

# Linear Algebra Notes

Nikhil R.

## Contents

<b>1</b>	<b>Vector Spaces</b>	<b>3</b>
1.1	Vector Space . . . . .	3
1.2	Subspaces . . . . .	4
1.3	Linear Combinations and Systems of Linear Equations . . . . .	4
1.4	Span and Linear Dependence . . . . .	5
1.5	Basis and Dimension . . . . .	6
<b>2</b>	<b>Linear Transformations</b>	<b>8</b>
2.1	Linear Transformations . . . . .	8
2.2	Matrix Representation of an LT . . . . .	9
2.3	Composition of Linear Transformations . . . . .	10
2.4	Invertibility and Isomorphisms . . . . .	11
<b>3</b>	<b>RREF Theory</b>	<b>13</b>
3.1	Elementary Row Operations . . . . .	13
3.2	Column Correspondence Principle . . . . .	13
3.3	Uniqueness of RREF . . . . .	13
3.4	Some Further Results . . . . .	14
<b>4</b>	<b>Determinants</b>	<b>15</b>
4.1	Construction . . . . .	15
4.2	Properties . . . . .	16
4.3	Cramer's Rule . . . . .	16
<b>5</b>	<b>Diagonalization</b>	<b>17</b>
5.1	Eigenvalues and Eigenvectors . . . . .	17
5.2	Diagonalizability . . . . .	17
5.3	Sums and Direct Sums . . . . .	18
5.4	Invariant Subspaces and Cayley-Hamilton . . . . .	19
<b>6</b>	<b>Inner Product Spaces over <math>\mathbb{C}</math></b>	<b>22</b>
6.1	Inner Products and Norms . . . . .	22
6.2	Orthogonality . . . . .	23
6.3	Orthogonal Complements and Projections . . . . .	24

6.4	Adjoint of a Linear Operator . . . . .	25
6.5	Unitary, Normal, and Self-Adjoint Operators . . . . .	25
6.6	Singular Value Decomposition . . . . .	27
<b>7</b>	<b>Jordan Canonical Form</b>	<b>29</b>
7.1	Motivation . . . . .	29
7.2	Generalized Eigenvectors & Eigenspaces . . . . .	29
7.3	Jordan Canonical Basis Construction . . . . .	29
7.4	The Jordan Matrix Form . . . . .	30

# 1 Vector Spaces

## 1.1 Vector Space

**Definition.** A **vector space**  $V$  over a field  $\mathbb{F}$  (a set with addition and multiplication) is a set with operations  $(+) : V^2 \rightarrow V, x + y, +(x, y) \mapsto x + y$  and  $(\times) : \mathbb{F} \times V \rightarrow V, \times(a, v) \mapsto av$  such that the following axioms hold:

1. Commutativity:  $(\forall x, y \in V) x + y = y + x$
2. Associativity:  $(\forall x, y, z \in V) (x + y) + z = x + (y + z)$
3. Additive Identity:  $\exists \vec{0} \in V$  s.t.  $(\forall x \in V) x + \vec{0} = x$
4. Additive Inverses:  $(\forall x \in V) \exists y \in V$  s.t.  $x + y = \vec{0}$
5. Multiplicative Identity:  $(\forall x \in V) 1x = x$  (as 1 is defined on  $\mathbb{F}$ )
6. Associativity:  $(\forall a, b \in \mathbb{F}, x \in V) (ab)x = a(bx)$
7. Distributivity:  $(\forall a \in \mathbb{F}, x, y \in V) a(x + y) = ax + ay$
8. Distributivity:  $(\forall a, b \in \mathbb{F}, x \in V) (a + b)x = ax + bx$

The elements of  $\mathbb{F}$  are called **scalars** and the elements of  $V$  are called **vectors**. The building block example is  $V = \mathbb{R}^n$  over  $\mathbb{F} = \mathbb{R}$ , but these axioms hold for more abstract cases too.

**Example.**

- $\mathbb{F}^n$  over  $\mathbb{F}$ , with addition and multiplication defined component-wise.
- $M_{m \times n}(\mathbb{F})$  over  $\mathbb{F}$ , addition and multiplication defined component-wise.
- The set of all functions from  $S \neq \emptyset \rightarrow \mathbb{F}$ , denoted  $\mathcal{F}(S, \mathbb{F})$ .
- The set of polynomials with coefficients in  $\mathbb{F}$ .
- The set of sequences  $\{(a_n) : \mathbb{N} \rightarrow \mathbb{F}\}$  (could be infinite); Subcase of  $\mathcal{F}(S, \mathbb{F})$ .

**Proposition (Cancellation).** If  $x, y, z \in V$ , and if  $x + z = y + z$  or  $z + x = z + y$ , then  $x = y$ .

*Proof.*  $(x + z) + z_0 = (y + z) + z_0 \implies x + (z + z_0) = x + \vec{0} = x = y + (z + z_0) = y + \vec{0} = y$ . The other side follows from commutativity.  $\square$

**Proposition.** Suppose  $x \in V$ . Additive identities and inverses are unique.

*Proof.* Let  $\vec{0}_1, \vec{0}_2$  be additive identities. Then  $x + \vec{0}_1 = x = x + \vec{0}_2 \implies \vec{0}_1 = \vec{0}_2$  by cancellation. Let  $x_1, x_2$  be additive inverses. Then  $x + x_1 = \vec{0} = x + x_2$ . Then by cancellation,  $x_1 = x_2$ .  $\square$

**Proposition.**

- $(\forall x \in V) 0x = \vec{0}$

*Proof.*  $0x = (0 + 0)x = 0x + 0x$ . By cancellation,  $0x = \vec{0}$ .  $\square$

- $(\forall x \in V) (-1)x = -(1x) = -x$

*Proof.*  $1x + (-1)x = (1 + (-1))x = 0x = \vec{0}$  from above. Thus  $(-1)x$  is the unique inverse.  $\square$

- $(\forall a \in \mathbb{F}) a\vec{0} = \vec{0}$

*Proof.*  $a(\vec{0}) = a(\vec{0} + \vec{0}) = a\vec{0} + a\vec{0}$ . By cancellation, we get the result  $a\vec{0} = \vec{0}$ .  $\square$

## 1.2 Subspaces

**Definition.** If  $V$  is a vector space over  $\mathbb{F}$ ,  $W$  is a **subspace** of  $V$  if  $W \subseteq V$  and  $W$  forms a vector space using the same rules of addition/scalar multiplication as  $V$ .

**Proposition.**  $W$  is a subspace if and only if the following hold:

1.  $\vec{0} \in W$  (demands that  $W \neq \emptyset$ )
2.  $(\forall x, y \in W)(\forall a, b \in \mathbb{F}) ax + by \in W$
3. We inherit the other axioms 1, 2, 5, 6, 7, 8 and addition, multiplication from  $V$ .

**Proposition (Intersection).** Suppose  $\mathcal{C}$  is a collection of subspaces  $W_1, \dots, W_n$ . Then  $W := \bigcap_{i=1}^n W_i$  is a subspace also.

*Proof.*  $\vec{0}$  is contained in each  $W_i$  and so in the intersection.  $(\forall x, y \in W, \forall a, b \in \mathbb{F})$ ,  $ax + by \in W_i$  so  $ax + by \in W$ .  $\square$

## 1.3 Linear Combinations and Systems of Linear Equations

**Definition.** Suppose  $V$  is a vector space over  $\mathbb{F}$ . Let  $S \subseteq V$ . Then  $v \in V$  is called a **linear combination** of vectors in  $S$  if  $\exists$  a finite number of vectors  $u_i \in S$  and scalars  $a_i \in \mathbb{F}$  s.t.  $v = \sum_{i=1}^n a_i u_i$ .

**Remark.** Given a system of equations  $A\vec{x} = \vec{b}$ , we write the augmented matrix  $[A|\vec{b}]$  and bring it to **Reduced Row Echelon Form (RREF)** using elementary row operations. Note that because elementary row operations are invertible, the solution to the RREF matrix (which happens to be unique) is equivalent to the original system.

## 1.4 Span and Linear Dependence

**Definition.** Let  $S \subseteq V$ . The **span** of  $S$ , denoted by  $\text{span}(S)$ , is the set of all linear combinations of vectors in  $S$ . We define  $\text{span}(\emptyset) := \{\vec{0}\}$ . If  $\text{span}(S) = V$ , we say  $S$  (or the vectors in  $S$ ) **span** or **generate** the vector space  $V$ .

**Theorem (Minimal Subspace).**  $\text{span}(S)$  is a subspace containing  $S$  and any subspace of  $V$  that contains  $S$  also contains  $\text{span}(S)$ , meaning  $\text{span}(S)$  is the **smallest subspace** that contains  $S$ .

*Proof.*  $S = \emptyset$  is a trivial result so suppose  $S \neq \emptyset$ . We must show that  $\text{span}(S)$  is a subspace of  $V$ :

- $\vec{0} \in \text{span}(S)$ ; use  $0\vec{s} := \vec{0}$  for  $\vec{s} \in S \neq \emptyset$ .
- $(\forall a, b \in \mathbb{F})(\forall x, y \in \text{span}(S)) ax + by \in \text{span}(S)$ . If  $x, y \in \text{span}(S)$ ,  $\exists u_i, v_i \in S$  and  $a_i, b_i \in \mathbb{F}$  s.t.  $x = \sum a_i u_i$  and  $y = \sum b_i v_i$ . Then  $ax + by = \sum (aa_i u_i + bb_i v_i) \in \text{span}(S)$ .
- The other half remains to show that any subspace  $W$  of  $V$  containing  $S$  also contains  $\text{span}(S)$ . Suppose  $W$  is such a subspace and let  $x \in \text{span}(S)$ . Then  $x = \sum c_i s_i$  for vectors  $s_i \in S$  and  $c_i \in \mathbb{F}$ . Since  $W \supseteq S$ ,  $s_i \in W$ , thus  $x \in W$ .

□

**Definition.** Let  $S \neq \emptyset \subseteq V$ . Then  $S$  is said to be **linearly dependent** if  $\exists$  finite distinct  $\vec{s}_1, \dots, \vec{s}_n \in S$  and scalars  $a_1, \dots, a_n$  (not all 0) s.t.  $\sum_{i=1}^n a_i \vec{s}_i = \vec{0}$ . If all scalars were 0, the sum would also be trivially 0.

**Definition.** A set that is not linearly dependent is called **linearly independent**. Using DeMorgan's contradiction laws, a set is linearly independent if  $\forall$  finite selections of  $u_1, \dots, u_n$  (distinct)  $\in S$ ,  $\sum a_i u_i = 0$  implies that all  $a_i$  are 0. As a result, the vector space must be infinite dimensional (defined later) in order to contain an infinite linearly independent subset  $S$ .

**Proposition.** Suppose  $V$  over  $\mathbb{F}$  is given; let  $S_1 \subseteq S_2 \subseteq V$ .

- If  $S_1$  is linearly dependent, then so is  $S_2$ .
- If  $S_2$  is linearly independent, then so is  $S_1$ .

