Disclaimer: this is a personal note compiled by D.Dai based on lectures of MATH 245, 2019, Spring with Professor ***B.Madill***. This note is only meant to be a complementary material rather than a fully comprehensive and self-contained notes (it has ***a lot of typos*** and I almost omitted the entire section about Exterior ALgebra.). Thus, you should always cross-referencing with your own notes.

This note ***cannot be used for monetized purposes in any sense***. This note should only be used for personal purposes. It is consented to post this note online with this first page intact and attached to the entire note if no legal dispute will occur in the possible future.

# Contents

# First day

**Definition 0.0.1** (Professor)**.** Blake Madill, MC 5423

**Definition 0.0.2** (Office Hours)**.** M,W 2-3 afternoon or by appt

**Definition 0.0.3** (Outline)**.**

1. Ring Theory (2 Weeks)
2. Jordan Canonical Form (2-3 Weeks)
3. Inner Product Spaces (A long time)
4. Additional Topics

**Remark 0.0.4** (Textbook)**.** The same from MATH 146

**Remark 0.0.5.**

1. Assignments: 15%
2. Midterms: 30%
3. Final: 55%

**Remark 0.0.6** (Midterm)**.** June 11, 4:30-6:20 p.m., RCH 305.

Five questions.

Q1: Short answer, 4 parts. Example: find minimal polynomial.

Q2: Find the JCF, 4 parts. It may be more than one possibility.

Q3: One proof (8 marks) and one extra (2 marks).

Q4: Proof, out of 10 marks.

Q5: Computation.

Two basic questions, two proofs from lecture (every proof is fair game), and one hard question. It covers up to the end of Jordan Canonical Form.

# Chapter 1

# Ring Theory

## 1.1 Only The Basics

**Definition 1.1.1.** A ring is a set $R$ with two operations, $+$ and $\times$, such that

1. (Associativity, $+$) For all $a, b, c \in R$, $a + (b + c) = (a + b) + c$
2. (Assoc, $\times$) $(ab)c = a(bc)$
3. (Commutativity, $+$) $\forall a, b \in R$, $a + b = b + a$
4. (Additive identity) $\exists 0 \in R$ such that $0 + a = a + 0 = a$ for all $a \in R$
5. (Additive Inverse) $\forall a \in R$, $\exists b \in R$ such that $a + b = 0$, we write $b = -a$
6. (Distributivity) $\forall a, b, c \in R$, $a(b + c) = ab + ac$

**Remark 1.1.2.** We will not insist on the existence of multiplicative identity, as this is a special type of ring

**Definition 1.1.3.** Some notations:

1. $+(a, b) \equiv a + b$
2. $\times(a, b) \equiv a \cdot b \equiv ab$
3. By assoc., $a + b + c$ and $abc$ are well-defined
4. For $n \in \mathbb{N}$, $a^n = \prod_{i=1}^n a$ and $na = \sum_{i=1}^n a$
5. $a + (-b) \equiv a - b$

**Remark 1.1.4.** $\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$ are rings with normal operations

**Example 1.1.5.** If $R, R_1, R_2$ is a ring, we can form the following rings:

1. $R[x] = \{\sum_{i=0}^n a_i x^i : a_i \in R, i \in \mathbb{N} \cup \{0\}\}$
2. $M_n(R) \to n \times n$, the $R-$matrices
3. $R_1 \oplus R_2 = \{(a, b) : a \in R_1, b \in R_2\}$ forms a ring component wise

**Example 1.1.6.** The following are not rings

1. $A = \{2k + 1, k \in \mathbb{Z}\}$, missing 0
2. $B = \mathbb{N} \cap \{0\}$ is not a ring, we do not have inverses except 0
3. $C(\mathbb{R}) = \{f : \mathbb{R} \to \mathbb{R} : \text{f is continuous}\}$ is not a ring, it is not distributive.

**Definition 1.1.7.** Let $R$ be a ring.

1. We say a ring $R$ is ***commutative*** if $ab = ba$ for all $a, b \in R$.
2. We say a ring $R$ is ***unital*** if $\exists 1 \in R$ such that $1a = a1 = a$ for all $a \in R$.

**Example 1.1.8.**

1. $2\mathbb{Z} = \{2k : k \in \mathbb{Z}\}$ is non-unital and commutative
2. $M_n(2\mathbb{Z})$ is non-unital and non-commutative

**Definition 1.1.9.** We say $R = \{0\}$ is a ring, called ***trivial ring***

**Remark 1.1.10.** In this class, only the trivial ring is allowed to have the trivial multiplication, i.e. $ab = 0$ for all $a, b \in R$. Thus, $(\mathbb{Z}, +, \star)$ with $a \star b = 0$ for all $a, b \in \mathbb{Z}$, is not a ring.

Moreover, we do not regard the trivial ring as being unital.

**Definition 1.1.11.** Let $R$ be a commutative ring, we say $a \in R$ is a ***zero divisor*** if $a \neq 0$ and there exists $b \neq 0$ such that $ab = 0$

**Example 1.1.12.** $a \in \mathbb{Z}_n$, then $a$ is a zero divisor iff $gcd(a, n) \neq 1$ and $a \neq 0$

**Definition 1.1.13.** A ring $R$ is an ***integral domain*** if $R$ is commutative, unital, and has no zero divisors

**Example 1.1.14.**

1. $\mathbb{Z}, \mathbb{Z}_p, \mathbb{Q}, \mathbb{R}, \mathbb{C}$, and $R[x]$ with $R$ is above are all integral domains.
2. The following are not integral domains: $\mathbb{Z}_n$, $n$ not a prime; $M_n(\mathbb{R})$ (not commutative); $2\mathbb{Z}$ (not unital); $\mathbb{R} \oplus \mathbb{R}$, as it contains zero divisor.

**Proposition 1.1.15.** *Let $R$ be commutative, and $a$ is not a zero divisor. If $a, b, c \in R$, $a \neq 0$ and $ab = ac$ then $b = c$*

*Proof.* Suppose $ab = ac$

$$ac - ab = 0$$

$$a(c - b) = 0$$

Since $a \neq 0$ and a is not a zero divisor, we have $c - b = 0$, hence $b = c$.

Note if $R$ is an integral domain, we can say this to all $x \in R$. ♡

**Definition 1.1.16.** Let $R$ be commutative, unital, we say $a \in R$ is a ***unit*** or ***invertible*** if $\exists b \in R$ such that $ab = 1$, we call $b$ the ***inverse*** of $a$ and write $b = a^{-1}$.

Moreover, we denote the group of units of $R$ by $R^\times$ or $U(R)$

**Example 1.1.17.** We have:

1. $U(\mathbb{Z}_p) = \mathbb{Z}_p^\times = \mathbb{Z}_p \backslash \{0\}$
2. $\mathbb{Z}_n^\times = \{a \in \mathbb{Z}_n : gcd(a, n) = 1\}$

**Remark 1.1.18.** If $a \in R^\times$ then $a$ is not a zero divisor. This is because $a \neq 0$

**Definition 1.1.19.** A ring $F$ is a field if $F$ is commutative, unital, and every non-zero element of $F$ is a unit.

**Example 1.1.20.** The following are fields: $\mathbb{Z}_p, \mathbb{Q}, \mathbb{R}, \mathbb{C}, \mathbb{Q}(\sqrt{2})$ and

$$F(x) = \{\frac{f(x)}{g(x)} : f(x), g(x) \in F[x], g(x) \neq 0\}$$

In addition, $Z$ and $F[x]$ are integral domains, but not fields.

**Proposition 1.1.21.** *Every field is an integral domain*

**Definition 1.1.22.** Let $R$ is unital, we define the ***characteristic*** of $R$ to be the least positive integer $n$ such that $n = 0$ in $R$, i.e. $n = n \cdot 1 = 0$. If no such $n$ exists, we say $R$ has characteristic zero.

In addition, we write $char(R) = n$ or $char(R) = 0$

**Example 1.1.23.** We have:

1. $char(\mathbb{Z}_4[x]) = 4$
2. $char(R) = 0$ means the set $R$ is infinite.

**Proposition 1.1.24.** *If $R$ is an integral domain, then $char(R)$ is $0$ or prime*

*Proof.* If $char(R) = 0$, we are done.

Say $char(R) = n \neq 0$, suppose for contradiction, let $n = ab$ where $a, b < n$. However, $n$ is the least element such that $n \cdot 1 = 0$, so $a \cdot 1, b \cdot 1 \neq 0$, so $ab \in R \neq 0$ as $R$ is an integral domain. ♡

**Example 1.1.25.** $char(\mathbb{Z}_p(x)) = p$ where $\mathbb{Z}_p(x)$ is rational functions over $\mathbb{Z}_p$

**Definition 1.1.26.** Let $(R, +, \cdot)$ be a ring, we say $S \subseteq R$ is a ***subring*** of $R$ if $(S, +, \cdot)$ forms a ring.

**Example 1.1.27.** $\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{Q}(\sqrt{2})$ are all subrings of $\mathbb{C}$

**Remark 1.1.28.** Let $R$ be a ring, and $S \neq \emptyset \subseteq R$. Then $S$ must have associativity of additions and multiplication, and Commutativity of addition, and distributivity.

To show $S$ is a subring of $R$, we must show:

1. Closed under operations
2. $0 \in S$
3. $0 \in S$ implies $-a \in S$

**Proposition 1.1.29** (Subring test)**.** *Let $R$ be a ring and let $S$ be a nonempty subset of $R$. Then $S$ is a subring of $R$ if and only if*

1. *for all $a, b \in S$, $a - b \in S$*
2. *for all $a, b \in S$, $ab \in S$*

*Proof.* Easy to see from Remark 1.1.28 where $a - b \in S$ covers all the things needed for addition ♡

**Example 1.1.30.** $\mathbb{Q}(\sqrt{2})$ is a subfield of $\mathbb{R}$

*Solution.* Let $a + b\sqrt{2}$ and $c + d\sqrt{2}$ be in $\mathbb{Q}(\sqrt{2})$. Then

$$(a + b\sqrt{2}) - (c + d\sqrt{2}) = (a - c) + (b - d)\sqrt{2} \in \mathbb{Q}(\sqrt{2})$$

Moreover, $(a + b\sqrt{2})(c + d\sqrt{2}) = (ac + 2bd) + (ad + bc)\sqrt{2} \in \mathbb{Q}(\sqrt{2})$. Thus $\mathbb{Q}(\sqrt{2})$ is a subring of $\mathbb{R}$.

Since $\mathbb{Q}(\sqrt{2}) \subseteq \mathbb{R}$ so it is commutative and unital as $1 = 1 + 0\sqrt{2} \in \mathbb{Q}(\sqrt{2})$. Take $0 \neq a + b\sqrt{2} \in \mathbb{Q}(\sqrt{2})$, since $\sqrt{2} \notin \mathbb{Q}$, we have $a^2 - 2b^2 \neq 0$ so

$$(a + b\sqrt{2})^{-1} = \frac{a}{a^2 - 2b^2} + \frac{-b}{a^2 - 2b^2}\sqrt{2} \in \mathbb{Q}(\sqrt{2})$$

Hence $\mathbb{Q}(\sqrt{2})$ is a subfield of $\mathbb{R}$ ♠

**Definition 1.1.31.** Let $R$ be a ring, a subring $I$ of $R$ is an ***ideal*** of $R$ iff $\forall a \in I, \forall r \in R$, $ar, ra \in I$

**Example 1.1.32.** $n\mathbb{Z} = \{nk : k \in \mathbb{Z}\}$ is an ideal of $\mathbb{Z}$.

Let $R = C(\mathbb{R})$ be the set of continuous functions over $\mathbb{R}$, let $I = \{f \in R : f(2) = 0\}$, then $I$ is an ideal of $C(\mathbb{R})$.

Let $R = \mathbb{C}$, then $\mathbb{R}$ is a subring and it is not an ideal of $\mathbb{C}$.

**Example 1.1.33.** Let $F$ be a field, then the only ideals we have are: $\{0\}$ and $F$.

Lets say we take a non-zero ideal of $F$, say $I$, suppose we have $0 \neq a \in I$, then $aa^{-1}$ must be in $I$ and hence $1 \in I$. Therefore $I$ must contain the entire field

**Example 1.1.34.** Let $\mathbb{F}$ ve a field and let $V$ be finite dimensional $\mathbb{F}$-vector space. Let $T : V \to V$ be linear operator. If $f(x) = \sum_{i=0}^{n} a_i x^i \in F[x]$ then we define $f(T) = \sum_{i=0}^{n} a_i T^i$ where $T^0 = id$ is the identity mapping. Prove that $I = \{f(x) \in \mathbb{F}[x] : f(T) = 0\}$ is an ideal of $\mathbb{F}[x]$.

*Solution.* Let $I = \{f(x) \in \mathbb{F}[x] : f(T) = 0\}$. Let $f(x), g(x) \in I$ so that $f(T) = g(T) = 0$. We note, for all $v \in V$, $(f(T) - g(T))(v) = f(T)(v) - g(T)(v) = 0 - 0 = 0$. Moreover, if $f(x) \in I$ and $h(x) \in \mathbb{F}[x]$, then for all $v \in V$, we have $(f(T) \circ h(T))(v) = f(T)(h(T)(v)) = 0$, thus $f(x)h(x) \in I$ and so $I$ is an ideal of $\mathbb{F}[x]$. ♠

**Definition 1.1.35.** Let $R$ be commutative and unital, and let $x \in R$, then the ***principal ideal of R generated by x*** is

$$\langle x \rangle = \{rx : r \in R\}$$

**Proposition 1.1.36** (Division Algorithm)**.** *Let $F$ be a field. For all $f(x), g(x) \in F[x]$, $g(x) \neq 0$, there exists unique $q(x), r(x) \in F[x]$, such that $f(x) = g(x)q(x) + r(x)$ where $r(x) = 0$ or $deg(r) < deg(g)$*

*Proof.* Easy ♡

**Proposition 1.1.37.** *Let $F$ be a field, every ideal of $F[x]$ is principal*

*Proof.* Let $I$ be an ideal of $F[x]$. If $I = \{0\}$, $I = \langle 0 \rangle$. Suppose $I \neq \{0\}$. Let $g(x) \in I$ be non-zero polynomial of minimal degree in $I$.

We claim that $I = \langle g(x) \rangle$. Clearly $\langle g(x) \rangle \subseteq I$.

Let $f(x) \in I$, by division algorithm, there exists $q(x), r(x)$ with $r(x) = 0$ or $deg(r) < deg(g)$ such that $f(x) = g(x)q(x) + r(x)$. But $r(x) = f(x) - g(x)q(x)$ where $f(x), q(x)g(x) \in I$. Thus $r(x) \in I$ and by minimality, we must have $r(x) = 0$. Thus $f(x) = g(x)q(x) \in \langle g(x) \rangle$ ♡

# Chapter 2

# Minimal Polynomials and Jordan Form

## 2.1 Characteristic and Minimal Polynomials

**Definition 2.1.1.** Let $A \in M_n(F)$, the ***characteristic polynomial*** is $det(A - xI)$. Let $T : V \to V$ be linear, then the characteristic polynomial is $det([T]_\beta - xI)$

**Remark 2.1.2.** Unless otherwise stated, in this note, $F$ is always a field(not just $\mathbb{R}$ or $\mathbb{C}$, but in the inner product section, we will only consider $F$ to be real or complex in the standard materials), and $V$ is always a finite dimensional vector space over $F$.

**Definition 2.1.3.** Let $T : V \to V$ be linear, we say a subspace $W \leq V$ is ***T-invariant*** if $T(W) \subseteq W$

**Remark 2.1.4.** If $W$ is T-invariant, then $T_W : W \to W$, i.e. $T_W(x) = T(x)$, is well-defined.

**Example 2.1.5.** Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ to be $T(x, y) = (x + 2y, 4y - x)$. Find an interesting T-invariant of $T$.

*Solution.* Let $W = \{(x, x) : x \in \mathbb{R}^2\}$, we have $T(x, x) = (3x, 3x) \in W$, so $W$ is T-invariant. ♠

**Example 2.1.6.** Let $T : V \to V$ be linear, let $\lambda$ be eigenvalue and $E_\lambda = \{v \in V : T(v) = \lambda v\}$, show it is T-invariant.

*Solution.* Let $v \in E_\lambda$, $T(T(v)) = T(\lambda v) = \lambda T(v)$, hence $T(v) \in E_\lambda$ and $E_\lambda$ is T-invariant. ♠

**Definition 2.1.7.** Let $T : V \to V$ be linear, $0 \neq x \in V$. The subspace $W_{T,x} = span(x, T(x), T^2(x), T^3(x), ...)$ is called the ***T-cyclic subspace generated by x***.

**Remark 2.1.8.** $W_{T,x}$ is the smallest T-invariant subspace of $V$ containing $x$.

**Proposition 2.1.9.** *Let $T : V \to V$ be linear, $W \leq V$ be $T$-invariant. Then the characteristic polynomial of $T_W$ divides the characteristic polynomial of $T$*

*Proof.* Let $\beta = (v_1, v_2, ..., v_m)$ be an ordered basis for $W$. Say $[T]_\beta = A$. Extend $\beta$ to an ordered basis $\gamma = (v_1, v_2, ..., v_m, v_{m+1}, ..., v_n)$ for $V$. Say $[T]_\gamma = B$.

So,

$$B = [T(v_1), T(v_2), T(v_3), ..., T(v_n)] = [T_W(v_1), ..., T_W(v_m), T(v_{m+1}), ..., T(v_n)]$$

Hence

$$B = \begin{array}{cc} & \overbrace{\hphantom{AAA}}^{m} \\ m\left\{\right. & \left[\begin{array}{c|c} A & \star \\ \hline 0 & A' \end{array}\right] \end{array}$$

So, $det(B - xI) = det(A - xI)det(A' - xI)$ ♡

**Example 2.1.10.** Let $\mathbb{F}$ be a field and $a_0, ..., a_{n-1} \in \mathbb{F}$. Try to determine the characteristic polynomial of

$$A = \begin{bmatrix} 0 & 0 & ... & 0 & -a_0 \\ 1 & 0 & ... & 0 & -a_1 \\ 0 & 1 & ... & 0 & -a_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & ... & 1 & -a_{n-1} \end{bmatrix}$$

*Solution.* We have

$$A - tI = \begin{bmatrix} -t & 0 & ... & 0 & -a_0 \\ 1 & -t & ... & 0 & -a_1 \\ 0 & 1 & ... & 0 & -a_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & ... & 1 & -t - a_{n-1} \end{bmatrix}$$

With elementary column operations, namely add $-a_i$ times the ith column to the last column for $1 \leq i \leq n-1$, we get a matrix

$$B = \begin{bmatrix} -t & 0 & ... & 0 & -a_1 t - a_0 \\ 1 & -t & ... & 0 & -a_2 t \\ 0 & 1 & ... & 0 & -a_3 t \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & ... & -t & -a_{n-1} t \\ 0 & 0 & ... & 1 & -t \end{bmatrix}$$

with $|A - tI| = |B|$. Thus, it suffice to calculate the determinant of $B$.

We claim $det(A - tI) = det(B) = (-1)^n(t^n + \sum_{i=0}^{n-1} a_i t^i)$. We use induction on $n$. When $n = 2$, we indeed have $\begin{vmatrix} -t & -a_1t - a_0 \\ 1 & -t \end{vmatrix} = t^2 + a_1t + a_0$. Suppose it holds for $n$, we need to show our claim holds for $n+1$.

By co-factor expansion on the first row, we have

$$\begin{aligned}
B &= \begin{vmatrix} -t & 0 & \dots & 0 & -a_1t - a_0 \\ 1 & -t & \dots & 0 & -a_2t \\ 0 & 1 & \dots & 0 & -a_3t \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & -t & -a_nt \\ 0 & 0 & \dots & 1 & -t \end{vmatrix} \\
&= (-t) \begin{vmatrix} -t & \dots & 0 & -a_2t - 0 \\ 0 & \dots & 0 & -a_3t \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & -t & -a_nt \\ 0 & \dots & 1 & -t \end{vmatrix} + (-1)^{n+1+1}(-a_1t - a_0) \begin{vmatrix} 1 & -t & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & -t \\ 0 & 0 & \dots & 1 \end{vmatrix} \\
&= (-t)(-1)^n(t^n + a_nt^{n-1} + a_{n-1}t^{n-2} + \dots + a_2t + 0) + (-1)^{n+1}(a_1t + a_0) \\
&= (-1)^{n-1}(t^{n+1} + a_nt^n + \dots + a_2t^2 + a_1t + a_0)
\end{aligned}$$

Thus, the induction hypothesis holds. ♠

**Proposition 2.1.11.** *Suppose $T : V \to V$ be linear and $W = W_{T,v}$ where $v \neq 0$. Say $dim(W) = k$.*

1. *$\{v, T(v), T^2(v), \dots, T^{k-1}(v)\}$ is a basis for $W$*
2. *If $f(x) = x^k + \sum_{i=1}^{k} a_{k-i}x^{k-i}$ is in $F[x]$ and $(f(T))(v) = 0 \in V$ then the characteristic polynomial for $T_W$ is $(-1)^k f(x)$*

*Proof.* 1. Let $j \in \mathbb{N}$ be maximal such that $\beta = \{v, T(v), T^2(v), \dots, T^{j-1}(v)\}$ is linear independent. We claim $j = k$. Let $U = span(v, \dots, T^{j-1}(v))$, we show $U = W$. Now, since $\{v, T(v), \dots, T^{j-1}(v), T^j(v)\}$ is linear dependent, and $\beta$ is linear independent. So $T^j(v) \in U$. Thus $U$ is T-invariant. By Remark 2.1.8 of $W$, $W \subseteq U$. However, $U \subseteq W$ so $U = W$ and $j = k$

2. From before, $\beta = \{v, T(v), \dots, T^{k-1}(v)\}$ is an ordered basis for $W$. Moreover, define $T^0(x) = x$, we have $T_k(v) + \sum_{i=0}^{k-1} a_i T^i(v) = 0$. Hence $T^k(v) =$

$\sum_{i=0}^{k-1} -a_i T^i(v)$. Thus,

$$[T_W]_\beta = \begin{bmatrix} 0 & 0 & ... & 0 & -a_0 \\ 1 & 0 & ... & 0 & -a_1 \\ 0 & 1 & ... & 0 & -a_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & ... & 1 & -a_{k-1} \end{bmatrix}$$

Hence, by Example 2.1.10, the characteristic polynomial of $T_w$ is $(-1)^k f(x)$.

♡

**Theorem 2.1.12** (Cayley-Hamilton)**.** *If $T : V \to V$ is a linear operator and $f(x) \in F[x]$ is its characteristic polynomial then $f(T) = 0$*

*Proof.* Since $f(T)$ is linear, $(f(T))(0) = 0$. Now let $v \neq 0 \in V$, we claim $f(T)(v) = 0$. Let $W = W_{T,v}$ and say $dim(W) = k$. Since $\{v, T(v), ..., T^{k-1}(v)\}$ is a basis for $W$, we have $\{v, ..., T^k(v)\}$ is linear dependent. Thus $\exists a_i \in F$, not all zero, such that $a_0 v + a_1 T(v) + ... + a_k T^k(v) = 0$ with $a_k \neq 0$. We can assume $a_k = 1$ as we can always multiply the equation by $a_k^{-1}$. Let $g(x) = x^k + a_{k-1}x^{k-1} + ... + a_1 x + a_0$ so that $g(T)(v) = 0$. Note $g$ is degree of $k$, hence the characteristic polynomial, $h(x)$, of $T_W$, is $(-1)^k g(x)$ by Proposition 2.1.11. Therefore, $h(T)(v) = (-1)^k 0 = 0$. Since $h(x)|f(x)$ by Proposition 2.1.9, so $f(T)(v) = 0$ and thus $f(T) = 0$. ♡

**Remark 2.1.13.** Let $T : V \to V$. Let $I = \{f(x) \in F[x] : f(T) = 0\}$ is an ideal of $F[x]$. From Proposition 1.1.37, we know $I = \langle g(x) \rangle$ for some $g(x) \in F[x]$. Note $a(x), b(x) \in F[x]$, we have $\langle a(x) \rangle = \langle b(x) \rangle$ if and only if $b(x) = ka(x)$ where $k \neq 0 \in F$, as we have $a(x)|b(x)$ and $b(x)|a(x)$. Thus there is only one ***monic*** (leading coeff=1) $m(x) \in F[x]$ such that $I = \langle m(x) \rangle$.

**Definition 2.1.14.** We call this $m(x)$ in the above remark the ***minimal polynomial for T***.

**Remark 2.1.15.** $f(x) \in F[x]$, $f(T) = 0$, thus we have

1. $m(x)|f(x)$
2. $m(x)|det([T] - xI)$ by Cayley-Hamilton

**Remark 2.1.16.** We similarly define the minimal polynomial of $A \in M_n(F)$. It is the minimal polynomial for $L_A : F^n \to F^n$ given by $L_A(v) = Av$.

**Example 2.1.17.** Calculate the minimal polynomials of the following linear operators:

1. $T : M_n(\mathbb{R}) \to M_n(\mathbb{R})$ where $T(A) = A^T$
2. $T : V \to V$ such that $T(v) = 0$ for all $v \in V$
3. $T : \mathbb{C}^4 \to \mathbb{C}^4$ such that $T(v) = Av$ where $A = \begin{bmatrix} 0 & i & 0 & i \\ i & 0 & i & 0 \\ 0 & i & 0 & i \\ i & 0 & i & 0 \end{bmatrix}$

*Solution.*

1. If the dimension is 1 then the minimal polynomial is $x-1$. If dimension is greater than 1, note we have $T(A)=A^T$ thus $T^2(A)=A=I(A)$ where $I(A)$ is the identical mapping. Hence, we have $m(x)$, the minimal polynomial of $T$, divide $x^2-1$. This is because $(x^2-1)(T)=T^2-I$. Thus, $m(x)$ must be one of $x^2-1, x+1, x-1$. Clearly $x+1$ and $x-1$ are impossible, as we have $A=\begin{bmatrix} 0 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{bmatrix}$ and $A^T-A, A^T+A \neq 0$.
2. Note we have $f(x)=x$ and $(f(T))(v)=T(v)=0$ for all $v \in V$. Thus the minimal polynomial must divide $f(x)$, so $m(x)=x$. We can also see $m(x)=x$ as $m(x)$ must be monic and divide $x$.
3. Note $A^3=-4\begin{bmatrix} 0 & i & 0 & i \\ i & 0 & i & 0 \\ 0 & i & 0 & i \\ i & 0 & i & 0 \end{bmatrix}$. Thus $T^3(v)=A^3v=-4T(v)$ and $T^3(v)+4T(v)=0$ for all $v \in \mathbb{C}^4$. So, we have $m(x)|x^3+4x$ and $m(x)$ can only be $x, x^2+4$, or $x^3+4x$.
   Note $T(v_1)\neq 0$ for $v_1=(1,1,1,1)$ so $m(x)\neq x$. Note $T^2(v_2)+4I(v_2)\neq 0$ for $v_2=(1,0,0,0)$ as $T^2(v_2)=(-2,0,-2,0)$ and $4I(v_2)=(4,0,0,0)$. Hence we must have $m(x)=x^3+4x$.

♠

**Example 2.1.18.** Find $A \in M_3(\mathbb{R})$ such that the minimal polynomial $m(x)$ for $A$ is $m(x)=x^3+2x^2+x+5$

*Solution.* Given a polynomial $f(x)$, we can generate a matrix with $f(x)$ as the characteristic polynomial, as we recall Example 2.1.10. Thus, let $A=\begin{bmatrix} 0 & 0 & -5 \\ 1 & 0 & -1 \\ 0 & 1 & -2 \end{bmatrix}$, we have $det(A-xI)=(-1)^3m(x)$, thus we must have the minimal polynomial of $A$, denoted as $m'(x)$, divides $m(x)$. If $m'(x)=x+r_1$, then $m'(A)=A+r_1I$ and there is no way this equal the zero matrix. If $m'(x)=x^2+r_2x+r_3$, then we have $m'(A)=\begin{bmatrix} r_3 & -5 & -5r_1+10 \\ r_2 & -1+r_3 & -r_1-3 \\ 1 & -2+r_1 & 3-2r_2+r_3 \end{bmatrix}$, which is impossible to equal zero as well. Hence the degree of $m'(x)$ is three and we have $m'(x)=m(x)$. Therefore we have our desired $A$. ♠

**Proposition 2.1.19.** *Let $T: V \to V$ be linear. Let $m(x)$ be the minimal polynomial for $T$, let $f(x)$ be the characteristic polynomial of $T$. Then $m(x)$ and $f(x)$ have the same roots in $F$.*

*Proof.* We first note that $m(x)|f(x)$ and so every root of $m(x)$ is a root of $f(x)$.

Now let $\lambda \in F$ be an eigenvalue of $T$, we must show $m(\lambda)=0$. Let $v \neq 0 \in E_\lambda$, so that $T(v)=\lambda v$. Thus $(m(T))v=m(\lambda)v$, so $m(\lambda)v=(m(T))(v)=0(v)=0$. This means $m(\lambda)=0$ as $v$ is non-zero. ♡

**Example 2.1.20.** Let $V = P_2(\mathbb{R})$. Let $T(g(x)) = g'(x) + 2g(x)$. Find the minimal polynomial of $T$.

*Solution.* Let $\beta = (1, x, x^2)$, we have $T(1) = 2$, $T(x) = 1+2x$ and $T(x^2) = 2x+2x^2$. Thus $A := [T]_\beta = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 2 \\ 0 & 0 & 2 \end{bmatrix}$. Thus the characteristic polynomial is $f(x) = (2-x)^3$. So the minimal polynomial is $m(x)$ must be one of $x - 2, (x-2)^2$, or $(x-2)^3$.

However, $A - 2I \neq 0$ and $(A - 2I)^2 \neq 0$, so we have $m(x) = (x-2)^3$. ♠

**Example 2.1.21.** Let $A := \begin{bmatrix} 3 & -1 & 0 \\ 0 & 2 & 0 \\ 1 & -1 & -2 \end{bmatrix}$, find the minimal polynomial of $A$.

*Solution.* We first find the characteristic polynomial of $A$, which is $f(x) = (3 - x)(2-x)^2$. Hence we must have $m(x) = (x-3)(x-2)$ or $m(x) = (x-3)(x-2)^2$. Note $((x-3)(x-2))(A) = 0$ so $m(x) = (x-3)(x-2)$. ♠

**Definition 2.1.22.** Let $T : V \to V$ be linear, we say that $V$ is ***T-cyclic*** if there exists $0 \neq v \in V$ such that $V = W_{T,v}$.

**Proposition 2.1.23.** *Let $T : V \to V$ be linear. If $dim(V) = n$ and $V$ is $T$-cyclic, then characteristic polynomial and the minimal polynomial for $T$ have the same degree. In particular, if $f(x)$ is the characteristic polynomial and $m(x)$ is the minimal polynomial, then $f(x) = (-1)^n m(x)$*

*Proof.* Say $V = W_{T,v}$ for $0 \neq v \in V$.

A basis for $V$ is then $\{v, T(v), T^2(v), ..., T^{n-1}(v)\}$. Let $g(x) = a_0 + a_1 x + ... + a_k x^k \in F[x]$ with $a_k \neq 0$ and $k < n$. Since $\{v, T(v), T^2(v), ..., T^k(v)\}$ is linear independent, we know $g(T) \neq 0$ since $a_k \neq 0$ and $a_0 v + a_1 T^1(v) + ... + a_k T^k(v) \neq 0$. Therefore, the degree of the minimal polynomial must be at least $n$. However, $m(x) \mid f(x)$, so $deg(m) \leq deg(f)$, so $deg(m) = n$. ♡

**Proposition 2.1.24.** *Let $T : V \to V$ and $m(x)$ be the minimal polynomial. Then $T$ is diagonalizable iff $m(x) = (x - \lambda_1)...(x - \lambda_k)$ where $\lambda_i$ are all distinct eigenvalues of $T$.*

*Proof.* ($\Rightarrow$): Suppose $T$ is diagonalizable. Let $\beta = (v_1, ..., v_n)$ be a basis of eigenvectors of $T$, for $V$. Let $p(x) = (x - \lambda_1)...(x - \lambda_k)$ where $\lambda_i$ are all of the distinct eigenvalues of $T$. We claim that $m(x) = p(x)$. From before, $p(x) \mid m(x)$ as $m(x)$ has all the eigenvalues as its roots. Let $v_i \in \beta$, so $T(v_i) = \lambda_j v_i$ for some $\lambda_j \in F$, in particular, $(T - \lambda_j I)(v_i) = 0$. Since $x - \lambda_j \mid p(x)$ where $p(x) = q_j(x)(x - \lambda_j)$, $q_j(x) \in F[x]$, we have

$$p(T)(v_i) = q_j(T) \circ (T - \lambda_j I)(v_j) = q_j(T)(0) = 0$$

Since $v_i \in \beta$ was arbitrary, we must have $p(T) = 0$, so $m(x) \mid p(x)$, so $m(x) = p(x)$.

($\Leftarrow$): Suppose $m(x) = (x - \lambda_1)...(x - \lambda_k)$ where $\lambda_i \in F$ are distinct. We will use induction on the dimension of $V$. The base case is easy. Let $W = Range(T - \lambda_k I)$, and $T_W : W \to W$, by induction, $T_W$ is diagonalizable and so there is a basis for $T_W$-eigenvectors for $W$: $\beta = (v_1, ..., v_m)$. Note $Null(T - \lambda_k I) = E_{\lambda_k}$ and let $\gamma = (w_1, ..., w_l)$ be a basis for $E_{\lambda_k}$. By the dimensional theorem, $dim(V) = n = m+l$. Let $y \in W$, thus $y = (T-\lambda_k I)(x)$ for some $x \in V$. Hence, $m(T)(x) = (T-\lambda_1)...(T-\lambda_{k-1})(y) = 0$ as $y$ can be arbitrary. Therefore, $(T_W - \lambda_1 I)...(T_W - \lambda_{k-1} I) = 0$ and so the minimal polynomial of $T_W$ divides $(x - \lambda_1)(x - \lambda_2)...(x - \lambda_{k-1})$. Hence, $\lambda_k$ is not an eigenvalue of $T_W$ as we assumed they are all different. This imply $(T_W - \lambda_k I)(v_i) = (T - \lambda_k I)(v_i) \neq 0$ for all $1 \leq i \leq m$ so $v_i \notin Null(T - \lambda_k I)$. Therefore, $\beta$ and $\gamma$ are disjoint.Therefore, $\beta \cup \gamma$ is linear independent and thus a basis of eigenvectors of $T$ for $V$. In particular, we have $T$ is diagonalizable. ♡

**Example 2.1.25.** Let $A \in M_n(\mathbb{C})$ such that $A^m = I$. Let $m(x)$ be the minimal polynomial for $A$. We know $m(x) \mid x^m - 1$. However, $x^m - 1$ has $m$ distinct roots, namely $1, \zeta, \zeta^2, ..., \zeta^{m-1}$, where $\zeta = e^{i(\frac{2\pi}{m})}$. Therefore, $m(x)$ splits over $\mathbb{C}$ and hence has distinct roots, by the above proposition, $A$ is diagonalizable.

