $T$  with respect to the basis  $\mathcal{B}$ . This matrix is now singular, so we must have

$$\det(\lambda I_n - A) = 0. \quad (2.6)$$

The equation  $\det(\lambda I_n - A)$  is called the *characteristic polynomial* of  $A$ , and also  $T(?)$ . The roots of this polynomial in  $\lambda$  which lie in  $\mathbb{F}$  are the eigenvalues of  $A$ , and  $T$  as well.

We would now like to show that similar matrices have the same eigenvalues, that is,

$$\det(\lambda I_n - P^{-1}AP) = \det(\lambda I_n - A). \quad (2.7)$$

This is simple to see as  $\det(\lambda I_n - P^{-1}AP) = \det(P^{-1}(\lambda I_n - A)P) = \det P^{-1} \cdot \det(\lambda I_n - A) \cdot \det P = \det(\lambda I_n - A)$ . The found out eigenvalues from this equation can then be put back and solved for  $v$  to get the corresponding eigenvectors.

**Proposition 2.6.** *Let  $\lambda_1, \dots, \lambda_r$  be distinct eigenvalues of  $T : V \rightarrow V$  and let  $v_1, \dots, v_r$  be the corresponding eigenvectors of  $T$ . Then  $(v_1, \dots, v_r)$  is a linearly independent set in  $V$ .*

*Proof.* We claim that this is true for  $r = 1, 2$ . Using a form of induction, we will assume the result for  $r - 1$ . Begin with

$$\begin{aligned} \alpha_1 v_1 + \dots + \alpha_r v_r &= 0 \\ \implies \alpha_1 T v_1 + \dots + \alpha_r T v_r &= 0 \\ \implies \alpha_1 \lambda_1 v_1 + \dots + \alpha_r \lambda_r v_r &= 0. \end{aligned} \quad (2.8)$$

Multiplying the first equation by  $\lambda_1$  and subtracting it from the current equation, we have

$$\begin{aligned} (\alpha_2 \lambda_2 - \alpha_2 \lambda_1) v_2 + (\alpha_3 \lambda_3 - \alpha_3 \lambda_1) v_3 + \dots + (\alpha_r \lambda_r - \alpha_r \lambda_1) v_r &= 0 \\ \implies \alpha_2 (\lambda_2 - \lambda_1) + \alpha_3 (\lambda_3 - \lambda_1) v_3 + \dots + \alpha_r (\lambda_r - \lambda_1) v_r &= 0. \end{aligned} \quad (2.9)$$

By hypothesis,  $\alpha_j (\lambda_j - \lambda_1) = 0$ . As the eigenvalues are distinct, we must have  $\alpha_j = 0$  for  $j = 2, 3, \dots, r$ . We are left with  $\alpha_1 v_1 = 0$ , which gives us  $\alpha_1 = 0$ . ■

When the  $n$  eigenvalues found of  $A$  are distinct, the corresponding eigenvectors  $v_1, \dots, v_n$  are linearly independent in  $\mathbb{F}^n$ , and hence  $\mathcal{B} = (v_1, \dots, v_n)$  is a basis of  $\mathbb{F}^n$ . The matrix  $P^{-1}AP$  is the matrix of the linear operator  $T_A : \mathbb{F}^n \rightarrow \mathbb{F}^n$  with respect to the basis  $\mathcal{B}$ , with the column of  $P$  being the eigenvectors  $v_1, \dots, v_n$ . As  $\mathcal{B}$  consists of only eigenvectors,  $P^{-1}AP$  is a diagonal matrix with the diagonal entries being the  $n$  eigenvalues.

We now define the determinant and trace for a linear operator. For such an operator  $T$ ,  $\text{tr} T = \text{tr} A$  where  $A$  is a matrix of  $T$  with respect to some arbitrary basis. Note that since  $\text{tr}(P^{-1}AP) = \text{tr}(APP^{-1}) = \text{tr} A$ , the choice of basis is not important. Similarly, we define  $\det T = \det A$ .

We can now have a closer look at the characteristic equation. To find the constant term of  $\det(xI - A)$ , we simply plug in  $x = 0$  to give us  $\det(-A) = (-1)^n \det A$ . The coefficient of  $x^{n-1}$  in  $\det(xI - A)$  is  $-\text{tr} A$  as the coefficients of  $x^{n-1}$  come solely from the expansion of  $(x - a_{11})(x - a_{22}) \cdots (x - a_{nn})$ . Clearly, we can conclude that the sum of the eigenvalues is  $\text{tr} A$  and the product of the eigenvalues is  $\det A$ .

### 2.2.1 Eigenspace

January 23rd.

For ease, let us denote  $\chi_T(x)$  to mean  $\det(xI - A)$ . The *eigenspace* for a given eigenvalue  $\lambda$  is defined as

$$E_\lambda = \{v \in V : Tv = \lambda v\}. \quad (2.10)$$

This is a subspace of the vector space  $V$ . The *geometric multiplicity* of  $\lambda$  is defined as the dimension of  $E_\lambda$ . This geometric multiplicity of  $\lambda$  is always less than or equal to its algebraic multiplicity in  $\chi_T(x)$ . For recall, the *algebraic multiplicity* of  $\lambda$  is the highest power of  $(x - \lambda)$  that divides  $\chi_T(x)$ .

**Theorem 2.7.** *Let  $\lambda$  be an eigenvalue of  $T : V \rightarrow V$ . Then the geometric multiplicity of  $\lambda$  is always less than or equal to its algebraic multiplicity.*

*Proof.* Let  $k$  be the geometric multiplicity of  $\lambda$ . Let  $(v_1, \dots, v_k)$  be an ordered basis of  $E_\lambda$ . Extend this to a basis  $\mathcal{B} = (v_1, \dots, v_k, u_1, \dots, u_{n-k})$  of  $V$ . The matrix of  $T$  with respect to the basis  $\mathcal{B}$  is of the form  $A = \begin{pmatrix} \lambda I_k & B \\ O & D \end{pmatrix}$ . Thus, the characteristic polynomial looks like

$$\chi_T(x) = \det(xI_n - A) = \det \begin{pmatrix} (x - \lambda)I_k & -B \\ O & xI_{n-k} - D \end{pmatrix} = (x - \lambda)^k \cdot \det(xI_{n-k} - D). \quad (2.11)$$

This shows that  $(x - \lambda)^k$  divides  $\chi_T(x)$ , so we must have an algebraic multiplicity greater than or equal to this  $k$ . ■

## 2.3 Diagonalizability

We first define what this means for a linear mapping from  $V$  to  $V$ .

**Definition 2.8.** A linear operator  $T : V \rightarrow V$  is said to be a *diagonalizable linear operator* if there exists a basis of  $V$  consisting of eigenvectors of  $T$ . This means that the matrix of  $T$  with respect to this basis is a diagonal matrix and the matrix of  $T$  with respect to any other basis is similar to this diagonal matrix.

A similar definition works for matrices.