*Proof.* We can use the assumed statements in addition to setting certain constants to 0. □

**Proposition.** Suppose  $S \subseteq V$  is linearly independent. Then given  $\vec{v} \in V$ , the set  $S \cup \{\vec{v}\}$  is linearly dependent if and only if  $\vec{v} \in \text{span}(S)$ .

*Proof.*  $(\Rightarrow)$  If dependent, there exist coefficients not all zero. One of the vectors must be  $\vec{v}$  with nonzero coefficient (otherwise  $S$  is dependent). Thus we can solve for  $\vec{v}$ .  $(\Leftarrow)$  Suppose  $\vec{v} \in \text{span}(S)$ . Then  $\vec{v} = \sum a_i u_i$ . Let the linear combination be  $\sum (-a_i) u_i + 1\vec{v} = \vec{0}$ . □

## 1.5 Basis and Dimension

**Definition.** A **basis**  $\beta$  for a vector space  $V$  over  $\mathbb{F}$  is a linearly independent subset that generates (spans)  $V$ .

**Theorem (Coordinate Representations).**  $\beta$  is a basis for  $V$  over  $\mathbb{F}$  iff every  $x \in V$  is a **unique** linear combination of basis vectors. We may then represent any  $x \in V$  as an ordered tuple of number of elements in the basis ( $\mathbb{F}^n$  for finite bases).

*Proof.*  $(\Rightarrow)$  Suppose  $\beta$  is a basis and  $x = \sum a_i \beta_i = \sum b_i \beta_i$ . Then  $0 = \sum (a_i - b_i) \beta_i \Rightarrow a_i = b_i$  due to independence.  $(\Leftarrow)$  Suppose  $\forall x \in V \exists$  unique  $a_i$  s.t.  $\sum a_i \beta_i = x$ . Clearly  $\text{span}(\beta) = V$ . Also for  $x = \vec{0}$ , the trivial solution  $a_i = 0$  must hold. This must then be the only linear combination since we assume uniqueness.  $\square$

**Proposition.** If  $S$  is a finite set that generates  $V$ , then  $\exists \beta \subseteq S$  that forms a basis.

*Proof.* Suppose  $V \neq \{\vec{0}\}$ . Assume a finite set  $S = \{s_1 \dots s_n\}$  generates  $V$ . Starting at  $s_1$ , choose linearly independent vectors and discard others, constructing  $\beta = \{\beta_1, \dots, \beta_k\}$ . Stop when you reach the end of  $S$ . It remains to show that  $S \subseteq \text{span}(\beta)$ . Suppose  $y \in S$ . If  $y \in \beta$ ,  $y \in \text{span}(\beta)$ . If  $y \notin \beta$ , then because  $\beta \cup \{y\}$  is linearly dependent, then  $y \in \text{span}(\beta)$ .  $\square$

**Theorem (Replacement).** Suppose  $V$  over  $\mathbb{F}$  is generated by  $G$  having  $n$  vectors. Let  $L \subseteq V$  have  $m$  vectors be independent. Then  $n \geq m$  and  $\exists H \subseteq G$  containing  $n - m$  vectors s.t.  $L \cup H$  spans  $V$ .

*Proof.* By induction.  $m = 0$ ,  $L = \emptyset$ ,  $H = G$ . Now assume the result holds for  $m$ . For  $m + 1$ , let  $L = \{v_1 \dots v_{m+1}\}$ . Then  $\{v_1 \dots v_m\}$  is also independent. By induction,  $\exists H' = \{u_1 \dots u_{n-m}\} \subseteq G$  s.t.  $v_{m+1} \in \text{span}(\{v_1 \dots v_m\} \cup H')$ . This tells us the set must at least contain one vector with nonzero coefficient to avoid contradictions. Call it  $u_k$ . Let  $H = H' \setminus \{u_k\}$ . By rearranging for  $u_k$ , we see that  $u_k \in \text{span}(\{v_1 \dots v_{m+1}\} \cup H)$ . Which means the span is  $V$ . By construction  $L$  has  $m + 1$  vectors and  $H$  has  $n - (m + 1)$  vectors.  $\square$

**Theorem (Invariance of Dimension).** Suppose  $V$  has a finite basis of size  $n$ . Then any basis for  $V$  has  $n$  elements.

*Proof.* Suppose  $\beta$  is a basis for  $V$  with  $n$  vectors. Then assume  $\gamma$  is another basis of  $V$  with  $m$  vectors. By replacement theorem:  $L = \beta, G = \gamma \Rightarrow n \leq m$  and  $L = \gamma, G = \beta \Rightarrow m \leq n$ . Thus  $m = n$ .  $\square$

**Definition.**  $\dim(V)$  describes the unique number of vectors in every basis for a vector space  $V$ .  $V$  is called **finite dimensional** if  $\dim(V)$  is finite and **infinite dimensional** otherwise.

**Corollary.**

- Any finite generating set contains  $\geq n$  elements.
- Any generating set with  $n$  vectors is a basis for  $V$ .

- Any independent subset with  $n$  vectors generates  $V$ .
- Every independent subset can be extended into a basis.

## 2 Linear Transformations

### 2.1 Linear Transformations

**Definition.** Let  $V, W$  be vector spaces over  $\mathbb{F}$ . The function  $T : V \rightarrow W$  is a **linear transformation** if  $\forall x, y \in V$  and  $c_1, c_2 \in \mathbb{F}$ ,

$$T(c_1x + c_2y) = c_1T(x) + c_2T(y)$$

**Proposition.**  $T(\vec{0}_V) = \vec{0}_W$  and  $T(\sum c_i \vec{x}_i) = \sum c_i T(\vec{x}_i)$ .

*Proof.*  $T(\vec{0}_V) = T(0_{\mathbb{F}} \vec{0}_V) = 0_{\mathbb{F}} T(\vec{0}_V) = \vec{0}_W$ . The second statement follows from induction.  $\square$

**Example.**

- $L_A : \mathbb{F}^n \rightarrow \mathbb{F}^m$  s.t.  $\vec{x} \mapsto A\vec{x}$  where  $A \in M_{m \times n}(\mathbb{F})$
- $I_V : V \rightarrow V$  s.t.  $\vec{x} \mapsto \vec{x}$ , Identity
- $T_0 : V \rightarrow W$  s.t.  $\vec{x} \mapsto \vec{0}_W$ , Zero transformation

**Definition.** Suppose  $T : V \rightarrow W$  is a LT.

- The **Range** (or Image):  $R(T) := \{T(\vec{x}) : \vec{x} \in V\}$ .
- The **Nullspace** (or Kernel):  $N(T) := \{\vec{x} : T(\vec{x}) = \vec{0}_W, \vec{x} \in V\}$ .

**Proposition.** The range is a subspace of  $W$  and the nullspace is a subspace of  $V$ .

*Proof.*  $T(\vec{0}_V) = \vec{0}_W$ , so  $\vec{0}_V \in N(T)$ . Also if  $x, y \in N(T)$  and  $a \in \mathbb{F}$  then  $T(ax + y) = aT(x) + T(y) = \vec{0}_W$ . Also  $T(\vec{0}_V) = \vec{0}_W \in R(T)$ . Then if  $x, y \in R(T)$ ,  $\exists u, v \in V$  s.t.  $x = T(u), y = T(v)$ . Then  $T(au + v) = ax + y \in R(T)$ .  $\square$

**Theorem (Transforming Basis).** Suppose  $V, W$  over  $\mathbb{F}$ , and  $T : V \rightarrow W$  linear is given.  $\beta = \{v_1 \dots v_n\}$  is a basis for  $V$ . Then  $R(T) = \text{span}(\{T(v_1) \dots T(v_n)\}) = \text{span}(T(\beta))$ .

*Proof.*  $T(v_i) \in R(T)$  since  $v_i \in V$ . Since  $R(T)$  is a subspace,  $\text{span}(T(\beta)) \subseteq R(T)$ . Now suppose  $w \in R(T)$ .  $\exists x \in V$  s.t.  $T(x) = w$ , and  $x$  can be uniquely written as  $\sum a_i v_i = x$ . Then  $w = T(\sum a_i v_i) = \sum a_i T(v_i) \in \text{span}(T(\beta))$ .  $\square$

**Definition (Rank and Nullity).** Suppose  $V$  and  $W$  are vector spaces over  $\mathbb{F}$ .  $T : V \rightarrow W$  linear with  $N(T)$  and  $R(T)$  being finite dimensional. Then  $\text{rank}(T) := \dim(R(T))$  and  $\text{nullity}(T) := \dim(N(T))$

**Theorem (Dimension).** Suppose  $V, W$  over  $\mathbb{F}$  and  $T : V \rightarrow W$  linear is given. If  $V$  is finite dimensional, then:

$$\text{rank}(T) + \text{nullity}(T) = \dim(V)$$

*Proof.* Suppose  $\dim(V) = n$  and  $\text{nullity}(T) = k$ . Let a basis for  $N(T)$  be  $\{v_1 \dots v_k\}$  which we extend to a basis for  $V$  with  $\{v_{k+1} \dots v_n\}$ . Then  $S = \{T(v_{k+1}) \dots T(v_n)\}$  is a basis for  $R(T)$ .

- $\text{span}(S) = R(T)$  (from Transforming Basis theorem).
- Independence:  $\sum_{k+1}^n b_i T(v_i) = 0 = T(\sum_{k+1}^n b_i v_i)$  means  $\sum b_i v_i \in N(T) = \sum_1^k c_i v_i$ . Since  $v_1 \dots v_n$  span  $V$ , independence implies  $b_i = 0$ .
- Thus  $\dim(V) = n, \text{rank}(T) = n - k, \text{nullity}(T) = k$ .

□

**Proposition.**  $T : V \rightarrow W$  is injective (one-to-one) if and only if  $N(T) = \{\vec{0}\}$ .

*Proof.*  $(\Rightarrow)$  Suppose  $N(T) = \{\vec{0}\}$  and  $\vec{x}, \vec{y} \in V$ . Assume  $T(\vec{x}) = T(\vec{y})$ . Then  $T(\vec{x}) - T(\vec{y}) = \vec{0} = T(\vec{x} - \vec{y})$ , so  $x - y \in N(T) \implies x = y$ .  $(\Leftarrow)$  Suppose  $T(\vec{x}) = T(\vec{y}) \implies \vec{x} = \vec{y}$ . Since  $T(\vec{0}) = \vec{0}$ , if  $T(\vec{y}) = \vec{0}$ ,  $\vec{y}$  must be  $\vec{0}$ . □

**Proposition.** Suppose  $V, W$  both have  $\dim(V) = \dim(W) = n$ , and that  $T : V \rightarrow W$  is linear. Then  $T$  is injective  $\iff T$  is surjective  $\iff \text{rank}(T) = \dim(V)$

*Proof.* Use the dimension theorem and that  $R(T)$  is a subspace of  $W$ .  $R(T) = W$  if and only if  $\text{rank}(T) = \dim(W) = \dim(V)$ . □

**Theorem (Uniqueness).** Suppose  $V, W$  are over  $\mathbb{F}$  and that  $\{v_1 \dots v_n\}$  is a basis for  $V$ . Given  $\{w_1 \dots w_n\} \subseteq W$ , there is a unique  $T$  s.t.  $T(v_i) = w_i$  for  $i = 1 \dots n$ .