**Example 2.1.26.** Describe all $A \in M_2(\mathbb{R})$ such that $A^2 - 7A + 10I = 0$

*Solution.* Suppose $A \in M_2(\mathbb{R})$ such that $A^2 - 7A + 10I = 0$ and let $m(x)$ be the minimal polynomial for $A$ over $\mathbb{R}$. Thus $m(x)|(x^2-7x+10)$ and so $m(x)|(x-2)(x-5)$.

If $m(x) = x - 2$ then $A = 2I$, if $m(x) = x - 5$ then $A = 5I$.

If $m(x) = (x - 2)(x - 5)$ then $A$ is diagonalizable by Proposition 2.1.24 and hence similar to $\begin{bmatrix} 2 & 0 \\ 0 & 5 \end{bmatrix}$. ♠

**Example 2.1.27** (Additional Materials)**.** Let $F, K$ be fields, we say $K$ is a ***field extension*** of $F$, and write as $K/F$, if $F$ is a subfield of $K$.

We say that a non-constant polynomial $f(x) \in F[x]$ is ***separable*** if $f(x)$ has no roots of multiplicity greater than one in any extension of $F$.

In addition, we say a non-constant polynomial $f(x) \in F[x]$ is ***irreducible*** over $F$ if whenever $g(x), h(x) \in F[x]$ such that $f(X) = g(x)h(x)$, then one of $g(x)$ or $h(x)$ is constant.

1. Prove that $f(x) \in F[x]$ is separable if and only if $f(x)$ and $f'(x)$ have no common factors in $F[x]$ of positive degree.
2. Prove that an irreducible polynomial $f(x) \in F[x]$ is separable if and only if $f'(x) \neq 0$.
3. Suppose $char(F) = p$, describe all irreducible polynomials in $F[x]$ which is non-separable.
4. Let $T : V \to V$ be linear operator. Prove that if the characteristic polynomial of $T$ is separable then it is a scalar multiple of the minimal polynomial of $T$.

*Solution.*

1. Suppose $f(x)$ and $f'(x)$ have no common factors in $F[x]$ of positive degree. By repeatedly applying the division algorithm, there exists $g(x), h(x) \in F[x]$ such that $1 = f(x)g(x) + h(x)f'(x)$. If $f(x)$ is not separable then there exists an extension $K/F$ such that $(x-\alpha)^2|f(x)$ for some $\alpha \in K$. But then $(x-\alpha)|f'(x)$ and so $f(\alpha) = f'(\alpha) = 0$. Hence $1 = 0 + 0 = 0$, a contradiction. Therefore, $f(x)$ is separable.
   Conversely, suppose $f(x)$ is separable. For contradiction, assume $\exists g(x) \in F[x]$ such that $deg(g) > 0$, where $g(x)|f(x), g(x)|f'(x)$. Let $K/F$ be an extension such that $g(x)$ has a root $\alpha \in K$ (See additional topic for a proof). Then $f(x) = (x-\alpha)h(x)$ and $f'(x) = (x-\alpha)k(x)$, where $h(x), k(x) \in K[x]$. However, $f'(x) = h(x) + (x-\alpha)h'(x)$ and so $0 = f'(\alpha) = h(\alpha)$ so that $(x-\alpha)|h(x)$ and $(x-\alpha)^2|f(x)$, a contradiction.
2. Note $f(x)$ is separable if and only if $f(x), f'(x)$ do not have a common factor of positive degree. Suppose $f(x)$ is irreducible, then the only divisor of $f(x)$ are $cf(x)$, where $c \in F\backslash\{0\}$. Since $deg(cf(x)) > deg(f'(x))$, we have $cf(x) \nmid f'(x)$ only when $f'(x) \neq 0$, hence $f'(x) \neq 0$. Conversely, if $f'(x) = 0$ then $f(x)$ is a common divisor of $f(x)$ and $f'(x)$, in which case $f(x)$ is not separable by part 1.
3. Suppose $char(F) = p$ and $f(x) \in F[x]$ is irreducible and non-separable. By part 2, this means $f'(x) = 0 \Leftrightarrow f'(x) = 0$. Say $f(x) = a_n x^n + a_{n-1}x^{n-1} + ... + a_1 x + a_0$. In particular, we have

$$\begin{aligned}
&f'(x) = 0 \\
\Leftrightarrow\ & n a_n x^{n-1} + (n-1)a_{n-1}x^{n-1} + ... + 2a_2 x + a_1 = 0 \\
\Leftrightarrow\ & i a_i = 0, i = 1, 2, ..., n \\
\Leftrightarrow\ & i a_i = p m_i a_i, i = 1, 2, ..., n, \text{ for some } m_i \in \mathbb{N} \\
\Leftrightarrow\ & f(x) = a_n x^{pm_n} + a_{n-1}x^{pm_{n-1}} + .. + a_2 x^{pm_2} + a_1 x^{pm_1} + a_0 \\
\Leftrightarrow\ & f(x) = g(x^p) \text{ where } g(x) \in F[x]
\end{aligned}$$

4. Let $A = [T]_\beta$, relative to some basis $\beta$ for $F$. Let $f(x)$ be the characteristic polynomial for $A$ and let $m(x)$ be the minimal polynomial for $A$ in $F[x]$. Suppose $f(x)$ is separable so that $f(x) = (x-\lambda_1)(x-\lambda_2)...(x-\lambda_k)$ for some distinct $\lambda_i \in K$ where $K$ is the splitting field of $f(x)$ on $F$. Let $m'(x)$ be the minimal polynomial for $A$ in $K[x]$. Then, by definition, we have $m'(x)|m(x)$ and $m(x)|f(x)$. However, $m'(x)$ has every eigenvalue $\lambda_i \in \mathbb{C}$ as a root and so $f(x)|m'(x)$. Therefore $m(x)|f(x)$ and $f(x)|m(x)$, in particular, $f(x) = cm(x)$, where $c \in F\backslash\{0\}$.

♠

**Remark 2.1.28.** The following are additional materials. Click here to jump to the next section with lecture materials.

### 2.1.1 Maximal Ideal and Quotient Ring

**Definition 2.1.29.** A left ideal of $R$ is a subring $I_L$ of $R$ such that $xa \in I_L$ for all $a \in I_L$ and $x \in R$. A right ideal of $R$ is a subring $I_R$ such that $ax \in I_R$ for all $a \in I_R$

and $x \in R$. A two sided ideal $I$ of $R$ is a subring that is left ideal and right ideal at the same time.

**Definition 2.1.30.** Let $R$ be a ring, a two-sided ideal $I$ of $R$ is called maximal if $I \neq R$ and no proper ideal of $R$ properly contains $I$.

**Definition 2.1.31.** Let $R$ be a ring and let $I$ be a two-sided ideal of $R$. If $a \in R$ then the coset of $I$ in $R$ determined by $a$ is defined by $a + I = \{a + r : r \in I\}$.

**Lemma 2.1.32.** *Let $a, b \in R$, and $I$ be a two sided ideal, then if $a - b \in I$, we have $a + I = b + I$. Moreover, $a - b \notin I$ then $a + I$ and $b + I$ are disjoint subsets of $R$.*

*Proof.* If $a-b \in I$, let $x \in a+I$. Then $x = a+m$ for $m \in I$. Thus $x = a-b+b+m = b + (a - b + m)$. Note $a - b + m \in I$, and so $x \in b + I$. Hence $a + I \subseteq b + I$. With the same argument, we get $b + I \subseteq a + I$ so $a + I = b + I$.

Now, suppose $a - b \notin I$ and let $c \in (a + I) \cap (b + I)$. Then $c = a + m_1 = b + m_2$ where $m_1, m_2 \in I$. It follows that $a - b = m_2 - m_1 \in I$, a contradiction. ♡

**Definition 2.1.33.** The set of cosets of $I$, a two sided ideal, in $R$ is denoted by $R/I$. We define addition and multiplication in $R/I$ as follows.

Let $a + I, b + I \in R/I$, we define their sum by $(a + I) + (b + I) = (a + b) + I$. We define their product to be $(a + I)(b + I) = ab + I$.

**Lemma 2.1.34.** *Addition and multiplication are well-defined in $R/I$.*

*Proof.* To show addition is well-defined, we need to show for $a+I, a_1+I, b+I, b_1+I \in R/I$, we have $a = a_1, b = b_1$ in $R$ then $(a + b) + I = (a_1 + b_1) + I$ in $R/I$. Suppose $a + I = a_1 + I$ and $b + I = b_1 + I$ for elements $a_1, b_1 \in R$. Then $a - a_1 \in I$ and $b - b_1 \in I$, by Lemma 2.1.32. Hence $(a - a_1) + (b - b_1) = (a + b) - (a_1 + b_1) \in I$, thus $(a + b) + I = (a_1 + b_1) + I$ by Lemma 2.1.32.

To show multiplication is well-defined, we need to show the same thing. Let $a + I, a_1 + I, b + I, b_1 + I \in R/I$ where $a + I = a_1 + I$ and $b + I = b_1 + I$. Then $a - a_1 \in I$ and $b - b_1 \in I$ by Lemma 2.1.32. We need to show $ab - a_1b_1 \in I$ by Lemma 2.1.32. Note $ab - a_1b_1 = ab - a_1b + a_1b - a_1b_1 = (a - a_1)b + a_1(b - b_1)$. In addition, since $I$ is a two sided ideal, we know $(a - a_1)b \in I$ and $a(b - b_1) \in I$. Thus $(a - a_1)b + a_1(b - b_1) = ab - a_1b_1 \in I$ as desired. ♡

**Proposition 2.1.35.** *$R/I$ is a ring with the addtion and multiplication we defined above.*

*Proof.* It follows from the fact that all the ring axioms are satisfied in $R$. ♡

**Theorem 2.1.36.** *Let $R$ be a unital commutative ring with identity, and let $M$ be an ideal of $R$. Then the quotient ring $R/M$ is a field if and only if $M$ is a maximal ideal of $R$.*

*Proof.* We first need to show that if $R/M$ is a field then $M$ is a maximal ideal. Then we need to show that if $M$ is a maximal ideal then every non-zero element in $R/M$ is a unit, since clearly we have $R/M$ to be unital and commutative where $R$ is.

Suppose $R/M$ is a field and let $I$ be an ideal of $R$ properly containing $M$. Let $a \in I, a \notin M$. Then $a+M$ is not the zero element of $R/M$, and so $(a+M)(b+M) = 1+M$ for some $b \in R$ as $R/M$ is a field. Then $ab-1 \in M$. Let $m = ab-1$. Now we have $1 = ab-m$ so $1 \in I$ since $a \in I$ and $m \in I$. It follows $I = R$ and so $M$ is a maximal ideal of $R$.

Suppose $M$ is a maximal ideal of $R$ and let $a+M$ be a non-zero element of $R/M$. We need to show there exists $b \in R$ such that $ab-1 \in M$ as that will imply $(a+M)(b+M) = ab+M = 1+M$. Let $M'$ be the set of elements of $R$ of the form $ar+s$, for some $r \in R$ and $s \in M$. Then $M'$ is an ideal of $R$. Let (note since $R$ is commutative we only need to show $M'$ is left ideal) $x \in R$ and $y \in M'$ we have $xy = x(ar_y + s_y) = xar_y + xs_y$, $xar_y \in R$ is obvious, and $xs_y \in M$ as $M$ is an ideal. Moreover, $x, y \in M'$ then $x - y = ar_x + s_x - ar_y - s_y = a(r_x - r_y) + (s_x - s_y)$, where $a(r_x - r_y) \in R$ by definition, and $s_x - s_y \in M$ since $M$ is a subring. Hence $M'$ is a subring hence an ideal. Then, $M'$ properly contains $M$ since $a \in M'$ and $a \notin M$. Then $M' = R$ since $M$ is a maximal ideal of $R$. In particular, $1 \in M'$ and $1 = ab+m$ for some $b \in R$ and $m \in M$. Then $ab-1 \in M$ as desired. So $a+M$ is a unit for all $a \in R$. ♡

### 2.1.2 Field Extension and Splitting Field

**Lemma 2.1.37.** *Let $E$, $F$ be fields and let $\sigma : F \to E$ be an embedding (or injective homomorphism). Then there exists field $K$ such that $F$ is a subfield of $K$ and $\sigma$ can be extended to an isomorphism of $K$ onto $E$.*

*Proof.* Let $S$ be a set whose cardinality is the same as that of $E\backslash\sigma(F)$ =the complement of $\sigma(F)$ in $E$, and disjoint from $F$. Let $f$ be a bijection from $S$ to $E\backslash\sigma(F)$. Set $K = F \cup S$. Then we extend $\sigma$ to a mapping $\sigma^* : K \to E$ as follows: $\sigma^*(a) = \sigma(a)$ if $a \in F$, $\sigma^*(a) = f(a)$ if $a \in S$. Clearly $\sigma^*$ is well-defined bijection. We now define a field on $K$. If $x, y \in E$, we define $x + y = (\sigma^*)^{-1}(\sigma^*(x) + \sigma^*(y))$ and $xy = (\sigma^*)^{-1}(\sigma^*(x)\sigma^*(y))$. Our definition of addition and multiplication coincide with the given addition and multiplication of elements of the original field $F$, and it is clear that $F$ is a subfield of $K$. Hence, we have the desired field $K$. ♡

**Remark 2.1.38.** Thus, if $\sigma$ is an embedding (or injective homomorphism) of $F$ into $E$, we can think $F$ in terms of the corresponding image $\sigma(F)$ in $E$ and can consider $E$ as an extension of $F$ as there is an isomorphism between a field extension of $F$ and $E$.

**Theorem 2.1.39.** *Let $p(x)$ be an irreducible polynomial in $F[x]$. Then there exists an extension $E$ of $F$ in which $p(x)$ has a root $\alpha$.*

*Proof.* Since $p(x)$ is irreducible in $F[x]$, $I =: \langle p(x) \rangle$ is a maximal ideal in $F[x]$. Let $I_1$ be an ideal of $F[x]$ containing $I$. Then $I_1 := \langle f(x) \rangle$ for some $f(x) \in F[x]$

as every ideal in $F[x]$ is principal. Since $p(x) \in I_1$, we have $p(x) = f(x)q(x)$ for some $q(x) \in F[x]$. Since $p(x)$ is irreducible this means that either $f(x)$ is of degree zero or $q(x)$ has degree zero. If $f(x)$ has degree zero then $f(x)$ is a unit in $F[x]$ and $I_1 = F[x]$. If $q(x)$ has degree zero than $p(x) = af(x)$ for $a \neq 0 \in F$, and $f(x) = a^{-1}p(x)$, then $f(x) \in I$ and $I_1 = I$. Thus either $I_1 = I$ or $I_1 = F[x]$, indeed $I$ is a maximal ideal in $F[x]$.

Thus, $E := F[x]/\langle p(x) \rangle$ is a field. We define $\sigma : F \to E$ as $\sigma(a) = a + \langle p(x) \rangle$, clearly this is an injective homomorphism. To show it is homomorphism, we have $\sigma(a+b) = (a+b) + I = \sigma(a) + \sigma(b) = (a+I) + (b+I)$ and $\sigma(ab) = ab + I$ and $\sigma(a)\sigma(b) = ab + I$ as desired by the way we defined quotient ring operation. It is injective as $a, b \in F$ with $a \neq b$ imply $a + I \neq b + I$. Thus $E$ is an extension field of $F$ as we mentioned in above remark. Let $p(x) = a_0 + a_1x + ... + a_nx^n$, we have $p(x) = 0$ in $E$. Thus, $x + I \in E$ is a root of $p(x)$ in $E$ where $E$ is an extension of $F$.

♡

**Theorem 2.1.40.** *Let $0 \leq f(x) \in F[x]$. Then there exists a field extension $K$ such that $f(x) = c\prod_{i=1}^{n}(x - c_i)$ where $c, c_1, ..., c_n \in K$ and $c \neq 0$. We say $f(x)$ splits over $K[x]$.*

*Proof.* Let $n = deg(f(x))$, we use induction on $n$. If $f(x) = cx + d$ and $c \neq 0$ then our field extension $K = F$ will do the job. Suppose $g(x)$ is a polynomial of degree $j \leq n - 1$ on some field $F_1$ and $g(x)$ splits over a field extension $G$ of $F_1$.

Then, we can write $f(x) = f_1(x)f_2(x)$ where $f_1(x)$ is irreducible in $F$. Hence, we have field extension $E$ of $F$ such that $\exists \alpha \in E$ and $f_1(\alpha_1) = 0$. Hence, on the extension $E$ of $F$, we have $f(x) = (x - \alpha)g(x) \in E[x]$, where $g(x)$ has degree $n - 1$. Hence, there exists an extension $E'$ of $E$ such that $g(x)$ splits over $E'[x]$, say $g(x) = c\prod_{i=2}^{n}(x - \alpha_i)$. Hence, the field $E'$ is an extension of $E$ and hence an extension of $F$, where $f(x) = c\prod_{i=1}^{n}(x - \alpha_i)$ splits over $E'[x]$. This closed the induction.

♡

## 2.2 Jordan Canonical Form

**Remark 2.2.1.** Our goal is to show that any operator $T : V \to V$ is almost diagonalizable.

**Remark 2.2.2.** How can $T$ fail to be diagonalizable?

1. characteristic polynomial doesn't split over $F$
2. Assume $det(A - xI)$ does split over $F$ where $A = [T]_\beta$ be $n \times n$. Say $\{v_1, ..., v_n\} \in V$ and each $v_i$ is a $\lambda_i$-eigenvector. Then let $P = [v_1, ..., v_n]$ and $D = diag(\lambda_1, ..., \lambda_n)$. We have $AP = [\lambda_1v_1, ..., \lambda_nv_n]$ and $PD = [\lambda_1v_1, ..., \lambda_nv_n]$. Thus, for all matrices, we have $AP = PD$. Thus, $A$ is diagonalizable iff $P$ is

invertible, $A = PDP^{-1}$ iff $\{v_1, ..., v_n\}$ is linear independent iff $\beta_1, ..., \beta_k$ is a basis of $V$ where $\beta_i$ are basis for $E_{\lambda_i}$.

**Definition 2.2.3.**

1. We say $A \in M_n(F)$ is a Jordan block if $A = \begin{bmatrix} \lambda & 1 & & & & \\ & \lambda & 1 & & & \\ & & \lambda & 1 & & \\ & & & \ddots & \ddots & \\ & & & & \lambda & 1 \\ & & & & & \lambda \end{bmatrix}$ In this note, I may write $J_r(n)$ to be a $n \times n$ Jordan block with $r$ on the diagonal, ***this is not standard notation in the lecture***.
2. We say $J \in M_n(F)$ is a Jordan matrix, when $J$ is a matrix with Jordan blocks on the diagonal.

**Example 2.2.4.** We have:

$$J = \begin{bmatrix} 2 & & & & & \\ & 2 & 1 & & & \\ & 0 & 2 & & & \\ & & & 3 & 1 & 0 \\ & & & 0 & 3 & 1 \\ & & & 0 & 0 & 3 \end{bmatrix}$$

Say $[T]_\beta = J$, where $\beta = \{v_1, ..., v_6\}$. Thus $T(v_1) = 2v_1$. $T(v_2) = 2v_2$ and $T(v_3) = v_2 + 2v_3$, therefore, $(T - 2I)(v_3) = v_2$ and $(T - 2I)^2(v_3) = 0$. Moreover, $T(v_4) = 3v_4$ and $T(v_5) = v_4 + 3v_5$ which imply $(T - 3I)^2(v_5) = 0$. $T(v_6) = v_5 + 3v_6$, which means $(T - 3I)^2(v_6) = (T - 3I)(v_5)$ and so $(T - 3I)^3(v_6) = 0$.

**Remark 2.2.5.** For this section, $T : V \to V$ be linear operator, then we assume the characteristic polynomial of splits over $F$. We always write $\lambda$ as an eigenvalue of $T$.

**Definition 2.2.6.** Let $T : V \to V$, and $\lambda$ be eigenvalues of $T$. We define $0 \neq v \in V$ to be a ***generalized eigenvector*** of $T$ if $(T - \lambda I)^p(v) = 0$ for some $p \in \mathbb{N}$.

Moreover, we define $K_\lambda = \{v \in V : \exists p \in \mathbb{N}((T - \lambda I)^p(v) = 0)\}$ to be the ***generalized $\lambda$ -eigenspace*** of $V$.

**Remark 2.2.7.** We have $K_\lambda = \bigcup_{i=1}^{\infty} Null(T - \lambda I)^i$

**Proposition 2.2.8.** *$K_\lambda$ is a T-invariant subspace of $V$ which contains $E_\lambda$.*

**Proposition 2.2.9.** *Say $T : V \to V$ and $\lambda \neq \mu$ be different eigenvalues. Then, $(T - \lambda I) : K_\mu \to K_\mu$ is one-to-one (bijection). In particular, $K_\lambda \cap K_\mu = \{0\}$.*

*Proof.* Suppose $0 \neq x \in Ker(T - \lambda I)$, where $Ker(T - \lambda I)$ is the kernel of the map and $x \in K_\mu$. Therefore, $x \in E_\lambda \cap K_\mu$. Let $p \in \mathbb{N}$ be minimal such that $(T - \mu I)^p(x) = 0$.

If $p = 1$, then $x \in E_\mu$ and $x \in E_\lambda$ so $x = 0$ as $E_\mu \cap E_\lambda = \{0\}$, a contradiction. Thus, suppose $p > 1$. Consider $y = (T - \mu I)^{p-1}(x) \neq 0$ so that $(T - \mu I)y = 0$ so $y \in E_\mu$. However, since $E_\lambda$ is $(T - \mu I)$-invariant (easy to verify) so we have $y \in E_\mu$ and $y \in E_\lambda$, so a contradiction.

Therefore, $Ker(T - \lambda I) = \{0\}$ and so $T - \lambda I$ is one-to-one. Thus, $(T - \lambda I)^p$ is one-to-one for all $p \in \mathbb{N}$. This means $K_\mu \cap Ker(T - \lambda I)^p = \{0\}$ for all $p \in \mathbb{N}$. Therefore, $K_\mu \cap K_\lambda = \{0\}$. ♡

**Proposition 2.2.10.** *Let $T : V \to V$ and $\lambda$ be eigenvalue. Suppose $\lambda$ has multiplicity $m$. Then $dim(K_\lambda)$ is at most $m$, and $K_\lambda = Null(T - \lambda I)^m$*

*Proof.* Let $W = K_\lambda$. Consider $T_W : W \to W$. Let $f(x)$ be the characteristic polynomial of $T$ and let $g(x)$ be the characteristic polynomial of $T_W$. Then $g(x)|f(x)$. Also, if $\mu \neq \lambda$ is an eigenvalue of $T$, then $(T-\mu I) : W \to W$ is one-to-one. Therefore, the only eigenvalue of $T_W$ is $\lambda$ as $Null((T-\mu I)_W) = \{0\}$. Hence $g(x) = (-1)^d(x-\lambda)^d$ where $d$ is the dimension of $W$. Since $g(x)|f(x)$, we have $d = dim(W) \leq m$.

It is clear that we have $W \supseteq Null(T - \lambda I)^m$. By Cayley-Hamilton, we have $(T_W - \lambda I)^d = 0, d \leq m$, where $g(x) = (x-\lambda)^d$ is the characteristic polynomial of $T_W$. Then let $w \in W$, we have $(T-\lambda I)^m(w) = (T-\lambda I)^{m-d}(T-\lambda I)^d(m) = (T-\lambda I)^{m-d}(0) = 0$. Therefore, $w \in Null(T - \lambda I)^m$ as desired. ♡

**Proposition 2.2.11.** *Let $T : V \to V$ and $\lambda_1, ..., \lambda_k$ be distinct. For every $x \in V$, there exists $v_i \in K_{\lambda_i}$ such that $x = v_1 + ... + v_k$.*

*Proof.* We use induction on $k$, the number of distinct eigenvalues.

Say $k = 1$, then the characteristic polynomial of for $T$ is $(-1)^d(x - \lambda)^d$ where $d = dim(V)$. By Cayley-Hamilton, we know $(T - \lambda I)^d = 0$, thus $K_\lambda = V$ and so the result follows.

Assume the result holds for operators with fewer than $k$ eigenvalues.

Suppose $\lambda_1, \lambda_2, ...\lambda_k$ are the distinct eigenvalues of $T : V \to V$. Let $m$ be the multiplicity of $\lambda_k$ and let $W = Range(T - \lambda_k I)^m$, we know this is T-invariant. Recall that for $i < k$, $(T - \lambda_k) : K_{\lambda_i} \to K_{\lambda_i}$ is bijection. In addition, we must also have $(T - \lambda_k I)^m : K_{\lambda_i} \to K_{\lambda_i}$ is also bijection as the power of bijection is bijection. In particular, we have

1. $(T - \lambda_k I)^m(K_{\lambda_i}) = K_{\lambda_i}$ and so
2. $K_{\lambda_i} \subseteq W$

Thus, $\lambda_1, \lambda_2, ..., \lambda_{k-1}$ are eigenvalues of $T_W : W \to W$. Suppose the minimal polynomial of $T$ to be $m(x) = (x - \lambda_1)^{m_1}(x - \lambda_2)^{m_2}...(x - \lambda_k)^m$. Let $t(x) = (x - \lambda_1)^{m_1}(x - \lambda_2)^{m_2}...(x - \lambda_{k-1})^{m_{k-1}}$ Then let $y \in W$, we have $y = (T - \lambda I)^m(x)$ where $x \in V$. Thus, note $m(T)(x) = 0 = t(T)y$ and so the minimal polynomial of $T_W$ must divide $t(x)$ which imply $\lambda_k$ is not an eigenvalue of $T_W$ as $\lambda_1, ..., \lambda_k$ are distinct (See diagonalizable theorem 2.1.24).

Let $x \in V$, by induction, we know $(T - \lambda_k I)^m(x) = w_1 + w_2 + ... + w_{k-1}$ where each $w_i$ is an element of $K_{\lambda_i}$. Since each $(T - \lambda_k I)^m$ restricted to $K_{\lambda_i}$ is bijection, for every $w_i$ there eixsts $v_i \in K_{\lambda_i}$ such that $(T - \lambda_k I)^m v_i = w_i$. Thus

$$(T - \lambda_k I)^m(x) = \sum_{i=1}^{k-1} (T - \lambda_k I)^m(v_i)$$

By linearity, we thus have

$$(T - \lambda_k I)^m(x - v_1 - v_2 - ... - v_{k-1}) = 0$$

Hence, let $v_k := x - v_1 - v_2 - ... - v_{k-1}$, we have $v_k \in K_{\lambda_k}$ and hence $x = v_1 + ... + v_k$. ♡

**Theorem 2.2.12.** *Let $T : V \to V$, assume the characteristic polynomial splits. Let $\lambda_1, ..., \lambda_k$ be distinct eigenvalues. Let $m_i$ be the multiplicity of $\lambda_i$. Let $\beta_i$ be ordered basis for $K_{\lambda_i}$. Then:*

1. *$\beta_i \cap \beta_j = \emptyset$ when $i \neq j$*
2. *$\beta := \bigcup_{i=1}^{k} \beta_i$ is an basis for $V$*
3. *$dim(K_{\lambda_i}) = m_i$*

*Proof.*

1. $K_{\lambda_i} \cap K_{\lambda_j} = \{0\}$ for $i \neq j$
2. By part 1, $\beta$ is linear independent. Also, $\beta$ spans $V$ by the previous proposition 2.2.11.
3. $dim(V) = |\beta| = |\beta_1| + |\beta_2| + ... + |\beta_k| \leq m_1 + m_2 + ... + m_k = dimV$. Moreover, note by proposition 2.2.10, we have $|\beta_i| \leq m_i$, thus we must have $|\beta_i| = m_i$.

♡

**Remark 2.2.13.** Now, the question is, how to find the Jordan Canonical Form of a matrix/operator?

Let $T : V \to V$ and $f(x) = \prod_{i=1}^{r}(x - \lambda_i)^{m_i}$ be the characteristic polynomial. Let $A = [T]_\sigma$ where $\sigma$ is the standard basis. We claim there exists Jordan matrix $J$ such that $A = PJP^{-1}$ where $P$ is an invertible matrix.

To compute it, first fix $\lambda = \lambda_1$, let $d_1 = dim(Null(A - \lambda I))$ and $\gamma_1$ be a basis for $Null(A - \lambda I)$, we have $d_1$ is the number of Jordan blocks in the form of $J_\lambda(k)$ where $k \in \mathbb{N}$.

Next, let $d_2 = dim(Null(A - \lambda I)^2)$, we have $d_2$ is the number of 2nd columns, thus $d_2 - d_1$ is the number of Jordan blocks in the form $J_\lambda(k)$ where $k \geq 2$. Then, we extend $\gamma_1$ to $\gamma_2$ be a basis for $Null(A - \lambda I)^2$ by solving $(A - \lambda I)x = v$ for all $v \in \gamma_1$.

Moreover, let $d_3 = dim(Null(A - \lambda I)^2)$, we have $d_3 - d_2$ would be the number of Jordan blocks $J_\lambda(k)$ where $k \geq 3$. Extend $\gamma_2$ to a basis $\gamma_3$ for $Null(A - \lambda I)^3$ by solving $(A - \lambda I)x = v$ for all $v \in \gamma_2 \backslash \gamma_1$.

We continue this way until $d_l = m_1 = dim(K_1)$, and we get a basis for $K_\lambda$, call it $\beta_1$. Then, we repeat the above procedure for $\lambda_i$ where $2 \leq i \leq r$, and for each $K_{\lambda_i}$, we would get a ordered basis $\beta_i$, let $\mathcal{B} = \bigcup_{i=1}^{r} \beta_i$. In this basis $\mathcal{B}$, we have $[T]_{\mathcal{B}} = J$ and therefore $A = PJP^{-1}$ where $P = [I]_{\mathcal{B}}^{\sigma}$.

Any $J$ computed in this way is called ***the*** Jordan Canonical Form of $T$ or $A$. This is unique up to block reordering.

**Example 2.2.14.** Calculate the Jordan Canonical Form of the following matrices:

$$A_1 = \begin{bmatrix} 2 & 3 & 2 \\ 1 & 8 & 2 \\ -2 & -14 & -3 \end{bmatrix} A_2 = \begin{bmatrix} 0 & 1 & 0 \\ -4 & 4 & 0 \\ -2 & 1 & 2 \end{bmatrix} A_3 = \begin{bmatrix} 1 & -3 & 3 \\ -2 & -6 & 13 \\ -1 & -4 & 8 \end{bmatrix}$$

$$A_4 = \begin{bmatrix} 3 & -1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 3 & 0 & 5 & -3 \\ 4 & -1 & 3 & -1 \end{bmatrix} A_5 = \begin{bmatrix} 1 & 1 & 1 & ... & 1 \\ 0 & 1 & 1 & ... & 1 \\ 0 & 0 & 1 & ... & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & ... & 1 \end{bmatrix}$$

*Solution.*

$A_1$**:** Note $f_1(x) = det(A_1 - xI) = (x-1)(x-3)^2$. Thus, we only have one Jordan block with diagonal $\lambda_1 = 1$. Next, we have

$$A_1 - \lambda_2 I = \begin{pmatrix} -1 & 3 & 2 \\ 1 & 5 & 2 \\ -1 & -14 & -6 \end{pmatrix} \Rightarrow \begin{pmatrix} -1 & 3 & 2 \\ 0 & 8 & 4 \\ 0 & 0 & 0 \end{pmatrix}$$

where $\lambda_2 = 3$. Thus, we have $rank(A_1 - \lambda_2 I) = 2$ and therefore, we have the number of Jordan blocks with 3 on the diagonal is $n - rank(A_1 - \lambda_2 I) = 3 - 2 = 1$. Hence the Jordan Canonical Form is $\begin{bmatrix} 1 & & \\ & 3 & 1 \\ & & 3 \end{bmatrix}$

$A_2$**:** Note $f_2(x) = det(A_2 - xI) = (x-2)^3$. Thus we calculate the rank $A_2 - 2I$ and obtain $rank(A - 2I) = 1$ and hence the number of Jordan blocks with 2 on the diagonal is $3 - 1 = 2$ and we get the Jordan Canonical Form is $\begin{bmatrix} 2 & & \\ & 2 & 1 \\ & & 2 \end{bmatrix}$

$A_3$**:** $f_3(x) = (x-1)^3$. Then $rank(A - I) = 2$ so the number of Jordan blocks with 1 on the diagonal is $3 - 2 = 1$ and hence the Jordan Canonical Form is $\begin{bmatrix} 1 & 1 & \\ & 1 & 1 \\ & & 1 \end{bmatrix}$

$A_4$**:** We have $f_4(x) = det(A_4 - xI) = (x-2)^4$. We have $rank(A-2I) = 2$. Thus the number of Jordan blocks with 2 on the diagonal is $4-2=2$. Next, we note $rank(A-2I)^2 + rank(A-2I)^0 - 2rank(A-2I) = 0+4-4 = 0$ (the trick we used here is covered in additional materials about Jordan form), thus the number of Jordan blocks with size 1 is 0 and hence the only possible Jordan matrix would be $\begin{bmatrix} 2 & 1 & & \\ & 2 & & \\ & & 2 & 1 \\ & & & 2 \end{bmatrix}$

$A_5$**:** Note $f_5(x) = det(A_5 - xI) = (x-1)^n$. Next, we have $rank(A-1I) = n-1$ and thus the number of Jordan blocks with 1 on the diagonal is $n-(n-1) = 1$ and hence the Jordan Canonical Form is $J_1(n)$. ♠

**Example 2.2.15.** Let $T : M_2(\mathbb{R}) \to M_2(\mathbb{R})$ be $T(A) = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} A$. Let $\sigma = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$. Then $[T]_\sigma = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$. Thus we have $n - rank(A-I) = 4-2$. Hence we have two Jordan blocks. Next, we have $n - rank(A-I)^2 = 4-0 = 4$ and thus we have $4-2$ many Jordan blocks at least size 2. Thus we have the Jordan Canonical Form $\begin{bmatrix} 1 & 1 & & \\ & 1 & & \\ & & 1 & 1 \\ & & & 1 \end{bmatrix}$

**Proposition 2.2.16.** *Let $T : V \to V$ be linear. Let the minimal polynomial be $m(x) = \prod_{i=1}^{k}(x-\lambda_i)^{m_i}$ where $\lambda_i \neq \lambda_j$. Then $m_i$ is the size of the largest $\lambda_i-$ Jordan block in the Jordan Canonical Form of $T$.*

*Proof.* Let $[T]_\sigma = A$, $A = PJP^{-1}$. Thus we must have $m(J) = 0$. Hence, we must have $m(J) = m(diag(J_1, J_2, ..., J_k)) \Leftrightarrow m(J_i) = 0$ for all Jordan blocks on the diagonal of $J$. Fix $\lambda = \lambda_i$ and let $J_i$ be a $\lambda-$Jordan block. Note $\mu \neq \lambda$, then $det((J_i - \mu I)) \neq 0$ and it must be invertible. Therefore,

$$m(J_i) = (J_i - \lambda I)^{m_i} \cdot \prod_{\mu \neq \lambda, t \neq i} (J_j - \mu I)^{m_t} \Rightarrow (J_i - \lambda I)^{m_i} = 0$$

We know $J_i - \lambda I$ is a matrix with one above diagonal, hence $m_i$ must be the size of the largest $\lambda$-Jordan blocks as it is the smallest positive integer makes $J$ equal zero. ♡

**Remark 2.2.17.** The following are additional materials ***not covered in lecture***. In addition, all the following proofs are from the book *Advanced Algebra* by Qiu Weisheng. Click here to the next chapter.