**Definition 2.9.** An  $n \times n$  matrix  $A$  over  $\mathbb{F}$  is said to be a *diagonalizable matrix* if  $A$  is similar to a diagonal matrix. Equivalently,  $\mathbb{F}^n$  then admits a basis consisting of eigenvectors of  $A$ , thinking of  $T_A : \mathbb{F}^n \rightarrow \mathbb{F}^n$  as a linear operator.

Now let us suppose that  $T$  is diagonalizable. Let  $\lambda_1, \lambda_2, \dots, \lambda_k$  be the distinct eigenvalues of  $T$ . There then exists an ordered basis consisting of eigenvectors of  $T$  and with respect to this basis, the matrix of  $T$  is a diagonal matrix with diagonal entries consisting solely of  $\lambda_1, \lambda_2, \dots, \lambda_k$ .

If  $\lambda_i$  is of algebraic multiplicity  $d_i$ , then the matrix of  $T$  looks like  $\begin{pmatrix} \lambda_1 I_{d_1} & & & \\ & \lambda_2 I_{d_2} & & \\ & & \dots & \\ & & & \lambda_k I_{d_k} \end{pmatrix}$ .

Thus, the characteristic polynomial then looks like  $(x - \lambda_1)^{d_1} (x - \lambda_2)^{d_2} \dots (x - \lambda_k)^{d_k}$ .

The geometric multiplicity of  $\lambda_i$  is the dimension of  $E_{\lambda_i}$ , that is, the nullity of the operator  $(\lambda_i I_n - A)$ . But here,  $\ker(\lambda_i I_n - A) = d_i$ , which is just the algebraic multiplicity of  $\lambda_i$ . Hence, if  $T$  is diagonalizable, then each eigenvalue of it has the same algebraic multiplicity and geometric multiplicity.

**Proposition 2.10.** *If  $E_{\lambda_1}, \dots, E_{\lambda_k}$  are the eigenspaces corresponding to the distinct eigenvalues, say,  $\lambda_1, \dots, \lambda_k$  of  $T$ , then  $E = E_{\lambda_1} + \dots + E_{\lambda_k}$  is a direct sum.*

*Proof.* It is enough to show that  $E_{\lambda_1}, \dots, E_{\lambda_k}$  are independent. Let  $v_1 + v_2 + \dots + v_k = 0$ , where  $v_i \in E_{\lambda_i}$ . As  $v_1, v_2, \dots, v_k$  come from distinct eigenspaces, they are linearly independent, and our equation must imply that  $v_1 = \dots = v_k = 0$ . ■

**Proposition 2.11.** *If  $T$  is a diagonalizable operator, and if  $\lambda_1, \dots, \lambda_k$  are the distinct eigenvalues of  $T$ , then*

$$V = E_{\lambda_1} \oplus \dots \oplus E_{\lambda_k}. \quad (2.12)$$

*Proof.* As  $T$  is diagonalizable, the algebraic and geometric multiplicities are equal for all the eigenvalues  $\lambda_i$ . Denote  $\dim E_{\lambda_i} = d_i$ . As  $\chi_T(x)$  completely factors into linear factors, due to  $T$  being diagonalizable, we have  $n = d_1 + \dots + d_k$ . Also,  $E_{\lambda_1} + \dots + E_{\lambda_k}$  is a direct sum, that is,

$$\dim(E_{\lambda_1} + \dots + E_{\lambda_k}) = \dim E_{\lambda_1} + \dots + \dim E_{\lambda_k} = n. \quad (2.13)$$

This direct sum is a subspace of  $V$  and has the dimension as  $V$ . This must mean that the direct sum is exactly  $V$ . ■

**Theorem 2.12.** *Let  $T$  be a linear operator on a finite dimensional vector space  $V$ , and let  $\lambda_1, \dots, \lambda_k$  be the distinct eigenvalues of  $T$ . Also let  $E_{\lambda_i}$  be the eigenspace of  $\lambda_i$ . Then, the following are equivalent.*

- $T$  is diagonalizable,
- $\chi_T(x) = (x - \lambda_1)^{d_1} \dots (x - \lambda_k)^{d_k}$  and  $\dim E_{\lambda_i} = d_i$ ,
- $V = E_{\lambda_1} \oplus \dots \oplus E_{\lambda_k}$ .

## 2.4 Polynomials

*January 28th.*

Let  $\mathbb{F}[x]$  denote the set of all polynomials with coefficients coming from the field  $\mathbb{F}$ . With respect to the addition, it is an Abelian group. The multiplication here is associative, commutative, and distributive; there also exists a multiplicative identity. This makes  $\mathbb{F}[x]$  into a commutative ring. Note that  $\mathbb{F}[x]$  is also an infinite dimensional vector space over  $\mathbb{F}$ , since scalar multiplication is also defined. Together, these combine to form an algebra over the field.

**Definition 2.13.** Let  $d \in \mathbb{F}[x]$  with  $d \neq 0$ . For  $f \in \mathbb{F}[x]$ , we say that  $d$  divides  $f$  if there exists a  $q \in \mathbb{F}[x]$  such that  $f = dq$  in  $\mathbb{F}[x]$ .

**Corollary 2.14.** *For  $f \in \mathbb{F}[x]$ ,  $f(c) = 0$  if and only if  $x - c$  divides  $f(x)$ .*

**Corollary 2.15.** *A polynomial  $f \in \mathbb{F}[x]$  of degree  $n$  has at most  $n$  roots in  $\mathbb{F}$ .*

*Proof.* The proof is by induction. Note that this is true for  $n = 0, 1$ . If  $\alpha$  is a root, then  $f(x) = (x - \alpha) \cdot q(x)$ . As  $q(x)$  is of degree  $n - 1$ , and all roots of  $q(x)$  are root of  $f(x)$ , this follows by hypothesis. ■

**Definition 2.16.** An *ideal* of  $\mathbb{F}[x]$  is a subspace of  $\mathbb{F}[x]$ , say  $I$ , such that if  $f \in I$  and  $g \in \mathbb{F}[x]$ , then  $fg \in I$ .

**Example 2.17.** Let  $f \in \mathbb{F}[x]$ . Define  $I_f = \langle f \rangle = \{fg : g \in \mathbb{F}[x]\}$ . Note that  $I_f$  is called a *principal ideal*, that is, it is an ideal generated by a single element.

**Theorem 2.18.**  *$\mathbb{F}[x]$  is a principal ideal domain, that is, every ideal in  $\mathbb{F}[x]$  is a principal ideal.*

*Proof.* Let  $d$  be a polynomial of least degree in the ideal  $I$ , where  $I$  is a non-zero ideal. Let, without loss of generality,  $d$  be monic (if not, simply multiply it by a suitable scalar).

Let  $f \in I$ . Then there exists  $q, r \in \mathbb{F}[x]$  such that  $f = dq + r$  and either  $r = 0$  or  $\deg r < \deg d$ . Note that since  $f, d \in I$ ,  $dq \in I$ , so  $f - dq \in I \implies r \in I$ . As  $d$  was of minimal degree in  $I$ , we must have  $r = 0$ . Thus,  $f = dq$  and, thus,  $I = \langle d \rangle$ . ■

If  $I$  is an ideal of  $\mathbb{F}[x]$ , then there exists a unique polynomial  $d \in I$  such that  $I = \langle d \rangle$ .