*Proof.*  $T(v_i) = w_i$  is defined and we see that it is unique since if there is a  $U : V \rightarrow W$  defined in the same way, then  $\forall x \in V, T(x) = U(x) \iff T = U$ . □

## 2.2 Matrix Representation of an LT

**Definition (Coordinate Vector).** Suppose  $V$  over  $\mathbb{F}$  is finite dimensional and that  $\beta = \{v_1 \dots v_n\}$  is an ordered basis. The unique column vector of scalars  $a_i$  s.t.  $\sum a_i v_i = x$  defines  $[x]_\beta$ , the **coordinate representation** of  $x \in V$ .

**Definition (Matrix of a Transformation).** Suppose  $V$  and  $W$  over  $\mathbb{F}$  are finite dimensional with basis  $\beta = \{v_1 \dots v_n\}$  for  $V$  and  $\gamma = \{w_1 \dots w_m\}$  for  $W$ . Then  $\exists$  unique scalars  $a_{ij} \in \mathbb{F}$  s.t.

$$T(v_j) = \sum_{i=1}^m a_{ij} w_i \implies [A]_{ij} = a_{ij}$$

Then, the matrix representation  $[T]_\beta^\gamma = A$ . Each column of  $A$  is computed by taking the corresponding ordered basis vector from  $\beta$ , applying  $T$  to it, and writing down the result in coordinates of  $\gamma$ .

**Proposition.** Suppose  $T, U : V \rightarrow W$ . Then the following hold

- $U + T : V \rightarrow W \equiv (U + T)(x) = U(x) + T(x)$

- $aT : V \rightarrow W \equiv (aT)(x) = aT(x)$
- The collection of all LT's from  $V \rightarrow W$ , denoted  $\mathcal{L}(V, W)$ , is itself a vector space over  $\mathbb{F}$ .
- $[aT + U]_\beta^\gamma = a[T]_\beta^\gamma + [U]_\beta^\gamma$ .

## 2.3 Composition of Linear Transformations

**Proposition.** Suppose  $V, W, Z$  are vector spaces over  $\mathbb{F}$  and that  $T : V \rightarrow W$ ,  $U : W \rightarrow Z$  is linear. Then  $UT \equiv U \circ T$  is also linear.

*Proof.*  $(UT)(ax + by) = U(T(ax + by)) = U(aT(x) + bT(y)) = a(UT)(x) + b(UT)(y)$ .  $\square$

**Remark (Properties).** Suppose  $T_1, T_2 \in \mathcal{L}(V, W)$  and  $U_1, U_2 \in \mathcal{L}(W, Z)$ .

- $U_1(T_1 + T_2) = U_1T_1 + U_1T_2$
- $(U_1 + U_2)T_1 = U_1T_1 + U_2T_1$
- $a(U_1T_1) = U_1(aT_1) = (aU_1)T_1$
- Associativity holds.

**Theorem (Main).** Suppose  $V, W, Z$  over  $\mathbb{F}$  with ordered bases  $\alpha, \beta, \gamma$ . Let  $A = [T]_\alpha^\beta$  and  $B = [U]_\beta^\gamma$ . Then  $[UT]_\alpha^\gamma = [U]_\beta^\gamma[T]_\alpha^\beta = BA$ .

*Proof.*  $(UT)(v_j) = U(\sum B_{kj}w_k) = \sum B_{kj}U(w_k) = \sum B_{kj} \sum A_{lk}z_l = \sum_l (\sum_k A_{lk}B_{kj})z_l$ . This matches matrix multiplication.  $\square$

**Definition.** Algebra on matrices is defined in the following way (this is consistent with what we expect for compositions of linear transformations)

- $A(B + C) = AB + AC$
- $(D + E)A = DA + EA$
- $a(AB) = (aA)B = A(aB)$
- $I_m A = A = A I_n$
- $[I_V]_\beta = I_n$

**Remark (Row and Column Picture).**

- $(AB)_j = A(B_j)$ : Column picture (each column of  $AB$  is a linear combination of columns of  $A$ ).
- Similarly, each row of  $AB$  is a linear combination of the rows of  $B$ , with the coefficients corresponding to the row vectors in  $A$ .

**Proposition.**  $(AB)^T = B^T A^T$ .

*Proof.*  $(AB)_{ij}^T = (AB)_{ji} = \sum_k A_{jk} B_{ki} = \sum_k (B^T)_{ik} (A^T)_{kj} = (B^T A^T)_{ij}$ .  $\square$

**Theorem.** Suppose  $V, W$  finite dimensional with bases  $\beta, \gamma$ .  $T : V \rightarrow W$  linear. Then  $\forall u \in V$ ,  $[T(u)]_\gamma = [T]_\beta^\gamma [u]_\beta$ .

*Proof.*  $[T(u)]_\gamma = [T(\sum a_i v_i)]_\gamma = \sum a_i [T(v_i)]_\gamma = \sum a_i [T]_\beta^\gamma e_i = [T]_\beta^\gamma \sum a_i e_i = [T]_\beta^\gamma [u]_\beta$ .  $\square$

## 2.4 Invertibility and Isomorphisms

**Proposition.** Suppose  $T : V \rightarrow W$  is an invertible linear function. Then  $T^{-1}$  is linear.

*Proof.* Let  $x_1$  and  $x_2$  be such that  $T(x_1) = y_1$  and  $T(x_2) = y_2$ . Then, by linearity,  $T^{-1}(ay_1 + y_2) = T^{-1}(aT(x_1) + T(x_2)) = T^{-1}(T(ax_1 + x_2)) = ax_1 + x_2 = aT^{-1}(y_1) + T^{-1}(y_2)$ .  $\square$

**Proposition.** Suppose  $T : V \rightarrow W$  is invertible linear.  $V$  is finite dimensional iff  $W$  is finite dimensional. Further  $\dim(V) = \dim(W)$ .

*Proof.* Choose a basis for  $V$ . The image of the basis is a basis for  $W$ .  $\square$

**Proposition.** If  $T$  is invertible,  $[T^{-1}]_\gamma^\beta = ([T]_\beta^\gamma)^{-1}$ .

*Proof.* Recall composition of transformations corresponds to matrix multiplication.  $\square$

**Definition.**  $V$  is **isomorphic** to  $W$  if  $\exists T : V \rightarrow W$  s.t.  $T$  is linear and invertible. We then call  $T$  an **isomorphism**. This is an equivalence relation.

*Proof.* Reflexivity is given by the identity transformation, symmetry is given by the inverse transformation, and transitivity is given since compositions of invertible linear transformations are also invertible linear transformations.  $\square$

**Proposition.** Suppose  $V, W$  are finite dimensional. They are isomorphic iff  $\dim(V) = \dim(W)$ .

*Proof.* Choose a basis  $\beta$  for  $V$  and  $\gamma$  for  $W$ . We know they have the same number of elements. Consider any linear transformation that is an injection from  $\beta \rightarrow \gamma$ . This is an isomorphism.  $\square$

**Theorem.** The following maps are isomorphism (assuming a basis has been decided for both)

- $\Phi : \mathcal{L}(V, W) \rightarrow M_{m \times n}(\mathbb{F})$  defined by  $T \mapsto [T]_\beta^\gamma$ .
- $\Phi_\beta : V \rightarrow \mathbb{F}^n$  defined by  $v \mapsto [v]_\beta$

**Corollary.**  $Q := [I_V]_{\beta'}^\beta$  is invertible and  $[v]_\beta = Q[v]_{\beta'}$ .

**Theorem.** Suppose  $V, W$  over  $\mathbb{F}$  are given with  $\dim(V) = n$  and  $\dim(W) = m$  and respective bases  $\beta$  and  $\gamma$ . Let  $T : V \rightarrow W$  be linear and  $L_A : \mathbb{F}^n \rightarrow \mathbb{F}^m$  be the corresponding matrix transformation. There are equivalent two paths from  $V \rightarrow \mathbb{F}^m$ :

1. Map  $v \rightarrow w$  via  $T$ , then to coordinate space via  $\Phi_\gamma$ .
2. Map  $v \rightarrow [v]_\beta$  via  $\Phi_\beta$ , then transform via  $L_A$ .

**Theorem (Linear Transformations in Different Bases).** Suppose  $T : V \rightarrow V$ . Then  $[T]_\beta = Q[T]_{\beta'}Q^{-1}$ .

*Proof.*  $Q[T]_{\beta'} = [I_V T]_{\beta'}^\beta = [T I_V]_{\beta'}^\beta = [T]_\beta [I_V]_{\beta'}^\beta = [T]_\beta Q$

*Special Case:* If  $A \in M_{n \times n}$  and  $\gamma$  is a basis for  $\mathbb{F}^n$ , let  $\beta$  be standard basis. Then  $[L_A]_\gamma = Q^{-1}AQ$ . □

**Definition.**  $A, B$  are **similar** if  $\exists$  invertible  $Q$  s.t.  $B = Q^{-1}AQ$ . This is clearly an equivalence relation. In particular, if  $T : V \rightarrow V$ , expressed as  $A = [T]_\beta$  and  $B = [T]_{\beta'}$  then  $A$  and  $B$  are similar.

## 3 RREF Theory

### 3.1 Elementary Row Operations

**Definition.** Given  $A \in M_{m \times n}(\mathbb{F})$ , there are three types of **elementary row operations**:

1. Interchanging two rows.
2. Multiplying a row by a nonzero scalar.
3. Adding a nonzero scalar multiple of a row to another row.

**Remark.** When considering a system of linear equations, performing these three operations never changes the solution space of the system.

**Theorem (Invertibility of Elementary Matrices).** An elementary matrix  $E$  is obtained by performing one of the elementary operations on  $I_n$ . Any elementary row operation on a matrix  $A$  can be performed by left multiplying  $A$  by the corresponding elementary matrix ( $B = EA$ ). Each elementary matrix  $E$  is invertible, and its inverse is of the same type. Thus, a product of elementary matrices is invertible.

### 3.2 Column Correspondence Principle

**Theorem (Column Correspondence).** Suppose  $A, B \in M_{m \times n}(\mathbb{F})$  are **row equivalent**, meaning they can be obtained from each other by an invertible sequence of elementary row operations. Let  $A = [\vec{a}_1 \dots \vec{a}_n]$  and  $B = [\vec{b}_1 \dots \vec{b}_n]$ . Then

$$\sum_{j=1}^n c_j \vec{a}_j = \vec{0} \iff \sum_{j=1}^n c_j \vec{b}_j = \vec{0}$$

*Proof.*  $N(A) = N(B)$ , since  $A$  and  $B$  are obtained from each other by an invertible matrix. Namely,  $A = EB$  implies  $A\vec{x} = \vec{0} \iff EB\vec{x} = \vec{0} \iff B\vec{x} = \vec{0}$  (since  $N(E) = \{\vec{0}\}$ ).  $\square$

### 3.3 Uniqueness of RREF

**Theorem.** Suppose  $A, R, S \in M_{m \times n}(\mathbb{F})$  and that  $R, S$  are in Reduced Row Echelon Form (RREF) and are row equivalent to  $A$ . Then  $R = S$ .