### 2.2.1 Quotient Space

**Definition 2.2.18.** Let $V$ be vector space and $W \leq V$. We define

$$a \sim_W b \Leftrightarrow a - b \in W$$

be a binary relation. It is easy to verify that it is a equivalence relation.

Let $a \in V$, we define the ***coset*** of $W$ in $V$ containing $a$ to be $\overline{a} = \{b \in V : b \sim_W a\}$ and write $a + W := \overline{a}$. We say $a$ is a representative of $\overline{a}$.

**Definition 2.2.19.** Let $V$ be a vector space over $\mathbb{F}$ and $W$ be a subspace. We define $V/W := \{a + W : a \in V\}$ and say $V/W$ is a quotient set of $V$ by $W$.

**Remark 2.2.20.** Then, we will define vector addition and scalar operations on this quotient set to form a vector space. It is intuitive to think that $\overline{a} + \overline{b} = \overline{a+b}$ and $\overline{ca} = c\overline{a}$ where $a, b \in V/W$ and $c \in \mathbb{F}$. The readers are encouraged to show indeed this two operations are well-defined and it forms a vector space over $\mathbb{F}$.

In the following, we will say $V/W$ is a ***quotient space*** of $V$ by $W$.

**Remark 2.2.21.** It is easy to see that, for all $w \in W$, we have $\overline{w} = \overline{0} = 0 + W = W$ in $V/W$.

**Theorem 2.2.22.** *Let $V$ be finite dimensional vector space over $\mathbb{F}$ and $W \leq V$. We have*

$$dim(V/W) = dim(V) - dim(W)$$

*Proof.* Let $(\alpha_1, ..., \alpha_s)$ be a basis for $W$, extend it to a basis for $V$, say

$$(\alpha_1, ..., \alpha_s, \alpha_{s+1}, ..., \alpha_n)$$

Then we have $dim(V) - dim(W) = n - s$.

Next, we try to find a basis for $V/W$. Let $\overline{v} \in V/W$. Say $v = \sum_{i=1}^{n} b_i\alpha_i$, then we have

$$\begin{aligned}
\overline{v} &= \overline{\sum_{i=1}^{n} b_i\alpha_i} \\
&= \overline{0} + \overline{0} + ... + \overline{0} + \sum_{i=s+1}^{n} b_i\overline{\alpha_i} \qquad \text{By Remark 2.2.21} \\
&= \sum_{i=s+1}^{n} b_i\overline{\alpha_i}
\end{aligned}$$

Next, suppose there exists $k_1, ..., k_{n-s} \in \mathbb{F}$ where $\sum_{i=1}^{n-s} k_i\overline{\alpha_{s+i}} = W = 0_{V/W}$. This imply

$$\overline{\sum_{i=1}^{n-s} k_i\alpha_{s+i}} = W \Rightarrow \sum_{i=1}^{n-s} k_i\alpha_{s+i} \in W$$

Then there exists $l_1, ..., l_s \in \mathbb{F}$ such that

$$\sum_{i=1}^{n-s} k_i \alpha_{s+i} = \sum_{i=1}^{s} l_i \alpha_i \Rightarrow \sum_{i=1}^{n-s} k_i \alpha_{s+i} - \sum_{i=1}^{s} l_i \alpha_i = 0 \in V$$

Since $(\alpha_1, ..., \alpha_n)$ is linear independent as it is a basis, we must have

$$l_1 = l_2 = ... = l_s = k_1 = ...k_{n-s} = 0$$

Thus $\overline{\alpha_{s+1}}, \overline{\alpha_{s+2}}, ..., \overline{\alpha_n}$ is a basis for $V/W$ as every vector in $V/W$ can be written as linear combination of them and they are linear independent. The proof follows as $dim(V/W) = |\{\overline{\alpha_{s+1}}, \overline{\alpha_{s+2}}, ..., \overline{\alpha_n}\}| = n - s = dim(V) - dim(W)$ ♡

**Definition 2.2.23.** Let $V$ be vector space over $\mathbb{F}$ and $W \leq V$ then we say a subspace $W^c$ is a ***complement space*** of $W$ if $W \oplus W^c = V$.

**Definition 2.2.24.** Let $V$ be over $\mathbb{F}$ and $W \leq V$. Then there exists a ***canonical mapping*** from $V$ to $V/W$, namely,

$$\pi : \alpha \mapsto \alpha + W$$

**Remark 2.2.25.** The canonical mapping is surjective. If $W$ is not the zero subspace, then $\pi$ is not injective. It is easy to see that the fiber $\sigma^{-1}(x)$ (or inverse image) of $\overline{x} \in V/W$ in $V$ is the coset $x + W$.

In addition, it is easy to show that $\pi$ is a homomorphism, that is, $\forall x, y \in V, k \in \mathbb{F}$, $\pi(x + y) = \pi(x) + \pi(y)$ and $\pi(kx) = k\pi(x)$. By this, we also deduce that $\mathbb{S} = \{\overline{\alpha_i} : \alpha_i \in V, i \in I\}$ is linear independent in $V/W$ then $S = \{\alpha_i \in V : i \in I\}$ is linear independent in $V$, where $I$ is an index set.

**Proposition 2.2.26.** *Let $V$ be a vector space over $\mathbb{F}$, finite dimensional or infinite, then every $W \leq V$ has a complement space*

*Proof.* Consider the quotient space $V/W$, let a basis for $V/W$ be $\mathbb{S} = \{\overline{\alpha_i} : \alpha_i \in V, i \in I\}$ where $I$ is a index set. Thus, we must have $S = \{\alpha_i \in V : i \in I\}$ to be linear independent by Remark 2.2.25. Let $U$ be the subspace span by $S$. Thus $S$ is a basis for $U$. We will show $U \oplus W = V$.

Let $v \in V$, since $\mathbb{S}$ is a basis for $V/W$, we have

$$v + W = \sum_{j=1}^{t} l_j(\alpha_{i_j} + W) = (\sum_{j=1}^{t} l_j \alpha_{i_j}) + W$$

where $\alpha_{i_j} \in S$ for all $1 \leq j \leq t$. Thus we have $v - \sum_{j=1}^{t} l_j \alpha_{i_j} \in W$. Let $\gamma = \sum_{j=1}^{t} l_j \alpha_{i_j}$, we have $\gamma \in U$, and therefore $v - \gamma \in W$. Hence, there exists $\delta \in W$ such that $v - \gamma = \delta$, that is, $v = \gamma + \delta$. Since $v$ was arbitrary, we have $V = W + U$.

Now, we just need to show $W \cap U = \{0\}$.

Let $\beta \in W \cap U$, we have $\beta = \sum_{j=1}^{r} k_j \alpha_{i_j}$ as $\beta \in U$. In addition, we have $\beta \in W$ thus

$$W = \beta + W = \sum_{j=1}^{r} k_j \alpha_{i_j} + W = \sum_{j=1}^{r} k_j(\alpha_{i_j} + W)$$

However, since $\mathbb{S}$ is linear independent, we must have $k_1 = k_2 = ... = k_r = 0$. Therefore $\beta = 0$ and the proof follows. ♡

### 2.2.2 Jordan Canonical Form for Nilpotent Transformation

**Definition 2.2.27.** Let $V$ be vector space over $\mathbb{F}$ and $B : V \to V$. If $\exists \eta \in V$ and $t \in \mathbb{N}$ such that $B^{t-1}\eta \neq 0$ and $B^t = 0$, then we call $\langle \eta, B^1\eta, B^2\eta, ..., B^{t-1}\eta \rangle$ the ***strong B-cyclic subspace generated by*** $\eta$

**Definition 2.2.28.** Let $B : V \to V$ and $B^l = 0$ for some $l \in \mathbb{N}$ then we say $B$ is ***niloptent transformation***. In addition, the smallest $m \in \mathbb{N}$ such that $B^m = 0$ is called the ***index*** of $B$.

**Remark 2.2.29.** Note we will write $span(S) = \langle S \rangle$ where $S$ is a set of vectors.

**Theorem 2.2.30.** *Let $W$ be a vector space over $\mathbb{F}$ with $dim(W) = r$, and let $B : W \to W$ be niloptent transformation with index $m$. Then $W$ can be written as a direct sum of $k$ strong B-cyclic subspaces where $k = dim(Null(B))$.*

*Proof.* We use induction on the dimension of $W$. If $r = 1$ then we must have $m = 1$. Therefore, $B = 0$ and take any $x \neq 0 \in W$, we have $Bx = 0$, thus $\langle x \rangle$ is a strong B-cyclic subspace where $W = \langle x \rangle$.

Suppose the claim holds for all vector space with dimension less than $r$.

Let $dim(W) = r$. Since $B$ is niloptent, 0 must be one of $B's$ eigenvalue. Thus, $E_0 = W_0 \neq \{0\}$ since it must contain at least one non-zero eigenvector of 0. Thus, $dim(W/W_0) = dim(W) - dim(W_0) < r$. Denote the linear transformation in $W/W_0$ induced by $B$ to be $\tilde{B}$. That is, we define $\tilde{B}(\overline{x}) = B(x) + W_0$ and $c\tilde{B}(\overline{x}) = cB(x) + W_0$. Then, for all $\overline{v} \in W/W_0$, we have

$$\tilde{B}^m(\overline{v}) = B^m(v) + W_0 = 0 + W_0 = W_0$$

Therefore, $\tilde{B}^m = 0$ and $\tilde{B}$ is niloptent transformation in $W/W_0$.

Hence, by our inductive hypothesis, we have $W/W_0$ is a sum of $s$ many strong $\tilde{B}$-cyclic subspaces. Say

$$\begin{aligned} W/W_0 &= \langle \tilde{B}^{t_1-1}(\overline{\xi_1}), \tilde{B}^{t_1-2}(\overline{\xi_1}), ..., \tilde{B}^1(\overline{\xi_1}), \overline{\xi_1} \rangle \\ &\quad \oplus \langle \tilde{B}^{t_2-1}(\overline{\xi_2}), \tilde{B}^{t_2-2}(\overline{\xi_2}), ..., \tilde{B}^1(\overline{\xi_2}), \overline{\xi_2} \rangle \\ &\quad \vdots \\ &\quad \oplus \langle \tilde{B}^{t_s-1}(\overline{\xi_s}), \tilde{B}^{t_s-2}(\overline{\xi_s}), ..., \tilde{B}^1(\overline{\xi_s}), \overline{\xi_s} \rangle \\ &= \bigoplus_{i=1}^{s} \langle \tilde{B}^{t_i-1}(\overline{\xi_i}), \tilde{B}^{t_i-2}(\overline{\xi_i}), ..., \tilde{B}^1(\overline{\xi_i}), \overline{\xi_i} \rangle \end{aligned}$$

where $\tilde{B}^{t_j}(\overline{\xi}) = W_0$ for all $1 \le j \le s$ and $s$ is the dimension of the 0-eigenspace of $\tilde{B}$ in $W/W_0$. Thus, we have

$$(\tilde{B}^{t_1-1}(\overline{\xi_1}), ..., , \overline{\xi_1}, \tilde{B}^{t_2-1}(\overline{\xi_2}), ..., \overline{\xi_2}, ..., \tilde{B}^{t_s-1}(\overline{\xi_s}), ..., \overline{\xi_s})$$

is a basis for $W/W_0$.

Let

$$U = \langle B^{t_1-1}(\xi_1), ..., , \xi_1, B^{t_2-1}(\xi_2), ..., \xi_2, ..., B^{t_s-1}(\xi_s), ..., \xi_s\rangle$$

As we showed in the proof of Proposition 2.2.26, we have $W = U \oplus W_0$ and

$$B^{t_1-1}(\xi_1), ..., , \xi_1, B^{t_2-1}(\xi_2), ..., \xi_2, ..., B^{t_s-1}(\xi_s), ..., \xi_s$$

is a basis for $U$.

Because $B^{t_j}(\xi_j) + W_0 = \tilde{B}^{t_j}(\overline{xi_j}) = W_0$, we have $B^{t_j}(\xi_j) \in W_0$ for all $1 \le j \le s$.

Let $c_1, ..., c_s \in \mathbb{F}$, and suppose

$$\sum_{i=1}^{s} c_i B^{t_i}(\xi_i) = 0$$

then we must have

$$B(\sum_{i=1}^{s} c_i B^{t_i-1}(\xi_i)) = 0$$

Therefore, we have $\sum_{i=1}^{s} c_i B^{t_i-1}(\xi_i) \in W_0 = E_0$. Thus

$$\sum_{i=1}^{s} c_i(\tilde{B}^{t_i-1}(\overline{\xi_i})) = W_0$$

However, $(\tilde{B}^{t_1-1}(\overline{\xi_1}), ..., \tilde{B}^{t_s-1}(\overline{\xi_s}))$ is a basis of $W/W_0$, and hence linear independent, we must have $c_1 = c_2 = ... = c_s = 0$.

Therefore, because $\sum_{i=1}^{s} c_i B^{t_i}(\xi_i) = 0$ imply $c_1 = ... = c_s = 0$, we must have $\{B^{t_1}(\xi_1), ..., B^{t_s}(\xi_s)\}$ is linear independent in $W_0$. Extent this linear independent set $\{B^{t_1}(\xi_1), ..., B^{t_s}(\xi_s)\}$ to a basis of $W_0$, say

$$(B^{t_1}(\xi_1), ..., B^{t_s}(\xi_s), \eta_1, ..., \eta_q)$$

we immediately get

$$W = \langle B^{t_1}(\xi_1), ..., \xi_1\rangle + ... + \langle B^{t_s}(\xi_s), ..., \xi_s\rangle + ... + \langle \eta_1\rangle + ... + \langle \eta_q\rangle$$

Because $B^{t_j}(\xi_j) \ne 0$ by the definition of linear independence, and $B^{t_j}(\xi_j) \in W_0$, we have $B(B^{t_j}(\xi_j)) = 0 = B^{t_j+1}(\xi_j)$ for all $1 \le j \le s$. Therefore, for all $1 \le j \le s$, we have $\langle B^{t_1}(\xi_1), ..., \xi_1\rangle$ is strong $B-$cyclic subspace. Note $B\eta_i = 0$, each of $\langle \eta_i\rangle$ is strong $B-$cyclic subspace as well.

Hence, recall that $W = U \oplus W_0$, where $K_1 := \{B^{t_1}(\xi_1), B^{t_2}(\xi_2)..., B^{t_s}(\xi_s), \eta_1, ..., \eta_q\}$ is a basis of $W_0$ and $K_2 := \{B^{t_1-1}(\xi_1), ..., , \xi_1, B^{t_2-1}(\xi_2), ..., \xi_2, ..., B^{t_s-1}(\xi_s), ..., \xi_s\}$ is a basis for $U$, we must have $K_1 \cup K_2$ is a basis for $W$ and hence they are linear independent, and hence

$$W = \langle B^{t_1}(\xi_1), ..., \xi_1 \rangle \oplus ... \oplus \langle B^{t_s}(\xi_s), ..., \xi_s \rangle \oplus ... \oplus \langle \eta_1 \rangle \oplus ... \oplus \langle \eta_q \rangle \tag{2.1}$$

From $K_1$ is a basis of $W_0$, we know $W$ is indeed the direct sum of $dim(W_0)$ many strong $B$-cyclc subspaces. $\heartsuit$

**Theorem 2.2.31.** *Let $W$ be vector space over $\mathbb{F}$ with $dim(W) = r$, and $B : W \to W$ be niloptent with index $l$. Then there exists a basis $\beta$ of $W$ such that $[B]_\beta$ is a Jordan matrix. In particular,*

1. *every Jordan block is in the form of $J_0(k)$ where $k \leq l$*
2. *the number of Jordan blocks is $dim(Null(B)) = r - rank(B)$*
3. *the number of $J_0(t)$, written as $N(t)$, is equal to $rank(B^{t+1}) + rank(B^{t-1}) - 2rank(B^t)$*

*Proof.* By Theorem 2.2.30, we have $W$ can be decompose into the direct sum of $dim(W_0)$ many strong $B-$cyclic subspaces. In every strong $B$-cyclic subspaces, the restriction of $B$ in that subspace with the basis mentioned in the previous proof (recall (2.1)) would be a Jordan block with zero on the diagonals and $J_0(k)$ must have $k \leq l$. Let $\beta$ be the basis of $W$ described in (2.1), then $[B]_\beta$ is a Jordan matrix with zero on the diagonal. In addition, the number of Jordan blocks is $dim(W_0) = dim(Null(B)) = r - rank(B)$ as we can see from the above argument as well.

Next, we calculate $N(t)$ where $t \leq l$. Note, when $m < n$, we have

$$J_0(n)^m = \begin{array}{cc} \begin{array}{l} \quad\overbrace{\hphantom{0 \quad \cdots \quad 0}}^{\text{m column}} \end{array} & \\ \begin{bmatrix} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & 1 & \cdots & 0 \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 1 \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \end{bmatrix} & \begin{matrix} \\ \\ \\ \\ \left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} \text{m row} \end{matrix} \end{array}$$

and when $m \geq n$, we have $J_0(n)^m = 0$. Thus

$$rank(J_0(n)^m) = \begin{cases} n - m \text{ when } m < n \\ 0 \text{ otherwise} \end{cases}$$

This imply that, for all given $t \in \mathbb{N}$, in order to calculate $N(t)$, we consider $B^{t-1}$, where all Jordan blocks with size smaller than $t$ will reduce to zero matrix. Therefore,

$$\begin{aligned}
rank(B^{t-1}) &= N(t)(t-(t-1)) + N(t+1)((t+1)-(t-1)) + ... + N(l)(l-(t-1)) \\
&= N(t) + 2N(t+1) + ... + (l-t)N(l)
\end{aligned} \tag{2.2}$$

Next, consider $B^t$, and obtain

$$rank(B^t) = N(t+1) + 2N(t+2) + ... + (l-t)N(l) \tag{2.3}$$

Therefore, (2.2) subtract (2.3), we get

$$rank(B^{t-1}) - rank(B^t) = N(t) + N(t+1) + ... + N(l) \tag{2.4}$$

Substitute $t$ to be $t+1$, when $t \leq l-1$, from (2.4) we get

$$rank(B^t) - rank(B^{t+1}) = N(t+1) + ... + N(l) \tag{2.5}$$

Therefore, when $t \leq l-1$, (2.4) subtract (2.5) imply $rank(B^{t-1}) + rank(B^{t+1}) - 2rank(B^t) = N(t)$

Obviously when $t = l$ we have our claim holds. ♡

### 2.2.3 Jordan Canonical Form, Second Proof

**Theorem 2.2.32.** *Let $V$ be a vector space over $\mathbb{F}$ and $T: V \to V$. Suppose the minimal polynomial of $A$, $m(x)$, splits over $F$ and $m(x) = \prod_{i=1}^{s}(x-\lambda_i)^{l_i}$. Then there exists a basis $\beta$ of $V$ such that $A := [T]_\beta$ is Jordan matrix. In particular, we have*

1. *the diagonal of $A$ contains only the eigenvalues of $T$*
2. *the number of Jordan blocks with $\lambda_j$ on the diagonal, write as $N_j$, is $N_j = n - rank(T-\lambda_j I)$*
3. *the number of $J_{\lambda_j}(t)$, write as $N_j(t)$, is*

$$N_j(t) = rank(T-\lambda_j I)^{t+1} + rank(T-\lambda_j I)^{t-1} - 2rank(T-\lambda_j I)^t$$

*where $t \leq l$ and $j = 1, 2, ..., s$*

*Proof.* It is easy to see $\lambda_1, ..., \lambda_s$ are all the eigenvalues of $T$. Thus, from Theorem 2.2.12, we have

$$V = K_{\lambda_1} \oplus K_{\lambda_2} \oplus ... \oplus K_{\lambda_s}$$

For all $1 \leq j \leq s$, denote $W_j = K_{\lambda_j}$, and let $B_j = T|W_j - \lambda_j I$ where $T|W_j$ is $T$ restricted to $W_j$. Note we have $B_j$ is niloptent transformation with index $l_j$, therefore, by Theorem 2.2.31, there exists a basis $\beta_j$ in $W_j$ such that $[B_j]_{\beta_j}$ is Jordan matrix. Hence $A_j := [T|W_j]_{\beta_j}$ is equal to $\lambda_j I + B_j$ and therefore each $A_j$ is a Jordan matrix with $\lambda_j$ on the diagonal. In particular, the number of Jordan blocks of with $\lambda_j$ on the diagonal is the number of Jordan blocks in $B_j$. Thus

$$\begin{aligned} N_j &= dim(Null(B_j)) = dim(Null(T|W_j - \lambda_j I)) \\ &= dim(Null(T - \lambda_j I)) = n - rank(T - \lambda_j I) \end{aligned}$$

Note here we used a fact that $Null(T|W_j - \lambda_j I) = Null(T - \lambda_j I)$.

The number of $J_{\lambda_j}(t)$ Jordan blocks, $N_j(t)$ is equal to the number of $t \times t$ Jordan blocks in $B_j$ and therefore, we have (3) checked as well by Theorem 2.2.31.

Let $\beta$ be the union of $\beta_j$, then we have found our desired basis for $V$. ♡

# Chapter 3

# Inner Product Spaces

## 3.1 Intro

**Remark 3.1.1.** Let $V$ be vector space over $\mathbb{F}$, where $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$, possiblely infinite dimensional.

**Definition 3.1.2.** An ***inner product*** on a vector space $V$ is a map $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{F}$ such that for all $x, y, z \in V$, $\sigma \in \mathbb{F}$, we have

1. $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$
2. $\langle \sigma x, y \rangle = \sigma \langle x, y \rangle$
3. $\langle y, x \rangle = \overline{\langle x, y \rangle}$
4. $\langle x, x \rangle \in \mathbb{R}$, then $\langle x, x \rangle \geq 0$ and $\langle x, x \rangle = 0$ iff $x = 0$

**Remark 3.1.3.** We have

1. $\langle x, x \rangle = \overline{\langle x, x \rangle} \Rightarrow \langle x, x \rangle \in \mathbb{R}$
2. $\langle x, y + z \rangle = \overline{\langle y + z, x \rangle} = \overline{\langle y, x \rangle + \langle z, x \rangle} = \overline{\langle y, x \rangle} + \overline{\langle z, x \rangle} = \langle x, y \rangle + \langle x, z \rangle$
3. $\langle x, \sigma y \rangle = \overline{\langle \sigma y, x \rangle} = \overline{\sigma \langle y, x \rangle} = \overline{\sigma} \langle x, y \rangle$
4. $\langle x, 0 \rangle = \langle x, 0 \cdot 0_V \rangle = \overline{0} \langle x, 0_V \rangle = 0$
5. $\langle 0, x \rangle = 0$

**Definition 3.1.4.** If $V$ is a vector space with an inner product, we call it an ***Inner Product Space***

**Proposition 3.1.5.** *Let $V$ be inner product space. If $y, z \in V$ and $\forall x \in V$, $\langle x, y \rangle = \langle x, z \rangle$ then $y = z$. In particular, if $\langle x, y \rangle = 0$ for all $x \in V$, then $y = 0$.*

*Proof.* Note $\langle x, y \rangle = \langle x, z \rangle$ then $\langle x, y \rangle - \langle x, z \rangle = 0$ then $\langle x, y - z \rangle$. When $x = y - z$, we have $\langle y - z, y - z \rangle = 0$ if and only if $y - z = 0$ if and only if $y = z$. ♡

**Definition 3.1.6.** Let $V = \mathbb{F}^n$, $v = (v_1, ..., v_n)$ and $w = (w_1, ..., w_n)$, we have the ***standard inner product*** be $\langle v, w \rangle = \sum_{i=1}^{n} v_i \overline{w_i}$.

**Example 3.1.7.** Let $V$ be a vector space and $W$ be inner product space with the inner product $\langle \cdot, \cdot \rangle$, both over $F$. Prove $\langle \cdot, \cdot \rangle_T : V \times V \to F$ given by $\langle v, w \rangle_W = \langle T(v), T(w) \rangle$ is an inner product on $V$ if and only if $T$ is injective.

*Solution.* If $T$ is injective, then $T(v) = T(u) \Rightarrow u = v$. We show $\langle v, w\rangle_T$ is an inner product by verifying the axioms. The other axioms could be easily checked, we only show that $\langle v, v\rangle_T \geq 0$. Since $T$ is injective, the kernel is trivial, and if $v \neq 0$, we have $T(v) \neq 0$, then

$$\langle v, v\rangle_T = \langle T(v), T(v)\rangle > 0$$

In particular, if $v = 0$, then $T(v) = 0$ and thus $\langle v, v\rangle_T = \langle 0, 0\rangle = 0$ and thus in general we have $\langle v, v\rangle \geq 0$ and $\langle v, v\rangle = 0$ if and only if $v = 0$.

Conversely, suppose $T$ is not injective, thus the kernel is not trivial. Say $v \neq 0 \in V$ so that $T(v) = 0$, then suppose $\langle \cdot, \cdot\rangle_T$ is an inner product. Therefore, $\langle v, v\rangle_T > 0$ as $v \neq 0$. However, $\langle v, v\rangle_T = \langle T(v), T(v)\rangle = \langle 0, 0\rangle = 0$, a contradiction. Hence we are finished. ♠

**Definition 3.1.8.** Let $A = (a_{ij}) \in M_n(\mathbb{F})$, the ***adjoint*** (or ***conjugate transpose***) of $A$ is the matrix $A^* \in M_n(\mathbb{F})$ defined by $A^* := (\overline{a_{ji}})$.

**Definition 3.1.9.** Let $A \in M_n(F)$, we say $A$ is ***positive definite*** if $A = A^*$ and $x^*Ax > 0$ for all $0 \neq x \in F^n$.

**Example 3.1.10.** Prove that $\langle \cdot, \cdot\rangle : F^n \times F^n \to F$ is an inner product on $F^n$ if and only if there exists positive definite matrix $A \in M_n(F)$ such that $\langle x, y\rangle = y^*Ax$ for all $x, y \in F^n$.

*Solution.* Let $V = F^n$. We first show that if $A = A^*$ and $x^*Ax > 0$ for all $0 \neq x \in V$, then $y^*Ax$ defines a inner product on $F^n$. We first note that $y^*Ax \in F$, thus $\overline{y^*Ax} = (y^*Ax)^* = x^*A^*y$ as the transpose of a complex number is just itself.

Let $x, y, z \in V, k \in F$.

$$\langle x + y, z\rangle = z^*A(x + y) = z^*Ax + z^*Ay = \langle x, z\rangle + \langle y, z\rangle$$

$$\langle kx, y\rangle = y^*A(kx) = k(y^*Ax) = k\langle x, y\rangle$$
$$\overline{\langle x, y\rangle} = \langle x, y\rangle^* = (y^*Ax)^* = x^*A^*y = x^*Ay = \langle y, x\rangle$$

If $x = 0$ then $x^*Ax = 0$, if $x \neq 0$ then first note $x^*Ax \in \mathbb{R}$, and

$$\langle x, x\rangle = x^*Ax > 0$$

Thus $\langle x, x\rangle \geq 0$ and $\langle x, x\rangle = 0 \Leftrightarrow x = 0$. Hence, $V$ does have an inner product defined by $\langle x, y\rangle = y^*Ax$.

Next, we suppose $V$ is an inner product space. Then, let $x, y \in V$, and $\sigma = \{e_1, ..., e_n\}$ be the standard basis. We have $x = \sum a_ie_i$ and $y = \sum b_ie_i$. Then,

$$\langle x, y\rangle = \langle \sum a_ie_i, \sum b_je_j\rangle = \sum_{i=1}^{n}\sum_{j=1}^{n} \bar{b_j}\langle e_i, e_j\rangle a_i$$

Define $A = (a_{ij})$ where $a_{ij} = \langle e_i, e_j \rangle$, then we have $\langle x, y \rangle = y^* A x$. In addition, we note $\langle e_i, e_j \rangle = \overline{\langle e_j, e_i \rangle}$ so $A^* = A$. Moreover, note $\langle x, x \rangle > 0$ if $x \neq 0$ which is the same as $x^* A x > 0$ if $x \neq 0$. Therefore, we indeed have $A$ is positive definite as desired. ♠

**Example 3.1.11.** Show that all finite dimensional vector spaces $V$ over $\mathbb{F}$ may be equipped with an inner product.

*Solution.* Since $V$ is finite dimensional, it is isomorphic to $\mathbb{F}^n$ where $n = dim(V)$. Thus, by the above example, we are done, as $\mathbb{F}^n$ is an inner product space. ♠

**Definition 3.1.12.** Let $V = M_n(\mathbb{F})$, then we define the ***Frobenius inner product*** to be $\langle A, B \rangle = tr(B^* A)$.

**Remark 3.1.13.** Let $A = (a_{ij})$ and $B = (b_{ij})$. Then we have $(B^* A)_{ii} = \sum_{k=1}^n \overline{b_{ki}} a_{ki}$. Thus $tr(B^* A) = \sum_{i=1}^n \sum_{k=1}^n \overline{b_{ki}} a_{ki} = \sum_{i,k} a_{ki} \overline{b_{ki}} = \langle v, w \rangle$ where $v, w \in \mathbb{F}^{n^2}$.

**Example 3.1.14.**

1. Let $V = \mathbb{F}^n$, we also have $\langle v, w \rangle' := r \langle v, w \rangle$ be an inner product where $r > 0 \in \mathbb{R}$.
2. Let $V = C[a, b]$ and let $\langle f, g \rangle = \frac{1}{b-a} \int_a^b f(x) g(x) dx$. It is an inner product. Also note $\langle 1, 1 \rangle = 1$.
3. Let $V = M_n(\mathbb{F})$, we have the Frobenius inner product is an inner product by Remark 3.1.13
4. Let $l^2(\mathbb{F}) := \{(x_n)_{n=1}^\infty : x_i \in \mathbb{F}, \sum |x_i| < \infty\}$, then we have $\langle (x_n), (y_n) \rangle = \sum x_i \overline{y_i}$ is an inner product.

**Definition 3.1.15.** Let $V$ be any vector space over $\mathbb{F}$. A ***norm*** (or ***metric***) of $V$ is a map $\|\cdot\| : V \to \mathbb{F}$ such that:

1. For all $v \in V$, $\|v\| \geq 0$
2. For all $\alpha \in \mathbb{F}$, $v \in V$, we have $\|\alpha v\| = |\alpha| \cdot \|v\|$
3. For all $v, w \in V$, we have $\|v + w\| \leq \|v\| + \|w\|$

We call it a normed vector space.

**Theorem 3.1.16** (Cauchy-Schwarz inequality)**.** *Let $V$ be an inner product space. Define $\|x\| = \sqrt{\langle x, y \rangle}$. For all $x, y \in V$, we have*

$$|\langle x, y \rangle| \leq \|x\| \cdot \|y\|$$

*Proof.* If $y = 0$ then the result is trivial. Assume $y \neq 0$, so that $\langle y, y \rangle > 0$. Let

$\alpha = \frac{\langle x,y \rangle}{\langle y,y \rangle} \in F$, we have

$$\begin{aligned}
0 &\le \|x - \alpha y\|^2 \\
&= \langle x - \alpha y, x - \alpha y \rangle \\
&= \langle x, x \rangle - \overline{\alpha}\langle x, y \rangle - \alpha \langle y, x \rangle + \alpha\overline{\alpha}\langle y, y \rangle \\
&= \langle x, x \rangle - \frac{\langle y, x \rangle}{\langle y, y \rangle}\langle x, y \rangle - \frac{\langle x, y \rangle}{\langle y, y \rangle}\langle y, x \rangle + \frac{\langle x, y \rangle}{\langle y, y \rangle}\frac{\langle y, x \rangle}{\langle y, y \rangle}\langle y, y \rangle \\
&= \langle x, x \rangle - 2\frac{\langle y, x \rangle\langle x, y \rangle}{\langle y, y \rangle} + \frac{\langle y, x \rangle\langle x, y \rangle}{\langle y, y \rangle} \\
&= \langle x, x \rangle - \frac{\langle y, x \rangle\langle x, y \rangle}{\langle y, y \rangle}
\end{aligned}$$

Thus, we have

$$\begin{aligned}
&\langle x, x \rangle \ge \frac{\langle x, y \rangle\langle y, x \rangle}{\langle y, y \rangle} \\
&\Rightarrow \langle x, x \rangle\langle y, y \rangle \ge \langle x, y \rangle\overline{\langle x, y \rangle} \\
&\Rightarrow \|x\|^2 \|y\|^2 \ge |\langle x, y \rangle|^2 \\
&\Rightarrow \|x\| \|y\| \ge |\langle x, y \rangle|
\end{aligned}$$

♡

**Proposition 3.1.17.** *Let $V$ be an inner product space. By setting $\|x\| := \sqrt{\langle x, x \rangle}$ for all $x \in V$, we have this defining an norm on $V$.*

*Proof.* (1) and (2) as homework.