### 2.4.1 Interaction with Linear Operators

Let  $f \in \mathbb{F}[x]$ , and let  $T : V \rightarrow V$  be a linear mapping. If

$$f(x) = a_0 + a_1x + \dots + a_kx^k$$

with  $a_k \neq 0$ , we define

$$f(T) = a_0I_n + a_1T + \dots + a_kT^k.$$

Note that  $f(T)$  is also a linear mapping from  $V$  to  $V$ . Let  $I$  be the set of all  $f \in \mathbb{F}[x]$  such that  $f(T)$  is the zero operator. All such polynomials are called *annihilators*.  $I$  satisfies the properties of a vector space; it is a subspace of the space of all polynomials.  $I$  is also an ideal of  $\mathbb{F}[x]$ .

**Definition 2.19.** The *minimal polynomial* of the linear operator  $T : V \rightarrow V$  is the generator of the ideal of annihilators.

Denote the minimal polynomial by  $m_T(x)$ . So,  $m_T(x)$  is

1. monic,
2. of least degree among all annihilators of  $T$ .

If  $A$  is a  $n \times n$  matrix, the minimal polynomial of  $A$  is defined as the unique monic polynomial  $m_A(x)$  of least degree such that  $m_A(A) = O_{n \times n}$ . It can be verified that if  $A$  is the matrix of a linear operator  $T : V \rightarrow V$  and if  $f \in \mathbb{F}[x]$ , then the matrix of the operator  $f(T) : V \rightarrow V$  is  $f(A)$  with respect to the same basis. It follows that the minimal polynomial of  $T$  is same as the minimal polynomial of a matrix of  $T$ .

Note that  $T$  belongs to  $\text{Hom}_{\mathbb{F}}(V, V)$ , which is of dimension  $n^2$ . Thus,  $I, T, T^2, \dots, T^{n^2}$  is a linearly dependent set and there exist scalars  $a_0, a_2, \dots, a_{n^2}$  such that

$$a_0 I + a_1 T + a_2 T^2 + \dots + a_{n^2} T^{n^2} = O. \quad (2.14)$$

So, an annihilator of  $T$  is

$$f(x) = a_0 + a_1 x + a_2 x^2 + \dots + a_{n^2} x^{n^2}$$

and we must have  $\deg m_T(x) \leq n^2$ .

**Theorem 2.20.** Let  $T : V \rightarrow V$  with  $n$  the dimension of the space  $V$ . The characteristic polynomial of  $T$  and the minimal polynomial of  $T$  have the same roots, except (possibly) for the multiplicities.

*Proof.* We claim that  $m_T(c) = 0$  if and only if  $c$  is an eigenvalue. Let  $m_T(c) = 0$ . Thus,  $m_T(c) = (x - c) \cdot q(x)$ , with  $q \in \mathbb{F}[x]$  and  $\deg q < \deg m$ . Also,  $q(T)$  is *not* the zero operator. So, there exists a  $u \in V$  (non-zero vector) such that  $q(T)(u) = v \neq 0$ . Then,

$$0 = m(T)(u) = (T - cI) \cdot q(T)(u) = (T - cI)v \quad (2.15)$$

which shows that  $v$  is an eigenvector of  $T$  with eigenvalue  $c$ . So all roots of  $m_T(x)$  are roots of the characteristic polynomial.

Conversely, let  $c$  be an eigenvalue of  $T$ . Say,  $Tv = cv$  for some  $v \neq 0$ . Thus,  $m_T(T)(v) = m(c)(v)$ . But  $m_T(T) = 0$  must mean that  $0 = m(c)(v)$ , and  $m(c) = 0$ . So every root of the characteristic polynomial is a root of the minimal polynomial. ■

January 30th.

**Proposition 2.21.** If  $\lambda$  is an eigenvalue of  $T$ , then  $f(\lambda)$  is an eigenvalue of  $f(T)$  for  $f \in \mathbb{F}[x]$ .

*Proof.* The proof is left as an exercise to the reader. ■

**Proposition 2.22.** Let  $T : V \rightarrow V$  be a diagonalizable operator. The minimal polynomial is the product of distinct linear factors, that is, if

$$\chi_T(x) = \prod_{i=1}^k (x - \lambda_i)^{d_i}$$

where the  $\lambda_i$ 's are the distinct eigenvalues, then

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

*Proof.* As  $T$  is a diagonalizable operator, there exists a basis of  $V$  consisting of eigenvectors of  $T$ , say  $\mathcal{B} = (v_1, v_2, \dots, v_n)$ . Note that  $m_T(T)v_i = 0$  for all valid  $i$ . For each  $v_i \in \mathcal{B}$ , there exists a  $\lambda_i$  such that  $(T - \lambda_i I)v_i = 0$ , which tells us

$$m_T(T) = (T - \lambda_1 I)(T - \lambda_2 I) \cdots (T - \lambda_k I)v_i = 0. \quad (2.16)$$

Hence,  $m_T(x)$  is an annihilator for  $T$ , and it is of minimal degree by the above theorem. ■

**Theorem 2.23** (*Cayley-Hamilton theorem*). *Let  $T : V \rightarrow V$  be a linear operator on a finite dimensional vector space  $V$ . If  $\chi_T(x)$  is the characteristic polynomial of  $T$ , then  $\chi_T(T) = 0$ , that is, the characteristic polynomial annihilates  $T$ . Hence, the minimal polynomial of  $T$  divides the characteristic polynomial.*

*Proof.* Let  $\mathcal{B} = (v_1, v_2, \dots, v_n)$  be a basis of  $V$ , and let  $A = (a_{ij})$  be the matrix of  $T$  with respect to the basis  $\mathcal{B}$ . We have

$$\begin{aligned} a_{1j}v_1 + a_{2j}v_2 + \dots + a_{nj}v_n &= Tv_j \\ \implies -a_{1j}v_1 - a_{2j}v_2 - \dots + (T - a_{jj})v_j - a_{(j+1)(j)}v_{j+1} - \dots - a_{nj}v_n &= 0. \end{aligned} \quad (2.17)$$

This system of equations can be written as

$$B_{n \times n} \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} \quad (2.18)$$

where

$$B = \begin{pmatrix} T - a_{11}I & -a_{21} & \cdots & -a_{n1} \\ -a_{12} & T - a_{22}I & \cdots & -a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ -a_{1n} & -a_{2n} & \cdots & T - a_{nn}I \end{pmatrix}. \quad (2.19)$$

Therefore  $\det B = \chi_T(T)$ . It is enough to show that  $\det B = 0$  as an operator, that is, to show  $\det B(b_i) = 0$  for all  $v_i \in \mathcal{B}$ . Let  $(\text{adj} B)_{ij} = c_{ij}$ , and  $(B)_{ij} = b_{ij}$ . Note that

$$\sum_{k=1}^n c_{ik} b_{kj} = \begin{cases} \det B & \text{if } i = j, \\ 0 & \text{if otherwise.} \end{cases}$$