*Proof.* Write  $R = [\vec{r}_1 \dots \vec{r}_n]$  and  $S = [\vec{s}_1 \dots \vec{s}_n]$ . By the Column Correspondence Principle, linear dependencies are preserved.

- If  $\vec{r}_j$  is a linear combination of previous columns,  $\vec{s}_j$  is the same linear combination of the corresponding columns of  $S$ .
- If a column is independent of previous columns, it is a pivot column.
- Working from  $j = 1$  to  $n$ :

$$\text{– 1st column: } \vec{r}_1 = \vec{0} \iff \vec{s}_1 = \vec{0} \text{ and } \vec{r}_1 = \hat{e}_1 \iff \vec{s}_1 = \hat{e}_1.$$

- Assume first  $k - 1$  columns are the same. For the  $k$ -th column, if it is independent of previous columns,  $\vec{r}_k = \vec{s}_k = \hat{e}_l$  where  $l = (\# \text{ of pivots}) + 1$ ; if it is dependent, the coefficients are uniquely determined by the previous pivot columns, since the pivot columns are independent and span the same as all the previous  $k - 1$  columns.

- By induction,  $R = S$ .

□

### 3.4 Some Further Results

**Corollary.** In  $\mathbb{F}^n$ , every finite independent set has  $\leq n$  elements.

*Proof.* Let  $\vec{u}_1 \dots \vec{u}_k \in \mathbb{F}^n$  be independent. Let  $A = [\vec{u}_1 \dots \vec{u}_k]$ . Then  $R = \text{RREF}(A) = \begin{bmatrix} I_k \\ 0 \end{bmatrix}$ . For this to exist inside an  $n \times k$  matrix, clearly  $k \leq n$ . □

**Corollary.** In  $\mathbb{F}^n$ , every finite generating set has  $\geq n$  elements.

*Proof.* Let  $\vec{v}_1 \dots \vec{v}_k$  generate  $\mathbb{F}^n$ . Let  $A = [\vec{v}_1 \dots \vec{v}_k]$  and  $R = \text{RREF}(A)$ . Since the columns of  $A$  generate  $\mathbb{F}^n$ ,  $R$  must have  $n$  pivots, meaning  $k \geq n$  (since  $R$  has  $k$  columns). □

**Corollary.** In  $\mathbb{F}^n$ , every basis has exactly  $n$  elements.

*Proof.* A basis is linearly independent ( $\leq n$ ) and generating ( $\geq n$ ), so it must have  $n$  elements. □

**Corollary.** If  $V$  has one finite basis containing  $n$  elements, then so does every basis.

*Proof.* There exists a unique isomorphism  $T : V \rightarrow \mathbb{F}^n$ . An isomorphism must take independent sets to independent sets and generating sets to generating sets. Thus the results for  $\mathbb{F}^n$  carry over to  $V$ . □

## 4 Determinants

### 4.1 Construction

**Definition.** Let the function  $\delta : M_{n \times n}(\mathbb{F}) \rightarrow \mathbb{F}$  have the following properties:

1. Multilinear: Linear in any one row when all others are held fixed.
2. Sign: If two rows are identical, then  $\delta(A) = 0$ . This is an equivalent statement to saying that if two rows are interchanged, then  $\delta$  changes sign.

*Proof.*  $\delta \begin{pmatrix} r_1+r_2 \\ r_1+r_2 \end{pmatrix} = \delta \begin{pmatrix} r_1 \\ r_1 \end{pmatrix} + \delta \begin{pmatrix} r_1 \\ r_2 \end{pmatrix} + \delta \begin{pmatrix} r_2 \\ r_1 \end{pmatrix} + \delta \begin{pmatrix} r_2 \\ r_2 \end{pmatrix}$ . Since terms with identical rows are 0, we get  $0 = 0 + \delta \begin{pmatrix} r_1 \\ r_2 \end{pmatrix} + \delta \begin{pmatrix} r_2 \\ r_1 \end{pmatrix} + 0 \implies \delta \begin{pmatrix} r_1 \\ r_2 \end{pmatrix} = -\delta \begin{pmatrix} r_2 \\ r_1 \end{pmatrix}$ .  $\square$

3. Normalization:  $\delta(I_n) = 1$ .

**Theorem.** Such a function exists and is unique. It is called the **determinant**.

*Proof.*

- Uniqueness: Let  $\delta$  be a function satisfying the three properties (multilinear, alternating, normalized). We can write each row of  $A$  as a linear combination of standard basis vectors  $e_1, \dots, e_n$ . By linearity, we can expand  $\delta(A)$  into a sum of terms of the form  $c \cdot \delta(M)$ , where  $M$  is a matrix with exactly one standard basis vector in each row. If  $M$  has a zero row or two identical rows,  $\delta(M) = 0$ . Thus, the only nonzero terms come from matrices  $M$  that are permutation matrices (columns are a permutation of  $e_1, \dots, e_n$ ). We can swap rows of such an  $M$  to transform it into  $I_n$ . Each swap changes the sign. Since  $\delta(I_n) = 1$ , the value of  $\delta(M)$  is determined solely by the number of swaps (the sign of the permutation). Therefore,  $\delta(A)$  is uniquely determined by the entries of  $A$ .
- Existence: We can inductively define the determinant using the **Cofactor Expansion** along the first row:

$$\det(A) = \sum_{j=1}^n (-1)^{1+j} A_{1j} \det(\tilde{A}_{1j})$$

It can be verified that this function satisfies the three axioms (linearity, alternating, normalization).  $\square$

**Proposition.** Left multiplying by elementary matrices has the following effects

- Swap: Interchanging two rows changes the sign of the determinant.
- Scaling: Multiplying a row by a scalar multiplies the determinant by that scalar.
- Row Replacement: Adding a multiple of one row to another leaves the determinant unchanged.

## 4.2 Properties

**Proposition.**  $\det(A) = 0 \iff A$  is non-invertible.

*Proof.* Follows from RREF computation.  $R = \text{RREF}(A) = I_n \iff A$  has rank  $n$ , meaning  $A$  is invertible. But the RREF was found by multiplying  $A$  by elementary matrices, which from above, never takes a matrix with a non-zero determinant into one with zero determinant.  $\square$

**Proposition.**  $\det(AB) = \det(A)\det(B)$ .

*Proof.* If  $AB$  is non-invertible, then either  $A$  or  $B$  is non-invertible (so both sides are 0). Otherwise, for two elementary matrices  $E_1, E_2$ , we verify that  $\det(E_1 E_2) = \det(E_1)\det(E_2)$ . Then,  $\det(AB) = \det((\prod E_i)B) = \prod \det(E_i)\det(B) = \det(\prod E_i)\det(B) = \det(A)\det(B)$ .  $\square$

**Corollary.** If  $A$  is invertible,  $\det(A^{-1}) = (\det(A))^{-1}$ .

**Proposition.**  $\det(A^T) = \det(A)$ .

*Proof.* We verify that  $\det(E) = \det(E^T)$  for any elementary matrix  $E$ .  $A$  is invertible if and only if  $A^T$  is invertible, and we note that  $A$  can be then written as a product of elementary matrices, meaning once the transpose is taken the product order switches, and each element is transposed.  $\square$

**Proposition.** Suppose  $T : V \rightarrow V$ , and there are two bases  $\beta$  and  $\beta'$  for  $V$ . We define the **determinant** of  $T$  as  $\det(T) = \det([T]_\beta)$ . In particular, the choice of basis does not matter.

*Proof.*  $[T]_\beta = Q[T]_{\beta'}Q^{-1}$  so  $\det([T]_\beta) = \det(Q[T]_{\beta'}Q^{-1}) = \det(Q)\det([T]_{\beta'})\det(Q^{-1}) = \det([T]_{\beta'})$   $\square$

## 4.3 Cramer's Rule

**Theorem (Cramer's Rule).** Suppose  $A$  is invertible and let  $A\vec{x} = \vec{b}$  and let  $A_k$  be the matrix  $A$  with the  $k$ -th column replaced by  $\vec{b}$ . Then,

$$x_k = \frac{\det(A_k)}{\det(A)}$$

*Proof.* Note that  $A_k = AX_k$  where  $X_k$  is the identity with the  $k$ -th column replaced by  $\vec{x}$ . Then  $\det(A_k) = \det(A)\det(X_k)$ . By cofactor expansion,  $\det(X_k) = x_k$ . So  $x_k = \det(A_k)/\det(A)$ .  $\square$

## 5 Diagonalization

### 5.1 Eigenvalues and Eigenvectors

**Definition.** A linear operator  $T : V \rightarrow V$  is **diagonalizable** if  $\exists$  a basis  $\beta$  s.t.  $[T]_\beta$  is diagonal. This happens if  $T(\beta_i) = \lambda_i \beta_i$  for  $\lambda_i \in \mathbb{F}$ .

**Definition.** Suppose  $T : V \rightarrow V$  is a linear operator. For  $v \neq 0 \in V$ , if  $Tv = \lambda v$  for  $\lambda \in \mathbb{F}$ , then  $v$  is called an **eigenvector** and  $\lambda$  is called an **eigenvalue**.

**Remark.** In summary,  $T$  is diagonalizable  $\iff V$  has a basis of eigenvectors of  $T$ . The diagonal matrix  $D = [T]_\beta$  has eigenvalues  $D_{jj}$  corresponding to the eigenvectors in  $\beta$ .

**Proposition.** Suppose  $A \in M_{n \times n}(\mathbb{F})$ .  $\lambda$  is an eigenvalue if and only if  $\det(A - \lambda I_n) = 0$ .

*Proof.*  $\lambda$  is an eigenvalue  $\iff \exists v \neq 0$  s.t.  $Av = \lambda v \iff (A - \lambda I_n)v = \vec{0}_v \iff (A - \lambda I_n)$  is not injective  $\iff \det(A - \lambda I_n) = 0$ .  $\square$

**Definition.** We may define the **characteristic polynomial** of a linear operator  $T : V \rightarrow V$  as  $f_T(t) := \det(T - tI_V)$ .

*Proof.*  $\det([T]_\beta - tI_n) = \det(Q[T]_{\beta'}Q^{-1} - tI_n) = \det(Q([T]_{\beta'} - tI_n)Q^{-1}) = \det([T]_{\beta'} - tI_n)$ . This shows the polynomial  $f(t)$  is well defined regardless of choice of basis.  $\square$

**Remark.** Suppose  $A \in M_{n \times n}(\mathbb{F})$  or that  $A = [T]_\beta$  with  $\dim(V) = n$ . Then  $f(t)$  is an  $n$ -degree polynomial with the coefficient for the  $n$ -th term being  $(-1)^n$ . By the fundamental theorem of algebra,  $f_T$  has at most  $n$  distinct eigenvalues.