To check (3), let $x, y \in V$, we have

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

♡

**Example 3.1.18.**

1. Let $v = (-1, i, 2 + i) \in \mathbb{C}^3$, then we have $\|v\| = \sqrt{7}$.
2. Let $v = \begin{bmatrix} -1 & 3-i \\ 4 & i \end{bmatrix}$, then we have $\|v\| = \sqrt{tr(\begin{bmatrix} -1 & 4 \\ 3+i & -i \end{bmatrix}\begin{bmatrix} -1 & 3-i \\ 4 & i \end{bmatrix})} = \sqrt{28}$
3. Let $f(x) = e^x \in C[0, 1]$, then we have $\|f(x)\| = \sqrt{\int_0^1 e^{2x} dx} = \sqrt{\frac{e^2-1}{2}}$

**Example 3.1.19.** Let $V$ be inner product space. Try to show for all $x, y \in V$, we have $\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2$

*Solution.* Let $x, y \in V$, then

$$\begin{aligned}\|x+y\|^2 + \|x-y\|^2 &= \langle x+y, x+y\rangle + \langle x-y, x-y\rangle \\ &= \langle x,x\rangle + \langle x,y\rangle + \langle y,x\rangle + \langle y,y\rangle + \langle x,x\rangle \\ &\quad - \langle x,y\rangle - \langle y,x\rangle + \langle y,y\rangle \\ &= 2\langle x,x\rangle + 2\langle y,y\rangle = 2\|x\|^2 + 2\|y\|^2\end{aligned}$$

♠

## 3.2 Orthogonality

**Remark 3.2.1.** Let $V = \mathbb{R}^2$, let $x, y \in V$ be non-zero. Recall that we have, by Cosine Law that $c^2 = a^2 + b^2 - 2ab\cos c$. This is the same as $\|x-y\|^2 = \langle x,x\rangle - \langle x,y\rangle - \langle y,x\rangle + \langle y,y\rangle = \|x\|^2 + \|y\| - 2\|x\|\|y\|\cos\theta$. Thus we have $\cos\theta = \frac{\langle x,y\rangle}{\|x\|\|y\|}$. Motivated by this, we say $x, y$ are perpendicular iff $\cos\theta = 0$ iff $\langle x,y\rangle = 0$.

**Definition 3.2.2.** Let $V$ be inner product space, we say $u, v \in V$ are ***orthogonal*** if $\langle u,v\rangle = 0$.

In addition, we say $S \subseteq V$ is ***orthogonal***, if whenever $x, y \in S$, $x \neq y$, then $x, y$ are orthogonal.

We say $S \subseteq V$ is ***orthonormal*** if $S$ is orthogonal and $\forall x \in S$, $\|x\| = 1$.

**Example 3.2.3.** Let $\beta = \{e_1, ..., e_n\}$ be the standard basis for $\mathbb{F}^n$. We have $\beta$ is orthonormal basis.

**Example 3.2.4.** Let $S = \{1, x, x^2\} \subseteq C[0,1]$, we have $S$ is not orthogonal as $\int_0^1 1 \cdot x dx = 1/2 \neq 0$.

However, let $S = \{1, x, x^2\} \subseteq P_2(\mathbb{R})$, we define $\langle ax^2 + bx + c, a'x^2 + b'x + c'\rangle = aa' + bb' + cc'$ as the inner product, then we have $S$ is orthonormal.

**Remark 3.2.5.** Let $V$ be inner product space, and say $S \subseteq V\backslash\{0\}$ is orthogonal. Say $S = \{v_1, v_2, v_3, ...\}$, then $S' = \{\frac{1}{\|v_1\|}v_1, \frac{1}{\|v_2\|}v_2, ...\}$ is orthonormal.

**Example 3.2.6.** Let $H = \{f \in C([0, 2\pi] \to \mathbb{C})\}$ be the continuous complex functions with domain $[0, 2\pi]$. Let $\langle f, g\rangle = \frac{1}{2\pi}\int_0^{2\pi} f(t)\overline{g(t)}dt$. Recall that for $f \in H$ we have $f(x) = u(x) + iv(x)$ where $u(x), v(x)$ are continuous real functions, and we define $\int f(t)dt = \int u(t)dt + i\int v(t)dt$.

Next, we let $S \subseteq H$ to be $S = \{f_n : n \in \mathbb{Z}\}$, where $f_n(t) = e^{int} = cos(nt) + isin(nt)$. Then, we have $S$ is orthonormal.

**Proposition 3.2.7.** *Let $V$ be an inner product space, and let $S = \{v_1, ..., v_k\} \subseteq V$ be orthogonal, and $\forall i, v_i \neq 0$. If $y \in span(S)$ and $y = \sum_{i=1}^k c_i v_i$, then*

$$c_i = \frac{\langle y, v_i\rangle}{\|v_i\|^2} = \frac{\langle y, v_i\rangle}{\langle v_i, v_i\rangle}$$

*In particular, if $S$ is orthonormal then $c_i = \langle y, v_i\rangle$.*

*Proof.* We have $\langle y, v_i\rangle = \langle\sum_{j=1}^k c_j v_j, v_i\rangle = \sum_{j=1}^k c_j\langle v_j, v_i\rangle = c_i\langle v_i, v_i\rangle$ and hence $c_i = \frac{\langle y, v_i\rangle}{\langle v_i, v_i\rangle}$ ♡

**Corollary 3.2.7.1.** *Let $V$ be inner product space and $S \subseteq V$ is orthogonal with non-zero vectors. Then $S$ is linear independent.*

*Proof.* Take $v_1, v_2, ..., v_k \in S$ and suppose $\sum c_i v_i = 0$. Then $c_i = \frac{\langle 0, v_i\rangle}{\langle v_i, v_i\rangle} = 0$ and hence linear independent. ♡

**Proposition 3.2.8.** *Suppose $A \in M_n(\mathbb{F})$ and $A = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{bmatrix}$ and $\{r_1, ..., r_n\}$ is orthogonal. Then $AA^*$ is the diagonal. Moreover, if $\{r_1, ..., r_n\}$ is orthonormal then $AA^* = I$.*

*Proof.* Note $AA^* = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{bmatrix} [\overline{r_1}, \overline{r_2}, ..., \overline{r_n}] = (a_{ij})$. Then we have $a_{ij} = \langle r_i, r_j\rangle \Rightarrow a_{ij} = 0$ if $i \neq j$ and $a_{ij} = \|r_i\|^2$ if $i = j$. The proof follows. ♡

**Remark 3.2.9** (Gram-Schmidt process)**.** Let $\{w_1, ..., w_k\} \subseteq V$ be linear independent. We want to find a orthogonal set $\{v_1, ..., v_k\}$ in $V$ such that $span(w_1, ..., w_k) = span(v_1, ..., v_k)$.

1. Take $v_1 = w_1$.
2. Next, we want $span(w_1, w_2) = span(v_1, v_2)$, in order to make the two span the same, we must have $v_2 = w_2 - \alpha v_1$ for some scalar $\alpha$. Thus, to make $\langle v_1, v_2\rangle = 0$, we must have $\alpha = \frac{\langle w_2, v_1\rangle}{\|v_1\|^2}$. Hence, $v_2 = w_2 - \frac{\langle w_2, v_1\rangle}{\|v_1\|^2} v_1$.
3. Then, we have $span(w_1, w_2, w_3) = span(v_1, v_2, w_3) = span(v_1, v_2, w_3 - \alpha v_1 - \beta v_2)$. Thus, we must have $\langle w_3 - \alpha v_1 - \beta v_2, v_1\rangle = 0$ and $\langle w_3 - \alpha v_1 - \beta v_2, v_2\rangle = 0$. This happen if and only if $\langle w_3 - \alpha v_1, v_1\rangle = 0$ and $\langle w_3 - \beta v_2, v_2\rangle = 0$. Thus $\alpha = \frac{\langle w_3, v_1\rangle}{\|v_1\|^2}$ and $\beta = \frac{\langle w_3, v_2\rangle}{\|v_2\|^2}$. Hence $v_3 = w_3 - \frac{\langle w_3, v_1\rangle}{\|v_1\|^2} v_1 - \frac{\langle w_3, v_2\rangle}{\|v_2\|^2} v_2$.
4. Continue this way by induction, we get
$$v_i = w_i - \sum_{j=1}^{j-1} \frac{\langle w_i, v_j\rangle}{\|v_j\|^2} v_j$$
for $2 \leq i \leq k$

**Theorem 3.2.10** (Gram-Schmidt)**.** *Let $V$ be inner product space, $S = \{w_1, ..., w_n\}$ be linear independent in $V$. Then $S' = \{v_1, ..., v_n\}$ defined recursively by $v_1 = w_1$ and $v_k = w_k - \sum_{j=1}^{k-1} \frac{\langle w_k, v_j\rangle}{\|v_j\|^2} v_j$ for $k \geq 2$ is an orthogonal set of nonzero vectors such that $span(S) = span(S')$.*

**Corollary 3.2.10.1.** *If $V$ is a finite dimension inner product space then $V$ have an orthonormal basis.*

**Remark 3.2.11.** Let $\{v_1, ..., v_n\} \subseteq V$ be orthogonal, then we have $\|v_1 + ... + v_n\| = \|v_1\|^2 + ... + \|v_n\|^2$.

**Example 3.2.12.** Let $W = span\{[1, 1, 0], [0, 2, 1]\} \subseteq \mathbb{R}^3$ and let $w_1 = [1, 1, 0]$ and $w_2 = [0, 2, 1]$. Then, we have $v_1 = w_1$ and $v_2 = w_2 - \frac{\langle w_2, v_1 \rangle}{\langle v_1, v_1 \rangle} v_1 = [0, 2, 1] - \frac{2}{2}[1, 1, 0] = [-1, 1, 1]$. Thus $\{v_1, v_2\}$ is an orthogonal basis for $W$ and thus $\{\frac{1}{\sqrt{2}} v_1, \frac{1}{\sqrt{3}} v_2\}$ is an orthonormal basis for $W$.

**Example 3.2.13.** Find an orthogonal basis for $P_2(\mathbb{R}) \subseteq C[0, 1]$

*Solution.* Let $\beta = \{1, x, x^2\} = \{w_1, w_2, w_3\}$ be the standard basis for $P_2(\mathbb{R})$. Then, we take $v_1 = w_1$. We have

$$\begin{aligned} v_2 &= w_2 - \frac{\langle w_2, v_1 \rangle}{\|v_1\|^2} v_1 \\ &= x - \frac{\int_0^1 x dx}{\int_0^1 1 dx} \cdot 1 \\ &= x - \frac{1}{2} \end{aligned}$$

Next, we have

$$\begin{aligned} v_3 &= w_3 - \frac{\langle w_3, v_2 \rangle}{\|v_2\|^2} v_2 - \frac{\langle w_3, v_1 \rangle}{\|v_1\|^2} v_1 \\ &= x^2 - \frac{\int_0^1 x^3 - 0.5x dx}{\int_0^1 (x - 0.5)^2 dx}(x - \frac{1}{2}) + \frac{\int_0^1 x^2 dx}{\int_0^1 1^2 dx}(1) \\ &= x^2 - x + \frac{1}{2} - \frac{1}{3} = x^2 - x + \frac{1}{6} \end{aligned}$$

♠

## 3.3 Orthogonal Projection

**Example 3.3.1.** Find the closest point on $x = \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 1 \\ -1 \end{bmatrix} t$ to the point $(3, 3)$.

*Solution.* Note the range of $\begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 1 \\ -1 \end{bmatrix} t$ is a affine space of $span(\begin{bmatrix} 1 \\ -1 \end{bmatrix})$. Thus, it suffice to find the closest point on $\begin{bmatrix} 1 \\ -1 \end{bmatrix} t$ to the point $(3, 3) - (1, 2) = (2, 1)$, then transform them back.

Let $v = (1, -1)$ and $u = (2, 1)$. Note $proj_v(u) = \frac{\langle u, v \rangle}{\|v\|^2} v = \frac{1}{2} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 1/2 \\ -1/2 \end{bmatrix}$. Therefore the closest point is $\begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 1/2 \\ -1/2 \end{bmatrix} = \begin{bmatrix} 3/2 \\ 3/2 \end{bmatrix}$ ♠

**Definition 3.3.2.** Let $A, B \leq V$ where $V$ is a subspace. We say $V$ is a ***direct sum*** of $A$ and $B$, and write $A \oplus B = V$, if $A + B := \{a + b : a \in A, b \in B\} = V$ and $A \cap B = \{0\}$.

**Proposition 3.3.3.** *Suppose $V = A \oplus B$,*

1. *every $v \in V$ can be uniquely written as $v = a + b$ where $a \in A, b \in B$,*
2. *if $\alpha$ is a basis for $A$ and $\beta$ is a basis for $B$ then $\alpha \cup \beta$ is a basis for $V$. In particular, $dim(V) = dim(A) + dim(B)$.*

*Proof.*

1. It suffice to show the uniqueness. $a + b = a' + b' \Rightarrow a - a' = b - b' \Rightarrow a - a' = 0 = b - b'$
2. Let $\alpha = \{v_1, ..., v_n\}, \beta = \{u_1, ..., u_m\}$ then since $V = A + B$, we have $\alpha \cup \beta$ spans $V$. Suppose $\sum c_i v_i + \sum d_j u_j = 0$, then $\sum c_i v_i = -\sum d_j w_j = 0$ and by linear independence, we have $c_i = 0$ and $d_j = 0$.

♡

**Definition 3.3.4.** Let $V$ be inner product space and let $\emptyset \neq S \subseteq V$, the ***orthogonal complement*** of $S$ is defined to be $S^\perp = \{x \in V : \forall v \in S, \langle v, x \rangle = 0\}$

**Remark 3.3.5.** $S^\perp$ is a subspace of $V$ even if $S$ is not a subspace of $V$.

**Theorem 3.3.6.** *Let $V$ be Inner Product Space and if $W$ is finite dimensional subspace of $V$, then $V = W \oplus W^\perp$.*

*Proof.* Let $v \in V$, and let $\beta = \{v_1, ..., v_k\}$ be an orthonormal basis for $W$. Let $u = \sum_{i=1}^k \langle v, v_i \rangle v_i \in W$ and $z = v - u$. Now, for every $v_j \in \beta$, we have $\langle z, v_j \rangle = \langle v - u, v_j \rangle = \langle v, v_j \rangle - \langle u, v_j \rangle = \langle v, v_j \rangle - \langle v, v_j \rangle \cdot \langle v_j, v_j \rangle = 0$.

Hence, $z \in W^\perp$ and $v = u + z$ so that $V = W + W^\perp$.

If $x \in W \cap W^\perp$, then we have $x$ is perpendicular to itself, and hence $\langle x, x \rangle = 0$ and hence $x = 0$. ♡

**Definition 3.3.7.** The vector $u$ in the previous proof is called the ***orthogonal projection*** of $v$ onto $W$. We write $u = proj_W(v)$.

**Theorem 3.3.8.** *Let $W$ be a finite dimensional subspace of an inner product space $V$. Let $v \in V$, thus $\exists$ unique $u \in W$ and $z \in W^\perp$ such that $v = u + z$. For any $x \in W$, $\|v - x\| \geq \|v - u\|$. We have equality if and only if $x = u$.*

*Proof.* Note

$$\begin{aligned}
\|v - x\|^2 &= \|u + z - x\|^2 \\
&= \|u - x + z\|^2 = \|u - x\|^2 + \|z\|^2 \\
&\geq \|z\|^2
\end{aligned}$$

Thus we have

$$\|v - x\| \geq \|z\| = \|v - u\|$$

♡

**Example 3.3.9.** Let $W = span\{(i,0,1+i),(0,-i,1)\} \subseteq \mathbb{C}^3$. We note $dim(W) = 2$ and therefore $dim(W^\perp) = 3-2 = 1$. Clearly, we have $(1-i,1,-i)$ is orthogonal to both $(i,0,1+i)$ and $(0,-i,1)$. Hence, $W^\perp = span\{(1-i,1,-i)\}$.

**Example 3.3.10.** Let $V = C[0,1]$ and $W = P_1(\mathbb{R})$. Find the closest element of $W$ to $f(x) = e^x \in V$.

*Solution.* We are looking for $proj_W(f(x))$. We know from Example 3.2.13 that $\beta = \{1, x - \frac{1}{2}\}$ is orthogonal basis for $W$. Thus, $\left\{1, \sqrt{12}\left(x-\frac{1}{2}\right)\right\}$ is orthonormal basis.

Therefore,

$$\begin{aligned}
\mathrm{proj}_W(f(x)) &= \langle e^x, 1\rangle + \left\langle e^x, \sqrt{12}\left(x-\frac{1}{2}\right)\right\rangle \left(\sqrt{12}\left(x-\frac{1}{2}\right)\right) \\
&= \int_0^1 e^x \,\mathrm{d}x + \left[\int_0^1 e^x \left(\sqrt{12}x - \sqrt{3}\right)\,\mathrm{d}x\right]\left(\sqrt{12}x-\sqrt{3}\right) \\
&= e - 1 + \left(e(\sqrt{12}-\sqrt{3}) + \sqrt{3} - \sqrt{12}(e-1)\right)\left(\sqrt{12}x-\sqrt{3}\right) \\
&= e - 1 + (3-e)\sqrt{3}\left(\sqrt{12}x - \sqrt{3}\right) \\
&= e - 1 + (3-e)(2x-3) \\
&= (6-2e)x + (4e-10).
\end{aligned}$$

♠

**Example 3.3.11.** Let $A \in M_2(\mathbb{R})$, find the closest symmetric matrix to $A$.

*Solution.* Let $W = \{x \in M_2(\mathbb{R}) : x = x^T\}$. We see $\beta' = \{\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix}, \begin{bmatrix}0 & 0\\0 & 1\end{bmatrix}, \begin{bmatrix}0 & 1\\1 & 0\end{bmatrix}\}$ is a basis for $W$. Moreover, we note $\beta = \{\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix}, \frac{1}{\sqrt{2}}\begin{bmatrix}0 & 0\\0 & 1\end{bmatrix}, \begin{bmatrix}0 & 1\\1 & 0\end{bmatrix}\}$ is an orthonormal basis for $W$.

Next, we have

$$\begin{aligned}
Proj_W(A) &= \langle A, v_1\rangle v_1 + \langle A, v_2\rangle v_2 + \langle A, v_3\rangle v_3 \\
&= tr(\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix}\begin{bmatrix}a & b\\c & d\end{bmatrix})\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix} + tr(\frac{1}{\sqrt{2}}\begin{bmatrix}0 & 1\\1 & 0\end{bmatrix}\begin{bmatrix}a & b\\c & d\end{bmatrix})\frac{1}{\sqrt{2}}\begin{bmatrix}0 & 1\\1 & 0\end{bmatrix} \\
&\quad + tr(\begin{bmatrix}0 & 0\\0 & 1\end{bmatrix}\begin{bmatrix}a & b\\c & d\end{bmatrix})\begin{bmatrix}0 & 0\\0 & 1\end{bmatrix} \\
&= \begin{bmatrix}a & \frac{b+c}{2}\\ \frac{b+c}{2} & d\end{bmatrix}
\end{aligned}$$

♠

## 3.4 The Adjoint

**Definition 3.4.1.** Let $V$ be vector space over $F$, then $T : V \to F$ is a ***linear functional*** if $T$ is linear.

We call $V^*$, the collection of all linear functionals of $V$, to be the ***dual space***.

**Example 3.4.2.** Let $V$ be inner product space over $F$. Let $v \in V$ be fixed, then $v$ induces a linear function of $V$. Namely, $T_v : V \to F$ given by $T_v(x) = \langle x, v\rangle$ is linear functional.

**Theorem 3.4.3.** *[Riesz Representation Theorem] Let $V$ be finite dimensional inner product space and let $T : V \to F$ be a linear functional. Then, there exists a unique $y \in V$ such that $T(x) = \langle x, y\rangle$ for all $x \in V$.*

*Proof.* Let $\{v_1, ...v_n\}$ be a orthonormal basis for $V$, and let $v = \sum_{i=1}^{n} T(v_i)v_i$. Note $\langle v_i, v_j\rangle = 0$ when $i \neq j$. Then, we have $y \in V$ then $y = \sum_{i=1}^{n} a_i v_i$. Then we have

$$\begin{aligned}
\langle y, v\rangle &= \langle y, \sum_{j=1}^{n} T(v_j)v_j\rangle \\
&= \langle \sum_{i=1}^{n} a_i v_i, \sum_{j=1}^{n} T(v_j)v_j\rangle = \sum_{i=1}^{n} \langle a_i v_i, \sum_{j=1}^{n} T(v_j)v_j\rangle \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} \langle a_i v_i, T(v_j)v_j\rangle \\
&= \sum_{i=1}^{n} \langle a_i v_i, T(v_i)v_i\rangle = \sum_{i=1}^{n} T(v_i)\langle a_i v_i, v_i\rangle \\
&= \sum_{i=1}^{n} T(\langle a_i v_i, v_i\rangle v_i) = \sum_{i=1}^{n} T(a_i\langle v_i, v_i\rangle v_i) \\
&= T(\sum_{i=1}^{n} a_i\langle v_i, v_i\rangle v_i) = T(\sum_{i=1}^{n} a_i v_i) = T(y)
\end{aligned}$$

Hence, we now only need to show uniqueness. Suppose $T(x) = \langle x, v_1\rangle$ and $T(x) = \langle x, v_2\rangle$ for all $x \in V$. Then we have $\langle x, v_1\rangle = \langle x, v_2\rangle$ for all $x \in V$ and thus $\langle x, v_1 - v_2\rangle = 0$ for all $x \in V$ which imply $v_1 - v_2 = 0$ and thus $v_1 = v_2$. Hence, we indeed have an unique $v \in V$ such that $T(x) = \langle x, v\rangle$ ♡

**Proposition 3.4.4.** *Let $V$ be finite dimensional inner product space, and let $T : V \to V$ be linear. There exists a unique function $T^* : V \to V$ such that for all $x, y \in V$, $\langle T(x), y\rangle = \langle x, T^*(y)\rangle$. Moreover, $T^*$ is linear.*

*Proof.* Fix $y \in V$, then $u_y : V \to F$ by $u_y(x) = \langle T(x), y\rangle$ is a linear functional. By the Riesz representation theorem 3.4.3, there exists unique $y' \in V$ such that $u_y(x) = \langle x, y'\rangle$. Thus, define $T^* : V \to V$ by $T^*(y) = y'$.

Thus, for any $x, y \in V$, $\langle T(x), y\rangle = \langle x, T^*(y)\rangle$.

Let $x, y_1, y_2 \in V$ and $c \in F$. Then $\langle x, T^*(cy_1+y_2)\rangle = \langle T(x), cy_1+y_2\rangle = \overline{c}\langle T(x), y_1\rangle + \langle T(x), y_2\rangle = \overline{c}\langle x, T^*(y)\rangle + \langle x, T^*(y_2)\rangle = \langle x, cT^*(y_1)+T^*(y_2)\rangle$. Therefore, for all $y_1, y_2$, $T^*(cy_1 + y_2) = cT^*(y_1) + T^*(y_2)$. ♡

**Definition 3.4.5.** In the previous proposition, we call $T^*$ the ***adjoint*** of $T$. Namely, $T^*$ is the adjoint of $T$ if and only if $\langle T(x), y\rangle = \langle x, T^*(y)\rangle$ for all $x, y \in V$.

**Proposition 3.4.6.** *Let $V$ be finite dimensional inner product space, let $\beta$ be orthonormal basis, and $T : V \to V$ be linear. Then $[T^*]_\beta = [T]^*_\beta$*

*Proof.* With Gram-Schmidt, we can find $\beta$ where $\beta$ is orthonormal. Let $A = [T]_\beta$ and $B = [T^*]_\beta$. We claim $A^* = B$. Note $b_{ij} = \langle T^*(v_j), v_i\rangle = \overline{\langle v_i, T^*(v_j)\rangle} = \overline{\langle T(v_i), v_j\rangle} = \overline{a_{ij}}$ and hence we are done. ♡

**Remark 3.4.7.** Let $A \in M_n(\mathbb{C})$, and $\sigma$ be the standard basis for $\mathbb{C}^n$. Then,

$$\begin{aligned} &A = [L_A]_\sigma \\ &\Rightarrow A* = [L_A^*]_\sigma = [L_{A^*}]_\sigma \\ &\Rightarrow \langle Ax, y\rangle = \langle x, A^*y\rangle \end{aligned}$$

**Example 3.4.8.** Let $T : P_2(\mathbb{R}) \to P_2(\mathbb{R})$ where $T(f(x)) = f'(x)$ and $P_2(\mathbb{R})$ is inner product space with dot product. Let $\sigma = \{x^2, x, 1\}$ be orthonormal. Then $[T]_\sigma = \begin{bmatrix} 0 & 0 & 0 \\ 2 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$. Then, $[T^*]_\sigma = [T]^*_\sigma = \begin{bmatrix} 0 & 2 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$. Thus $T^*(x^2) = 0, T^*(x) = 2x^2$ and $T^*(1) = x$ and so $T^*(ax^2 + bx + c) = 2bx^2 + cx$.

**Proposition 3.4.9.** *Let $T, U : V \to V$ be two linear operators. Let $V$ be finite dimensional inner product space, let $\alpha \in F$,*

1. $(T + U)^* = T^* + U^*$
2. $(\alpha T)^* = \bar{\alpha}T^*$
3. $(T \circ U)^* = U^* \circ T^*$
4. $(T^*)^* = T$
5. $I^* = I$

*Proof.* We prove 3.

For all $x, y \in V$, we have

$$\begin{aligned} \langle (T \circ U)(x), y\rangle &= \langle T(U(x)), y\rangle \\ &= \langle U(x), T^*(y)\rangle = \langle x, U^*(T^*(y))\rangle \\ &= \langle x, (U^* \circ T^*)(y)\rangle \end{aligned}$$

By uniqueness, we have $(T \circ U)^* = U^* \circ T^*$. ♡

**Remark 3.4.10.** We may similarly define the adjoint of any $A \in M_{m\times n}(F)$ by conjugate-trasnpose.

**Lemma 3.4.11.** *Let $A \in M_{m\times n}(F), x \in F^n$, and $y \in F^m$. Then $\langle Ax, y\rangle = \langle x, A^*y\rangle$.*

*Proof.* Note $\langle Ax, y\rangle = y^*Ax = (A^*y)^*x = \langle x, A^*y\rangle$. ♡

**Lemma 3.4.12.** *$A \in M_{m\times n}(F)$, then $rank(A) = rank(A^*A)$. If $rank(A) = n$ then $A^*A$ is invertible.*

*Proof.* We show $Null(A) = Null(A^*A)$. $Null(A) \subseteq Null(A^*A)$ is clear. Let $x \in Null(A^*A)$, then $\|Ax\|^2 = \langle Ax, Ax\rangle = \langle x, A^*Ax\rangle = \langle x, 0\rangle = 0$. ♡

**Remark 3.4.13** (Least Square Approximation)**.** The motivation for least square approximation is to approximate a set of data with linear functions to minimize the error.

Let $(t_i, y_i)$ be a set of data where $1 \le i \le m$. We wish to find a line that the vertical distances from the points to the line is minimized.

The line $y = cx + d$ which best fits this data will be called the least squared line if the error term $E = \sum_{i=1}^{m}(ct_i + d - y_i)^2$ is minimal.

Let $A = \begin{bmatrix} t_1 & 1 \\ \vdots & \vdots \\ t_m & 1 \end{bmatrix}$, $x = \begin{bmatrix} c \\ d \end{bmatrix}$, and $y = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$, we have $E = \|Ax - y\|^2$. Thus, we must find $x_0$ so that $\|Ax_0 - y\|^2$ is minimal.

To minimize $\|Ax - y\|$, we let $W = range(A)$, and we first find $y_0 = proj_W(y) \in W$, then $y_0 = Ax_0$, then $\|Ax_0 - y\|$ is minimal.

In practice, $y - y_0 \in W^\perp$, which means $y - Ax_0 \in W^\perp$ and thus $\langle Ax, y - Ax_0\rangle = 0$ for all $x \in F^n$. This would mean that $\langle x, A^*(y - Ax_0)\rangle = 0$ for all $x \in F^n$. Thus, $A^*(y - Ax_0) = 0$ imply

$$A^*y = A^*Ax_0$$

Moreover, if $rank(A) = n$, then we have

$$x_0 = (A^*A)^{-1}A^*y$$

In practice, we would always have $rank(A) = 2$ and therefore $x_0$ always exists.

**Example 3.4.14.** The last 4 years MATH 245 final exam averages is

$$(1, 75), (2, 82), (3, 60), (4, 70)$$

Then, we have $x_0 = \begin{bmatrix} -3.7 \\ 81 \end{bmatrix}$ and thus $y = -3.7x + 81$.

Note there may be one question on the final to calculate least square of line and quadratic.

**Example 3.4.15.** Find the least square line and quadratic of best fit for the data set:

$$\{(1, -1), (3, 2), (5, 1)(7, 0)\}$$

*Solution.* First, we calculate the least square line.

Let $A = \begin{bmatrix} 1 & 1 \\ 3 & 1 \\ 5 & 1 \\ 7 & 1 \end{bmatrix}$ and $y = \begin{bmatrix} -1 \\ 2 \\ 1 \\ 0 \end{bmatrix}$. Let $y = ax + b$, then let $x = \begin{bmatrix} a \\ b \end{bmatrix}$, we have

$$\begin{aligned} x &= (A^*A)^{-1}A^*y \\ &= \begin{bmatrix} 84 & 16 \\ 16 & 4 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 3 & 5 & 7 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} -1 \\ 2 \\ 1 \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} 84 & 16 \\ 16 & 4 \end{bmatrix}^{-1} \begin{bmatrix} 10 \\ 2 \end{bmatrix} \\ &= \frac{1}{80} \begin{bmatrix} 4 & -16 \\ -16 & 84 \end{bmatrix} \begin{bmatrix} 10 \\ 2 \end{bmatrix} = \frac{1}{80} \begin{bmatrix} 8 \\ 8 \end{bmatrix} \\ &= \begin{bmatrix} 0.1 \\ 0.1 \end{bmatrix} \end{aligned}$$

Therefore, $y = 0.1x + 0.1$ is the line of best fit. Apparently, this is not any best fit of the data set.

We calculate the least square quadratic next.

Let $A = \begin{bmatrix} 1^2 & 1^1 & 1 \\ 3^2 & 3^1 & 1 \\ 5^2 & 5^1 & 1 \\ 7^2 & 7^1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 9 & 3 & 1 \\ 25 & 5 & 1 \\ 49 & 7 & 1 \end{bmatrix}$ and $y = \begin{bmatrix} -1 \\ 2 \\ 1 \\ 0 \end{bmatrix}$, then $A^*A = \begin{bmatrix} 3108 & 496 & 84 \\ 496 & 84 & 16 \\ 84 & 16 & 4 \end{bmatrix}$

Hence, let $y = ax^2 + bx + c$ be the quadratic of best fit, we have

$$\begin{aligned} \begin{bmatrix} a \\ b \\ c \end{bmatrix} &= (A^*A)^{-1}A^* \begin{bmatrix} -1 \\ 2 \\ 1 \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} -1/4 \\ 21/10 \\ -53/20 \end{bmatrix} \end{aligned}$$

Thus, we have $y = -\frac{1}{4}x^2 + \frac{21}{10}x - \frac{53}{20}$ is the quadratic of best fit.

Note the best fit of this data set is $f(x) = \frac{1}{12}x^3 - \frac{5}{4}x^2 + \frac{65}{12}x - \frac{21}{4}$. ♠

**Remark 3.4.16.** This works for degree $n$ polynomial of best fit as well. Let $x = (a_n, a_{n-1}, ..., a_0)^T$, $y = y_1, y_2, ..., y_m$, and let

$$A = \begin{bmatrix} t_1^n & t_1^{n-1} & ... & t_1 & 1 \\ t_2^n & t_2^{n-1} & ... & t_2 & 1 \\ \vdots & \vdots & ... & \vdots & \vdots \\ t_m^n & t_m^{n-1} & ... & t_m & 1 \end{bmatrix}$$

Then, we have $x_0 = (A^*A)^{-1}A^*y$

**Proposition 3.4.17.** *Let $T : V \to V$ be linear where $V$ is finite dimensional inner product space $V$. Then $V = Range(T^*) \oplus Null(T)$.*

*Proof.* We first note that $rank(T) = rank(T^*)$.

For all $x \in Null(T)$ and for all $y \in V$, we have $\langle T(x), y\rangle = \langle x, T^*(y)\rangle = 0$ as $T(x) = 0$. Note $T^*(y) \in Range(T^*)$ and thus, for all $x \in Null(T)$, we have $\langle x, T^*(y)\rangle = 0$, where $y$ is arbitrary as well. Hence, denote $W = Null(T)$, we have $Range(T^*) \subseteq W^\perp$ where $W^\perp \oplus W = V$.