Now,

$$\begin{aligned} \sum_{j=1}^n b_{kj} v_j &= 0 \text{ for all } 1 \leq k \leq n \\ \implies \sum_{j=1}^n b_{kj} v_j &= 0. \end{aligned}$$

Summing over all rows,

$$\begin{aligned} \sum_{k=1}^n \left( \sum_{j=1}^n c_{ik} b_{kj} v_j \right) &= 0 \\ \implies \sum_{j=1}^n \left( \sum_{k=1}^n c_{ik} b_{kj} \right) v_j &= 0. \end{aligned} \quad (2.20)$$

The left hand side is zero except for when  $i = j$ , in which case it is  $\det B$ —

$$0 = \sum_{j=1}^n \left( \sum_{k=1}^n c_{ik} b_{kj} \right) v_j = (\det B) v_i \quad (2.21)$$

which implies that the operator  $\det B$  is zero on all the basis vectors, and hence it is the zero vector. Thus, since  $\chi_T(T) = \det B$ ,  $\chi_T(T)$  is also the zero operator. ■

*February 4th.*

**Proposition 2.24.** *If the minimal polynomial  $m_T(x) \in \mathbb{F}[x]$  of a linear operator  $T : V \rightarrow V$  splits into distinct linear factors, then  $T$  is diagonalizable.*

*Proof.* Let  $m_T(x) = (x - \lambda_1)(x - \lambda_2) \cdots (x - \lambda_k)$  where the  $\lambda_i$ 's are distinct. We are to show that  $V = E_{\lambda_1} \oplus \dots \oplus E_{\lambda_k}$ . We wish to find polynomials  $h_1(x), \dots, h_k(x)$  such that



1.  $h_1(x) + \dots + h_k(x) = 1$ ,
2.  $(x - \lambda_i) \cdot h_i(x)$  is divisible by  $m_T(x)$  for all  $1 \leq i \leq k$ .

The second condition implies that  $(T - \lambda_i I) \cdot h_i(T)$  is the zero operator. The first condition implies that  $\sum_{i=1}^k h_i(T)(v) = v$ . But the second condition again implies that  $h_i(T)(v)$  is an eigenvector corresponding to  $\lambda_i$ , that is,  $h_i(T)(v) \in E_{\lambda_i}$ . If we can find these  $h_i$ 's satisfying the two conditions then we can say that  $V$  is the direct sum of the eigenspaces.

For  $1 \leq i \leq k$ , let  $f_i(x) = \frac{m_T(x)}{(x - \lambda_i)} = \prod_{j \neq i} (x - \lambda_j)$ . As the  $\lambda_i$ 's are distinct, the  $f_i$ 's are relatively prime, so there exist  $g_1, \dots, g_k$  such that

$$f_1(x)g_1(x) + \dots + f_k(x)g_k(x) = 1. \quad (2.22)$$

Let  $h_i(x) = f_i(x)g_i(x)$  for all  $1 \leq i \leq k$ . Both the conditions hold, and the result follows.  $\blacksquare$

**Corollary 2.25.** *Let  $T : V \rightarrow V$  be a linear operator on a finite dimensional complex vector space such that  $T^m = I$  for some positive integral  $m$ . Then  $T$  is diagonalizable.*

**Proposition 2.26.** *Let  $T : V \rightarrow V$  be linear operator, and let  $U$  be an invariant subspace of  $T$ , that is,  $T(U) \subseteq U$  (or equivalently,  $T(u) \in U$  for all  $u \in U$ ). The minimal polynomial  $m_{T|_U}(x)$  of the operator  $T|_U : U \rightarrow U$  divides the minimal polynomial  $m_T(x)$  of the operator  $T : V \rightarrow V$  in  $\mathbb{F}[x]$ .*

*Proof.* Note that  $m_T(T)(u) = 0$  for all  $u \in U$  as  $U \subseteq V$ . Thus,  $m_T(T) = 0$  on the subspace  $U$ . So  $m_T(x)$  annihilates  $T|_U$ . So, as  $m_{T|_U}(x)$  is the minimal polynomial of  $T|_U$ , it should divide all annihilators of  $T|_U$  and thus divides  $m_T(x)$  in  $\mathbb{F}[x]$ .  $\blacksquare$

## 2.5 Triangularizability

A similar definition works as in the case of diagonalizability.

**Definition 2.27.** A linear operator  $T : V \rightarrow V$  is said to be a *triangularizable linear operator* if there exists a basis of  $V$  with respect to which the matrix of  $T$  is a triangular matrix, be it upper or lower.

If our basis is  $\mathcal{B} = (v_1, v_2, \dots, v_n)$ , then we can show that  $Tv_k \in \text{span}(v_1, \dots, v_k)$ .

**Theorem 2.28.** *A linear operator  $T : V \rightarrow V$  is triangularizable if and only if the minimal polynomial splits into linear factors.*

*Proof.* Let  $T : V \rightarrow V$  be triangularizable, that is, there exists a basis with respect to which the matrix of  $T$  is a triangular matrix, with diagonal entries  $\lambda_1, \dots, \lambda_n$ , say. Then the characteristic polynomial of  $T$  is  $(x - \lambda_1)(x - \lambda_2) \cdots (x - \lambda_n)$  where the  $\lambda_i$ 's are not necessarily distinct. But  $m_T(x)$  divides  $\chi_T(x)$ , hence is again a product of linear factors.

Conversely, let  $m_T(x) = (x - \lambda_1)(x - \lambda_2) \cdots (x - \lambda_k)$  where the  $\lambda_i$ 's are not necessarily distinct. We prove by induction on the number of factors of  $m_T(x)$ . If  $k = 1$ , then  $m_T(x) = x - \lambda_1$ ; as  $m_T(T) = 0$ ,  $T = \lambda_1 I$ , matrix of  $T$  is the scalar matrix. Now let  $k > 1$ , and let the result hold for smaller positive integers. Let  $U = \text{Im}(T - \lambda_k I)$ . We find that  $U$  is a proper subspace of  $V$ . Note that  $U$  is an invariant subspace of  $T$ ; if we let  $u = (T - \lambda_k I)(v)$  for some  $v \in V$ , then

$$T(u) = T(T - \lambda_k I)(v) = (T - \lambda_k I)T(v) \in U. \quad (2.23)$$

The minimal polynomial  $m_{T|_U}(x)$  of  $T|_U$  divides  $m_T(x)$ , and hence  $m_{T|_U}(x) = (x - \alpha_1) \cdots (x - \alpha_l)$ , where  $l \leq k$ , and  $\alpha_1, \dots, \alpha_l \in \{\lambda_1, \dots, \lambda_k\}$ . By hypothesis,  $T|_U$  is triangularizable. So there exists a basis of  $U$ , say  $(u_1, \dots, u_m)$  with respect to which the matrix of  $T|_U$  is a triangular matrix. So  $T|_U(u_k) \in \text{span}(u_1, \dots, u_k)$ . Extend this to a basis  $\mathcal{B} = (u_1, \dots, u_m, v_{m+1}, \dots, v_n)$ . If we rewrite  $Tv_j = (T - \lambda_k I)v_j + \lambda_k Iv_j$ , we see that  $(T - \lambda_k I)v_j \in U = \text{span}(u_1, \dots, u_m)$ , and  $Tv_j \in \text{span}(u_1, \dots, u_m, v_j)$ ; the matrix of  $T$  with respect to  $\mathcal{B}$  is a triangular matrix.  $\blacksquare$