**Remark.** Suppose  $T : V \rightarrow V$  and let  $\lambda$  be an eigenvalue of  $T$ .  $v \neq 0$  is an eigenvector corresponding to  $\lambda$  if and only if  $v \in N(T - \lambda I_V)$ . Thus for an eigenvalue  $\lambda$ , we call the nullspace of  $T - \lambda I_V$  the **eigenspace** corresponding to  $\lambda$ , denoted  $E_\lambda$ . Note that if  $\lambda$  is not an eigenvalue then  $T - \lambda I$  is invertible and  $N(T - \lambda I_V) = \{\vec{0}_V\}$ .

### 5.2 Diagonalizability

**Theorem (Independence of Eigenvectors).** Suppose  $T : V \rightarrow V$  and  $\lambda_1 \dots \lambda_k$  are distinct eigenvalues. Let  $v_1 \dots v_k$  be eigenvectors corresponding to the  $\lambda_i$ 's. Then the set  $\{v_1 \dots v_k\}$  is linearly independent.

*Proof.* For  $k = 1$ , since  $v_1 \neq 0$  is an eigenvector,  $\{v_1\}$  is independent. Now suppose the result holds for  $\{v_1 \dots v_{k-1}\}$ . Consider the equation  $\sum_{i=1}^k a_i v_i = 0$ . Then,  $(T - \lambda_k I_V)(\sum_{i=1}^k a_i v_i) = \vec{0}_V = \sum_{i=1}^k a_i (\lambda_i - \lambda_k) v_i = \sum_{i=1}^{k-1} a_i (\lambda_i - \lambda_k) v_i$ . By the induction hypothesis,  $a_i (\lambda_i - \lambda_k) = 0$  and since the eigenvalues are distinct,  $a_i = 0$  for up to  $k - 1$ , but this implies  $a_k = 0$  as well.  $\square$

**Corollary.** If  $T : V \rightarrow V$  for  $\dim(V) = n$  has  $n$  distinct eigenvalues, then  $T$  is diagonalizable.

**Definition (Splitting).**  $T : V \rightarrow V$  **splits** if  $f_T(t) = c(t - a_1) \dots (t - a_n)$  where  $c, a_i \in \mathbb{F}$  and need not be distinct. Note that the scalars must be from  $\mathbb{F}$ . (This is always guaranteed in  $\mathbb{C}$ , but not in all fields such as  $\mathbb{R}$ .)

**Proposition.** If  $T$  is diagonalizable, then  $f(t)$  splits over  $\mathbb{F}$ .

*Proof.* Consider a basis  $\beta$  for which  $[T]_\beta = D$ . Then  $\det(T - \lambda I) = \det(D - tI_n) = (-1)^n \prod_{i=1}^n (\lambda_i - t)$ . The converse need not be true.  $\square$

**Definition.** The maximum value  $k$  for which  $(\lambda - t)^k$  is a factor of  $f(t)$  is called the **algebraic multiplicity** of  $\lambda$ , while  $\dim(E_\lambda)$ , the maximum number of linearly independent eigenvectors corresponding to  $\lambda$ , is called the **geometric multiplicity** of  $\lambda$ .

**Theorem.**  $1 \leq \text{GM} \leq \text{AM}$ .

*Proof.* Let  $\{v_1 \dots v_k\}$  be a basis for  $E_\lambda$  and extend it to a basis  $\beta = \{v_1 \dots v_k, v_{k+1} \dots v_n\}$  for  $V$ . Notice  $[T]_\beta$  has the form  $\begin{bmatrix} \lambda I_k & B \\ 0 & C_{n-k} \end{bmatrix}$ . Then  $f_T(t) = \det(T - tI_V) = \det(\lambda I_k - tI_k) \det(C - tI_{n-k}) = (\lambda - t)^k \det(C - tI_{n-k})$ . Depending on  $C$ , the algebraic multiplicity  $\geq k$  which is the geometric multiplicity.  $\square$

**Theorem (Linear Independence of Eigenspaces).** Suppose  $T : V \rightarrow V$  and  $\lambda_1 \dots \lambda_k$  are distinct eigenvalues. Let  $S_i$  be a set of independent eigenvectors with eigenvalue  $\lambda_i$ . Then  $S = \bigcup_{i=1}^k S_i$  is linearly independent.

*Proof.* Let  $S_i = \{v_{i1} \dots v_{in_i}\}$  and consider  $\sum_{i=1}^k \sum_{j=1}^{n_i} a_{ij} v_{ij} = 0$ . Let  $w_i = \sum_{j=1}^{n_i} a_{ij} v_{ij}$ . Because  $E_{\lambda_i}$  is a subspace,  $w_i \in E_{\lambda_i}$ . So  $\sum_{i=1}^k w_i = 0$ . Since  $w_i$  are eigenvectors of distinct eigenvalues, they are independent. Thus  $w_i = 0$  for all  $i$ . Then  $\sum_{j=1}^{n_i} a_{ij} v_{ij} = 0$ . Since  $v_{ij}$  were independent by the setup,  $a_{ij} = 0$ .  $\square$

**Corollary.** Suppose  $T : V \rightarrow V$  with distinct eigenvalues  $\lambda_1 \dots \lambda_k$ . Assume  $f(t)$  splits.

- $T$  is diagonalizable iff Algebraic Multiplicity = Geometric Multiplicity for every  $\lambda_i$ . (In other words,  $\text{nullity}(T - \lambda_i I) = \text{AM of } \lambda_i$ .)
- $T$  is diagonalizable if and only if for  $\lambda_1, \dots, \lambda_k$  distinct, with  $\beta_i$  is a basis for  $E_{\lambda_i}$ ,  $\beta = \bigcup_{i=1}^k \beta_i$  is an eigenbasis for  $V$ .

## 5.3 Sums and Direct Sums

**Definition.** Suppose  $W_1 \dots W_k$  are subspaces of  $V$ . Their **sum** is defined as  $\sum_{i=1}^k W_i := \{\sum_{i=1}^k x_i : x_i \in W_i\}$ . The sum is a **Direct Sum**, denoted  $V = W_1 \oplus \dots \oplus W_k$ , if  $\sum W_i = V$  and  $W_j \cap (\sum_{i \neq j} W_i) = \{\vec{0}\}$ .

**Proposition.** The following statements are equivalent

1.  $V = W_1 \oplus \dots \oplus W_k$

2.  $V = \sum W_i$  and for  $v_i \in W_i$ ,  $\sum_{i=1}^k v_i = \vec{0} \implies v_i = \vec{0}$  for all  $i$ .
3.  $(\forall v \in V)$  there is a unique linear combination of  $v_i \in W_i$  summing to  $v$  (i.e.,  $v = \sum v_i$  is unique).
4. If  $\gamma_i$  is a basis for  $W_i$ , then  $\gamma = \bigcup_{i=1}^k \gamma_i$  is a basis for  $V$ .

*Proof.*

- $(1 \Rightarrow 2)$ : Assume (1). By definition,  $V = \sum W_i$ . Now suppose  $\sum_{i=1}^k v_i = \vec{0}$ . For any  $j$ , we can write  $-v_j = \sum_{i \neq j} v_i$ . The LHS is in  $W_j$ , and the RHS is in  $\sum_{i \neq j} W_i$ . Since the intersection of these sets is  $\{\vec{0}\}$  (by definition of direct sum), we must have  $-v_j = \vec{0} \implies v_j = \vec{0}$  for all  $j$ .
- $(2 \Rightarrow 3)$ : Existence follows from  $V = \sum W_i$ . For uniqueness, suppose  $v = \sum v_i = \sum v'_i$ . Then  $\sum (v_i - v'_i) = \vec{0}$ . Since  $W_i$  is a subspace,  $u_i = (v_i - v'_i) \in W_i$ . By (2),  $\sum u_i = \vec{0} \implies u_i = \vec{0} \implies v_i = v'_i$ .
- $(3 \Rightarrow 4)$ : Let  $\gamma_i$  be a basis for  $W_i$ . Any  $v \in V$  can be uniquely written as  $\sum w_i$  with  $w_i \in W_i$ . Each  $w_i$  is a linear combination of  $\gamma_i$ . Thus  $v$  is a linear combination of  $\bigcup \gamma_i$ , so  $\gamma$  spans  $V$ . Suppose a linear combination of  $\gamma$  is  $\vec{0}$ . Grouping terms from the same subspace,  $\sum_{i=1}^k (\text{lin comb of } \gamma_i) = \vec{0}$ . But each linear combination of  $\gamma_i$  corresponds to some  $y_i \in W_i$  meaning  $\sum y_i = \vec{0}$ . By the uniqueness in (3) we must have  $y_i = \vec{0}$  for all  $i$  since that combination is already known to work. Since  $\gamma_i$  is a basis,  $y_i = \vec{0}$  implies all coefficients are 0. Thus  $\gamma$  is independent.
- $(4 \Rightarrow 1)$ :  $V = \text{span}(\gamma) = \text{span}(\bigcup \gamma_i) = \sum \text{span}(\gamma_i) = \sum W_i$ . Let  $v \in W_j \cap \sum_{i \neq j} W_i$ . Since  $v \in W_j$ ,  $v$  is a linear combination of  $\gamma_j$ . Since  $v \in \sum_{i \neq j} W_i$ ,  $v$  is a linear combination of  $\bigcup_{i \neq j} \gamma_i$ . Then  $v - v = \vec{0}$  gives a linear dependence relation among disjoint basis vectors in  $\gamma$ . Since  $\gamma$  is a basis, all coefficients are 0, so  $v = \vec{0}$ . Thus the intersection is trivial.

□

**Corollary.**  $T$  is diagonalizable if and only if for distinct eigenvalues  $\lambda_i$ ,  $V = E_{\lambda_1} \oplus \cdots \oplus E_{\lambda_k}$ .

## 5.4 Invariant Subspaces and Cayley-Hamilton

**Definition (Invariant Subspace).** Suppose  $W$  is a subspace of vector space  $V$ . If  $T : V \rightarrow V$  is a linear transformation, then  $W$  is a **T-invariant subspace** if  $T(W) \subseteq W$  (i.e.,  $\forall w \in W, T(w) \in W$ ).

**Example.**  $\{\vec{0}\}$ ,  $V$ ,  $R(T)$ ,  $N(T)$ ,  $E_\lambda$ .

**Remark.** Given  $W$  a  $T$ -invariant, we may define the restriction  $T_W : W \rightarrow W$ .

**Proposition.** If  $W$  is a  $T$ -invariant subspace of finite dimensional  $V$ , then  $f_{T_W}(t)$  divides  $f_T(t)$ .

*Proof.* Suppose  $\gamma = \{w_1 \dots w_k\}$  is a basis for  $W$  and extend it to  $\beta = \{w_1 \dots w_k, v_{k+1} \dots v_n\}$  for  $V$ . Then  $[T]_\beta = \begin{bmatrix} [T_W]_\gamma & B_1 \\ 0 & B_2 \end{bmatrix}$  so  $f(t) = \det([T_W]_\gamma - tI_k) \det(B_2 - tI_{n-k}) = g(t) \det(B_2 - tI_{n-k})$ .  $\square$

**Definition.** The  $T$ -cyclic subspace generated by  $x \neq \vec{0}$  is  $W_x := \text{span}(\{x, T(x), T^2(x), \dots\})$ .