Let $rank(T) = rank(T^*) = l = dim(Range(T^*))$ and $dim(Null(T)) = dim(W) = k$, we have $k + l = n$ where $n = dim(V)$. However, since $W^\perp \cap W = \{0\}$ and $W^\perp \oplus W = V$, we must have $dim(W^\perp) + dim(W) = n$, and thus

$$dim(W^\perp) = dim(Range(T^*)) \Rightarrow W^\perp = Range(T^*)$$

Hence, $Range(T^*) \oplus Null(T) = V$ as desired. ♡

**Example 3.4.18.** Let $V = span(1, x, x^2)$ be a subspace of $C[0, 1]$, and define $T : V \to V$ be $T(f(x)) = f'(x)$. Then, note $1 \in Range(T)$ as $T(x) = 1$, and $T(1) = 0$ and thus $1 \in Null(T)$. Therefore, $Range(T) \cap Null(T) \neq \{0\}$ and hence $V \neq Range(T) \oplus Null(T)$.

**Remark 3.4.19.** The following are additional materials that are not covered in lecture.

## 3.5 Bilinear Forms

### 3.5.1 Intro

**Remark 3.5.1.** Note in this note, the terms are slightly different from what they are normally been called. In particular, in this note, we would say $f$ is a "bilinear function", but the more common term for it is "bilinear form" (to specify it is a bilienar function maps to the scalar field instead of another vector space). We defined "bilinear form" differently, but as we can see from Remark 2.8 that they (our definition of bilinear function and bilinear form) are actually the same.

**Definition 3.5.2.** Let $V$ be finite dimensional $F$ vector space, $f : V \times V \to F$ is a ***bilinear function*** if, for all $x_1, x_2, y_1, y_2, x, y \in V$, $k_1, k_2 \in F$, we have

1. $f(k_1x_1 + k_2x_2, y) = k_1f(x_1, y) + k_2f(x_2, y)$
2. $f(x, k_1y_1 + k_2y_2) = k_1f(x, y_1) + k_2f(x, y_2)$

**Remark 3.5.3.** Let $V$ be finite dimensional $F$ vector space and $f : V \times V \to F$ be bilinear, then, for fixed $\alpha, \beta \in V$, we have $\alpha_L(x) = f(\alpha, x)$ and $\beta_R(x) = f(x, \beta)$ are linear mappings from $V \to V$.

**Example 3.5.4.**

1. The standard inner products on $\mathbb{R}^n$ and $\mathbb{C}^n$ are bilinear functions
2. Let $V = M_n(F)$ then $f(A, B) = tr(AB)$ is bilinear.
3. Let $V = C[a, b]$ be real continuous function on $[a, b]$, then $f(g(x), h(x)) = \int_a^b g(x)h(x)dx$ is bilinear.

**Definition 3.5.5.** Let $V$ be $n$ dimensional $F$ vector space where $\infty > n \in \mathbb{N}$, let $v = \{v_1, ..., v_n\}$ be a basis for $V$, and then for all $x, y \in V$, we let $[x]_v = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ and $[y]_v = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$. Suppose $f : V \times V \to V$ is bilinear, then

$$[f]_v = A = \begin{bmatrix} f(v_1, v_1) & f(v_1, v_2) & ... & f(v_1, v_n) \\ f(v_2, v_1) & f(v_2, v_2) & ... & f(v_2, v_n) \\ \vdots & \vdots & \ddots & \vdots \\ f(v_n, v_1) & f(v_n, v_2) & ... & f(v_n, v_n) \end{bmatrix}$$

is the ***matrix representation of f under v***.

**Remark 3.5.6.** Note, for a fixed basis of $V$, we have $f(x, y) = [x]_v^T [f]_v [y]_v$. Indeed, note $f(x, y) = f(\sum_{i=1}^n x_i v_i, \sum_{j=1}^n y_j v_j) = \sum_{i=1}^n \sum_{j=1}^n x_i y_j f(v_i, v_j) = [x]_v^T A[y]_v$.

In addition, given a matrix $A$ and a fixed basis $v$ for $V$, define $f_A(x, y) = [x]_v^T A[y]_v$, then $f_A : V \times V \to F$ and we can see $f_A$ is indeed bilinear. Moreover, the matrix representation of $f_A$ under $v$ is exactly $A$.

**Definition 3.5.7.** For arbitrary $A \in M_n(F)$, let $X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ and $Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$ be variables, then $X^T AY$ is called a ***bilinear form*** of $x_1, ..., x_n$ and $y_1, ..., y_n$.

**Remark 3.5.8.** Let $V$ be $n$ dimensional $F$ vector space, $x, y \in V$ and $v$ be a basis for $V$. Then every bilinear form of $x_1, ..., x_n$ and $y_1, ..., y_n$ could induce a bilinear function on $V$, namely consider $X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = [x]_v$ and $Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = [y]_v$ and then compute $X^T AY$.

**Theorem 3.5.9.** *Let $V$ be $n$ dimensional $F$ vector space, let $v = (v_1, ..., v_n), w = (w_1, ..., w_n)$ be two bases for $V$. Then, let $P = [I]_v^w$ (recall $I$ is the identical mapping on $V$), $A = [f]_w$ and $B = [f]_v$, we have*

$$B = P^T AP$$

*Proof.* Let $x, y \in V$ be arbitrary, and suppose $x = \sum_{i=1}^n x_i w_i = \sum_{i=1}^n x_i' v_i$ and $y = \sum_{i=1}^n y_i w_i = \sum_{i=1}^n y_i' v_i$ where $X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = [x]_w$, $X_0 = \begin{bmatrix} x_1' \\ x_2' \\ \vdots \\ x_n' \end{bmatrix} = [x]_v$, $Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = [y]_w$ and $Y_0 = \begin{bmatrix} y_1' \\ y_2' \\ \vdots \\ y_n' \end{bmatrix} = [y]_v$. Then, we have $f(x, y) = X^T AY = X_0^T BY_0$. Next, note $X = PX_0$ and $Y = PY_0$, indeed, $P$ is the change of coordinate matrix from $v$ to $w$, and so

$$X^T AY = (PX_0)^T A(PY_0) = X_0^T (P^T AP)Y_0$$

Since $X^T AY = X_0^T BY_0$, we have $X_0^T (P^T AP)Y_0 = X_0 BY_0$ for all $X_0, Y_0 \in F^n$, and thus we must have $P^T AP = B$(indeed, consider $e_i^T (P^T AP)e_j$ for all $1 \leq i, j \leq n$, we have $(i, j)$ entry of $P^T AP$ is the same as the $(i, j)$ entry of $B$). ♡

**Definition 3.5.10.** Let $f$ be bilinear over $V$ where $V$ is $n$ dimensional $F$ vector space. The ***matrix rank*** (or ***rank***) of $f$ is $rank([f]_v)$ for arbitrary basis $v$ of $V$.

**Remark 3.5.11.** Note this rank is unique by our above theorem. Indeed, recall $B = P^T AP$ where $det(P) \neq 0$ then we say $B$ and $A$ are congruent and congruent matrices have the same rank. This tell us the matrix representation of $f$ under arbitrary basis all have the same rank.

**Definition 3.5.12.** Let $f$ be bilinear over $V$, then $rad_L(V) = \{a \in V : \forall b \in V, f(a, b) = 0\}$ is called the ***left radical*** of $f$ in $V$. Similarly, $rad_R(V) = \{b \in V : \forall a \in V, f(a, b) = 0\}$ is called the ***right radical*** of $f$ in $V$.

**Remark 3.5.13.** It is easy to see $rad_L(V), rad_R(V)$ are subspaces of $V$ for arbitrary bilinear function $f$ on $V$.

**Definition 3.5.14.** Let $f$ be bilinear over $V$. Then we say $f$ is ***non-degenerate*** if $rad_L(V) = rad_R(V) = \{0\}$.

**Theorem 3.5.15.** *Let $V$ be $n$ dimensional $F$ vector space and $f$ is bilinear. Then $f$ is non-degenerate if and only if there exists a basis $v$ of $V$ such that $[f]_v$ is invertible matrix.*

*Proof.* Let $A = [f]_v$ where $v = \{v_1, ..., v_n\}$. Let $x, y \in V$ be arbitrary and $x = \sum_{i=1}^n x_i v_i, y = \sum_{i=1}^n y_i v_i$. Let $X = [x]_v, Y = [y]_v$ and $\epsilon = \{e_1, ..., e_n\}$ be the standard basis for $F^n$. Then

$$\begin{aligned}
&x \in rad_L(V) \\
&\Leftrightarrow \forall y \in V, f(x,y) = 0 \\
&\Leftrightarrow \forall Y \in F^n, X^T AY = 0 \\
&\Leftrightarrow X^T Ae_i = 0, \forall 1 \leq i \leq n \\
&\Leftrightarrow X^T A[e_1, e_2, ..., e_n] = 0 \\
&\Leftrightarrow X^T AI = 0 \\
&\Leftrightarrow A^T X = 0 \\
&\Leftrightarrow X \in Null(A^T)
\end{aligned}$$

Thus, $rad_L(V) = \{0\}$ if and only if $Null(A^T) = \{0\}$ and thus $rank(A^T) = n = rank(A)$ as desired.

Similarly, we can show $rad_R(V) = \{0\}$ if and only if $rank(A) = n$ and the proof follows. ♡

### 3.5.2 Symmetric and Skew-Symmetric Bilinear Function

**Definition 3.5.16.** Let $f$ be bilinear over $V$. If $f(x,y) = f(y,x)$ for all $x, y \in V$ then we say $f$ is ***symmetric bilinear function***(or ***symmetric***).

**Definition 3.5.17.** Let $f$ be bilinear over $V$. If $f(x,y) = -f(y,x)$ for all $x, y \in V$, then we say $f$ is ***skew-symmetric bilinear function*** (or ***skew-symmetric***).

**Remark 3.5.18.** Note $f$ is symmetric if and only if $[f]_v$ is symmetric matrix for arbitrary basis $v$ of $V$. Similarly, $f$ is skew-symmetric if and only if $[f]_v$ is skew-symmetric matrix ($A = -A^T$) for arbitrary basis $v$ of $V$.

**Theorem 3.5.19.** *Let $V$ be $n$ dimensional $F$ vector space with $char(F) \neq 2$. Let $f$ be symmetric bilinear on $V$, then there exists a basis of $V$ such that $[f]_v$ is diagonal.*

*Proof.* We use induction on $n$.

If $n = 1$ then let $v = \{v_1\}$ be a basis of $V$, we have $[f]_v = \big[f(v_1, v_1)\big]$. This is indeed diagonal matrix.

Suppose it holds for $n - 1$. If $f = 0$ then $[f]_v = 0$ for all basis $v$ of $V$ and thus $f$ is diagonal. Now, we suppose $f \neq 0$. Thus, there must exists $0 \neq \alpha_1 \in V$ such that $f(\alpha_1, \alpha_1) \neq 0$. Indeed, suppose $f(x,x) = 0$ for all $x \in V$, we have $\forall x, y \in V$, $0 = f(x+y, x+y) = f(x,x) + f(x,y) + f(y,x) + f(y,y) = 2f(x,y)$. However, $char(F) \neq 2$, and thus $f(x,y) = 0$ and since $x, y$ was arbitrary, we have $f = 0$, a contradiction.

We extend $\alpha$ to a basis of $V$, $\alpha_1, ..., \alpha_n$ and let

$$\tilde{\alpha}_i = \alpha_i - \frac{f(\alpha_i, \alpha_1)}{f(\alpha_1, \alpha_1)}\alpha_1, i = 2, ..., n$$

Then, for $i = 2, ..., n$, we have

$$\begin{aligned} f(\alpha_1, \tilde{\alpha}_i) &= f(\alpha_1, \alpha_i - \frac{f(\alpha_i, \alpha_1)}{f(\alpha_1, \alpha_1)}\alpha_1) \\ &= f(\alpha_1, \alpha_i) - \frac{f(\alpha_i, \alpha_1)}{f(\alpha_1, \alpha_1)} f(\alpha_1, \alpha_1) = 0 \end{aligned}$$

Moreover, note[1] $span(\alpha_1, ..., \alpha_n) = span(\alpha_1, \tilde{\alpha}_2, ..., \tilde{\alpha}_n)$ and thus $\alpha_1, \tilde{\alpha}_2, ..., \tilde{\alpha}_n$ is a basis for $V$. Let $W = span(\tilde{\alpha}_2, ..., \tilde{\alpha}_n)$, then we have $V = span(\alpha_1) \oplus W$. Let $f_W$ be the bilinear function obtained by restrict $f$ on $W$, we have $f_W$ is symmetric. Note $dim(W) = n - 1$, and so there exists basis $\eta = \{\eta_2, ..., \eta_n\}$ for $W$ such that $[f_W]_\eta$ is diagonal. In particular, this imply $[f_W]_\eta = diag\{f(\eta_2, \eta_2), f(\eta_3, \eta_3), ..., f(\eta_n, \eta_n)\}$. Since $span(\alpha_1) \oplus W = V$, we have $\beta = \{\alpha_1, \eta_2, ..., \eta_n\}$ is a basis for $V$, and since $f(\alpha_1, \tilde{\alpha}_i) = 0$, we have $f(\alpha_1, \eta_i) = 0$ for $i = 2, ..., n$. Thus, $[f]_\beta$ is diagonal and this concluded the induction. ♡

**Theorem 3.5.20.** *Let $V$ be $n$ dimensional $F$ vector space with $char(F) \neq 2$. Let $f$ be skew-symmetric bilinear on $V$. Then, there exists a basis*

$$v = \{\delta_1, \delta_{-1}, ..., \delta_r, \delta_{-r}, \eta_1, ..., \eta_s\}$$

*where $0 \leq r \leq \frac{n}{2}, s = n - 2r$ and so*

$$[f]_v = diag\{\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, ..., \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, 0, ..., 0\}$$

*Sketch Proof.* The proof is similar to the proof of Theorem 2.18. We use induction on $dim(V)$ and $n = 1$ is easy to check. Suppose it holds for all dimension less than or equal to $n - 1$. Suppose $f \neq 0$. There must exists linear independent $\delta_1, \alpha_2 \in V$ so that $f(\delta_1, \alpha_2) \neq 0$. Indeed, suppose all lienar independent paris $x, y \in V$, we have $f(x, y) = 0$. Next, note for all $x \in V$ and $k \in F$, we have $f(x, kx) = kf(x, x) = -kf(x, x)$ and so $2kf(x, x) = 0$, since $char(F) \neq 2$, $f(x, x) = 0$ for all $x \in V$. Thus $f(x, y) = 0$ for all $x, y \in V$.

Let $\delta_{-1} = f(\delta_1, \alpha_2)^{-1}\alpha_2$, then we have $f(\delta_1, \delta_{-1}) = 1$. Next, extend $\delta_1, \delta_{-1}$ to a basis of $V$, say $\delta_1, \delta_{-1}, \beta_3, ..., \beta_n$. Then let

$$\tilde{\beta}_i = \beta_i - f(\beta_i, \delta_{-1})\delta_1 + f(\beta_i, \delta_1)\delta_{-1}, i = 3, ..., n$$

We check that $f(\delta_1, \tilde{\beta}_i) = f(\delta_{-1}, \tilde{\beta}_i) = 0$ for $i = 3, ..., n$ and $span(\delta_1, \delta_{-1}, \beta_3, ..., \beta_n) = span(\delta_1, \delta_{-1}, \tilde{\beta}_3, ..., \tilde{\beta}_n)$.

Then, let $W = span(\tilde{\beta}_3, ..., \tilde{\beta}_n)$ and we have $V = span(\delta_1, \delta_{-1}) \oplus W$. Thus, $f_W$ is skew-symmetric bilinear on $W$ and by induction hypothesis, we have a basis $w$ of $W$ so $[f_W]_w = diag\{\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, ..., \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, 0, ..., 0\}$. Thus, check $\delta_1, \delta_{-1} \cup w$ is the basis we are looking for and this should conclude the proof. ♡

[1]to see this, recall how we defined $\tilde{\alpha}_i$

**Definition 3.5.21.** Let $V$ be finite dimensional $F$ vector space. Let $T_2(V)$ be the set of all bilinear functions. We call $T_2(V)$ the ***bilinear function space*** with function addition and scalar multiplication(easy to check it is a vector space).

**Remark 3.5.22.** Note $T_2(V) \cong M_n(F) \cong Hom(V, V)$. We can construct an isomorphism from $T_2(V)$ to $M_n(F)$ by simply consider the matrix representation of $f$. We already know $M_n(F) \cong Hom(V, V)$ by consider the matrix representation of $T \in Hom(V, V)$.

However, we can also construct isomorphism between $T_2(V)$ and $Hom(V, V)$ without consider matrices. Let $f$ be non-degenerate, we can show $\forall g \in T_2(V)$, there exists unique $G \in Hom(V, V)$ such that $g(x, y) = f(Gx, y)$. Moreover, we have $\sigma : g \mapsto G$ is isomorphism.

**Definition 3.5.23.** Let $V$ be finite dimensional $F$ vector space. Then, $S_2(V)$ is the set of all symmetric bilinears. $\Lambda_2(V)$ is the set of all skew-symmetric bilinear.

**Theorem 3.5.24.** *Let $V$ be $F$ vector space with $char(F) \neq 2$. We have*

$$T_2(V) = S_2(V) \oplus \Lambda_2(V)$$

*Proof.* Let $f \in T_2(V)$ be arbitrary, we have $g(x, y) = (2e)^{-1}(f(x, y) + f(y, x)) \in S_2(V)$ and $h(x, y) = (2e)^{-1}(f(x, y) - f(y, x)) \in \Lambda_2(V)$ where $e \in F$ is the multiplicative identity. Then, we have $f(x, y) = g(x, y) + h(x, y)$. Thus $T_2(V) = S_2(V) + \Lambda_2(V)$.

Next, we show $S_2(V) \cap \Lambda_2(V) = \{0\}$. Note $f \in S_2(V) \cap \Lambda_2(V)$ imply $f(x, y) = f(y, x) = -f(y, x)$ and so $2f(y, x) = 0$ for all $x, y \in V$. In particular, since $char(F) \neq 2$, we have $f(y, x) = 0$ for all $x, y \in V$, thus $f = 0$. The proof follows. ♡

### 3.5.3 Inner Product Space On Arbitrary Field

**Remark 3.5.25.** In an arbitrary field $F$, it may not be the case that $F$ has a total order relation. Thus, an inner product on arbitrary field does not nessary have the positive definite property, i.e. $\langle x, x\rangle \geq 0$, since it is possible that we cannot even define the relation $\geq$. However, we still want to insist orthogonality of the inner product, as that's one of the big reason why we study inner product at all. Hence, we want $x \perp y$ and $y \perp x$ at the same time, that is, $f(x, y) = \langle x, y\rangle = 0$ if and only if $f(y, x) = \langle y, x\rangle = 0$.

Moreover, we want to be able to perform scalar multiplication and vector addition on both side, and thus we want $f$, the inner product on $F$, to be bilinear.

**Theorem 3.5.26.** *Let $f(x, y)$ be bilinear on $V$. Then $\forall x, y \in V$, $f(x, y) = 0 \Leftrightarrow f(y, x) = 0$ if and only if $f$ is symmetric or skew-symmetric.*

*Proof.* First, ntoe $f$ is skew-symmetric if and only if $f(x, x) = 0$ for all $x \in V$. (try to prove this)

Suppose $f$ is symmetric, then $f(x,y) = f(y,x)$ and so $f(x,y) = 0 \Leftrightarrow f(y,x) = 0$. Suppose $f$ is skew-symmetric, then $f(x,y) + f(y,x) = 0$ for all $x, y \in V$. Thus $f(x,y) = 0 \Leftrightarrow f(y,x) = 0$.

Conversely, suppose $\forall x, y \in V, f(x,y) = 0 \Leftrightarrow f(y,x) = 0$. Let $x, y, z \in V$ be arbitrary. Let $w = f(x,y)z - f(x,z)y$, then $f(x,w) = 0 = f(w,x)$ is the same as $f(x,y)f(z,x) = f(y,x)f(x,z)$ for all $x, y, z \in V$. Let $x = y$, we get $f(x,x)(f(x,z) - f(z,x)) = 0$ for all $x, z \in V$.

We claim $f(x,z) = f(z,x)$ or $f(x,x) = 0$. Suppose otherwise for a contradiction. Then there exists $u, v \in V$ such that $f(u,v) \neq f(v,u)$ and a vector $w \in V$ such that $f(w,w) \neq 0$. Then, note $f(u,u)(f(u,v) - f(v,u)) = 0$ but $f(u,v) - f(v,u) \neq 0$ which imply $f(u,u) = 0$. Similarly, we have $f(v,v) = 0$. Moreover, note $f(w,u) = f(u,w)$ and $f(w,v) = f(v,w)$ as

$$f(w,w)(f(w,u) - f(u,w)) = 0 = f(w,w)(f(w,v) - f(v,w))$$

and $f(w,w) \neq 0$ imply $f(w,u) - f(u,w) = 0 = f(w,v) - f(v,w)$. Since $f(u,v) \neq f(v,u)$, we have $f(u,v)f(w,u) = f(v,u)f(u,w)$ and thus we must have $f(w,u) = 0$ as $f(u,w) = f(w,u)$. Similarly, we must have $f(w,v) = 0$.

Then, $f(u, w+v) = f(u,v) \neq f(v,u) = f(w+v, u)$. Hence, we have $f(w+v, w+v) = 0$ as $f(w+v, w+v)(f(w+v, u) - f(u, w+v)) = 0$. However, $f(w+v, w+v) = f(w,w) + f(w,v) + f(v,w) + f(v,v) = f(w,w) \neq 0$. Thus, we have our contradiction and the proof follows. ♡

**Remark 3.5.27.** Thus, in order to define an inner product with the orthogonality property, we must use symmetric bilinear or skew-bilinear functions. Moreover, if the inner product on $V$ is symmetric, we call this inner product space ***orthogonal space***(this is a term defined only in the book) and if the inner product is skew-symmetric, we call this inner product space ***symplectic space***.

**Remark 3.5.28.** Note the normal inner products we are working with always have the positive definite property. However, in real life applications, there are cases where we do not use positive definite bilinear functions, but non-degenerate bilinear functions as the inner product.

**Example 3.5.29** (Lorentz Transformation and Minkowski Space)**.** All physical quantities(force, distance, velocity, etc.) can be represented by a set of numbers, and this set of numbers usually depends on the coordinate system (frame of reference) you are choosing.

In particular, if a frame of reference is idle or moving with constant speed along one straight line, we say this frame of reference is inertial frame of reference, otherwise we say it is non-inertial frame of reference.

Consider a frame of reference $Oxyz$ to be an inertial and another frame of reference $O'x'y'z'$ is moving with constand speed $v$ along the direction of $x$ axis of $Oxyz$. The two origin $O$ and $O'$ overlap at time $t = t' = 0$. One point $P$ in this system can be described as a vector in $\mathbb{R}^4$, with one component as the time, and the rest to be the $x, y, z$ coordinate to describe the location of the point in space. We call this spacetime coordinate.

In particular, we write $P$ with respect to the frame of reference $Oxyz$ at time $t$ to be $\begin{bmatrix} t \\ x \\ y \\ z \end{bmatrix}$ and with respect to $O'x'y'z'$ to be $\begin{bmatrix} t' \\ x' \\ y' \\ z' \end{bmatrix}$. When $v$ is much much smaller than the speed of light, we have the spacetime transformation of the point $P$ from $Oxyz$ to $O'x'y'z'$ to be

$$\begin{cases} t' = t \\ x' = x - vt \\ y' = y \\ z' = z \end{cases}$$

This is called the Galilean transformation. It is easy to verify that if Newton's law of motion holds for one frame of reference, then it holds for another. This is called the Galilean covariance in Newtonian mechanics, or the principle of relativity for Newtonian mechanics.

In the late 19th century, the fundamental principles of electromagnetism has been found, namely, the Maxwell's Equations. However, this set of equations is not covariant to Galilean transformation, i.e. with different frame of reference, the result is different. It is with Einstein's theory of special relativity, that this problem has been fixed.

In particular, Einstein introduced a new spacetime transformation, called ***Lorentz transformation***, that is,

$$\begin{cases} t' = \frac{t - \frac{v}{c^2}x}{\sqrt{1-\frac{v^2}{c^2}}} \\ x' = \frac{x - vt}{\sqrt{1-\frac{v^2}{c^2}}} \\ y' = y \\ z' = z \end{cases}$$

where $c$ is the speed of light.

Einstein showed that Maxwell's equation to Lorentz transformation is covariant, and this imply that principle of relativity applies to both mechanics and electromagnetism. Moreover, Einstein also generalized this result as "all fundamental physical relations should have the same form in arbitrary inertial frame of reference". This is called the ***special principle of relativity***.

Next, note the Lorentz transformation is given by $\sigma = M_4(\mathbb{R})$ where

$$\sigma = \begin{bmatrix} \frac{1}{\sqrt{1-\frac{v^2}{c^2}}} & -\frac{v/c^2}{\sqrt{1-\frac{v^2}{c^2}}} & 0 & 0 \\ -\frac{v}{\sqrt{1-\frac{v^2}{c^2}}} & \frac{1}{\sqrt{1-\frac{v^2}{c^2}}} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Indeed, we have $\begin{bmatrix} t' \\ x' \\ y' \\ z' \end{bmatrix} = T \begin{bmatrix} t \\ x \\ y \\ z \end{bmatrix}$.

Hence, Lorentz transformation is a linear mapping on $\mathbb{R}^4$. Now, consider $x = (t_1, x_1, y_1, z_1)^T$ and $y = (t_2, x_2, y_2, z_3)^T$ in the inertial frame of reference $Oxyz$. Next, we want to define a non-degenerate bilinear function $f$ as the inner product of $\mathbb{R}^4$. Just like $\mathbb{R}^4$ with standard inner product, we define the distance between $x, y \in \mathbb{R}^4$ as $\sqrt{f(x-y, x-y)} =: d(x, y)$, and thus, we call $f(x-y, x-y)$ the square of the spacetime interval between $x$ and $y$. Moreover, by the special principle of relativity, we should have $f(x, y) = f(\sigma(x), \sigma(y))$ as the two points in an arbitrary inertial frame of reference should have the same form.

Let $f : \mathbb{R}^4 \times \mathbb{R}^4 \to \mathbb{R}$ to be in the form

$$f(x, y) = -c^2 t_1 t_2 + x_1 x_2 + y_1 y_2 + z_1 z_2$$

where $c$ is the speed of light. We see $f$ is non-degenerate bilinear function, and we take this $f(x, y)$ as the inner product, we have $(\mathbb{R}^4, f)$ is a ***Minkowski space***, where Minkowski space is where the spacetime interval between any two events is independent of the inertial frame of reference in which they are recorded.

We can indeed show $f(x, x) = f(\sigma(x), \sigma(x))$ for all $x \in \mathbb{R}^4$, and thus $(\mathbb{R}^4, f)$ is indeed a ***Minkowski space***. However, if we use the standard inner product of $\mathbb{R}^4$, we would fail to obtain a inner product space with the property $\langle \sigma(x), \sigma(y) \rangle = \langle x, y \rangle$. This is one example where we use non-degenerate bilinear function instead of positive definite bilinear function as the inner product.

**Definition 3.5.30.** Let $V$ be $F$ vector space with symmetric bilinear function $f$, then we say $f$ is an ***inner product*** of $V$, and say $V$ is orthogonal space. Moreover, if $f$ is non-degenerate, we say $(V, f)$ is ***normal space*** and otherwise ***non-normal space***.

**Definition 3.5.31.** In orthogonal space $(V, f)$, if $f(x, y) = 0$ then we say $x$ and $y$ are ***orthogonal*** and write $x \perp y$.

**Definition 3.5.32.** In orthogonal space $(V, f)$, a non-zero vector $x \in V$ is said to be ***isotropic*** if $f(x, x) = 0$, otherwise we say $x$ is ***anisotropic***.

**Definition 3.5.33.** An orthogonal space $(V, f)$ is ***isotropic*** if there is at least one isotropic vector $x$ in $V$. Otherwise, we say $(V, f)$ is ***anisotropic***. If every non-zero vector in $V$ is isotropic, we say $(V, f)$ is ***totally isotropic***.

**Proposition 3.5.34.** *If $(V, f)$ is anisotropic, then $(V, f)$ is a normal space.*

*Proof.* If $f$ is degenerate, then $rad_L(V) \neq 0$ and thus there exists $0 \neq a \in rad_L(V)$ such that $f(a, a) = 0$, a contradiction. ♡

**Proposition 3.5.35.** *If $(V, f)$ is totally isotropic and $char(F) \neq 2$ where $F$ is the underlying field, then $f = 0$.*

*Proof.* Let $x, y \in V$, since $(V, f)$ is totally isotropic, $0 = f(x+y, x+y) = 2f(x, y)$, since $char(F) \neq 0$, $f(x, y) = 0$. Since $x, y \in V$ was arbitrary, $f = 0$. ♡

**Remark 3.5.36.** Let $(V, f)$ be orthogonal space, let $W$ be a subspace of $V$, then $f_W$, the restriction of $f$ on $W$ is a symmetric bilinear function on $W$. Thus, $(W, f_W)$ is an orthogonal space. However, even if $(V, f)$ is normal, $(W, f_W)$ may not be normal.

**Example 3.5.37.** Let $f(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}, \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}) = x_1y_1 - x_2y_2 - x_3y_3 - x_4y_4$, then it is easy to see $(\mathbb{R}^4, f)$ is normal. Moreover, let $W = span \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}$, then $\langle W, f_W \rangle$ is non-normal as $f(\begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}) = 0$.

**Definition 3.5.38.** Suppose $V$ is a vector space, and $f$ is symmetric bilinear or skew-symmetric bilinear, $S \subseteq V$ be a subset of $V$, then the set $S^\perp = \{a \in V : \forall b \in V, f(a, b) = 0\}$ is called the orthogonal complement of $S$.

**Remark 3.5.39.** It is easy to see $S^\perp$ is a subspace of $V$.

**Proposition 3.5.40.** *Let $V$ be $n$ dimensional $F$ vector space, let $f$ be non-degenerate symmetric or skew-symmetric bilinear function, let $W$ be a subspace of $V$, let $W^\perp = \{a \in V : \forall b \in W, f(a, b) = 0\}$, then we have*

1. $dim(W) + dim(W^\perp) = dim(V) = n$
2. $(W^\perp)^\perp = W$

*Proof.* Let $w_1, ..., w_m$ be a basis of $W$, extend it to a basis of $V$, say

$$w = \{w_1, ..., w_m, w_{m+1}, ..., w_n\}$$

Suppose $A = [f]_w$, since $f$ is non-degenerate, we have $[f]_w$ is invertible. Thus, let $x \in V$ be arbitrary, then $x = \sum_{i=1}^n c_i w_i$, denote $X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ and $\{\epsilon_1, ..., \epsilon_n\}$ be the standard basis for $F^n$(note $[w_i]_w = \epsilon_i$. Moreover, we denote $U = [\epsilon_1, ..., \epsilon_m]$) and we have

$$\begin{aligned} x \in W^\perp &\Leftrightarrow \forall y \in W, f(x, y) = 0 \\ &\Leftrightarrow f(x, w_i) = 0, i = 1, ..., m \\ &\Leftrightarrow X^T A \epsilon_i = 0, i = 1, ..., m \\ &\Leftrightarrow X^T A[\epsilon_1, ..., \epsilon_m] = 0 \\ &\Leftrightarrow U^T A^T X = 0 \\ &\Leftrightarrow X \in Null(U^T A^T) \end{aligned}$$

Let $S = Null(U^T A^T)$, we have $x \in W^\perp$ if and only if $[x]_w \in S$. Since $[]_w : V \to F^n$ is an isomorphism, we have the image of $W^\perp$ under $[]_w$ is $U$, and hence we have $dim(W^\perp) = dim(U) = n - rank(U^T A^T) = n - rank(U^T) = n - m = dim(V) - dim(W)$. Thus $dim(W) + dim(W^\perp) = dim(V)$.

Next, we show the second claim. Let $x \in W$, then for all $y \in W^\perp$, we have $f(x, y) = 0$. Since $f$ is symmetric or skew-symmetric, we have $f(y, x) = 0$. Thus, $x \in (W^\perp)^\perp$. Thus $W \subseteq (W^\perp)^\perp$. Apply part one of this proposition, we have $dim(W^\perp) + dim((W^\perp)^\perp) = dim(V)$ and so $dim(W) = dim((W^\perp)^\perp)$ and hence $W = (W^\perp)^\perp$. ♡

**Theorem 3.5.41.** *Let $(V, f)$ be normal orthogonal space, let $W$ be a subspace of $V$, then $dim(W) + dim(W^\perp) = dim(V)$ and $(W^\perp)^\perp = W$.*

*Proof.* Immediately by the above proposition. ♡

**Definition 3.5.42.** Let $(V, f)$ be $n$ dimensional $F$ vector space. A basis $\alpha_1, ..., \alpha_n$ is called ***orthogonal basis*** of $V$ if $f(\alpha_i, \alpha_j) = 0$ for all $i \neq j$.

**Theorem 3.5.43.** *Let $(V, f)$ be $n$ dimensional $F$ vector space with $char(F) \neq 2$. Then $V$ has an orthogonal basis.*

*Proof.* Since $f$ is symmetric bilinear, by Theorem 3.5.19, there exists a basis $v = \{v_1, ..., v_n\}$ such that $[f]_v$ is diagonal. In particular, this imply $f(v_i, v_j) = 0$ for all $i \neq j$. Thus $v$ is an orthogonal basis. ♡

**Remark 3.5.44.** Note that if $(V, f)$ is normal, then $[f]_v$ in the above proof is invertible. Thus $f(v_i, v_i) \neq 0$ for all $v_i \in v$ and hence $v_i$ is anisotropic. Therefore, if $(V, f)$ is normal, then there exists an orthogonal basis with anisotropic vectors. Moreover, every vector in an orthogonal basis for normal orthogonal space $(V, f)$ is anisotropic.

**Definition 3.5.45.** Let $(V, f)$ be $n$ dimensional $F$ orthogonal space, then an orthogonal basis $v_1, ..., v_n$ is called ***orthonormal basis*** if $f(v_i, v_i) = 0$ or $\pm 1$.