**Corollary 2.29.** *Every operator  $T : V \rightarrow V$ , where  $V$  is a complex finite dimensional vector space, is triangularizable.*

### 2.5.1 Determinant of Partitioned Matrices

**Proposition 2.30.** *Let  $\Gamma = \begin{pmatrix} A & O \\ O & I \end{pmatrix}$  or  $\Gamma = \begin{pmatrix} I & O \\ O & A \end{pmatrix}$ , where  $A$  is a square matrix. Then we necessarily have  $\det \Gamma = \det A$ .*

*Proof.* Let  $\Gamma$  be of order  $(n+1) \times (n+1)$  and  $A$  be of order  $n \times n$ . By definition,

$$\det \Gamma = \sum_{\sigma \in S_{n+1}} \epsilon(\sigma) g_{1\sigma(1)} \cdots g_{(n+1)\sigma(n+1)}. \quad (2.24)$$

Note that  $g_{(n+1)\sigma(n+1)}$  is 1 if  $\sigma(n+1) = n+1$ , and 0 otherwise. Also,  $\epsilon(\sigma)$  remains the same when  $\sigma$  is considered to be an element of  $S_n$ . Thus,

$$\det \Gamma = \sum_{\sigma \in S_n} \epsilon(\sigma) g_{1\sigma(1)} \cdots g_{n\sigma(n)} = \det A. \quad (2.25)$$

Iterating this, we get the desired result. A similar proof works for the other type of matrix stated. ■

**Proposition 2.31.** *Let  $\Gamma = \begin{pmatrix} A & B \\ O & D \end{pmatrix}$  where  $A$  and  $D$  are square matrices. Then we necessarily have  $\det \Gamma = \det A \cdot \det D$ .*

*Proof.* Let the orders be  $A_{k \times k}$ ,  $D_{l \times l}$ ,  $B_{k \times l}$  and  $O_{l \times k}$ . Note that  $\Gamma$  can be broken up as

$$\Gamma = \begin{pmatrix} I_k & O_{k \times l} \\ O_{l \times k} & D_{l \times l} \end{pmatrix} \begin{pmatrix} I_k & B_{k \times l} \\ O_{l \times k} & I_l \end{pmatrix} \begin{pmatrix} A_{k \times k} & O_{k \times l} \\ O_{l \times k} & I_l \end{pmatrix}. \quad (2.26)$$

The determinant is multiplicative, so  $\det \Gamma = \det D \cdot \det A$  as the determinant of the middle matrix can be shown to be 1. ■

**Proposition 2.32.** *Let  $\Gamma = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$  where  $A$  and  $D$  are square matrices. If  $A$  is invertible, then we necessarily have  $\det \Gamma = \det A \cdot \det(D - CA^{-1}B)$ .*

*Proof.* Again, we break down  $\Gamma$ .

$$\Gamma = \begin{pmatrix} I & O \\ CA^{-1} & I \end{pmatrix} \begin{pmatrix} A & O \\ O & D - CA^{-1}B \end{pmatrix} \begin{pmatrix} I & A^{-1}B \\ O & I \end{pmatrix}. \quad (2.27)$$

From here, it is clear that  $\det \Gamma = \det A \cdot \det(D - CA^{-1}B)$ . ■

**Proposition 2.33.** *Let  $\Gamma = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$  where  $A$  and  $D$  are square matrices. If  $D$  is invertible, then we necessarily have  $\det \Gamma = \det D \cdot \det(A - BD^{-1}C)$ .*

*Proof.* Yet again, we break down  $\Gamma$ .

$$\Gamma = \begin{pmatrix} I & BD^{-1} \\ O & I \end{pmatrix} \begin{pmatrix} A - BD^{-1}C & O \\ O & D \end{pmatrix} \begin{pmatrix} I & O \\ D^{-1}C & I \end{pmatrix}. \quad (2.28)$$

From here, it is clear that  $\det \Gamma = \det D \cdot \det(A - BD^{-1}C)$ . ■

## 2.6 On the Characteristic and Minimal Polynomials

February 6th.

**Theorem 2.34.** *Let  $A$  be a  $m \times n$  matrix and  $B$  be a  $n \times m$  matrix with  $m \leq n$ . Then,*

$$\chi_{BA}(x) = x^{n-m} \chi_{AB}(x). \quad (2.29)$$

*Proof.* Note that there exist non-singular matrices  $P$  and  $Q$  such that  $PAQ = \begin{pmatrix} I_r & O \\ O & O \end{pmatrix}$ , where  $r$  is the rank of  $A$ . Partition  $Q^{-1}BP^{-1}$  as  $\begin{pmatrix} C & D \\ E & G \end{pmatrix}$  where  $C$  is of order  $r \times r$ , with the other submatrices being of appropriate order. Then,

$$PABP^{-1} = PAQQ^{-1}BP^{-1} = \begin{pmatrix} I_r & O \\ O & O \end{pmatrix} \begin{pmatrix} C_{r \times r} & D \\ E & G \end{pmatrix} = \begin{pmatrix} C & D \\ O & O \end{pmatrix}. \quad (2.30)$$

Similarly,

$$Q^{-1}BAQ = Q^{-1}BP^{-1}PAQ = \begin{pmatrix} C_{r \times r} & D \\ E & G \end{pmatrix} \begin{pmatrix} I_r & O \\ O & O \end{pmatrix} = \begin{pmatrix} C & O \\ E & O \end{pmatrix}. \quad (2.31)$$

Thus,

$$\chi_{AB}(x) = \chi_{PABP^{-1}}(x) = \det \begin{pmatrix} xI_r - C & -D \\ O & xI \end{pmatrix} = x^{m-r} \det(xI - C) \quad (2.32)$$

and

$$\chi_{BA}(x) = \chi_{Q^{-1}BAQ}(x) = \det \begin{pmatrix} xI - C & O \\ -E & xI \end{pmatrix} = x^{n-r} \det(xI - C) \quad (2.33)$$

which tells us that  $\chi_{BA}(x) = x^{n-m} \chi_{AB}(x)$ . ■

1. Suppose  $T : V \rightarrow V$  and  $U \subseteq V$ .
  - (a) If  $U \subseteq \ker T$ , then  $U$  is  $T$ -invariant.
  - (b) If  $T(V) \subseteq U$ , then  $U$  is  $T$ -invariant.
2. If  $V_1, \dots, V_m$  are  $T$ -invariant subspaces, then  $V_1 + \dots + V_m$  is  $T$ -invariant.
3. Let  $P : V \rightarrow V$  be a linear operator such that  $P^2 = P$ ; then the eigenvalues of  $P$  are 0 or 1.
4. Let  $T : V \rightarrow V$  be an invertible operator. Then  $\lambda$  is an eigenvalue of  $T$  if and only if  $\frac{1}{\lambda}$  is an eigenvalue of  $T^{-1}$ .
5. Let  $T : V \rightarrow V$ , with  $0 \neq v \in V$ . Then  $W = \text{span}(v, Tv, T^2v, \dots)$  is  $T$ -invariant, and is the smallest  $T$ -invariant subspace of  $V$  containing  $v$ .
6. Let  $T : V \rightarrow V$  and  $\text{rank} T = k \leq n$ , where  $n$  is the dimension of  $V$ . Then  $T$  has at most  $k + 1$  distinct eigenvalues.
- 7.