- This is  $T$ -invariant. In particular, if  $w \in W_x$ , then  $w = \sum a_i T^i(w)$ , and  $T(w) = \sum a_i T^{i+1}(w)$  which is still in  $W_x$
- Any  $T$ -invariant subspace  $W$  containing  $x$  must contain  $W_x$  (in particular, it is the smallest  $T$ -invariant subspace containing  $x$ ), since if  $x \in W$ , then  $T(x) \in W$ , and thus by induction  $T^k(x) \in W \forall k \in \mathbb{N}$ . Thus so is the span of these.

**Proposition.** Suppose  $\dim(W_x) = k$ . Then  $\{x, T(x), \dots, T^{k-1}(x)\}$  is a basis for  $W_x$ . Furthermore, if  $\sum_{i=0}^k a_i T^i(x) = 0$  with  $a_k = 1$ , then the characteristic polynomial for  $T_{W_x}$  is  $f(t) = (-1)^k (\sum_{i=0}^{k-1} a_i t^i + t^k)$ .

*Proof.* Since  $x \neq 0$ , consider the longest independent chain  $\beta = \{x, T(x), \dots, T^{j-1}(x)\}$ . Let  $Z = \text{span}(\beta)$ . Then  $T^j(x) \in \text{span}(\beta)$ . Thus for any  $w \in Z$ ,  $T(w) \in Z$ , so  $Z$  is  $T$ -invariant and thus  $W_x \subseteq Z$ . But by construction  $\beta$  is a basis for  $Z$  so must be a basis for  $W_x$ . Using this basis and assuming  $\sum_{i=0}^k a_i T^i(x) = 0$ , this means

$$[T_{W_x}]_\beta = \begin{bmatrix} 0 & 0 & \dots & -a_0 \\ 1 & 0 & \dots & -a_1 \\ \vdots & \ddots & & \vdots \\ 0 & \dots & 1 & -a_{k-1} \end{bmatrix}$$

. The characteristic polynomial follows by induction. In particular, computing the determinant  $\det([T_{W_x}]_\beta - tI)$  by cofactor expansion along the first row yields  $-t$  multiplied by the determinant of the  $(k-1) \times (k-1)$  submatrix (which by the induction hypothesis is  $(-1)^{k-1}(t^{k-1} + \dots + a_1)$ ) plus a term involving  $a_0$ . The second term is  $(-1)^{1+k}(-a_0)$  multiplied by the determinant of a triangular matrix with 1s on the diagonal, which is 1. Summing these contributions gives  $(-1)^k(t^k + a_{k-1}t^{k-1} + \dots + a_1t + a_0)$ . For the base case, we note that for a 1x1 matrix, the determinant is indeed  $(-1)^1(-a_0)$ .  $\square$

**Theorem (Cayley-Hamilton).** Suppose  $T : V \rightarrow V$  and  $f(t)$  is the characteristic polynomial. Then  $f(T) = T_0$  (the zero transformation).

*Proof.* Suppose  $v \neq 0 \in V$ . Consider  $W_v$  with  $\dim(W_v) = k$ . Consider the nontrivial equation  $\sum_{i=0}^{k-1} a_i T^i(v) + T^k(v) = 0$ . Let  $g(t) = (-1)^k (\sum_{i=0}^{k-1} a_i t^i + t^k)$ . Then  $g(T)(v) = (-1)^k (\sum a_i T^i(v) + T^k(v)) = 0$ . Since  $g(t)$  divides  $f(t)$ ,  $f(T)(v) = 0$ . But  $v$  was arbitrary meaning  $f(T) = T_0$ .  $\square$

**Theorem.** Suppose  $V = W_1 \oplus \dots \oplus W_k$  for  $T$ -invariant subspaces  $W_i$ . Let  $f_i(t)$  be the characteristic polynomial for  $T_{W_i}$ . Then,  $f_T(t) = \prod_{i=1}^k f_i(t)$ .

*Proof.* If we consider a basis for each  $W_i$  then take the union to form a basis for  $V$ , then the matrix representation of  $T$  in this basis will be block diagonal. Taking the determinant of such a matrix gives the products of the determinants of the blocks back.  $\square$

## 6 Inner Product Spaces over $\mathbb{C}$

### 6.1 Inner Products and Norms

**Definition.** An **inner product** on a vector space  $V$  over  $\mathbb{F}$  is a function  $\langle \cdot, \cdot \rangle : V^2 \rightarrow \mathbb{F}$  such that  $\forall x, y, z \in V$  and  $\forall c \in \mathbb{F}$  the following hold

1. Linearity in the first argument:  $\langle cx + y, z \rangle = c\langle x, z \rangle + \langle y, z \rangle$ .
2. Conjugate Symmetry:  $\langle x, y \rangle = \overline{\langle y, x \rangle}$ .
3. Positive Definiteness:  $\langle x, x \rangle \geq 0$  and  $\langle x, x \rangle = 0$  if and only if  $x = 0$ .

**Remark.** The inner product is linear in the first variable and conjugate linear in the second variable.

- $\langle x, cy \rangle = \bar{c}\langle x, y \rangle = \overline{\langle cy, x \rangle} = \overline{c\langle y, x \rangle}$
- $\langle x, y + z \rangle = \langle x, y \rangle + \langle x, z \rangle$ , follows by taking the conjugate twice
- $\langle x, 0 \rangle = \langle 0, x \rangle = 0$ , follows because of linearity in the first variable
- $\langle x, y \rangle = \langle x, z \rangle \forall x \in V \implies y = z$  since then  $\langle x, y - z \rangle = 0$  and we may freely choose  $x = y - z$

**Example.**

- **Standard Inner Product on  $\mathbb{F}^n$ :**  $\langle x, y \rangle = \sum_{i=1}^n x_i \bar{y}_i$  (simplifies to dot product on  $\mathbb{R}^n$ )
- **On  $C([0, 1])$  over  $\mathbb{R}$ :**  $\langle f, g \rangle = \int_0^1 f(t)g(t)dt$

**Definition (Norm).** If  $V$  is an inner product space, the **norm** is defined as  $\|x\| \equiv \sqrt{\langle x, x \rangle}$ .

**Theorem (Cauchy-Schwarz Inequality).** For all  $x, y \in V$ ,  $|\langle x, y \rangle| \leq \|x\| \cdot \|y\|$ .

*Proof.* If  $y = 0$ , trivial. Otherwise, consider  $0 \leq \|x - cy\|^2 = \langle x - cy, x - cy \rangle$ . Expanding,  $\langle x, x \rangle - \bar{c}\langle x, y \rangle - c\langle y, x \rangle + c\bar{c}\langle y, y \rangle$ . Letting  $c = \frac{\langle x, y \rangle}{\langle y, y \rangle}$ , we get  $\|x\|^2 - \frac{|\langle x, y \rangle|^2}{\|y\|^2} \geq 0$ , and we rearrange and take the square root.  $\square$

**Theorem (Triangle Inequality).**  $\|x + y\| \leq \|x\| + \|y\|$ .

*Proof.*

$$\begin{aligned}
 \|x + y\|^2 &= \langle x + y, x + y \rangle = \langle x, x \rangle + \langle x, y \rangle + \langle y, x \rangle + \langle y, y \rangle \\
 &= \|x\|^2 + \|y\|^2 + 2\operatorname{Re}(\langle x, y \rangle) \\
 &\leq \|x\|^2 + \|y\|^2 + 2|\langle x, y \rangle| \\
 &\leq \|x\|^2 + \|y\|^2 + 2\|x\| \cdot \|y\| = (\|x\| + \|y\|)^2
 \end{aligned}$$

$\square$

**Proposition.** For a linear transformation  $U : V \rightarrow V$ , if  $\langle x, Ux \rangle = 0$  for all  $x \in V$ , then  $U = T_0$ .

*Proof.* Since  $\langle v, Uv \rangle = 0$  for any vector  $v$ , we substitute arbitrary vectors  $x, y \in V$ . Consider  $\langle x + y, U(x + y) \rangle = 0$ . Expanding gives  $\langle x, Ux \rangle + \langle x, Uy \rangle + \langle y, Ux \rangle + \langle y, Uy \rangle = 0$ . The first and last terms are 0 by hypothesis, meaning  $\langle x, Uy \rangle + \langle y, Ux \rangle = 0$ . Instead, expanding  $\langle x + iy, U(x + iy) \rangle = 0$ , we get  $\langle x, Ux \rangle + \langle x, U(iy) \rangle + \langle iy, Ux \rangle + \langle iy, U(iy) \rangle = 0$ , in which case  $\langle x, Uy \rangle - \langle y, Ux \rangle = 0$ . Adding the two results,  $(\langle x, Uy \rangle + \langle y, Ux \rangle) + (\langle x, Uy \rangle - \langle y, Ux \rangle) = 0$ , and by cancellation  $\langle x, Uy \rangle = 0$ . Since this holds for all  $x$ , we can choose  $x = Uy$ , which yields  $\langle Uy, Uy \rangle = \|Uy\|^2 = 0$ . Therefore,  $Uy = 0$  for all  $y$ , so  $U = 0$ .  $\square$

**Remark.** The above is not true for inner product spaces over  $\mathbb{R}$ . For example, consider a linear transformation that rotates the plane by 90 degrees.

## 6.2 Orthogonality

**Definition.** Two vectors  $x, y$  are said to be **orthogonal** if  $\langle x, y \rangle = 0$ . This generalization comes from the fact that  $\frac{|\langle x, y \rangle|}{\|x\| \cdot \|y\|} \leq 1$ , matching the range of  $\cos(\theta)$ .

**Definition.** A subset  $S \subseteq V$  is **orthogonal** if any two distinct vectors are orthogonal. It is **orthonormal** if it is orthogonal and all elements have a norm of 1.

**Remark (Coordinate Representation in Orthogonal Basis).** Suppose  $S = \{s_1, \dots, s_k\}$  is an orthogonal subset of nonzero vectors. Then for any  $y \in \text{span}(S)$ :

$$y = \sum_{i=1}^k \frac{\langle y, s_i \rangle}{\langle s_i, s_i \rangle} s_i$$

*Proof.* Write  $y = \sum_{i=1}^k a_i s_i$ . Taking the inner product with  $s_j$ ,

$$\langle y, s_j \rangle = \langle \sum a_i s_i, s_j \rangle = \sum a_i \langle s_i, s_j \rangle = a_j \langle s_j, s_j \rangle$$

since  $\langle s_i, s_j \rangle = 0$  for  $i \neq j$ . Thus  $a_j = \frac{\langle y, s_j \rangle}{\langle s_j, s_j \rangle}$ .  $\square$

**Corollary.** An orthogonal set of nonzero vectors is linearly independent.

*Proof.* If we decompose the  $\vec{0}$  vector, we get all coefficients to be 0.  $\square$

**Corollary.** If  $\beta = \{v_1, \dots, v_n\}$  is an orthonormal basis for  $V$ , then we may express any  $x \in V$  as  $x = \sum_{j=1}^n \langle x, v_j \rangle v_j$ . The scalars  $\langle x, v_j \rangle$  are called the **Fourier coefficients** of  $x$  relative to  $\beta$ .