**Proposition 3.5.46.** *Let $(V, f)$ be $n$ dimensional normal orthogonal space over $F$. Let $v_1, ..., v_n$ be an orthogonal basis of $V$, then for all $x \in V$, we have*

$$x = \sum_{i=1}^{n} \frac{f(x, v_i)}{f(v_i, v_i)} v_i$$

*Proof.* Let $x \in \sum_{i=1}^{n} c_i v_i$, then $f(x, v_i) = c_i f(v_i, v_i)$ and the proof follows. ♡

**Theorem 3.5.47.** *Let $(V, f)$ be $F$ orthogonal space and $W$ is a finite dimensional subspace of $V$. Let $f_W$ be the restriction of $f$ on $W$. Then $V = W \oplus W^\perp$ if and only if $(W, f_W)$ is normal.*

*Proof.* Suppose $W$ is normal, $f_W$ is non-degenerate. Note since $f$ is symmetric, we have $rad_L(W) = rad_R(W) = \{0\}$ and so we just write $rad(W)$. Note $rad(W) = W \cap W^\perp$. Indeed, $rad(W) \subseteq W$ and $rad(W) \subseteq W^\perp$. Hence, $rad(W) \subseteq W \cap W^\perp$, and if $x \in W \cap W^\perp$, we have $f(x, y) = 0$ for all $y \in W$ and thus $x \in rad(W)$. This imply $rad(W) = W \cap W^\perp$ and thus $W$ and $W^\perp$ intersect trivially.

Since $(W, f_W)$ is finite dimensional normal orthogonal space, there exists an orthogonal basis $v_1, ..., v_m$ of $W$ consists of anisotropic vectors. For arbitrary $\beta \in V$, let $\beta_1 = \sum_{i=1}^m \frac{f(\beta, v_i)}{f(v_i, v_i)} v_i$, we have $\beta_1 \in W$. Let $\beta_2 = \beta - \beta_1$, we have $f(\beta_2, v_i) = 0$ for all $i = 1, ..., m$. Thus $\beta_2 \in W^\perp$ and $\beta = \beta_1 + \beta_2$ and we have $V = W \oplus W^\perp$.

Conversely, suppose $V = W \oplus W^\perp$, then $W \cap W^\perp = \{0\}$. Note $rad(W) = W \cap W^\perp$, and hence we have $f_W$ is non-degenerate, and thus $W$ is normal. ♡

**Remark 3.5.48.** The slightly generalized version of the above theorem is that $W$ is finite dimensional normal orthogonal subspace with $f_W$ then $V = W \oplus W^\perp$. Let $W$ be a subspace (finite or infinite) of $V$, then $V = W \oplus W^\perp$ imply $(W, f_W)$ is normal.

**Definition 3.5.49.** Let $W_1, W_2$ be subspaces of orthogonal space $(V, f)$, if for all $x \in W_1, y \in W_2$ we have $f(x, y) = 0$ then we say $W_1$ and $W_2$ are orthogonal.

**Definition 3.5.50.** Let $(V, f)$ be $F$ orthogonal space, if $T \in L(V, V)$ satisfying $f(x, y) = f(Tx, Ty)$, then we say $T$ is an ***orthogonal operator***.

**Proposition 3.5.51.** *Let $(V, f)$ be normal finite dimensional orthogonal space on $F$ with $char(F) \neq 2$, $T$ is an orthogonal operator if and only if $f(Tx, Tx) = f(x, x)$ for all $x \in V$.*

*Proof.* First, we have $T$ is orthogonal imply $f(Tx, Tx) = f(x, x)$.

Conversely, suppose $f(Tx, Tx) = f(x, x)$. Since $(V, f)$ is normal, we have anisotropic orthogonal basis $v = \{v_1, ..., v_n\}$. In particular, we have $\forall x, y \in V$, $x = \sum_{i=1}^n a_i v_i$ and $y = \sum_{i=1}^n b_i v_i$. Moreover, we have

$$f(Tx, Ty) = \sum_{i=1}^n a_i b_i f(Tv_i, Tv_i) = \sum_{i=1}^n a_i b_i f(v_i, v_i) = f(x, y)$$

Thus $T$ is orthogonal. ♡

# Chapter 4

# Special Classes of Operators

## 4.1 Normal Operators

**Remark 4.1.1.** Let $A \in M_n(F)$, suppose the column of $A$ form an orthonormal basis for $F$, then $AA^* = I \Rightarrow A^{-1} = A^*$.

**Lemma 4.1.2.** *Let $V$ be a finite inner product space, $T : V \to V$ be linear. If $T$ has an eigenvector then $T^*$ has an eigenvector.*

*Proof.* Let $v \neq 0$ in $V$ such that $T(v) = \lambda v$, where $\lambda \in F$. In particular, $(T - \lambda I)(v) = 0$ and for all $x \in V$, we have $\langle (T - \lambda I)(v), x \rangle = \langle v, (T - \lambda I)^*(x) \rangle = 0$ where $(T - \lambda I)^* = T^* - \lambda^* I^* = T^* - \bar{\lambda} I$. Thus, there is a non-zero element $v \in Range(T^* - \bar{\lambda} I)^{\perp}$ and therefore $Range(T^* - \bar{\lambda} I) \neq V$. Hence $Null(T^* - \bar{\lambda} I) \neq \{0\}$. ♡

**Theorem 4.1.3** (Schur). *Let $V$ be finite dimensional inner product space, let $T : V \to V$ be linear. Assume the characteristic polynomial of $T$ splits. Then, there exists an orthonormal basis $\beta$ for $V$ such that $[T]_\beta$ is upper triangular.*

*Proof.* We proceed by induction on $dim(V) = n$.

If $n = 1$ we are done. Assume the result hold for all inner product spaces $V$ with $dim(V) < n$.

Let $dim(V) = n$. Since the characteristic polynomial splits, thus $T$ has an eigenvector. By Lemma 4.1.2, $T^*$ has an eigenvector as well. Suppose $0 \neq v \in V$ and $\lambda \in F$ such that $T^*(v) = \lambda v$. Take $W = span\{v\}$ and thus we have $V = W \oplus W^{\perp}$.

We may assume $\|v\| = 1$, and we claim $W^{\perp}$ is T-invariant. Let $y \in W^{\perp}$, and then $\langle T(y), v \rangle = \langle y, T^*(v) \rangle = \langle y, \lambda v \rangle = \bar{\lambda} \langle y, v \rangle = 0$. Hence, $T(y) \in W^{\perp}$ and so $W^{\perp}$ is T-invariant.

However, $dim(W^{\perp}) = n - 1$ and we can use induction on $W^{\perp}$ as the characteristic polynomial for $T$ restricted to $W^{\perp}$, $T_{W^{\perp}}$, divides the characteristic polynomial of $T$

and thus $T_{W^\perp}$ splits. Hence, we have orthonormal basis $\gamma = \{\gamma_1, ..., \gamma_{n-1}\}$ for $W^\perp$ such that $[T_{W^\perp}]_\gamma$ is upper triangular.

Then, $\beta = \gamma \cup \{v\}$ is an orthonormal basis for $V$. Moreover,

$$[T]_\beta = [T_{W^\perp}(\gamma_1), ..., T_{W^\perp}(\gamma_{n-1}), T(v)]$$

$$\left[\begin{array}{c|c} [T_{W^\perp}]_\gamma & T(v) \\ \hline 0\ 0\ 0 \cdots 0 & \end{array}\right]$$

This closed the induction. ♡

**Corollary 4.1.3.1.** *Let $A \in M_n(F)$ and the characteristic polynomial splits. Then, there exists $U, B$ in $M_n(F)$ such that $U^{-1} = U^*$, $B$ is upper triangular, and $A = UBU^*$.*

*Proof.* We know $[L_A]_\beta = B$ where $B$ is upper triangular. Also, we have $[L_A]_\sigma = A$ and $[I]_\beta^\sigma = U$ and indeed $A = UBU^*$ as $[I]_\beta^\sigma$ is orthogonal matrix. ♡

**Definition 4.1.4.** Let $V$ be finite dimensional inner product space and $T : V \to V$ be linear. We say $T$ is ***normal*** if $TT^* = T^*T$. In addition, we say $A \in M_n(F)$ is ***normal*** if $AA^* = A^*A$.

**Proposition 4.1.5.** *Let $V$ be finite dimensional inner product space, and $T : V \to V$ be normal. Then,*

1. *$\forall x \in V, \|T(x)\| = \|T^*(x)\|$,*
2. *Every $\lambda-$eigenvector of $T$ is a $\bar{\lambda}-$eigenvector of $T^*$,*
3. *If $x$ is a $\lambda-$eigenvector of $T$ and $y$ is a $\mu-$eigenvector of $T$, $\lambda \neq \mu$, then $x, y$ are orthogonal.*

*Proof.*

1. $\|T(x)\|^2 = \langle T(x), T(x)\rangle = \langle x, T^*T(y)\rangle = \langle x, TT^*(x)\rangle = \langle T^*(x), T^*(x)\rangle = \|T^*(x)\|^2$
2. Let $T(v) = \lambda v$ and $\lambda \neq 0$. Consider $U = T - \lambda I$, then $U^* = T^* - \bar{\lambda} I$ and thus $UU^* = U^*U$. Then $\|U^*(v)\| = \|U(v)\| = 0$ and hence $U^*(v) = 0$ which imply $T^*(v) = \bar{\lambda} v$.

3. Let $T(x) = \lambda x$ and $T(y) = \mu y$, $\lambda \neq \mu$, $x, y \neq 0$. Then, $\lambda\langle x, y\rangle = \langle \lambda x, y\rangle = \langle T(x), y\rangle = \langle x, T^*(y)\rangle = \langle x, \bar{\mu} y\rangle = \mu\langle x, y\rangle$. This, we have $(\lambda - \mu)\langle x, y\rangle = 0$, which force $\langle x, y\rangle = 0$ as $\lambda - \mu \neq 0$

♡

**Theorem 4.1.6.** *Let $V$ be finite dimensional inner product space over $\mathbb{C}$, and $T : V \to V$ is linear. Then $T$ is normal if and only if there exists an orthonormal basis $\beta$ for $V$, consisting of eigenvectors of $T$.*

*Proof.* ($\Rightarrow$) Assume $T$ is normal. By Schur's theorem, there exists an orthonormal basis of $V$, $\beta = \{v_1, ..., v_n\}$ , such that $[T]_\beta$ is upper triangular.

We use induction on $v_k$, where $k \leq n$. We claim $v_1, ..., v_{k-1}$ are eignevectors for $T$ then $v_k$ is an eigenvector of $T$. We first show the base case. Say $A = [T]_\beta = (a_{ij})$, then we have $T(v_1) = a_{11}v_1$. Thus, $v_1$ is an eigenvector of $T$.

Next, suppose it holds for all $v_i$ where $1 < i < k \leq n$. We show $v_1, ..., v_{k-1}$ are eigenvectors imply $v_k$ is eigenvector of $T$. Now, $T(v_k) = a_{1k}v_1 + a_{2k}v_2 + ... + a_{kk}v_k$ since $A$ is upper triangular. However, for $i < k$, we have

$$a_{ik} = \langle T(v_k), v_i\rangle = \langle v_k, T^*(v_i)\rangle = \langle v_k, \bar{\lambda} v_i\rangle = \lambda\langle v_k, v_i\rangle = 0$$

where we used Proposition 4.1.5.(2) in the intermediate step. Hence, $T(v_k) = a_{kk}v_k$ and so $\beta$ is an orthonormal basis of eigenvectors.

($\Leftarrow$) Assume there exists an orthonormal basis $\beta$ of eigenvectors of $T$. Then, $[T]_\beta$ is diagonal. However, since $\beta$ is orthonormal, we have $[T^*]_\beta = [T]^*_\beta$ and therefore $[T^*]_\beta$ is diagonal. Therefore, we have $[TT^*]_\beta = [T^*T]_\beta$ as they are diagonal. Therefore, $T^*T = TT^*$ as desired. ♡

**Corollary 4.1.6.1.** *Let $A \in M_n(\mathbb{C})$, then $A$ is normal if and only if there exists $U$ and $D$ in $M_n(\mathbb{C})$ such that $U^{-1} = U^*$, $D$ is diagonal, and $A = UDU^*$.*

*Proof.* Let $A = [L_A]_\sigma$, $D = [L_A]_\beta$ as stated in Theorem 4.1.6, then $U = [I]^\sigma_\beta$. ♡

**Example 4.1.7.** Let $A = \begin{bmatrix} 0 & 1 \\ 0 & -1 \end{bmatrix} \in M_2(\mathbb{R})$, then $AA^* = A^*A = I$, but $f(x) = x^2 + 1$ which imply $A$ is not diagonalizable.

## 4.2 Hermitian Operators and Unitary

**Definition 4.2.1.** Let $T$ be an operator on a finite dimensional inner product space $V$. We say $T$ is ***Hermitian*** if $T = T^*$. We say $A \in M_n(F)$ is ***Hermitian*** if $A = A^*$.

**Proposition 4.2.2.** *Let $V$ be finite dimensional inner product space. Let $T$ be Hermitian linear operator. Then*

1. *Every eigenvalue of $T$ is real,*
2. *The characteristic polynomial of $T$ splits.*

*Proof.* Note 1 imply 2 and thus we only need to show 1.

Note $T$ is Hermitian, and thus $T$ is normal. Let $\lambda$ be an eigenvalue of $T$ with eigenvector $0 \neq x$. Then $T(x) = \lambda x$ and thus $T^*(x) = \bar{\lambda}x$. Therefore, we have $\lambda = \bar{\lambda}$ which imply $\lambda \in \mathbb{R}$. ♡

**Theorem 4.2.3.** *Let $V$ be finite dimensional inner product space over $\mathbb{R}$ and $T$ is linear. Then $T$ is Hermitian if and only if there exists an orthonormal basis $\beta$ consisting of eigenvectors of $T$.*

*Proof.* ($\Rightarrow$) Assume $T$ is Hermitian. By the proposition, its characteristic polynomial splits over $\mathbb{R}$. By Schur's theorem, there exists an orthonormal basis $\beta$ such that $[T]_\beta$ is upper triangular. Thus, $[T]_\beta^* = [T^*]_\beta = [T]_\beta$. Thus, $[T]_\beta$ is symmetric and hence diagonal.

($\Leftarrow$) Assume there exists an orthonormal basis $\beta$ of $T$. Therefore, $[T^*]_\beta = [T]_\beta^*$ where $[T]_\beta^* = [T]_\beta^T$ is diagonal, and thus $[T]_\beta^T = [T]_\beta$ and therefore we have $[T^*]_\beta = [T]_\beta$ which imply $T = T^*$ ♡

**Corollary 4.2.3.1.** *Let $A \in M_n(\mathbb{R})$. Then $A$ is Hermitian iff $\exists$ a symmetric matrix $U \in M_n(\mathbb{R})$ and a diagonal $D \in M_n(\mathbb{R})$ such that $A = UDU^T$.*

**Remark 4.2.4.** Note the similarity between Theorem 4.2.3 and Theorem 4.1.6. The difference is that Theorem 4.2.3 is over $\mathbb{R}$ where Theorem 4.1.6 is over $\mathbb{C}$.

**Proposition 4.2.5.** *Let $V$ be finite dimensional inner product space and $T : V \to V$ be linear. Suppose $W \leq V$ is $T-$invariant.*

1. *$T$ is Hermitian then $T_W$ is Hermitian,*
2. *$W^\perp$ is $T^*-$invariant,*
3. *if $W$ is both $T$ and $T^*$ invariant then $(T_W)^* = (T^*)_W$,*
4. *if $W$ is both $T$ and $T^*$ invariant and $T$ is normal, then $T_W$ is normal.*

*Proof.*
1. Suppose $T$ is Hermitian. Let $x, y \in W$. then $\langle T_W(x), y\rangle = \langle x, T^*(y)\rangle = \langle x, T(y)\rangle = \langle x, T_W(y)\rangle$. By uniqueness, $T_W^* = T_W$ and $T_W$ is Hermitian.
2. Let $x \in W^\perp$ and $w \in W$, then $\langle T^*(x), w\rangle = \langle x, T(w)\rangle = 0$ so $T^*(x) \in W^\perp$.
3. Suppose $W$ is also $T^*-$invariant. Let $x, y \in W$, then
$$\langle T_W(x), y\rangle = \langle T(x), y\rangle = \langle x, T^*(y)\rangle$$
Note $T^*(y) \in W$, so by uniqueness of the adjoint, we have $(T_W)^* = (T^*)_W$.
4. Suppose $W$ is $T^*-$invariant. Assume $T$ is normal. Then, for all $w \in W$, we have
$$\begin{aligned}(T_W)(T_W)^*(w) &= (T_W)(T^*)_W(w) = TT^*(w)\\ &= T^*T(w) = (T^*)_W(T_W)(w) = (T_W)^*(T_W)(w)\end{aligned}$$
Thus $(T_W)(T_W)^* = (T_W)^*(T_W)$ and $T_W$ is normal.

♡

**Definition 4.2.6.** Let $V$ be finite dimensional inner product space over $F$, and $T : V \to V$ be linear. If $T^{-1} = T^*$, we say $T$ is ***orthogonal*** if $F = \mathbb{R}$ and ***unitary*** if $F = \mathbb{C}$. We define the same thing for matrices.

**Remark 4.2.7.** The matrix $A$ is unitary/orthogonal iff $L_A$ is unitary/orthogonal.

**Proposition 4.2.8.** *Let $V$ be finite dimensional inner product space, $T : V \to V$ be linear. Then the following are equivalent:*

1. *$T$ is unitary/orthogonal*
2. *$\langle T(x), T(y)\rangle = \langle x, y\rangle$ for all $x, y \in V$*
3. *If $\beta$ is an orthonormal basis for $V$ then $T(\beta)$ is an orthonormal basis for $V$*
4. *There exists an orthonormal basis $\beta$ for $V$ such that $T(\beta)$ is an orthonormal basis for $V$.*
5. *For all $x \in V$, $\|T(x)\| = \|x\|$.*

*Proof.* $(1 \Rightarrow 2)$ Assume $T^{-1} = T^*$. Then, $\langle T(x), T(y)\rangle = \langle x, T^*T(y)\rangle = \langle x, y\rangle$.

$(2 \Rightarrow 3)$ Assume $\beta = \{v_1, ..., v_n\}$ is an orthonormal basis for $V$. We first show $T$ is injective. Let $x \in Ker(T)$, then $0 = \|T(x)\|^2 = \langle T(x), T(x)\rangle = \langle x, x\rangle$ and thus $x = 0$. Hence $Ker(T) = \{0\}$ and $T$ is injective as desired. Therefore, $T(\beta)$ is a basis for $V$.

For $v_i, v_j \in \beta$, $i \neq j$, then $\langle T(v_i), T(v_j)\rangle = \langle v_i, v_j\rangle = 0$. Similarly, we have $\|T(v_i)\|^2 = \|v\|^2 = 1$. Thus $T(\beta)$ is orthonormal.

$(3 \Rightarrow 4)$ trivial

$(4 \Rightarrow 5)$ Assume (4), and let $\beta$ be an orthonormal basis for $V$ such that $T(\beta)$ is an orthonormal basis. Let $\beta = \{v_1, ...v_n\}$. Let $x \in V$ so that $x = \sum_{i=1}^n a_i v_i$ where $a_i \in F$. Then, $T(x) = \sum_{i=1}^n a_i T(v_i)$. Then, $\|x\|^2 = \langle \sum_{i=1}^n a_i v_i, \sum_{i=1}^n a_i v_i\rangle$, and thus

$$\|x\|^2 = \sum_{i=1}^n \langle a_i v_i, a_i v_i\rangle = \sum_{i=1}^n a_i \overline{a_i} \langle v_i, v_i\rangle = \sum_{i=1}^n a_i \overline{a_i} = \sum_{i=1}^n |a_i|^2$$

Similarly, we have

$$\|T(x)\|^2 = \langle \sum_{i=1}^n a_i T(v_i), \sum_{i=1}^n a_i T(v_i)\rangle = \sum_{i=1}^n |a_i|^2$$

$(5 \Rightarrow 1)$ Assume (5). Then, for all $x \in V$, we have $\langle T(x), T(x)\rangle = \langle x, x\rangle$ which imply $\langle x, T^*T(x)\rangle = \langle x, x\rangle$. Therefore, $\langle x, (T^*T - I)(x)\rangle = 0$. Denote $U = T^*T - I$, we need to show $U = 0$.

Note $U$ is Hermitian, and then thus there exists an orthonormal basis of eigenvectors of $U$ for $V$. Let $0 \neq x$ be an $\lambda-$eigenvector of $U$. Then, we have $\langle x, U(x)\rangle = \langle x, \lambda x\rangle = \overline{\lambda}\langle x, x\rangle = 0$, and so we must have $\overline{\lambda} = \lambda = 0$ as $x \neq 0$. Hence, the only possible eigenvalue of $U$ is zero. Therefore, since $U$ is diagonalizable as $U$ is Hermitian, we must have $U = 0$ as the eigenvalues are all zero. Hence, $T^*T - I = 0$ and $T^*T = I$. ♡

**Corollary 4.2.8.1.** *Let $V$ be finite dimensional inner product space, if $T$ is orthogonal/unitary then every eigenvalues of $T$ has absolute value* 1*.*

*Proof.* $|\lambda| \, \|x\| = \|\lambda x\| = \|T(x)\| = \|x\|$. Hence $\lambda = 1$ ♡

**Corollary 4.2.8.2.** *Let $V$ be finite dimensional real inner product space, and $T : V \to V$ be linear. Then $T$ is orthogonal and Hermitian if and only if there exists an orthonormal basis for $V$ consisting of $\pm 1$-eigenvectors of $T$.*

*Proof.* Left as an exercise. ♡

**Corollary 4.2.8.3.** *Let $V$ be a finite dimensional $\mathbb{C}$ inner product space, and $T : V \to V$ be linear. Then $T$ is unitary if and only if there exists an orthonormal basis consisting of eigenvectors of $T$ where all eigenvectors have eigenvalue on the unit circle.*

**Definition 4.2.9.** Let $A, B \in M_n(F)$, we say $A, B$ are ***orthogonally/unitarily equivalent***, if there exists an orthogonal/unitary matrix $U \in M_n(F)$ such that $A = UBU^*$. If $B$ is diagonal, we say $A$ is ***orthogonally/unitarily diagonalizable***.

**Example 4.2.10.** Let $A, B \in M_n(F)$. Prove that if $A = (a_{ij})$ and $B = (b_{ij})$ are unitarily/orthogonally equivalent then

$$\sum_{i,j} |a_{ij}|^2 = \sum_{i,j} |b_{ij}|^2$$

*Solution.* Suppose $A, B \in M_n(F)$ are unitarily/orthogonally equivalent. Then there exists unitary/orthogonal matrices $U \in M_n(F)$ such that $A = UBU^*$. Then

$$\begin{aligned} \sum |a_{ij}|^2 &= Tr(A^*A) = Tr(UB^*U^*UBU^*) \\ &= Tr(UB^*BU^*) = Tr(B^*B) \\ &= \sum |b_{ij}|^2 \end{aligned}$$

♠

**Example 4.2.11.** Show that $A = \begin{bmatrix} 1 & 2 \\ 2 & i \end{bmatrix}$ and $B = \begin{bmatrix} i & 4 \\ 1 & 1 \end{bmatrix}$ are not unitarily equivalent.

*Solution.* Note $\sum |a_{ij}| = 10$ and $\sum |b_{ij}|^2 = 19$, so they cannot be unitarily equivalent by Example 4.2.10. ♠

**Example 4.2.12.** Find all $A \in M_2(\mathbb{R})$ such that the characteristic polynomial of $A$ splits and $A$ is orthogonally equivalent to its Jordan canonical form.

*Solution.* A few remarks. We note $A^*A = I$ if and only if $AA^* = I$.

We note $A$ is orthogonally equivalent to $B$ if and only if there exists $U$ such that $U^*U = I$ and $A = U^*BU$ if and only if there exists $M$ such that $M^*M = I$ and $A = MBM^*$.

There are only two possible cases for Jordan canonical form. Namely, either $A$ has Jordan canonical form with one size 2 Jordan block, or two size 1 Jordan block.

***Case One***: If $A$ has Jordan canonical form is $J = \begin{bmatrix} x & 0 \\ 0 & y \end{bmatrix}$, then $A$ is orthogonally equivalent to $J$ if and only if there exists matrix $U$ such that $A = U^*JU$ where $U^* = U^{-1}$ ,or equivalently, $U^*U = I$. Let $U = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, then $U^*U = I$ if and only if $\begin{cases} a^2 + c^2 = 1 \\ b^2 + d^2 = 1 \\ ab + cd = 0 \end{cases}$ Note since $a^2 + c^2 = 1$ and $b^2 + d^2 = 1$, they are on the unit circle and thus there exists $0 \leq \theta, \alpha \leq 2\pi$ so $a = \cos\theta$, $c = \sin\theta$, $b = \cos\alpha$ and $d = \sin\alpha$. In particular, since $ab + cd = 0$, we have $\cos\theta\cos\alpha + \sin\theta\sin\alpha = \cos(\theta - \alpha) = 0$ and thus $\theta - \alpha = \frac{\pi}{2}$ or $\theta - \alpha = \frac{3\pi}{2}$. Hence, $U$ is orthogonal if and only if there exists $0 \leq \theta, \alpha \leq 2\pi$ such that $U = \begin{bmatrix} \cos\theta & \cos\alpha \\ \sin\theta & \sin\alpha \end{bmatrix}$ and $\theta - \alpha = \frac{\pi}{2}$ or $\theta - \alpha = \frac{3\pi}{2}$.

Therefore, $A$ must be in the form

$$\begin{aligned} U^*JU &= \begin{bmatrix} \cos\theta & \sin\theta \\ \cos\alpha & \sin\alpha \end{bmatrix} \begin{bmatrix} x & 0 \\ 0 & y \end{bmatrix} \begin{bmatrix} \cos\theta & \cos\alpha \\ \sin\theta & \sin\alpha \end{bmatrix} \\ &= \begin{bmatrix} \cos\theta^2 x + \sin\theta^2 y & \cos\theta\cos\alpha x + \sin\theta\sin\alpha y \\ \cos\theta\cos\alpha x + \sin\theta\sin\alpha y & \cos\alpha^2 x + \sin\alpha^2 y \end{bmatrix} \end{aligned}$$

***Case Two***: If $A$ has Jordan canonical form $J = \begin{bmatrix} x & 1 \\ 0 & x \end{bmatrix}$. Then, $A$ is orthogonally equivalent to $J$ if and only if there exists $U$ such that $U^*U = I$ and $A = UJU^*$. With the same argument, we see $U$ must be in the form $\begin{bmatrix} \cos\theta & \cos\alpha \\ \sin\theta & \sin\alpha \end{bmatrix}$ where $0 \leq \theta, \alpha \leq 2\pi$ and $\theta - \alpha = \frac{\pi}{2}$ or $\theta - \alpha = \frac{3\pi}{2}$. Therefore, we have $A$ must be in the form

$$\begin{aligned} U^*JU &= \begin{bmatrix} \cos\theta & \sin\theta \\ \cos\alpha & \sin\alpha \end{bmatrix} \begin{bmatrix} x & 1 \\ 0 & x \end{bmatrix} \begin{bmatrix} \cos\theta & \cos\alpha \\ \sin\theta & \sin\alpha \end{bmatrix} \\ &= \begin{bmatrix} x + \cos\theta\sin\theta & \cos\theta\sin\alpha \\ \cos\alpha\sin\theta & x + \cos\alpha\sin\alpha \end{bmatrix} \end{aligned}$$

♠

**Remark 4.2.13.** To summary the operators, we have

1. $A \in M_n(\mathbb{C})$, then $A$ is normal if and only if $A$ is unitarily diagonalizable
2. $A \in M_n(\mathbb{R})$, then $A$ is Hermitian if and only if $A$ is orthogonally diagonalizable
3. $A \in M_n(\mathbb{F})$, then $A$ is unitarily/orthogonally equivalent to an upper-triangular matrix.

**Example 4.2.14.** Let $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$, then $A$ is not symmetric so it is not orthogonally diagonalizable. However, $A$ is unitary and hence is unitarily diagonalizable in $\mathbb{C}$.

Indeed, $det(A - xI) = f(x) = (x - i)(x + i)$. Then we have $u_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} i \\ 1 \end{bmatrix}$ is an $i$-eigenvector and $u_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ i \end{bmatrix}$ is an $-i$-eigenvector. Then, we have $u_1, u_2$ are orthonormal basis of $\mathbb{C}^2$. Thus, we have $A = UDU^*$ where $U = [u_1, u_2]$, $D = diag(i, -i)$.

**Example 4.2.15.** Let $A = \begin{bmatrix} 0 & 2 & 2 \\ 2 & 0 & 2 \\ 2 & 2 & 0 \end{bmatrix}$. Since $A$ is symmetric, then $A$ is orthogonally diagonalizable.

Indeed, $det(A-xI) = f(x) = (x+2)^2(x-4)$. Thus $\lambda_1 = -2, \lambda_2 = 4$ are the only two eigenvectors. Then, we have $u_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$ and $u_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$ are $\lambda_1$-eigenvectors. Moreover, $u_3 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ is $\lambda_2$-eigenvector. Then, note $\{u_1, u_2, u_3\}$ is orthonormal basis for $V$ and thus $A = UDU^*$ where $U = [u_1, u_2, u_3]$ and $D = diag(-2, -2, 4)$.

**Example 4.2.16.** Let $V$ be $m$ dimensional complex inner product space and $T : V \to V$ be a normal linear operator. Prove that, for all $n \in N$, there exists a normal linear operator $U : V \to V$ such that $U^n = T$.

*Solution.* Since $T$ is normal and $\beta = (v_1, ..., v_m)$ be orthonormal ordered basis of $V$, we have $S = [T]_\beta \in M_m(\mathbb{C})$ and $S$ is normal. Thus, there exists unitary matrix $A \in M_m(\mathbb{C})$ and diagonal matrix $D \in M_m(\mathbb{C})$, $D = diag(d_1, ..., d_m)$, such that $S = A^*DA$. In particular, for any $n \in \mathbb{N}$, we have $D = (D_n)^n$ where $D_n = diag(d_1^{\frac{1}{n}}, ..., d_m^{\frac{1}{n}})$. Thus, let $W = A^*D_nA$, we claim $L_W$ is normal and $L_W^n = T$. Note $W$ is a normal matrix as $W = A^*D_nA$ where $D_n$ is diagonal and $A$ is unitary. Thus, we have $W^*W = WW^*$ and hence $[L_W L_W^*]_\beta = [L_W^* L_W]_\beta$ and thus we must have $L_W L_W^* = L_W^* L_W$. Next, we note $W^n = A^*D_nAA^*D_nA...A^*D_nA = A^*(D_n)^nA = S$, and thus $[L_W^n(x)]_\beta = [L_W^n]_\beta[x]_\beta = [L_W]^n[x]_\beta = W^n[x]_\beta = S[x]_\beta = [T]_\beta[x]_\beta = [T(x)]_\beta$. Hence. $L_W^n = T$ as $x$ was arbitrary. ♠

**Example 4.2.17.** Let $T : \mathbb{C}^2 \to \mathbb{C}^2$ be given by $T(x, y) = (x - iy, 3y + ix)$. Find a normal operator $U : \mathbb{C}^2 \to \mathbb{C}^2$ such that $U^3 = T$.

*Solution.* Consider $\sigma = (e_1, e_2)$, then $[T]_\sigma = \begin{bmatrix} 1 & -i \\ i & 3 \end{bmatrix}$ and thus $[T^*]_\sigma = [T]_\sigma^*$ as $\sigma$ is orthonormal. Hence, $[T^*]_\sigma = \begin{bmatrix} 1 & -i \\ i & 3 \end{bmatrix} = [T]_\sigma$. This imply $T = T^*$. Thus, we have $TT^* = TT = T^*T$ and thus normal as desired.

Note the characteristics polynomial of $T$ is $det([T]_\sigma - xI) = x^2 - 4x + 2$. A eigenvector for $\lambda_1 = 2 - \sqrt{2}$ would be $x_1 = \frac{1}{\sqrt{4-2\sqrt{2}}}\begin{bmatrix} 1 \\ i(1-\sqrt{2}) \end{bmatrix}$, then, we have $x_2 = \frac{1}{\sqrt{4+2\sqrt{2}}}\begin{bmatrix} 1 \\ i(1+\sqrt{2}) \end{bmatrix}$ be an eigenvector of $\lambda_2 = 2 + \sqrt{2}$. Thus, we note $\langle x_1, x_2 \rangle = 0$ and $\|x_1\| = \|x_2\| = 1$, and so they are orthonormal basis of $\mathbb{C}^2$. In particular, we note $W = \begin{bmatrix} \frac{1}{\sqrt{4-2\sqrt{2}}} & \frac{1}{\sqrt{4+2\sqrt{2}}} \\ \frac{i(1-\sqrt{2})}{\sqrt{4-2\sqrt{2}}} & \frac{i(1+\sqrt{2})}{\sqrt{4+2\sqrt{2}}} \end{bmatrix}$, $D = \begin{bmatrix} 2-\sqrt{2} & 0 \\ 0 & 2+\sqrt{2} \end{bmatrix}$ then we have $WW^* = I$ and $[T]_\sigma = WDW^*$. Let $a = (2-\sqrt{2})^{1/3}$ and $b = (2+\sqrt{2})^{1/3}$, we have

$$P = W\begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}W^* = \begin{bmatrix} \frac{b^3a+ba^2}{4} & \frac{\sqrt{2}i(a-b)}{4} \\ \frac{\sqrt{2}i(b-a)}{4} & \frac{-(1-\sqrt{2})a+(1+\sqrt{2})b}{2\sqrt{2}} \end{bmatrix}$$

Thus, we have $L_P^3 = T$. ♠

## 4.3 Rigid Motions

**Definition 4.3.1.** Let $V$ be finite dimensional real inner product space, $f : V \to V$ is rigid motion if $\|f(x) - f(y)\| = \|x - y\|$ for all $x, y \in V$.

**Example 4.3.2.**

1. For $\mathbb{R}^2$, the counter-clockwise rotation by $\theta$ is rigid motion
2. Let $v \in \mathbb{R}^2$ be fixed. $f : \mathbb{R}^2 \to \mathbb{R}^2$ given by $f(x) = x + v$ is rigid motion.
3. The reflection over $y = mx$ is a rigid motion.

**Definition 4.3.3.** Let $V$ be vector space, a ***translation*** is a function $f : V \to V$ given by $f(x) = x + v$ for some fixed $v \in V$.