## Chapter 3

# INNER PRODUCT SPACES

February 13th.

### 3.1 An Introduction

The function  $\langle, \rangle : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$  defined as

$$\langle X, Y \rangle = x_1 y_1 + x_2 y_2 + \dots + x_n y_n \quad (3.1)$$

is called an inner product. Specifically, this is the dot product on the vector space over the reals. It satisfies the following properties;

1.  $\langle X, X \rangle \geq 0$  for all  $X \in \mathbb{R}^n$ .
2.  $\langle X, X \rangle = 0$  if and only if  $X = 0$ .
3.  $\langle X, Y \rangle = \langle Y, X \rangle$ .
4.  $\langle X_1 + X_2, Y \rangle = \langle X_1, Y \rangle + \langle X_2, Y \rangle$ .
5.  $\langle \alpha X, Y \rangle = \alpha \langle X, Y \rangle$ .

In  $\mathbb{C}^n$ , we have the product

$$\langle Z, W \rangle = z_1 \overline{w_1} + z_2 \overline{w_2} + \dots + z_n \overline{w_n}. \quad (3.2)$$

This satisfies the properties—

1.  $\langle Z, Z \rangle \geq 0$ .
2.  $\langle Z, Z \rangle = 0$  if and only if  $Z = 0$ .
3.  $\langle Z, W \rangle = \overline{\langle W, Z \rangle}$ .
4.  $\langle Z_1 + Z_2, W \rangle = \langle Z_1, W \rangle + \langle Z_2, W \rangle$ .
5.  $\langle \alpha Z, W \rangle = \alpha \langle Z, W \rangle$ .

These properties are, respectively, called the positivity, the definiteness, the conjugate symmetry, the additivity, and the homogeneity of the inner product over the complex vector space. We now define a general inner product.

Let the underlying field be either  $\mathbb{R}$  or  $\mathbb{C}$ , and let  $V$  be a vector space over this field. An *inner product* on  $V$  is simply a function  $\langle, \rangle : V \times V \rightarrow \mathbb{F}$  such that it satisfies the following properties for all  $v, u, v_1, v_2 \in V$  and  $\alpha \in \mathbb{F}$ —

1.  $\langle v, v \rangle \geq 0$ ,
2.  $\langle v, v \rangle = 0$  if and only if  $v = 0$ ,

3.  $\langle v_1 + v_2, u \rangle = \langle v_1, u \rangle + \langle v_2, u \rangle$ ,
4.  $\langle \alpha v, u \rangle = \alpha \langle v, u \rangle$ , and
5.  $\langle v, u \rangle = \overline{\langle u, v \rangle}$ .

A vector space over  $\mathbb{F}$ ,  $\mathbb{F}$  being either  $\mathbb{R}$  or  $\mathbb{C}$ , is called an *inner product space* if  $V$  is equipped with a valid inner product. As seen earlier, on  $\mathbb{R}^n$ , the usual dot product makes  $\mathbb{R}^n$  an inner product space. As another example, if  $V$  is the space of all real valued continuous function  $f : (-1, 1) \rightarrow \mathbb{R}$ , then the inner product on here can be defined as

$$\langle f, g \rangle = \int_{-1}^1 f(x)g(x)dx. \quad (3.3)$$

On  $\mathbb{C}^{m \times n}$ , we can define the inner product as

$$\langle A, B \rangle = \text{tr}(B^* A). \quad (3.4)$$

Every inner product  $\langle u, v \rangle$  for any vector space  $V$  will look like

$$\langle u, v \rangle = Y^* A X, \quad (3.5)$$

where  $Y$  and  $X$  are the coordinate vectors of  $v$  and  $u$  with respect to some basis  $\mathcal{B}$ . This will be proved later.

For an inner product space  $V$ , the following properties may be derived from the basic properties;

1.  $\langle 0, v \rangle = 0$  for all  $v \in V$ .
2. Fix  $v \in V$ . Define  $f_v : V \rightarrow \mathbb{F}$  as  $u \mapsto \langle u, v \rangle$ . Then  $f_v$  is a linear mapping from the space  $V$  to the space  $\mathbb{F}$  for any  $v \in V$ .
3. Let  $v = \alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_k v_k$  and  $u = \beta_1 u_1 + \beta_2 u_2 + \dots + \beta_l u_l$  where  $u, v, u_j, v_i \in V$  and  $\alpha_i, \beta_j \in \mathbb{F}$ . Then,

$$\langle v, u \rangle = \sum_{i=1}^k \sum_{j=1}^l \alpha_i \beta_j \langle v_i, u_j \rangle. \quad (3.6)$$

## 3.2 The Notion of Length and Orthogonality

Let  $(V, \langle, \rangle)$  be an inner product space. We define the *norm* of a vector  $v \in V$ , denoted by  $\|v\|$ , as

$$\|v\| = \sqrt{\langle v, v \rangle}. \quad (3.7)$$

For  $V$  being either  $\mathbb{R}^n$  or  $\mathbb{C}^n$ , we can easily verify that the norm becomes the usual Euclidean length of a vector. Note that  $\|v\| = 0$  if and only if  $v$  is the zero vector in  $V$ . It can also be shown that  $\|\lambda v\| = |\lambda| \|v\|$  for some  $\lambda \in \mathbb{F}$ .

**Definition 3.1.** We say that two vectors  $v, w \in V$  are *orthogonal vectors* if  $\langle v, w \rangle = 0$ .

Note that the zero vector is orthogonal to every vector in the vector space, even itself; in fact, it is the only vector orthogonal to itself. We can also make sense of a Pythagorean theorem here. If  $u, v \in V$  are orthogonal, then we can show that

$$\|u + v\|^2 = \|u\|^2 + \|v\|^2. \quad (3.8)$$

Given two vectors  $u, v \in V$ , we can write  $u$  as the sum of a scalar multiple of  $v$ , say  $cv$ , and a vector  $w$  such that  $\langle w, v \rangle = 0$ . If we rewrite  $u$  as  $u = cv + (u - cv)$ , and impose that  $\langle u - cv, v \rangle = 0$ , then we get  $c = \frac{\langle u, v \rangle}{\|v\|^2}$  fulfilling our conditions.