**Theorem (Gram-Schmidt Orthogonalization Algorithm).** Let  $S = \{w_1, \dots, w_n\}$  be a linearly independent subset. We define  $S' = \{v_1, \dots, v_n\}$  as follows

$$\begin{cases} v_1 = w_1 \\ v_k = w_k - \sum_{j=1}^{k-1} \frac{\langle w_k, v_j \rangle}{\langle v_j, v_j \rangle} v_j \text{ for } 2 \leq k \leq n \end{cases}.$$

Then  $S'$  is an orthogonal set and  $\text{span}(S') = \text{span}(S)$ . We may then normalize these vectors to obtain an orthonormal set. (Note if we had started with a linearly dependent set, some of the vectors would have been changed into  $0_V$ , in which case we would have thrown them out.)

### 6.3 Orthogonal Complements and Projections

**Definition (Orthogonal Complement).** Suppose  $S \subseteq V, S \neq \emptyset$ . The orthogonal complement is defined as:

$$S^\perp = \{x \in V : (\forall y \in S) \langle x, y \rangle = 0\}$$

**Remark.**  $S^\perp$  is always a subspace of  $V$ .

*Proof.*  $0_V \in S^\perp$ . If  $x, y \in S^\perp, a \in \mathbb{C}, z \in S$ , then  $\langle z, ax + y \rangle = \bar{a}\langle x, y \rangle + \langle x, z \rangle = \bar{a}0 + 0 = 0$   $\square$

**Definition.** If  $W$  is a subspace of an inner product space  $V$ , given  $y \in V$ , the **orthogonal projection** is  $u = \sum_{i=1}^k \langle y, \beta_i \rangle \beta_i$  where  $\{\beta_1, \dots, \beta_k\}$  is an orthonormal basis for  $W$ .

**Remark.**  $y - u \in W^\perp$ .

*Proof.* By definition, the orthogonal projection  $u$  is given by  $u = \sum_{i=1}^k \langle y, \beta_i \rangle \beta_i$  where  $\{\beta_i\}$  is an orthonormal basis for  $W$ . We check the inner product of  $y - u$  with any basis vector  $\beta_j$

$$\begin{aligned} \langle y - u, \beta_j \rangle &= \langle y, \beta_j \rangle - \left\langle \sum_{i=1}^k \langle y, \beta_i \rangle \beta_i, \beta_j \right\rangle \\ &= \langle y, \beta_j \rangle - \sum_{i=1}^k \langle y, \beta_i \rangle \delta_{ij} = \langle y, \beta_j \rangle - \langle y, \beta_j \rangle = 0 \end{aligned}$$

$\square$

**Proposition.** The orthogonal projection is the vector that minimizes the norm to the subspace. In particular  $\|y - u\| \leq \|y - w\|$  for all  $w \in W$ .

*Proof.* Let  $u$  be the orthogonal projection of  $y$  onto  $W$ . Then  $y - u \in W^\perp$ . For any vector  $w \in W$ , we can write  $y - w = (y - u) + (u - w)$ . Notice that  $(u - w) \in W$  because  $W$  is a subspace. Since  $(y - u) \perp (u - w)$ ,  $\|y - w\|^2 = \|(y - u) + (u - w)\|^2 = \|y - u\|^2 + \|u - w\|^2$ . Since  $\|u - w\|^2 \geq 0$ , it follows that  $\|y - w\|^2 \geq \|y - u\|^2$ , and thus  $\|y - w\| \geq \|y - u\|$ . The minimum is achieved when  $\|u - w\| = 0$ , meaning  $u = w$ .  $\square$

**Theorem (Orthogonal Decomposition).** If  $W$  is a subspace of  $V$ , then  $V = W \oplus W^\perp$ . In particular, any  $y \in V$  can be uniquely written as  $y = x + z$  where  $x \in W$  and  $z \in W^\perp$ . As a result,  $\dim(W) + \dim(W^\perp) = \dim(V)$ .

*Proof.*  $V = W + W^\perp$  since for any vector  $v \in V$ , we can write  $v = u + (v - u)$  where  $u$  is the orthogonal projection on to  $W$ . For the intersection, suppose  $v \in W \cap W^\perp$ . Since  $v \in W^\perp$ ,

it is orthogonal to every vector in  $W$ . Since  $v \in W$ , it must be orthogonal to itself. Thus  $\langle v, v \rangle = 0$ , which implies  $v = 0$  by the positive definiteness of the inner product. Therefore, the sum is direct.  $\square$

## 6.4 Adjoint of a Linear Operator

**Theorem (Riesz Representation).** Suppose  $V$  is a finite-dimensional inner product space over  $\mathbb{F}$ . Let  $g : V \rightarrow \mathbb{F}$  be a linear functional. Then there exists a unique  $y \in V$  such that for all  $x \in V$ ,  $g(x) = \langle x, y \rangle$ .

*Proof.* Let  $\beta = \{v_1, \dots, v_n\}$  be an orthonormal basis. Define  $y = \sum_{i=1}^n \overline{g(v_i)} v_i$ . Then,  $\langle v_j, y \rangle = \sum_{i=1}^n g(v_i) \langle v_j, v_i \rangle = \sum g(v_i) \delta_{ji} = g(v_j)$ . Since this holds for the basis, it holds for all  $x \in V$  by linearity.  $\square$

**Definition.** Suppose  $T : V \rightarrow V$  is a linear operator on a finite-dimensional inner product space. There exists a unique linear operator  $T^* : V \rightarrow V$  that is called the **adjoint** of  $T$  such that  $(\forall x, y \in V) \langle T(x), y \rangle = \langle x, T^*(y) \rangle$

*Proof.* Existence: Fix  $y$ . The map  $h(x) = \langle T(x), y \rangle$  is a linear functional on  $V$ . By Riesz Representation, there is a unique  $y' \in V$  such that  $\langle T(x), y \rangle = h(x) = \langle x, y' \rangle$  for all  $x$ . Define  $T^*(y) = y'$ . We now show this is linear. Namely, for any  $x \in V$ ,

$$\begin{aligned} \langle x, T^*(cy_1 + y_2) \rangle &= \langle T(x), cy_1 + y_2 \rangle = \bar{c} \langle T(x), y_1 \rangle + \langle T(x), y_2 \rangle \\ &= \bar{c} \langle x, T^*(y_1) \rangle + \langle x, T^*(y_2) \rangle = \langle x, cT^*(y_1) + T^*(y_2) \rangle \end{aligned}$$

but since  $x$  is arbitrary,  $T^*(cy_1 + y_2) = cT^*(y_1) + T^*(y_2)$ .  $\square$

**Proposition.** Given an orthonormal basis  $\beta$ ,  $[T^*]_\beta = ([T]_\beta)^*$ , where  $A_{ij}^* = \overline{A_{ji}}$ . As a result,  $A^*$  is called the **adjoint** of a matrix. Since the matrices are isomorphic to linear transformations, we see immediately that  $(cT + U)^* = \bar{c}T^* + U^*$ ,  $(TU)^* = U^*T^*$ , and  $T^{**} = T$ .

*Proof.* Let  $\beta = \{v_1, \dots, v_n\}$  be an orthonormal basis and let  $A = [T]_\beta$ . By the definition of matrix representation with respect to an orthonormal basis, the entry  $A_{ij}$  is the coefficient of  $v_i$  in  $T(v_j)$ , which is given by  $A_{ij} = \langle T(v_j), v_i \rangle$ . Let  $B = [T^*]_\beta$ . Similarly, its entries are given by  $B_{ij} = \langle T^*(v_j), v_i \rangle$ . Then it follows that  $B_{ij} = \overline{A_{ji}}$ .  $\square$

## 6.5 Unitary, Normal, and Self-Adjoint Operators

**Definition.** An operator  $T : V \rightarrow V$  is **unitary** (over  $\mathbb{C}$ ) or **orthogonal** (over  $\mathbb{R}$ ) if  $\|T(x)\| = \|x\|$  for all  $x \in V$ . Then we note that all eigenvalues of a unitary operator  $|\lambda| = 1$ .

**Proposition.** The following statements are equivalent:

1.  $T$  is unitary
2.  $TT^* = T^*T = I$  (i.e.,  $T^* = T^{-1}$ )

3.  $\langle T(x), T(y) \rangle = \langle x, y \rangle$  (preserves inner products)
4.  $T$  maps an orthonormal basis to an orthonormal basis.

*Proof.*

- (1  $\implies$  2): Squaring both sides gives  $\langle T(x), T(x) \rangle = \langle x, x \rangle$ . Using the definition of the adjoint, we get  $\langle x, T^*T(x) \rangle = \langle x, x \rangle$ , which implies  $\langle x, (T^*T - I)x \rangle = 0$  for all  $x$ . Let  $U = T^*T - I$ . We have previously shown that  $\langle x, Ux \rangle = 0$  ( $\forall x \in V$ ) means  $U = T_0$  meaning  $T^*T = I$ . By symmetry,  $TT^* = I$  meaning  $T^{-1} = T^*$ .
- (2  $\implies$  3): For any  $x, y \in V$ ,  $\langle T(x), T(y) \rangle = \langle x, T^*T(y) \rangle = \langle x, I(y) \rangle = \langle x, y \rangle$
- (3  $\implies$  4): Let  $\beta = \{v_1, \dots, v_n\}$  be an orthonormal basis. Taking the inner product of the image,  $\langle T(v_i), T(v_j) \rangle = \langle v_i, v_j \rangle = \delta_{ij}$ , meaning  $\{T(v_1), \dots, T(v_n)\}$  is an orthonormal set of  $n$  vectors, it is a basis.
- (4  $\implies$  1): Let  $\{v_1, \dots, v_n\}$  be an orthonormal basis. For  $x \in V$ , we can write  $x = \sum_{i=1}^n c_i v_i$ . Then,  $\|x\|^2 = \langle \sum c_i v_i, \sum c_j v_j \rangle = \sum |c_i|^2$ . Then  $T(x) = \sum_{i=1}^n c_i T(v_i)$  and  $\|T(x)\|^2 = \langle \sum c_i T(v_i), \sum c_j T(v_j) \rangle = \sum |c_i|^2$  since  $\{T(v_1), \dots, T(v_n)\}$  is orthonormal by assumption.

□

**Definition.** A linear operator  $T$  (or matrix  $A$ ) is **normal** if  $TT^* = T^*T$  (or  $AA^* = A^*A$ ).

**Proposition.**

- $\|T(x)\| = \|T^*(x)\|$  for all  $x \in V$ .