**Remark 4.3.4.** Note for $W \leq V$ and $v \in V$, we say $W + v = \{u + v : u \in W\}$ be a affine space in $V$.

**Theorem 4.3.5.** *Let $V$ be finite dimensional real inner product space, let $f : V \to V$ be rigid motions, there exists a unique orthogonal operator $T : V \to V$ and a unique translation $g$ such that $f = g \circ T$.*

*Proof.* Define $T : V \to V$ by $T(x) = f(x) - f(0)$.

Claim: $T$ is linear and orthogonal.

1. For $x, y \in V$, then
$$\begin{aligned} \|T(x) - T(y)\| &= \|f(x) - f(0) - f(y) + f(0)\| \\ &= \|f(x) - f(y)\| = \|x - y\| \end{aligned}$$
Thus, $T$ is a rigid motion.
2. For $x \in V$, $\|T(x)\|^2 = \|f(x) - f(0)\|^2 = \|x - 0\|^2 = \|x\|^2$, thus if we can show $T$ is linear, then by Proposition 4.2.8.5, $T$ must be orthogonal.

3. For $x, y \in V$, we have $\|T(x)-T(y)\|^2 = \|T(x)\|^2 + \|T(y)\|^2 - 2\langle T(x), T(y)\rangle = \|x\|^2 + \|y\|^2 - 2\langle T(x), T(y)\rangle$ and $\|T(x)-T(y)\|^2 = \|x-y\|^2 = \|x\|^2 + \|y\|^2 - 2\langle x, y\rangle$. Thus, we have $\|x\|^2 + \|y\|^2 - 2\langle x, y\rangle = \|x\|^2 + \|y\|^2 - 2\langle T(x), T(y)\rangle$ which imply $\langle x, y\rangle = \langle T(x), T(y)\rangle$.
4. Let $x, y \in V$ and let $\alpha \in \mathbb{R}$. Then,

$$\begin{aligned}
\|T(x+\alpha y) - T(x) - \alpha T(y)\|^2 &= \|T(x+\alpha y) - T(x)\|^2 + \|\alpha T(y)\|^2 \\
&\quad - 2\langle T(x+\alpha y) - T(x), \alpha T(y)\rangle \\
&= \|x+\alpha y - x\|^2 + \alpha^2 \|y\|^2 \\
&\quad - 2\langle T(x+\alpha y), \alpha T(y)\rangle - 2\langle -T(x), \alpha T(y)\rangle \\
&= 2\alpha^2 \|y\|^2 - 2\alpha\langle x+\alpha y, y\rangle + 2\alpha\langle x, y\rangle \\
&= 2\alpha^2 \|y\|^2 - 2\alpha\langle x, y\rangle - 2\alpha^2 \|y\|^2 + 2\alpha\langle x, y\rangle \\
&= 0
\end{aligned}$$

Thus, $T$ is linear and by 2, we have $T$ is orthogonal.

For uniqueness, suppose $T, U$ are orthogonal, and $a, b \in V$ such that $f(x) = T(x) + a = U(x) + b$. Then, $f(0) = a = b$ and thus $U = T$. ♡

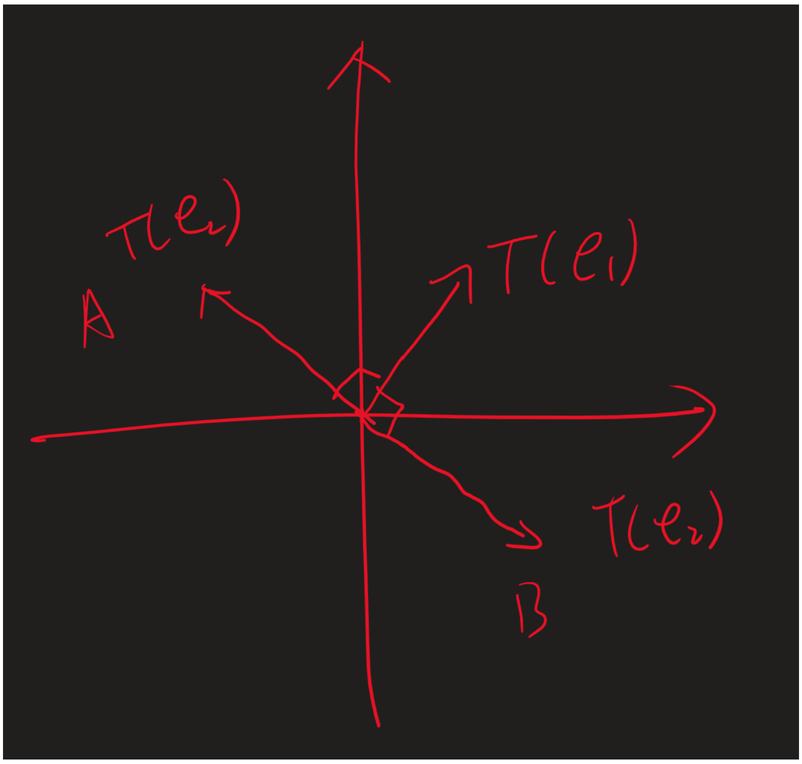
**Example 4.3.6.** Let $T_\theta : \mathbb{R}^2 \to \mathbb{R}^2$, counter-clockwise $\theta$ rotation is a rigid motion. Clearly, for all $x \in \mathbb{R}^2$, $\|T_\theta(x)\| = \|x\|$ and hence $T_\theta$ is orthogonal. Moreover, we have $T_\theta(e_1) = \begin{bmatrix}\cos\theta \\ \sin\theta\end{bmatrix}$ and $T_\theta(e_2) = \begin{bmatrix}-\sin\theta \\ \cos\theta\end{bmatrix}$. Therefore, $[T_\theta]_\sigma = \begin{bmatrix}\cos\theta & -\sin\theta \\ \sin\theta & \cos\theta\end{bmatrix}$

**Example 4.3.7.** Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a reflection over $y = mx$. Let $\alpha$ be the angle between the line and the $x$ axis. Then, let $v_1 = \begin{bmatrix}\cos\alpha \\ \sin\alpha\end{bmatrix}$ and $v_2 = \begin{bmatrix}-\sin\alpha \\ \cos\alpha\end{bmatrix}$, then $\beta = \{v_1, v_2\}$ form a orthonormal basis of $\mathbb{R}^2$. Thus, $[T]_\beta = \begin{bmatrix}1 & 0 \\ 0 & -1\end{bmatrix}$, and $[T]_\sigma = U[T]_\beta U^T$ where $U = \begin{bmatrix}\cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha\end{bmatrix}$. Therefore, $[T]_\sigma = \begin{bmatrix}\cos(2\alpha) & \sin(2\alpha) \\ \sin(2\alpha) & -\cos(2\alpha)\end{bmatrix}$.

**Proposition 4.3.8.** *Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be an orthogonal operator. Then $T$ is a rotation or a reflection over a line through the origin. In particular, if $f : \mathbb{R}^2 \to \mathbb{R}^2$ is a rigid motion, then $f$ is either a rotation or a reflection followed by a translation.*

*Proof.* Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be orthogonal.

Let $\sigma = \{e_1, e_2\}$ be the standard basis for $\mathbb{R}^2$. Since $T$ is orthogonal, we have $\sigma = \{T(e_1), T(e_2)\}$ is an orthonormal basis for $\mathbb{R}^2$. Then, $\|T(e_1)\| = 1$ and thus $T(e_1) = (\cos\theta, \sin\theta)$ for some $\theta$. Note $\langle T(e_1), T(e_2)\rangle = 0$ and $\|T(e_2)\| = 1$. Thus,




If $T(e_2)$ is case $A$, then $T(e_2) = (\cos(\theta + \pi/2), \sin(\theta + \pi/2)) = (-\sin\theta, \cos\theta)$. Then, $[T]_\sigma = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ and thus $T$ is a rotation.

If $T(e_2)$ is case $B$, then $T(e_2) = (\cos(\theta - \pi/2), \sin(\theta - \pi/2)) = (\sin\theta, -\cos\theta)$, then $[T]_\sigma = \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix}$ and thus $T$ is a reflection with $\alpha = \frac{\theta}{2}$. ♡

## 4.4 Spectral Theorem

**Definition 4.4.1.** Let $V$ be vector space, let $W_1, W_2$ be subspaces of $V$ such that $V = W_1 \oplus W_2$. Let $v \in V$, then there exists unique $x \in W_1, y \in W_2$ such that $v = x + y$. Then, the linear map $T : V \to V$ given by $T(v) = x$ is called the ***projection*** on $W_1$ along $W_2$.

**Remark 4.4.2.** Let $T$ be projection on $W_1$ along $W_2$, then $Range(T) = W_1$ and $Null(T) = W_2$. Note the difference between orthogonal projection and projection is that orthogonal projection must have $T^* = T$.

**Proposition 4.4.3.** *Let $V$ be a vector space. Then linear operator $T : V \to V$ is a projection if and only if $T = T^2$.*

*Proof.* ($\Rightarrow$) Assume $T$ is a projection. Thus, $T$ is the projection on $Range(T)$ along $Null(T)$. Let $v \in V$ such that $v = T(x) + z$ where $z \in Null(T)$. Thus, $T^2(v) = T(T(v)) = T(T(x) + z) = T(T(x)) + 0 = T(x) = T(v)$. Thus $T^2 = T$.

($\Leftarrow$) Suppose $T^2 = T$. We first claim $V = Range(T) \oplus Null(T)$. Let $x \in Range(T) \cap Null(T)$. Thus, $x = T(y)$ and $T(x) = 0$. Thus $0 = T(x) = T(T(y)) = T(y) = x$ and thus $Range(T) \cap Null(T) = \{0\}$. However, note $dim(Range(T) \oplus Null(T)) = dim(Range(T)) + dim(Null(T)) = dim(V)$. Thus $Range(T) \oplus Null(T) = V$.

Next, let $v \in V$, we have $v = T(x) + z$ where $x \in V$ and $z \in Null(T)$ as $Range(T) \oplus Null(T) = V$. Then, $T(v) = T(T(x)) = T(x)$. Thus, $T$ is the projection on $Range(T)$ along $Null(T)$. ♡

**Remark 4.4.4.** Let $V$ be finite dimensional inner product space and let $W$ be a subspace of $V$, then $V = W \oplus W^\perp$. We have $proj_W : V \to V$ is a orthogonal projection on $V$.

**Proposition 4.4.5.** *Let $V$ be finite dimensional inner product space and let $T : V \to V$ be linear. Then $T$ is an orthogonal projection if and only if $T^2 = T = T^*$.*

*Proof.* Let $T : V \to V$ be the orthogonal projection from $V$ to $W$ where $W \leq V$ is a subspace. We first note that $T$ is a linear operator(one can check this if they desire more works).

We first show that $T$ is orthogonal projection then $T^2 = T$ and $T^* = T$. Note $T(x) \in W$ so that $T(T(x)) = T(x)$ for all $x \in V$, hence $T^2 = T$ as desired. Next, note $T(x) \in W$ and for all $y \in V$, we have $y - T(y) \in W^\perp$, thus

$$\forall x, y \in V, \langle T(x), y - T(y)\rangle = 0$$

Hence, $\forall x, y \in V, \langle T(x), y\rangle = \langle T(x), T(y)\rangle$. On one hand, we have $\langle T(x), y\rangle = \langle x, T^*(x)\rangle$ for all $x, y \in V$. Thus, from $\langle T(x), y\rangle = \langle T(x), T(y)\rangle$, we have $\langle T(x), y\rangle = \langle x, T^*T(y)\rangle = \langle x, T^*(y)\rangle$ for all $x, y \in V$ and so $T^*T = T^*$. Next, note $(T^*)^* = T = (T^*T)^* = T^*T$, so $T = T^*$ as desired.

Conversely, suppose $T^2 = T$ and $T^* = T$, we will show $T(x) = proj_W(x)$ where $W = range(T)$. First, note $range(T^*) \oplus null(T) = V$ by Proposition 3.4.17, where $range(T^*) = range(T)$ as $T = T^*$, so we have $range(T) \oplus null(T) = V$ where $W = range(T)$. Thus, for all $x \in V$, we have $x = w_1 + w_2$ where $w_1 \in W$ and $w_2 \in null(T)$ and $T(x) = T(w_1 + w_2) = T(w_1) + 0 = T(w_1)$. However, note $w_1 \in W = range(T)$, so $w_1 = T(\eta)$ for some $\eta \in V$. Thus $T(w_1) = T(T(\eta)) = T(\eta) = w_1$, so we have $T(x) = w_1$, and hence $T$ is orthogonal projection as desired. ♡

**Definition 4.4.6.** The ***Kronecker delta*** is defined as $\delta_{ij} = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$.

**Theorem 4.4.7** (Spectral Theorem)**.** *Let $V$ be finite dimensional inner product space, let $T : V \to V$ be linear and $T$ have all the distinct eigenvalues to be $\lambda_1, ..., \lambda_k$. Assume $T$ is normal if $F = \mathbb{C}$ or Hermitian if $F = \mathbb{R}$. Let $W_i = E_{\lambda_i}$ and $T_i(x) = proj_{W_i}(x)$ for $1 \leq i \leq k$. Then,*

1. $V = W_1 \oplus W_2 \oplus ... \oplus W_k$
2. $W_i^\perp = \bigoplus_{j \neq i} W_j =: W_i'$
3. $T_i \circ T_j = \delta_{ij} T_i$

4. $I = \sum_{i=1}^n T_i$
5. $T = \sum_{i=1}^n \lambda_i T_i$, *note this is called the* ***spectral decomposition*** *of* $T$

*Proof.*

1. Since $T$ is normal/Hermitian, we have $T$ is diagonalizable and thus $V = \bigoplus_{i=1}^k W_i$ as $\bigoplus_{i=1}^k W_i \leq V$ and $dim(\bigoplus_{i=1}^k W_i) = dim(V)$.
2. Let $x = x_1 + ... + x_{i-1} + x_{i+1} + ... + x_k \in W_i'$ where $x_j \in W_j$ and let $y \in W_i$. THen, $\langle x, y\rangle = \sum_{j\neq i}\langle x_j, y\rangle = \sum 0 = 0$ since $T$ is normal. Thus $W_i' \subseteq W_i^\perp$, however, $dim(W_i') + dim(W_i) = n = dim(W_i^\perp) + dim(W_i)$ and thus $W_i' = W_i^\perp$.
3. We know $T_i \circ T_i = T_i$ since $T_i$ is a projection. Suppose $i \neq j$ and let $x \in V$. Then, $x = \sum_{l=1}^k x_l$ where $x_l \in W_l$. Therefore, $T_i \circ T_j(x) = \sum_{l=1}^k T_i(T_j(x_l)) = T_i(x_j) = 0$.
4. Let $x = \sum_{i=1}^k x_i \in V$, $x_i \in W_i$. Then $(\sum_{i=1}^k T_i)(x) = T_1(x) + T_2(x) + ... + T_k(x) = x_1 + x_2 + ... + x_k = x$.
5. Let $x = x_1 + ... + x_k$ where $x_i \in W_i$. We have $(\sum_{i=1}^k \lambda_i T_i)(x) = \sum_{i=1}^k \lambda_i x_i = \sum_{i=1}^k T(x_i) = T(x)$.

♡

**Definition 4.4.8.** Let $V$ be vector space, the set of eigenvalues of $T$ is called the ***spectrum*** of $T$ and we write $\sigma(T)$.

**Remark 4.4.9** (Lagrange interpolation)**.** Let $c_0, c_1, ..., c_n \in F$ be distinct. Let

$$f_i(x) = \frac{\prod_{i=0}^n (x - c_i)}{(c_i - c_0)(c_i - c_1)...(c_i - c_{i-1})(c_i - c_{i+1})...(c_i - c_n)}$$

We note $f_i(c_j) = 0$ and $f_i(c_i) = 1$.

Then, $\{f_0, ..., f_n\}$ is a basis for $P_n(F)$. Indeed, let $\sum_{i=0}^n a_i f_i = 0$, then $0 = (\sum_{i=0}^n a_i f_i)(c_i) = a_i$. Thus, let $g(x) \in P_n(F)$ with $g(x) = \sum_{j=0}^n b_j f_j(x)$, then $g(c_i) = b_i$.

**Remark 4.4.10.** Let $T$ be normal/Hermitian, and suppose the spectral decomposition is $T = \sum_{i=1}^n \lambda_i T_i$. Then $T^2 = \sum_{i=1}^n \lambda_i^2 T_i$. Similarly, we see $T^l = \sum_{i=1}^n \lambda_i^l T_i$.

Thus, let $f(x) = \sum_{j=1}^k a_j x^j \in F[x]$, we have $f(T) = \sum_{i=1}^n f(\lambda_i) T_i$ as

$$\begin{aligned} f(T) &= \sum_{j=1}^k a_j (T)^j = \sum_{j=1}^k a_j \sum_{i=1}^n \lambda_i^j T_i = \sum_{j=1}^k \sum_{i=1}^n a_j \lambda_i^j T_i \\ &= \sum_{i=1}^n \left( \sum_{j=1}^k a_j \lambda_i^j \right) T_i = \sum_{i=1}^n f(\lambda_i) T_i \end{aligned}$$

**Proposition 4.4.11.** *Let* $F = \mathbb{C}$ *and* $V$ *be finite dimensional* $\mathbb{C}$ *inner product space. Then* $T$ *is normal if and only if* $T^* = f(T)$ *for some* $f(x) \in F[x]$.

*Proof.* ($\Leftarrow$) It is clear.

($\Rightarrow$) Suppose $T$ is normal and $T = \sum_{i=1}^{k} \lambda_i T_i$ is the spectral decomposition of $T$. Using Lagrange interpolation, let $f(x) \in F[x]$ such that $f(\lambda_i) = \overline{\lambda_i}$.

Thus, $T^* = \sum_{i=1}^{k} \overline{\lambda_i} T_i^* = \sum_{i=1}^{k} \overline{\lambda_i} T_i = \sum_{i=1}^{k} f(\lambda_i) T_i = f(T)$. The last step is obtained by 4.4.10. ♡

**Proposition 4.4.12.** *Let $T$ be normal/Hermitian and $T = \sum_{i=1}^{k} \lambda_i T_i$ be the spectral decomposition. For each $i$, there exists $g_i(x) \in F[x]$ such that $g_i(T) = T_i$.*

*Proof.* Let $g_i(x) \in F[x]$ be $g_i(\lambda_i) = \delta_{ij}$, then by Remark 4.4.10 we are done . ♡

**Proposition 4.4.13.**

1. *Let $V$ is finite dimensional real inner product space. Let $U, T$ be Hermitian. Prove that there exists an orthonormal basis for $V$ consisting of eigenvectors of both $U$ and $T$ at the same time if and only if $UT = TU$.*
2. *Let $V$ is finite dimensional complex inner product space. Let $U, T$ be normal. Prove that there exists an orthonormal basis for $V$ consisting of eigenvectors of both $U$ and $T$ at the same time if and only if $UT = TU$.*

*Proof.*

1. Let $V$ be finite real inner product space. Assume $T, U$ are Hermitian. Suppose there exists an orthonormal basis $\beta$ for $V$ consisting eigenvectors of both $T$ and $U$. Then $[T]_\beta$ and $[U]_\beta$ are diagonal and $[TU]_\beta = [T]_\beta [U]_\beta = [UT]_\beta$.
   Conversely, suppose $TU = UT$. Since $T$ is Hermitian, $T$ is diagonalizable. Let $\lambda_1, ..., \lambda_k$ be distinct eigenvalues of $T$, let $W_i = E_{\lambda_i}$, we have $V = W_1 \oplus W_2 \oplus ... \oplus W_k$. Note $W_i$ is $T$-invariant. For $v \in W_i$, we have $T(U(v)) = U(T(v)) = \lambda_i U(v)$ so $U(v) \in W_i$ and $W_i$ is $U-$invariant. Thus, by Proposition 4.2.5, $U_{W_i}$ is Hermitian and so there exists an orthonormal basis $\beta_i$ for $W_i$ consisting of eigenvectors of $U$. Since $T$ is Hermitian, $\beta = \beta_i \cup \beta_2 \cup ... \cup \beta_k$ is an orthonormal basis for $V$ consisting of eigenvectors of both $U$ and $T$.
2. Let $V$ be finite dimensional complex inner product space. Assume $T, U : V \to V$ are normal. Note ($\Rightarrow$) is identical to the part one.
   Conversely, suppose $TU = UT$, since $T$ is normal, $T$ is unitarily diagonalizable and $V = \bigoplus W_i$ where $W_i = E_{\lambda_i}$ and $\lambda_i \in \sigma(T)$. Moreover, note $W_i$ is $U-$invariant just like in part one. We must show $W_i$ is $U^*-$invariant. Since $U$ is normal, there exists $f(x) \in \mathbb{C}[x]$ such that $U^* = f(U)$ by Proposition 4.4.11. Thus $U^*T = TU^*$ and so $W_i$ is $U^*$-invariant. Hence, $U_{W_i}$ is normal. Let $\beta_i$ be an orthonormal basis for $W_i$ consisting eigenvectors of $U$, then $\beta = \bigcup \beta_i$ would be the desired basis.

♡

**Proposition 4.4.14.** *Let $V$ be finite dimensional complex inner product space. Let $T$ be linear, then $T$ is Hermitian iff $T$ is normal and $\sigma(T) \subseteq \mathbb{R}$.*

*Proof.* ($\Rightarrow$) Done.

($\Leftarrow$) $T = \sum_{i=1}^{k} \lambda_i T_i$, $\lambda_i \in \mathbb{R}$. Then $T^* = \sum_{i=1}^{k} \overline{\lambda_i} T_i^* = \sum_{i=1}^{k} \lambda_i T_i = T$. ♡

**Proposition 4.4.15.** *Let $V$ be finite dimensional $\mathbb{C}$ inner product space. Let $T : V \to V$ be linear. Then $T$ is unitary iff $T$ is normal and $|\lambda| = 1$ for all $\lambda \in \sigma(T)$.*

*Proof.* ($\Rightarrow$) Done.

($\Leftarrow$) Suppose $T$ is normal and $\forall \lambda \in \sigma(T), |\lambda| = 1$. Say that $T = \sum_{i=1}^{k} \lambda_i T_i$, then $T^* = \sum_{i=1}^{k} \overline{\lambda_i} T_i$ then $T \circ T^* = \sum_{i=1}^{k} \lambda_i \overline{\lambda_i} T_i^2 = I$. ♡

## 4.5 Singular Value Decomposition

**Example 4.5.1.** [May be in Final] Let $T : V \to W$ be linear with $V, W$ be finite dimensional inner product space over $F$. A function $T^* : W \to V$ is an adjoint of $T$ if $\langle T(x), y\rangle = \langle x, T^*(y)\rangle$ for all $x \in V$ and $y \in W$. Show that:

1. $T^*$ exists and is unique and linear
2. $\beta, \gamma$ are orthonormal basis for $V$ and $W$, respectively. Then, $[T^*]_\gamma^\beta = ([T]_\beta^\gamma)^*$

*Solution.* We first show $T^*$ exists. Let $y \in W$ be given, we note $y$ induces an linear functional of $V$, namely $\mu_y(x) = \langle T(x), y\rangle_W$. Indeed, we have $\mu_y : V \to F$ and $\mu_y(ax_1 + x_2) = \langle T(ax_1 + x_2), y\rangle_W = \langle aT(x_1) + T(x_2), y\rangle_W = a\mu_y(x_1) + \mu_y(x_2)$. Thus, by Riesz's representation theorem 3.4.3, there exists unique $x \in V$ so that $\mu_y(v) = \langle T(v), y\rangle_W = \langle v, x\rangle_V$ for all $v \in V$. Let $T^* : W \to V$ be the mapping such that maps $y \in W$ into $x \in V$ where we described above. By the uniqueness of Riesz's representation theorem, this $T^*$ is well-defined.

Next, we show it is linear. Let $y_1, y_2 \in W$, suppose $x_1, x_2$ are the corresponding values after the mapping $T^*$, so $T^*(y_i) = x_i$. Then let $x_3 = T^*(ay_1 + y_2)$, we have $\langle T(v), ay_1 + y_2\rangle_W = \langle v, x_3\rangle$ for all $v \in V$. However, we note

$$\begin{aligned}\langle T(v), ay_1 + y_2\rangle_W &= \overline{a}\langle T(v), y_1\rangle_W + \langle T(v), y_2\rangle_W \\ &= \langle v, ax_1\rangle_V + \langle v, x_2\rangle_V = \langle v, ax_1 + x_2\rangle_V\end{aligned}$$

Thus by uniqueness we must have $ax_1 + x_2 = x_3$ and so $T^*(ay_1 + y_2) = aT^*(y_1) + T^*(y_2)$ as desired. To show uniqueness of $T^*$, let $T^*, T_1^*$ be so that $\langle T(x), y\rangle = \langle x, T^*(y)\rangle = \langle x, T_1^*(x)\rangle$ for all $x \in V, y \in W$, then we have $\langle x, (T^* - T_1^*)(y)\rangle = 0$. In particular, for arbitrary $y \in W$, we have all $x \in V$ and $\langle x, (T^* - T_1^*)(y)\rangle = 0$. Thus $(T^* - T_1^*)(y) = 0$ and so $T^*(y) = T_1^*(y)$ for arbitrary $y \in W$ and so $T^* = T_1^*$. Thus $T^*$ is unique.

Next, we show $[T^*]_\gamma^\beta = ([T]_\beta^\gamma)^*$. Let $\beta, \gamma$ be orthonormal basis for $V$ and $W$, respectively. Let $\beta = (v_1, ..., v_n)$ and $\gamma = (w_1, ..., w_m)$. Then

$$[T]_\beta^\gamma = \big[[T(v_1)]_\gamma, [T(v_2)]_\gamma, ..., [T(v_n)]_\gamma\big] \Rightarrow ([T]_\beta^\gamma)^* = \begin{bmatrix} [T(v_1)]_\gamma^* \\ [T(v_2)]_\gamma^* \\ \vdots \\ [T(v_n)]_\gamma^* \end{bmatrix}$$

In particular, since $\gamma$ is orthonormal, we have $[T(v_i)]_\gamma = \begin{bmatrix} \langle T(v_i), w_1 \rangle \\ \langle T(v_i), w_2 \rangle \\ \vdots \\ \langle T(v_i), w_m \rangle \end{bmatrix}$ so that

$$([T]_\beta^\gamma)^* = \begin{bmatrix} \langle T(v_1), w_1 \rangle & \langle T(v_2), w_1 \rangle & \vdots & \langle T(v_n), w_1 \rangle \\ \langle T(v_1), w_2 \rangle & \langle T(v_2), w_2 \rangle & \vdots & \langle T(v_n), w_2 \rangle \\ \langle T(v_1), w_3 \rangle & \langle T(v_2), w_3 \rangle & \vdots & \langle T(v_n), w_3 \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle T(v_1), w_m \rangle & \langle T(v_2), w_m \rangle & \vdots & \langle T(v_n), w_m \rangle \end{bmatrix}^*$$

$$= \begin{bmatrix} \langle v_1, T^*(w_1) \rangle & \langle v_2, T^*(w_1) \rangle & \vdots & \langle v_n, T^*(w_1) \rangle \\ \langle v_1, T^*(w_2) \rangle & \langle v_2, T^*(w_2) \rangle & \vdots & \langle v_n, T^*(w_2) \rangle \\ \langle v_1, T^*(w_3) \rangle & \langle v_2, T^*(w_3) \rangle & \vdots & \langle v_n, T^*(w_3) \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle v_1, T^*(w_m) \rangle & \langle v_2, T^*(w_m) \rangle & \vdots & \langle v_n, T^*(w_m) \rangle \end{bmatrix}^*$$

$$= \begin{bmatrix} \overline{\langle v_1, T^*(w_1) \rangle} & ... & \overline{\langle v_1, T^*(w_m) \rangle} \\ \vdots & \ddots & \vdots \\ \overline{\langle v_n, T^*(w_1) \rangle} & ... & \overline{\langle v_n, T^*(w_m) \rangle} \end{bmatrix}$$

Similarly, we have $[T^*(w_j)]_\beta = \begin{bmatrix} \langle T^*(w_j), v_1 \rangle \\ \vdots \\ \langle T^*(w_j), v_n \rangle \end{bmatrix}$ so that

$$[T^*]_\gamma^\beta = \begin{bmatrix} \langle T^*(w_1), v_1 \rangle & ... & \langle T^*(w_m), v_1 \rangle \\ \vdots & \ddots & \vdots \\ \langle T^*(w_m), v_n \rangle & ... & \langle T^*(w_m), v_n \rangle \end{bmatrix}$$

Thus, since $\overline{\langle v_i, T^*(w_j) \rangle} = \langle T^*(w_j), v_i \rangle$, we have $[T^*]_\gamma^\beta = ([T]_\gamma^\beta)^*$ as desired. ♠

**Example 4.5.2.** Let $V$ be finite dimensional inner product space and $T : V \to V$ be linear. We say $T$ is positive semidifinite if $T$ is Hermitian and $\langle T(x), x \rangle \geq 0$ for all $x \in V$. Show that:

1. $T$ is positive semidifinite iff $T = T^*$ and $\sigma(T) \subseteq [0, \infty)$.
2. $T$ is positive semidifinite iff $T = U^*U$ for some linear $U : V \to V$.
3. $T : V \to W$ be linear, $V, W$ be finite dimensional inner product spaces over $F$. Then $T^*T, TT^*$ are positive semidifinite with $rank(T^*T) = rank(T) = rank(TT^*)$.

*Solution.* 1. Suppose $T$ is positive semidifinite, then $T = T^*$ by definition. Moreover, suppose for a sake of contradiction $\exists \lambda \in \sigma(T)$ such that $\lambda < 0$, then there exists $0 \neq x \in V$ such that $T(x) = \lambda x$. Then $\langle T(x), x \rangle = \lambda \langle x, x \rangle$ where $\lambda < 0$

and $\langle x, x\rangle > 0$ so $T$ is not positive semidifinite, which is a contradiction. Conversely, suppose $T = T^*$ and $\sigma(T) \subseteq [0, \infty)$, we will show $\langle T(x), x\rangle \geq 0$. By spectral theorem, $V = \bigoplus_{\lambda_i \in \sigma(T)} E_{\lambda_i}$ and so $x \in V$ and there exists $x_i \in E_{\lambda_i}$ and $x = \sum x_i$. Hence, we have $\langle T(x), x\rangle = \langle \sum \lambda_i x_i, \sum x_i\rangle = \sum \lambda_i \langle x_i, x_i\rangle \geq 0$ as $\langle x_i, x_i\rangle \geq 0$ and $\lambda_i \geq 0$.

2. Let $T$ be positive semi-definite, thus $T$ is Hermitian. Let $\beta = (v_1, ..., v_n)$ be orthonormal basis consists of eigenvectors, then we have $[T]_\beta = D$ where $D = diag(\lambda_1, ..., \lambda_n)$ and note all of the eigenvalues are non-negative. Let $U = L_{\sqrt{D}}$ where $\sqrt{D} = diag(\sqrt{\lambda_1}, ..., \sqrt{\lambda_n})$, then for $x = \sum_{i=1}^n a_i v_i$, we have $U(x) = \sum_{i=1}^n a_i \sqrt{\lambda_i} v_i$. Next, note

$$U^*U(x) = \sum_{i=1}^n a_i \lambda_i v_i = \sum_{i=1}^n a_i T(v_i) = T(\sum_{i=1}^n a_i v_i) = T(x)$$

Conversely, suppose $T = U^*U$ for some linear operator $U : V \to V$. Then $T^* = (U^*U)^* = U^*U = T$, thus $T$ is Hermitian. Next, note $\langle T(x), x\rangle = \langle U^*U(x), x\rangle = \langle U(x), U(x)\rangle \geq 0$ for all $x \in V$, so we are done.

3. To see $T^*T$ is positive semidifinite, we note $U = T^*T$ and thus $U$ must be semidifinite with similar reasoning of our part two and Example 4.5.1. Similarly we can show $TT^*$ is positive semidifinite. To see $rank(T^*T) = rank(T)$, we show $Null(T^*T) = Null(T)$. Clearly $Null(T) \subseteq Null(T^*T)$. Let $x \in Null(T^*T)$, then $\langle T^*T(x), x\rangle = \langle 0, x\rangle = 0 = \langle T(x), T(x)\rangle$ so $T(x) = 0$ as desired. Thus we are done.

♠

**Definition 4.5.3.** In the below theorem, the $\sigma_i$'s are called the ***singular values*** of $T$. If $r$ is less than both $m, n$ then we also call $\sigma_{r+1} = ... = \sigma_k = 0$ with $k = \min\{m, n\}$, singular values of $T$.

**Theorem 4.5.4** (Singular Value Decomposition Theorem)**.** *Let $V$ be finite dimensional inner product spaces over $F$. Let $T : V \to W$ be linear with $rank(T) = r$. There exists orthonormal bases $\{v_1, ..., v_n\}$ and $\{u_1, ..., u_m\}$ for $V$ and $W$ respectively, and real scalars $\sigma_1 \geq \sigma_2 \geq ... \geq \sigma_r > 0$ such that $T(v_i) = \sigma_i u_i$ for $i \leq r$ and $T(v_i) = 0$ for $i > r$. Conversely, if the above conclusion holds (the existence of bases and $\sigma$'s) then each $v_i$ is a $\sigma_i^2$ eigenvectors of $T^*T$. In particular, the $\sigma_i$'s are uniquely determined.*

*Proof.* Let $V$, $W$ and $T$ be in the theorem. Consider the positive semidifinite operator $T^*T$ on $V$. Note $rank(T^*T) = r$. Sicne $T^*T$ is Hermitian, there exists an orthonormal basis $\{v_1, ..., v_n\}$ for $V$ consisting of eigenvectors of $T^*T$. Say $T^*T(v_i) = \lambda_i v_i$ where $\lambda_i \geq 0$. Let $\sigma_i = \sqrt{\lambda_i}$. We may assume $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_r > 0$ and $\lambda_{r+1} = ... = \lambda_n = 0$. Hence, we have $\sigma_1 \geq \sigma_2 \geq ... \geq \sigma_r > 0$ and $\sigma_i = 0$ for $i > r$.