We also have a *Cauchy-Schwarz* inequality. It says that for any  $u, v \in V$ , then

$$|\langle u, v \rangle| \leq \|u\| \|v\| \quad (3.9)$$

and equality holds if and only if one of the vectors is a scalar multiple of the other.

*Proof.* If either one of the vectors is the zero vector, both sides are just zero. Hence, assume that neither vector is zero, and note that we can write

$$u = \frac{\langle u, v \rangle}{\|v\|^2} v + w \quad (3.10)$$

where  $\langle w, v \rangle = 0$ . Thus,

$$\|u\|^2 = \left\langle \frac{\langle u, v \rangle}{\|v\|^2} v + w, \frac{\langle u, v \rangle}{\|v\|^2} v + w \right\rangle = \frac{\langle u, v \rangle \overline{\langle u, v \rangle}}{\|v\|^4} \langle v, v \rangle + \langle w, w \rangle = \frac{|\langle u, v \rangle|^2}{\|v\|^2} + \langle w, w \rangle \geq \frac{|\langle u, v \rangle|^2}{\|v\|^2}. \quad (3.11)$$

The inequality follows. The equality is left as an exercise to the reader. ■

Let  $V$  be an inner product space, and let  $\mathcal{B} = (v_1, \dots, v_n)$  be a basis of  $V$ . Let  $X = (x_1, \dots, x_n)^t$  and  $Y = (y_1, \dots, y_n)^t$  be the coordinate vectors of  $v, w \in V$ , respectively, with respect to the basis  $\mathcal{B}$ . Then,

$$\langle v, w \rangle = Y^* A X \text{ where } A = (a_{ij}) \text{ and } a_{ij} = \langle v_j, v_i \rangle. \quad (3.12)$$

This can be seen since

$$\langle v, w \rangle = \left\langle \sum_{i=1}^n x_i v_i, \sum_{j=1}^n y_j v_j \right\rangle = \sum_{i=1}^n \sum_{j=1}^n x_i \overline{y_j} \langle v_i, v_j \rangle. \quad (3.13)$$

Conversely, let  $V$  be a vector space of dimension  $n$  and  $\mathcal{B}$  be a basis of  $V$ . Then defining  $\langle v, w \rangle = Y^t A X$ , where  $A$  is of order  $n \times n$  satisfying  $A^* = A$ , gives an inner product

*March 4th.*

**Theorem 3.2** (The *triangle inequality*). For all  $v, w \in V$ ,  $\|v + w\| \leq \|v\| + \|w\|$  holds.

*Proof.* We square the left side to get

$$\begin{aligned} \|v + w\|^2 &= \langle v + w, v + w \rangle = \langle v, v \rangle + \langle v, w \rangle + \langle w, v \rangle + \langle w, w \rangle \\ &= \|v\|^2 + 2\operatorname{Re}\langle v, w \rangle + \|w\|^2 \leq \|v\|^2 + 2|\langle v, w \rangle| + \|w\|^2 \\ &\leq \|v\|^2 + 2\|v\|\|w\| + \|w\|^2 = (\|v\| + \|w\|)^2. \end{aligned} \quad (3.14)$$

■

**Theorem 3.3** (The *parallelogram law*). For all  $x, y \in V$ ,  $\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2$  holds.

*Proof.* The proof is left as an exercise to the reader. ■

### 3.3 Orthogonality and Orthonormality

Note that if  $v$  is orthogonal to  $w$  and  $z$ , then it is orthogonal to both  $w + z$  and  $\alpha w$  for  $v, w, z \in V$  and  $\alpha \in \mathbb{F}$ . Thus, if  $x \in \operatorname{span}\{w, z\}$ , then  $v$  is also orthogonal to  $x$ . The entire subspace spanned by  $w$  and  $z$  is orthogonal to  $v$ .

**Proposition 3.4.** A set of orthogonal vectors, say  $(S = \{v_1, v_2, \dots, v_n\})$  with  $\langle v_i, v_j \rangle = 0$  for all  $i \neq j$ , is a linearly independent set, provided that  $S$  does not contain the zero vector.

*Proof.* Let  $\alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_n v_n = 0$ . For any valid  $j$ , note that

$$0 = \langle 0, v_j \rangle = \left\langle \sum_{i=1}^n \alpha_i v_i, v_j \right\rangle = \alpha_j \langle v_j, v_j \rangle$$

which must imply that  $\alpha_j = 0$  since  $v_j$  is not the zero vector. ■

**Definition 3.5.** A set of orthogonal vectors is said to be an *orthonormal set of vectors* if the norm of every vector in the set is unity.

We note that if  $S = \{v_1, v_2, \dots, v_n\}$  is an orthogonal set, then  $S' = \{\frac{v_1}{\|v_1\|}, \frac{v_2}{\|v_2\|}, \dots, \frac{v_n}{\|v_n\|}\}$  is an orthonormal set, provided  $v_i$  is never the zero vector.

In the vector space  $\mathbb{R}^n$ , an orthonormal basis forms the columns of an invertible matrix  $A$  such that  $A^t A = I_n$ . We take this as our definition of an orthogonal matrix.

**Definition 3.6.** A invertible matrix  $A$  is said to be an *orthogonal matrix* if  $A^t = A^{-1}$ , that is,  $A^t A = AA^t = I_n$ .

**Proposition 3.7.** Let  $\mathcal{B} = \{v_1, \dots, v_n\}$  be an orthonormal basis of  $V$ , where  $V$  is an inner product space. Then for every  $v \in V$ ,  $v = \sum_{i=1}^n \langle v, v_i \rangle v_i$ .

*Proof.* We simply have  $\langle v, v_j \rangle = \langle \sum_{i=1}^n \alpha_i v_i, v_j \rangle = \alpha_j \langle v_j, v_j \rangle = \alpha_j$ , for some  $v = \alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_n v_n$  with  $\alpha_i \in \mathbb{F}$ . ■

*March 6th.* Here are some properties;

- Two vectors  $x$  and  $y$  in a real inner product space are orthogonal if and only if  $\|x + y\|^2 = \|x\|^2 + \|y\|^2$ .
- Two vectors  $x$  and  $y$  in a complex inner product space are orthogonal if and only if  $\|\alpha x + \beta y\|^2 = \|\alpha x\|^2 + \|\beta y\|^2$  for all pairs  $\alpha$  and  $\beta$  in  $\mathbb{C}$ .
- If  $x$  and  $y$  are vectors in a real inner product space, and if  $\|x\| = \|y\|$ , then  $x - y$  and  $x + y$  are orthogonal.