*Proof.*  $\|T(x)\|^2 = \langle T(x), T(x) \rangle = \langle T^*T(x), x \rangle = \langle TT^*(x), x \rangle = \langle T^*(x), T^*(x) \rangle = \|T^*(x)\|^2$ . □

- If  $Tx = \lambda x$ , then  $T^*x = \bar{\lambda}x$ .

*Proof.*  $\|(T - \lambda I)x\| = 0 = \|(T - \lambda I)^*x\| = 0 \implies \|(T^* - \bar{\lambda}I)x\| = 0$ . □

- Eigenvectors of distinct eigenvalues are orthogonal.

*Proof.* Let  $Tx_1 = \lambda_1 x_1$  and  $Tx_2 = \lambda_2 x_2$ . Then, by the previous result,  $\lambda_1 \langle x_1, x_2 \rangle = \langle T(x_1), x_2 \rangle = \langle x_1, T^*(x_2) \rangle = \langle x_1, \bar{\lambda}_2 x_2 \rangle = \bar{\lambda}_2 \langle x_1, x_2 \rangle = \lambda_2 \langle x_1, x_2 \rangle$ . Thus  $(\lambda_1 - \lambda_2) \langle x_1, x_2 \rangle = 0$ . Since  $\lambda_1 \neq \lambda_2$ ,  $\langle x_1, x_2 \rangle = 0$ . □

**Theorem (Spectral Theorem for Normal Operators).**  $T$  is normal iff there exists an orthonormal basis of  $V$  consisting of eigenvectors of  $T$ .

*Proof.*

- (  $\implies$  ) We proceed by induction on the dimension  $n$  of  $V$ . For  $n = 1$ , any operator on a 1-dimensional space has an eigenvector (Fundamental Theorem of Algebra). We can normalize it to length 1. This forms an orthonormal basis. Assume the theorem holds for all normal operators on complex inner product spaces of dimension  $n - 1$ . Now suppose  $\dim(V) = n$ . Since we are working over  $\mathbb{C}$ , the characteristic polynomial of  $T$  splits. There exists at least one eigenvalue  $\lambda_1$  and a corresponding unit eigenvector  $u_1$ . Let  $W = \text{span}\{u_1\}^\perp$ , then the dimension of  $W$  is  $n - 1$ . We show that  $W$  is invariant under  $T$ . Let  $w \in W$ . By definition,  $\langle w, u_1 \rangle = 0$ . Then  $\langle Tw, u_1 \rangle = \langle w, T^*u_1 \rangle = \langle w, \bar{\lambda}_1 u_1 \rangle = \bar{\lambda}_1 \langle w, u_1 \rangle = 0$ . Thus  $Tw$  is orthogonal to  $u_1$ , so  $Tw \in W$  as desired. By the same reasoning  $W$  is also  $T^*$ -invariant since  $T^*$  and  $T$  have the same eigenvectors. We may then define the restriction  $T|_W : W \rightarrow W$ , and similarly  $T^*|_W$  which are also normal. By the inductive hypothesis, there exists an orthonormal basis  $\{u_2, \dots, u_n\}$  for  $W$  consisting of eigenvectors of  $T|_W$ . Combining  $u_1$  with  $\{u_2, \dots, u_n\}$  gives an orthonormal basis for the entire space  $V$  consisting of eigenvectors of  $T$ .
- (  $\impliedby$  ) Suppose there exists an orthonormal basis  $\beta = \{u_1, \dots, u_n\}$  of  $V$  consisting of eigenvectors of  $T$ . Let  $\lambda_1, \dots, \lambda_n$  be the corresponding eigenvalues. The matrix representation of  $T$  with respect to this basis, denoted as  $M = [T]_\beta$ , is diagonal. And  $[T^*]_\beta = M^*$  is also diagonal. We know that diagonal matrices always commute. Thus the corresponding operators also commute.

□

**Corollary.** A matrix  $A$  is normal iff it is unitarily equivalent to a diagonal matrix ( $A = PDP^*$  where  $P$  is unitary).

**Definition.**  $T$  is **self-adjoint** (or **Hermitian**) if  $T = T^*$  (for matrices,  $A = A^*$ ). In particular Hermitian operators are a subset of normal operators.

**Remark.** All eigenvalues of a Hermitian operator are **real**.

**Remark.** A real symmetric matrix is Hermitian.

## 6.6 Singular Value Decomposition

**Theorem (Singular Value Decomposition).** For any real matrix  $A \in \mathbb{R}^{m \times n}$ , there is a way to write  $A = U\Sigma V^T$  where  $U \in \mathbb{R}^{m \times m}$  and  $V \in \mathbb{R}^{n \times n}$  are orthogonal and  $\Sigma \in \mathbb{R}^{m \times n}$  is rectangular diagonal with nonnegative entries.

*Proof.*

- Consider the matrix  $A^T A$ . It is an  $n \times n$  symmetric matrix. By the Spectral Theorem, we can orthogonally diagonalize it into  $A^T A = V\Lambda V^T$ . We note that for any  $x$ ,  $x^T(A^T A)x = (Ax)^T(Ax) = \|Ax\|^2 \geq 0$ , meaning  $\lambda_i \geq 0$ .
- We rearrange the eigenvalues so that  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r > 0$  and  $\lambda_{r+1} = \dots = \lambda_n = 0$  (may not be distinct). We define  $\sigma_i = \sqrt{\lambda_i}$  meaning  $\sigma_1 \geq \dots \geq \sigma_r > 0$  and the rest are 0.

- For  $i = 1, \dots, r$ , we will define  $u_i = \frac{1}{\sigma_i} Av_i$ . These vectors are orthonormal since  $u_i^T u_j = \left(\frac{1}{\sigma_i} Av_i\right)^T \left(\frac{1}{\sigma_j} Av_j\right) = \frac{1}{\sigma_i \sigma_j} v_i^T (A^T A) v_j = \frac{\sigma_j^2}{\sigma_i \sigma_j} (v_i^T v_j) = \delta_{ij}$  since the  $v_i$  are orthonormal by construction.
- We have  $r$  orthonormal vectors  $\{u_1, \dots, u_r\}$ . If  $m > r$ , we use the Gram-Schmidt process (or basis extension theorem) to extend this set to a full orthonormal basis for  $\mathbb{R}^m$ . Let  $U = [u_1 | \dots | u_m]$ . (The choice of extension does not matter.)
- We want to show  $A = U \Sigma V^T$ , equivalently  $AV = U \Sigma$ . For columns  $i = 1$  to  $r$ ,  $(AV)_i = Av_i$  and  $(U \Sigma)_i = \sigma_i u_i$ . But we defined  $u_i = \frac{1}{\sigma_i} Av_i$ . For columns  $i = r + 1$  to  $n$ , these correspond to the zero eigenvalues. If  $\lambda_i = 0$ , then  $A^T Av_i = 0$ . Then, the LHS is  $\|Av_i\|^2 = v_i^T A^T Av_i = v_i^T (0) = 0$ . Thus  $(AV)_i = Av_i = 0$ . For the RHS,  $\sigma_i = 0$ , so  $(U \Sigma)_i = 0 \cdot u_i = 0$ .

□

**Corollary.**  $A = \sum_{i=1}^r \sigma_i u_i v_i^T$  where each matrix in the sum has rank 1.

**Remark.** The columns of  $U$  are called **left singular vectors** and the columns of  $V$  are called **right singular vectors**. The entries of  $\Sigma$  are called the **singular values**. The SVD describes the geometry of linear transformations in three steps

1.  $V^T$  (Rotation/Reflection): Rotate the space  $\mathbb{R}^n$  to align the basis with the principal axes.
2.  $\Sigma$  (Scaling): Stretch or shrink along these axes by the factors  $\sigma_i$ . If  $m \neq n$ , this step also projects from dimension  $n$  to  $m$  (adding or removing zeros).
3.  $U$  (Rotation/Reflection): Rotate the result in the codomain space  $\mathbb{R}^m$  to the final orientation.

**Remark.** The SVD has the following properties for  $A$

- Row Space: Spanned by first  $r$  columns of  $V$  ( $v_1, \dots, v_r$ ).
- Null Space: Spanned by the last  $n - r$  columns of  $V$  ( $v_{r+1}, \dots, v_n$ ).
- Column Space (Image): Spanned by first  $r$  columns of  $U$  ( $u_1, \dots, u_r$ ).
- Column Space Complement: Spanned by the last  $m - r$  columns of  $U$  ( $u_{r+1}, \dots, u_m$ ).

## 7 Jordan Canonical Form

### 7.1 Motivation

Recall that if the characteristic polynomial of a linear transformation  $T$  splits and the algebraic multiplicity is equal to the geometric multiplicity for every eigenvalue, then  $T$  is diagonalizable.

When the algebraic multiplicity exceeds the geometric multiplicity, the matrix is not diagonalizable. To handle this, we introduce **Generalized Eigenvectors** to form the Jordan Canonical Form.

### 7.2 Generalized Eigenvectors & Eigenspaces

**Definition.** Given a vector space  $V$  and a linear transformation  $T$ , a non-zero vector  $x$  is called a **generalized eigenvector** corresponding to eigenvalue  $\lambda$  if for some  $p \in \mathbb{N}$ ,  $(T - \lambda I)^p x = 0$ . The **generalized eigenspace**  $K_\lambda$  is the  $T$ -invariant subspace containing all the generalized eigenvectors corresponding to  $\lambda$ .

**Theorem.** The generalized eigenspace has the following properties:

1. The dimension of the generalized eigenspace  $K_\lambda$  equals the algebraic multiplicity of  $\lambda$ .
2.  $K_\lambda = \text{Nul}((T - \lambda I)^m)$  where  $m$  is the algebraic multiplicity.
3. If we find a basis for each generalized eigenspace, their union is a basis for  $V$ .

### 7.3 Jordan Canonical Basis Construction

If  $\dim(E_\lambda) < \dim(K_\lambda) = m$  (where  $E_\lambda$  is the standard eigenspace), we must construct a basis for  $K_\lambda$  using **Jordan Chains**.

1. Start with an eigenvector  $v_k$  such that  $(A - \lambda I)v_k = 0$
2. Solve for the next vector in the chain,  $v_{k+1}$ , using the equation

$$(A - \lambda I)v_{k+1} = v_k$$

which implies  $(A - \lambda I)^2 v_{k+1} = 0$

3. Continue this process of finding linearly independent solutions until a basis for  $K_\lambda$  is found
4. Rearranging the chain equation, we find that  $Av_{k+1} = v_k + \lambda v_{k+1}$

## 7.4 The Jordan Matrix Form

Using this basis,  $A$  is similar to a matrix  $J$  in Jordan Form. The matrix  $J$  is block diagonal, composed of Jordan Blocks corresponding to the chains constructed above.

$$J = \begin{bmatrix} \lambda_1 & 1 & 0 & \dots \\ 0 & \lambda_1 & 1 & \dots \\ 0 & 0 & \ddots & \ddots \\ \dots & \dots & 0 & \lambda_k \end{bmatrix}$$

**Remark.** The matrix  $J$  is upper triangular and is **uniquely determined** from  $A$  up to the order of permuting the eigenvalues (blocks).