### 3.3.1 The Orthogonal Complement

**Definition 3.8.** In any inner product space, the *distance between two vectors*  $x$  and  $y$  is defined as

$$d(x, y) = \|x - y\| = \sqrt{\langle x - y, x - y \rangle}.$$

**Definition 3.9.** In any inner product space, the *angle between two vectors*  $x$  and  $y$  is defined to be  $\theta \in [0, \pi]$  such that

$$\cos \theta = \frac{\langle x, y \rangle}{\|x\| \|y\|}.$$

**Definition 3.10.** Given a subspace  $W \subseteq V$ , the *orthogonal complement* of  $W$  is defined as

$$W^\perp = \{v \in V : \langle v, w \rangle = 0 \text{ for all } w \in W\}.$$

**Proposition 3.11.** Let  $W$  be a subspace of  $V$ . Then every vector  $v \in V$  can be written as  $v = w + w^\perp$  where  $w \in W$  and  $w^\perp \in W^\perp$ . Also,  $w$  is characterize by the property that it is the nearest point in  $W$  to  $v$ .

*Proof.* Fix  $\{w_1, \dots, w_k\}$  be an orthonormal basis of  $W$ . Let  $w = \sum_{i=1}^k \langle v, w_i \rangle w_i \in W$ . Let  $w^\perp = v - w$ . Then, for some  $1 \leq j \leq k$ ,

$$\langle v - w, w_j \rangle = \langle v, w_j \rangle - \langle \sum_{i=1}^k \langle v, w_i \rangle w_i, w_j \rangle = \langle v, w_j \rangle - \langle v, w_j \rangle \langle w_j, w_j \rangle = \langle v, w_j \rangle - \langle v, w_j \rangle = 0. \quad (3.15)$$

The second part of the proposition is left as an exercise to the reader. ■

**Proposition 3.12.** Let  $V$  be an inner product space, with  $W$  as a subspace. The following hold true:

- $W^\perp$  is a subspace,
- $W \cap W^\perp = \{0\}$ ,
- $W + W^\perp = V$ ,
- $(W^\perp)^\perp = W$

*Proof.* The proof is left as an exercise to the reader. ■

**Proposition 3.13.** *Let  $U, W$  be subspaces of  $V$ . The following hold true:*

- *If  $U \subseteq W$ , then  $W^\perp \subseteq U^\perp$ ,*
- *$(U + W)^\perp = U^\perp \cap W^\perp$ ,*
- *$(U \cap W)^\perp = U^\perp + W^\perp$ .*

*Proof.* The proof is left as an exercise to the reader. ■

**Definition 3.14.** Let  $W$  be a subspace of an inner product space  $V$ , and let  $v \in V$  such that  $v = w + w^\perp$  where  $w \in W$  and  $w^\perp \in W^\perp$ . Then  $w$  is called the *orthogonal projection* of  $v$  into  $W$  along  $W^\perp$ .

Take  $V$  to be  $\mathbb{R}^n$  or  $\mathbb{C}^n$  with the standard inner product(s).

**Theorem 3.15.** *For any real matrix  $A$ ,  $\text{nullsp} A = (\text{rowsp} A)^\perp$ . For any complex matrix  $A$ ,  $\text{nullsp} A = (\text{colsp} A^*)^\perp$ .*

*Proof.* Note that it is enough to prove the second part. Let  $X \in (\text{colsp} A^*)^\perp$  for a complex matrix  $A$ .  $X$  is in this space if and only if  $\langle X, A^*Y \rangle = 0$  for all  $Y \in \mathbb{C}^n$ .

$$\Leftrightarrow (A^*Y)^*X = 0 \Leftrightarrow Y^*AX = 0 \Leftrightarrow AX = 0 \Leftrightarrow X \in \text{nullsp} A. \quad (3.16)$$

■

**Definition 3.16.** Let the ground field  $\mathbb{F}$  be  $\mathbb{C}$  or  $\mathbb{R}$ , and let  $W$  be a subspace of  $\mathbb{F}^n$ . The *orthogonal projector* into  $W$  is the matrix  $P_{n \times n}$  such that for all  $v \in \mathbb{F}^n$ ,  $Pv$  is the orthogonal projection of  $v$  into  $W$ . Conversely, a matrix  $Q_{n \times n}$  is said to be an orthogonal projector if it is the orthogonal projector into some subspace  $W$  of  $\mathbb{F}^n$ .

**Proposition 3.17.** *In the previous definition,  $\mathbb{F}^n$  can be written as  $\text{colsp}(Q) \oplus \text{colsp}(I - Q)$ .*

*Proof.* It is enough to show that  $\text{colsp}(I - Q) = \text{colsp} Q^\perp$ . ■

**Theorem 3.18.** *For an  $n \times n$  matrix  $Q$ , the following are equivalent;*

- *$Q$  is an orthogonal projector,*
- *$Q^*Q = Q$ ,*
- *$Q^*Q$  and  $Q^2 = Q$ .*

*Proof.* If we first assume that  $Q$  is an orthogonal projector, then  $v - Qv$  is orthogonal to every vector in  $\text{colsp} Q$  and any vector  $v \in \mathbb{F}^n$ . So, for all  $u, v \in \mathbb{F}^n$ ,

$$\langle Qu, (I - Q)v \rangle = 0 \implies v^*(I - Q)^*Qu = 0 \implies (I - Q)^*Q = 0 \implies Q = Q^*Q. \quad (3.17)$$

The converse is also true. So, (i) and (ii) imply each other. Showing that (ii) and (iii) are equivalent is trivial. ■

### 3.3.2 Orthogonal and Unitary Matrices

We have already seen orthogonal matrices.

**Definition 3.19.** A complex square matrix is called a *unitary matrix* such that  $AA^* = A^*A = I_n$ .

It can be shown that  $2 \times 2$  orthogonal matrices are of the form

$$\begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \text{ or } \begin{pmatrix} \cos \theta & \sin \theta \\ \sin \theta & -\cos \theta \end{pmatrix} \text{ for } \theta \in [0, 2\pi). \quad (3.18)$$





# Appendices



## Chapter A

# Appendix

Extra content goes here.



# Index

algebraic multiplicity, 10  
angle between two vectors, 20  
annihilator, 11  
  
Cauchy-Schwarz, 18  
Cayley-Hamilton theorem, 13  
characteristic polynomial, 9  
cofactor, 4  
coordinate vector, 7  
cycle, 1  
  
determinant, 2  
diagonalizable linear operator, 10  
diagonalizable matrix, 10  
distance between two vectors, 20  
  
eigenspace, 10  
eigenvalue, 8  
eigenvector, 8  
even permutation, 2  
  
geometric multiplicity, 10  
  
ideal, 11  
inner product, 17  
inner product space, 18

k-cycle, 1  
  
minimal polynomial, 12  
  
norm, 18  
  
odd permutation, 2  
orthogonal complement, 20  
orthogonal matrix, 20  
orthogonal projection, 21  
orthogonal projector, 21  
orthogonal vectors, 18  
orthonormal set of vectors, 19  
  
parallelogram law, 19  
principal ideal, 11  
  
similar matrices, 8  
symmetric group, 1  
  
transpositions, 1  
triangle inequality, 19  
triangularizable linear operator, 14  
  
unitary matrix, 21