For $i \leq r$, let $u_i = \frac{1}{\sigma_i} T(v_i)$. Then $\langle u_i, u_j\rangle = \langle \frac{1}{\sigma_i} T(v_i), \frac{1}{\sigma_j} T(v_j)\rangle = \frac{1}{\sigma_i \sigma_j} \langle T(v_i), T(v_j)\rangle = \frac{1}{\sigma_i \sigma_j} \langle T^*T(v_i), v_j\rangle = \frac{1}{\sigma_i \sigma_j} \langle \lambda_i v_i, v_j\rangle = \frac{\lambda_i}{\sigma_i \sigma_j} \langle v_i, v_j\rangle = \delta_{ij}$ as $i \neq j$ imply $\langle v_i, v_j\rangle = 0$ and note $i = j$ then $\sigma_i \sigma_i = \lambda_i$. Thus $\{u_1, u_2, ..., u_r\}$ is an orthonormal set. By the

Gram-Schmidt procedure, we may extend this to an orthonormal basis $\{u_1, ..., u_m\}$ for $W$.

For $i \le r$, we have $T(v_i) = \sigma_i u_i$ as we defined this way. For $i > r$, we have $T^*T(v_i) = 0$ as $T^*T(v_i) = \lambda_i v_i = 0$. This imply $\langle T^*T(v_i), v_i\rangle = 0 \Rightarrow \langle T(v_i), T(v_i)\rangle = 0 \Rightarrow T(v_i) = 0$.

Next, we show $\sigma_i$ is uniquely determined by eigenvalues of $T^*T$. Suppose we have such $u_i, v_i$ and $\sigma_i$. Then, $T^*(u_i) \in V$ and so

$$\begin{aligned} T^*(u_i) &= \sum_j \langle T^*(u_i), v_j\rangle_V v_j \\ &= \sum_j \langle u_i, T(v_j)\rangle_W v_j \\ &= \begin{cases} \sigma_i v_i & \text{if } i \le r \\ 0 & i > r \end{cases} \end{aligned}$$

Then, if $i \le r$, we have

$$\begin{aligned} T^*T(v_i) &= T^*(\sigma_i u_i) = \sigma T^*(u_i) \\ &= \sigma_i^2 v_i \end{aligned}$$

and for $i > r$, we have

$$T^*T(v_i) = T^*(0) = 0 = 0^2 v_i$$

♡

**Example 4.5.5.** Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be $T(x, y) = (x, x + y, x - y)$. Let $\sigma$ be standard basis for $\mathbb{R}^2$ and $\beta$ be standard basis for $\mathbb{R}^3$, we have

$$[T]_\sigma^\beta = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & -1 \end{bmatrix} = A$$

Then, $[T^*T]_\sigma^\beta = A^*A = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}$ and so $\sigma(T^*T) = \{3, 2\}$ (and we have $v_1 := e_1$ is $3-$eigenvector and $v_2 := e_2$ is $2-$eigenvector). Therefore, $\sigma$ is a orthonormal basis for $\mathbb{R}^2$ such that $T^*T(v_1) = 3v_1$ and $T^*T(v_2) = 2v_2$. Thus, we have $\sigma_1 = \sqrt{3}$ and $\sigma_2 = \sqrt{2}$.

Let $u_1 = \frac{1}{\sqrt{3}}T(v_1) = \frac{1}{\sqrt{3}}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ and $u_2 = \frac{1}{\sqrt{2}}T(v_2) = \frac{1}{\sqrt{2}}\begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}$, and by the proof, $\{u_1, u_2\}$ is orthonormal. Moreover, $u_3 = \frac{1}{\sqrt{1.5}}\begin{bmatrix} 1 \\ -\frac{1}{2} \\ -\frac{1}{2} \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{2}}{\sqrt{3}} \\ -\frac{\sqrt{2}}{2\sqrt{3}} \\ -\frac{\sqrt{2}}{2\sqrt{3}} \end{bmatrix}$ is orthonormal to

$u_1, u_2$ and hence we have our desired orthonormal basis of $\mathbb{R}^3$, say $\delta = \{u_1, u_2, u_3\}$. In particular, we have $[T]_\sigma^\beta = [I]_\delta^\beta [T]_\sigma^\delta ([I]_\sigma^\sigma)^* = UDV^*$ and $U = [u_1, u_2, u_3]$, $D = \begin{bmatrix} \sqrt{3} & 0 \\ 0 & \sqrt{2} \\ 0 & 0 \end{bmatrix}$ and $V = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

**Definition 4.5.6.** Let $A \in M_{m\times n}(F)$, the singular value of $A$ are the singular value of $L_A : F^n \to F^m$.

**Theorem 4.5.7** (Singular Decomposition for Matrix)**.** *Let $A \in M_{m\times n}(F)$ and $rank(A)$ to be $r$. Let the singular values be $\sigma_1 \geq \sigma_2 \geq ... \geq \sigma_r > 0$. There exists unitary $U \in M_m(F)$, unitary $V \in M_n(F)$ such that $A = UDV^*$ where* $d_{ij} = \begin{cases} \sigma_i & i = j \\ 0 & otherwise \end{cases}$

**Proposition 4.5.8.** *Let $A \in M_n(F)$ be positive definite. Then if $A = UDV^*$ is the singular value decomposition of $A$ then $U = V$.*

*Proof.* Suppose $A$ is positive definite, then $A$ is Hermitian and $y^*Ax > 0$ for all non-zero $y, x \in F^n$. Suppose $A = UDV^*$ is the singular value decomposition of $A$.

Next, suppose $\lambda \in \sigma(A)$, there exists $x \in F^n$ so that $Ax = \lambda x$, in particular, we have $x^*Ax = \lambda x^*x > 0$. Note $x^*x$ is always greater than zero so we must have $\lambda > 0$. Hence, all the eigenvalues of $A$ must be positive (and thus $A$ is invertible as $det(A) = det(U\Lambda U^{-1})$ where $\Lambda$ is the diagonal matrix with $A's$ eigenvalues). Therefore, $A$ has $n$ positive eigenvalues (with repetition) and note every eigenvalue's square of $A$ is an eigenvalue of $A^*A$. Indeed, let $\lambda \in \sigma(A)$, we have $\lambda$ is real and there exists $0 \neq x \in F^n$ so $Ax = \lambda x$, and $A^*Ax = A^*(\lambda x) = \lambda A^*(x) = \lambda^2(x)$ as every $\lambda-$eigenvector of $L_A$ is a $\overline{\lambda}-$eigenvector of $L_A^*$ imply every $\lambda-$eigenvector of $A$ is a $\overline{\lambda}-$eigenvector of $A^*$ and $\lambda = \overline{\lambda}$. Thus, $\lambda \in \sigma(A) \Rightarrow \lambda^2 \in \sigma(A^*A)$.

In particular, let $\lambda_1 > ... > \lambda_k > 0$ be all the distinct eigenvalues of $A$. Let $\lambda_i$ be an eigenvalue of $A$, denote the eigenspace of $\lambda_i$ of $A$ to be $W_i$. Then, we have $F^n = W_1 \oplus W_2 \oplus ... \oplus W_k$. In particular, we have $\beta_i$, an orthonormal basis for $W_i$, contains only $\lambda_i$ eigenvectors of $A$. Moreover, let $x, y \in \beta_i$, we have $A^*Ax = \lambda_i^2 x \neq 0$, $A^*Ay = \lambda_i^2 y \neq 0$, and $\langle A^*Ax, A^*Ay\rangle = \lambda_i^2\langle x, \lambda_i^2 y\rangle = \lambda_i^4\langle x, y\rangle = 0$(as $\lambda_i \in \mathbb{R}$). Thus, $A^*Ax$ and $A^*Ay$ are orthogonal and hence linear independent. Moreover, let $x \in W_i$ then $A^*Ax = \lambda_i^2 x \in W_i$ as $A(\lambda^2 x) = \lambda_i^2(\lambda_i x) = \lambda_i(\lambda_i^2 x)$, thus $W_i$ is $A^*A$ invariant.

Therefore, let $\beta_i = \{v_{i,1}, v_{i,2}, ..., v_{i,r_i}\}$, $\{A^*A(v_{i,1}), ..., A^*A(v_{i,r_i})\}$ is lienar independent and thus $W_i = span\{A^*A(v_{i,1}), ..., A^*A(v_{i,r_i})\}$. Thus every $\lambda_i$ eigenspace of $A$ is an $\lambda_i^2$ eigenspace of $A^*A$ and so $\lambda_1^2, ..., \lambda_k^2$ are all the eigenvalues of $A^*A$ as $F^n = W_1 \oplus ... \oplus W_k$ (indeed, suppose $\mu \notin \{\lambda_1^2, ..., \lambda_k^2\}$ is an eigenvalue of $A^*A$ then there exists $x = \sum_{i=1}^k t_i \in F^n$ where $t_i \in W_i$ such that $A^*Ax = \mu x$. In particular, we have $A^*A(x) = \sum_{i=1}^k \lambda_i^2 v_i - \sum_{i=1}^k \mu v_i \Rightarrow \sum_{i=1}^k(\lambda_i^2 - \mu)v_i = 0$. Since $\mu \notin \{\lambda_1^2, ..., \lambda_k^2\}$, $\lambda_i^2 - \mu \neq 0$ and so contradicting the fact that $v_1, ..., v_k$ are linear independent as $v_1, ..., v_k$ are orthogonal). Note this imply $F^n = R_1 \oplus R_2 \oplus ... \oplus R_k$ where $R_i$ is the $\lambda_i^2-$eigenspace of $A^*A$, which is equal to the $\lambda$-eigenspace of $A$. This imply every $\lambda_i^2$ eigenvector of $A^*A$ is a $\lambda$-eigenvector of $A$.

Thus, all the eigenvalues of $A^*A$ are $\lambda_1^2 > \lambda_2^2 > ... > \lambda_k^2 > 0$ and hence the singular values of $A$ are $\sigma_1 = \lambda_1, ..., \sigma_{r_1} = \lambda_1, \sigma_{r_1+1} = \lambda_2, ..., \sigma_{r_1+r_2} = \lambda_2, ..., \sigma_{\sum_{j=1}^{k-1} r_j+1} = \lambda_k, ..., \sigma_n = \lambda_k$ where $r_i = dim(W_i)$. Next, suppose $V = \begin{bmatrix} v_1 & ... & v_n \end{bmatrix}$ where $\beta = \{v_1, ..., v_n\}$ is the orthonormal basis of $A^*A$ consisting of eigenvectors of $A^*A$. We have $u_i = \frac{1}{\sigma_i} A v_i$, where $\sigma_i = \lambda_j$, so that $u_i = \frac{1}{\lambda_j}\lambda_j v_i = v_i$ as $v_i$ is an $\sigma_i^2-$eigenvector of $A^*A$ where $\sigma_i = v_j$ so that $v_i$ is an $v_j$-eigenvector of $A$. Hence, we have $u_i = v_i$ for all $1 \leq i \leq n$. Thus, $U = V$ as desired. ♡

**Definition 4.5.9.** Let $A \in M_n(F)$. The ***polar decomposition*** of $A$ is the decomposition of the form $A = UP$ where $U$ is unitary and $P$ is positive semidifinite Hermitian matrix.

**Remark 4.5.10.** The below proposition, Proposition 4.5.11.1, guarantee the existence of polar decomposition.

**Proposition 4.5.11.** *Let $A \in M_n(F)$,*

1. *There exists a unitary $U \in M_n(F)$ and a positive semidifinite Hermitian $P \in M_n(F)$ such that $A = UP$,*
2. *Let $A = UP$ as in part one, then $A$ is normal if and only if $UP^2 = P^2U$,*
3. *Let $A = UP$ as in part one, then $A$ is normal if and only if $UP = PU$.*

*Proof.* Let $A = U_1 D V_1^*$ be the singular value decomposition of $A$.

1. Note $A^*A = V_1 D^* U_1^* U_1 D V_1^* = V_1 D^2 V_1^*$ as $D$ contain the singular values of $A$ and it is always non-negative real numbers. Next, note $A = U_1 D V_1^* = U_1 V_1^* V_1 D V_1^*$, let $U = U_1 V_1^*$ and $P = V_1 D V_1^*$. We first remark $P^2 = A^*A$ as we will use it in part two. Clearly $U$ is unitary, and $P$ is positive semidifinite as $D$ contains non-negative entries on the diagonal. $P$ is Hermitian as $P^* = V_1 D^* V^* = V_1 D V^* = P$.
2. Suppose $A = UP$ as in part one. Note $P^2 = P^*P = PP^* = A^*A$. Thus, $A^*A = P^*U^*UP = P^*P = P^2$. Next, note $AA^* = UPP^*U^* = UP^2U^*$.
   Suppose $A$ is normal. Since $A^*A = AA^*$, we have $P^2 = UP^2U^*$ and so $P^2U = UP^2$ as desired.
   Next, suppose $P^2U = UP^2$, then $AA^* = UP^2U^* = UU^*P^2 = P^2 = A^*A$.
3. Let $A = UP$ as in part one, then $P$ is Hermitian. Let $P = \sum_{i=1}^k \lambda_i P_i$ be the spectral decomposition of $P$. Note $P^2 = \sum_{i=1}^k \lambda_i^2 P_i$ by the spectral decomposition and note there exists polynomial $h(x)$ so that $h(\lambda_i^2) = \lambda_i$ by Lagrange interpolation , we have $h(P^2) = \sum_{i=1}^k h(\lambda_i^2)P_i = \sum_{i=1}^k \lambda_i P_i = P$. Thus, since $A$ is normal, we have $UP^2 = P^2U$ and in particular, we note $Uh(P^2) = h(P^2)U$ as every term in the polynomial is some power of $P^2$ and thus $U$ commutes with them. In particular, this imply $UP = PH$ as $h(P^2) = P$.
   Conversely, suppose $UP = PU$, then $UP^2 = UPP = PUP = P(PU) = P^2U$ and so $A$ is normal by part two.

♡

**Proposition 4.5.12.** *Let $A \in M_{m\times n}(\mathbb{C})$ with positive singular values $\sigma_1 \geq \sigma_2 \geq ... \geq \sigma_r > 0$ where $r = rank(A)$. Then*

$$\sigma_1 = \max\{|\langle Ax, y\rangle| : x \in \mathbb{C}^n, y \in \mathbb{C}^m, \|x\| = \|y\| = 1\}$$

*Solution.* Let $A = UDV^*$, note both $U^*, V^*$ are unitary matrix, and thus $\|V^*x\| = \|x\|$ and $\|U^*y\| = \|y\|$. Moreover, since they are invertible (hence bijection), for all $0 \neq x \in \mathbb{C}^n$, we can find unique $v$ so that $0 \neq v = V^*x$, for all $0 \neq y \in \mathbb{C}^m$, we can find unique $u$ so that $0 \neq u = U^*y$.

Next, note $\langle Ax, y\rangle = \langle DV^*x, U^*y\rangle = \langle Dv, u\rangle$ and note if $\|x\| = \|y\| = 1$, we must have $\|u\| = \|v\| = 1$. Thus, it suffice to find the maximum of $\{\langle Dv, u\rangle : v \in \mathbb{C}^n, u \in \mathbb{C}^m, \|u\| = \|v\| = 1\}$. Indeed, let $a = \max\{\langle Ax, y\rangle : x \in \mathbb{C}^n, y \in \mathbb{C}^m, \|x\| = \|y\| = 1\}$ and $b = \max\{\langle Dv, u\rangle : v \in \mathbb{C}^n, u \in \mathbb{C}^m, \|u\| = \|v\| = 1\}$. Note $a = \langle Ax_a, y_a\rangle = \langle D(V^*x_a), U^*y_a\rangle \leq b$, and $b = \langle Dv_b, u_b\rangle = \langle D(V^*x_b), U^*y_b\rangle \leq a$. Thus $a = b$.

To find the maximum element[1] of $\mathcal{S} = \{\langle Dv, u\rangle : v \in \mathbb{C}^n, u \in \mathbb{C}^m, \|u\| = \|v\| = 1\}$, we note $\|v\| = 1$ imply each component of $v$ must have norm less than or equal one. Let $v = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}$, $u = \begin{bmatrix} u_1 \\ \vdots \\ u_m \end{bmatrix}$, we have $Dv = \begin{bmatrix} \sigma_1 v_1 \\ \vdots \\ \sigma_r v_r \\ 0 \\ \vdots \\ 0 \end{bmatrix}$, thus we have $\langle Dv, u\rangle = \sum_{i=1}^r \sigma_i \overline{u_i} v_i$.

Next, we note $\langle Dv, u\rangle$ is maximal when $v = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$ and $u = \begin{bmatrix} 1 \\ \vdots \\ 0 \end{bmatrix}$ as $\|u\|$ and $\|v\|$ are equal one so that $\langle Dv, u\rangle$ is a weighted average with $v_i, u_i$ weighted the most. Hence, the maximal is $\sigma_1$ as desired. ♠

## 4.6 Quotient Space

**Remark 4.6.1.** Recall the additional materials provided before the quotient space

**Example 4.6.2.** Let $V = M_2(\mathbb{R})$ and $W = \{A \in V : A = A^T\}$. We have $\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} + W = \overline{\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}}$.

**Proposition 4.6.3.** *Let $V$ be finite dimensional vector space over $F$. Let $W$ be a subspace of $V$, say $\{v_1, ..., v_m\}$ is a basis for $W$. Then, extend this basis to $\{v_1, ..., v_n\}$ to be a basis for $V$. Then $\{\overline{v_{m+1}}, ..., \overline{v_n}\}$ is a basis for $V/W$. In particular, we have $dim(V/W) = dim(V) - dim(W)$.*

*Proof.* Let $\overline{v} \in V/W$ be arbitrary. We have $v = \sum_{i=1}^n a_i v_n$ and so $\overline{v} = \sum_{i=m+1}^n a_i(v_i + W)$ as $v_i \in W$ for $1 \leq i \leq m$.

[1]This maximal must exists as the dot product is a continuous mapping and the unit circle is compact

Next, we need to show $\{\overline{v_{m+1}}, ..., \overline{v_n}\}$ is linear independent.

Consider $\sum_{i=m+1}^{n} a_i \overline{v_i} = 0$, then $\sum_{i=m+1}^{n} a_i v_i \in W$. Thus $\sum_{i=m+1}^{n} a_i v_i = \sum_{i=1}^{m} b_i v_i$. Hence, as $\{v_1, ..., v_n\}$ is a basis for $V$, we have $a_{m+1} = ... = a_n = b_1 = ... = b_m = 0$. ♡

Note in general, we have $(a, b) \neq (b, a) \in \mathbb{C}^2$. However, consider $S = \{(a, b) - (b, a) : a, b \in \mathbb{C}\}$ and $W = span(S)$. Then, the order does not matter in $V/W$. In particular, we have $\overline{(a, b) - (b, a)} = \overline{0}$ and so $\overline{(a, b)} = \overline{(b, a)}$.

# Chapter 5

# Tensor Product

## 5.1 Intro

**Remark 5.1.1.** The motivation of this chapter is to extend a vector space $V$ over $F$ into a ring with $(T(V), +, \otimes)$.

Thus, we want, for $x, y, z \in V$ and $\alpha \in F$,

1. $x \otimes (y + z) = x \otimes y + x \otimes z$ and $(y + z) \otimes x = y \otimes x + z \otimes x$
2. $\alpha(x \otimes y) = (\alpha x) \otimes y = x \otimes (\alpha y)$

**Definition 5.1.2.** Let $X$ be a set of symbols, we define the ***free vector space*** on $X$ by $V = Free(X) = \{\sum_{i=1}^{n} a_i x_i : a_i \in F, x_i \in X, n \in \mathbb{N}\}$ with addition defined to be

$$\sum \alpha_i x_i + \sum \beta_i x_i = \sum (\alpha_i + \beta_i) x_i$$

and $\alpha(\sum \alpha_i x_i) = \sum(\alpha \alpha_i x_i)$.

**Remark 5.1.3.** By construction, $X$ is a basis for $Free(X)$.

**Definition 5.1.4.** Let $V, W$ be fintie dimensional vector space over $F$, and let $X = V \times W$ to be a set of symbols. Let $S$ be the set of vectors in $Free(X)$ of the form

$$\begin{cases} (x + y, z) - (x, z) - (y, z) \\ (z, x + y) - (z, x) - (z - y) \\ \alpha(x, y) - (\alpha x, y) \\ \alpha(x, y) - (x, \alpha y) \end{cases}$$

Then, we define the ***tensor product*** of $V$ and $W$ to be $V \otimes W = Free(X)/span(S)$.

**Definition 5.1.5.** We define $v \otimes w := \overline{v, w} = (v, w) + span(S) = \overline{(v, w)}$ where $v \in V$ and $w \in W$ and call them ***pure tensor***.

**Remark 5.1.6.** First, note $(v + w) \otimes z - v \otimes z - w \otimes z = 0 \otimes 0 = 0$, and so $(v + w) \otimes z = v \otimes z + w \otimes z$. Also, $\alpha(v \otimes w) = (\alpha v) \otimes w = v \otimes (\alpha w)$.

A typical element of $V \otimes W$ looks like

$$\alpha_1(v_1 \otimes w_1) + ... + \alpha_n(v_n \otimes w_n)$$

**Example 5.1.7.** Consider $\mathbb{C}^2 \otimes \mathbb{C}^3$(or $\mathbb{C}^2 \otimes_{\mathbb{C}} \mathbb{C}^3$ where $\otimes_{\mathbb{C}}$ indicate the underlying field), consider

$$\begin{bmatrix} 1 \\ 2 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

Let the standard basis for $\mathbb{C}^2$ be $\sigma_2 = \{a_1, a_2\}$ and the standard basis for $\mathbb{C}^3$ be $\sigma_3 = \{b_1, b_2, b_3\}$. Then, we have

$$\begin{aligned}
\begin{bmatrix} 1 \\ 2 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} &= (a_1 + 2a_2) \otimes (b_1 + 2b_2 + 3b_3) \\
&= a_1 \otimes (b_1 + 2b_2 + 3b_3) + 2(a_2 \otimes (b_1 + 2b_2 + 3b_3)) \\
&= a_1 \otimes b_1 + 2(a_1 \otimes b_2) + 3(a_1 \otimes b_3) \\
&\quad + 2(a_2 \otimes b_1) + 4(a_2 \otimes b_2) + 6(a_2 \otimes b_3)
\end{aligned}$$

**Example 5.1.8.** Note $2 \otimes 2 = 2(1 \otimes 2) = 4(1 \otimes 1)$.

**Proposition 5.1.9.** *Let $V, W$ be finite dimensional $F$ vector space. Let $\{v_1, ..., v_n\}$ be a basis of $V$, $\{w_1, ..., w_m\}$ be a basis of $W$. A basis for $V \otimes_F W$ is*

$$\{v_i \otimes w_j : 1 \leq i \leq n, 1 \leq j \leq m\}$$

*In particular, we have $dim(V \otimes_F W) = nm = dim(V) \cdot dim(W)$.*

*Proof.* Let $x \in A$ and $y \in B$, then $x \otimes y \in A \otimes B$. In particular, note every element of $A \otimes B$ is the sum of some pure tensors. Thus, $x = \sum_{i=1}^n t_i a_i$ and $y = \sum_{j=1}^m l_j b_j$, we have

$$\begin{aligned}
x \otimes y &= (\sum_{i=1}^n t_i a_i) \otimes (\sum_{j=1}^m l_j b_j) \\
&= \sum_{i=1}^n \left( t_i a_i \otimes (\sum_{j=1}^m l_j b_j) \right) \\
&= \sum_{i=1}^n \left( \sum_{j=1}^m t_i a_i \otimes l_j b_j \right) \\
&= \sum_{i=1}^n \sum_{j=1}^m t_i l_j (a_i \otimes b_j) \in span(W)
\end{aligned}$$

Hence, $A \otimes B \subseteq span(W)$ as every pure tensor is in the span of $W$. Next, we show $W$ is linear independent.

Suppose $\sum_{i,j} d_{ij}(a_i \otimes b_j) = 0$. Let $f_k \in A^*$ be $f_k(a_k) = 1$ and $f_k(a_l) = 0$ for $1 \leq l \leq n$ and $k \neq l$. Let $F_k : A \times B \to B$ be $F_k(a, b) = f_k(a)b$, then $F_k$ is bilinear. Hence, there exists a linear mapping $T_k$ (by universal property) from $A \otimes B \to B$ such that $T_k(a \otimes b) = f_k(a)b$. Thus, for all $1 \le k \le n$,

$$\begin{aligned} 0 &= T_k(\sum_{i,j} d_{ij}(a_i \otimes b_j)) \\ &= \sum_{i,j} d_{ij} f_k(a_i) b_j \\ &= \sum_{j=1}^{n} d_{kj} b_j \end{aligned}$$

and since $\{b_1, ..., b_m\}$ is linear independent, we have $d_{kj} = 0$ for all $1 \le j \le m$. Since this hold for all $1 \le k \le n$, we have $d_{ij} = 0$ for all $1 \le i \le n$ and $1 \le j \le m$. Therefore, $\{(a_i \otimes b_j) : 1 \le i \le n, 1 \le j \le m\}$ is a basis of $A \otimes B$. ♡

**Lemma 5.1.10.** *Suppose $V$ and $W$ are vector spaces over $F$ and $T : V \to W$ is linear. Let $S$ be a subspace of $V$. Then there exists a linear transformation $\mathfrak{T} : V/S \to W$ such that $\mathfrak{T}(x + S) = T(x)$ for all $x \in V$ if and only if $T(s) = 0$ for all $s \in S$.*

*Moreover, if $\mathfrak{T}$ exists, it is unique, and every element in $L(V/S, W)$ arises in this way from a unique $T$.*

*Proof.* Suppose $\mathfrak{T}$ exists. Then for all $s \in S$, we have $T(s) = \mathfrak{T}(s + S) = \mathfrak{T}(0) = 0$.

Now, suppose $T(s) = 0$ for all $s \in S$. We must show that $\mathfrak{T}(x + S) = T(x)$ makes $\mathfrak{T}$ well-defined. In other words, we must show that $v, v' \in V$ are such that $v + S = v' + S$, then $\mathfrak{T}(v + S) = \mathfrak{T}(v' + S)$. Now, note $v + S = v' + S$ then $v - v' \in S$ and so $\mathfrak{T}((v - v') + S) = \mathfrak{T}(v + S) - \mathfrak{T}(v' + S) = 0$. Thus $\mathfrak{T}(v + S) = \mathfrak{T}(v' + S)$ and so $\mathfrak{T}$ is well-defined.

Next, we only need to show $\mathfrak{T}$ is lienar. Indeed, consider $x + S, y + S \in V/S$ and $a \in F$, we have $\mathfrak{T}(a(x+S)+(y+S)) = \mathfrak{T}((ax+y)+S) = T(ax+y) = aT(x)+T(y) = a\mathfrak{T}(x + S) + \mathfrak{T}(y + S)$. Hence $\mathfrak{T}$ is linear.

The final remarks are clear, since the statement $\mathfrak{T}(x+S) = T(x)$ uniquely determines either of $T$, $\mathfrak{T}$ from the other. Indeed, let $T, U$ be lienar and they induce the same linear transformation $\mathfrak{T}$ on $V/S \to W$, we see that $\mathfrak{T}(x + S) = T(x) = U(x)$ for all $x \in V$ and so $T = U$. ♡

**Theorem 5.1.11** (Universal Property of Tensor Product)**.** *Let $V, W, Z$ be $F$ vector spaces. Let $\phi : V \times W \to Z$ be bilinear, i.e. $\phi(\alpha x + y, z) = \alpha\phi(x, z) + \phi(y, z)$ and $\phi(x, \alpha z_1 + z_2) = \alpha\phi(x, z_1) + \alpha(x, z_2)$. Then, there exists a unique linear transformation $T : V \otimes W \to Z$ such that $T(v \otimes w) = \phi(v, w)$. Moreover, all linear transformation from $V \otimes W \to Z$ can be constructed in this way.*

*Proof.* Let $X = V \times W$. We first show $\phi$ induce a unique linear mapping $\Phi$ from $Free(X) \to Z$. Note $V \times W$ is a basis for $Free(X)$, thus, let $x = \sum_{i=1}^{n} a_i(v_i, w_i) \in Free(X)$ where $(v_i, w_i) \in X$. Define $\Phi(x) = \sum_{i=1}^{n} a_i\phi(v_i, w_i) \in Z$, first note this $\Phi$ is indeed unique (this $\Phi$ only depends on how $\phi$ maps all the elements of $V \times W$ to $Z$, and thus every $\phi$ uniquely induce the $\Phi$), then we will show it is linear. Let $x = \sum_{i=1}^n a_i(v_i, w_i)$ and $y = \sum_{i=1}^n b_i(v_i, w_i)$, and $k \in F$, we have $\Phi(kx+y) = \Phi(\sum_{i=1}^n (ka_i+b_i)(v_i, w_i)) = \sum_{i=1}^n (ka_i+b_i)\phi(v_i, w_i) = \sum_{i=1}^n ka_i\phi(v_i, w_i) + \sum_{i=1}^n b_i\phi(v_i, w_i) = k\Phi(x) + \Phi(y)$. Hence, $\Phi$ is indeed linear and since it is essentially the same as $\phi$, we will call it $\phi$ and specify it is a linear mapping from $Free(X) \to Z$(instead of bilinear function from $V \times W \to Z$).

Next, we will show $\phi(s) = 0$ for all $s \in span(S)$ where $S \subseteq Free(X)$ is the set of all vectors in the following form

$$\begin{cases} (x_1 + y_1, z_2) - (x_1, z_2) - (y_1, z_2) \\ (z_1, x_2 + y_2) - (z_1, x_2) - (z_1, y_2) \\ \alpha(x_1, y_2) - (\alpha x_1, y_2) \\ \alpha(x_1, y_2) - (x_1, \alpha y_2) \end{cases}$$

where $x_1, y_1, z_1 \in V, x_2, y_2, z_2 \in W, \alpha \in F$.

Let $\phi : V \times W \to Z$ be bilinear. Then $\phi(\alpha x_1 + y_1, z_2) = \alpha\phi(x_1, z_2) + \phi(y_1, z_2)$ and $\phi(z_1, \alpha x_2 + y_2) = \phi(z_1, x_2) + \alpha(z_1, y_2)$ for all $x_1, y_1, z_1 \in V, x_2, y_2, z_2 \in W, \alpha \in F$. In particular, it suffice to show each vector in the above form in $S$ is equal zero under the linear mapping $\phi$ as $span(S)$ is linear combinations of the above forms. Let $0_F \in F$, $0_V \in V$, and $0_W \in W$.

$$\begin{aligned} \phi((x_1 + y_1, z_2) - (x_1, z_2) - (y_1, z_2)) &= \phi(x_1 + y_1, z_2) - \phi(x_1, z_2) - \phi(y_1, z_2) \\ &= \phi(x_1 + y_1 - x_1, z_2) - \phi(y_1, z_2) \\ &= \phi(y_1, z_2) - \phi(y_1, z_2) \\ &= 0 \end{aligned}$$

$$\phi((z_1, x_2 + y_2) - (z_1, x_2) - (z_1, y_2)) = 0$$

$$\begin{aligned} \phi(\alpha(x_1, y_2) - (\alpha x_1, y_2)) &= \alpha\phi(x_1, y_2) - \phi(\alpha x_1, y_2) \\ &= \phi(\alpha x_1, y_2) - \phi(\alpha x_1, y_2) \\ &= 0 \end{aligned}$$

$$\phi(\alpha(x_1, y_2) - (x_1, \alpha y_2)) = 0$$

Thus, by the above Lemma, we have a linear transformation from $V \otimes W \to Z$ induced by $\phi$ as $V \otimes W = (Free(X))/S$ and $\phi$ is a linear mapping on $Free(X)$.

Then, we show the uniqueness. Let $\phi$ and $\psi$ be two bilinear functions on $V \times W$ to $Z$ and suppose they induce the same lienar transformation from $Free(X) \to Z$, say $T$. Then, since $T$ exists, it is uniquely determined by $\phi$ and by $\psi$. In particular, we must have $T(x + S) = \phi(x) = \psi(x)$ for all $x \in Free(V \times W)$ and so $\phi = \psi$. Thus, if $T$ exists, the bilinear function that induces $T$ is unqiue and every $T$ gives us an bilinear function on $V \times W$ to $Z$. ♡

**Remark 5.1.12.** Let $V$ be finite dimensional $F$ vector space, $V^* = \{T : V \to F : T \text{ is linear}\}$ is called the ***dual space of V***. Then, let $\{v_1, ..., v_n\}$ be a basis for $V$, then let $v_i^* \in V^*$ to be $v_i^*(v_j) = \delta_{ij}$, and $\{v_1^*, ..., v_n^*\}$ is a basis for $V^*$.

**Definition 5.1.13.** Let $V, W$ be finite dimensional $F$ vector space, we define

$$L(V,W) := \{T : V \to W : T \text{ is linear}\} =: Hom(V,W)$$

Note $L(V,W)$ is a $F$ vector space.

**Example 5.1.14.** Let $V, W$ be finite dimensional $F$ vector space, show $V^* \otimes_F W \cong L(V,W)$.

*Solution.* Define $\phi : V^* \times W \to L(V,W)$ by $\phi(f,w)(v) = f(v)w$. We can show $\phi$ is bilinear and well-defined. Thus, by the Universal property, there is a unique linear transformation $T : V^* \otimes W \to L(V,W)$ such that $T(f \otimes W) = \phi(f,w)$.

We will show $T$ is an isomorphism by finding the inverse. Define $U : L(V,W) \to V^* \otimes W$ by $U(F) = \sum_{j=1}^m w_i^* F \otimes w_i$ where $\{w_1, ..., w_m\}$ is a basis for $W$ and $\{w_1^*, ..., w_m^*\}$ is a basis for $W^*$.

Then, let $v \in V$ be arbitrary. Suppose $F(v) = \sum_{i=1}^m \alpha_i w_i$.

$$\begin{aligned}
U \circ U(F)(v) &= T(\sum w_i^* \circ F \otimes w_i)(v) \\
&= \sum T(w_i^* \circ F \otimes w_i)(v) \\
&= \sum w_i^* \circ F(v) \times w_i \\
&= \sum w_i^*(\alpha_1 w_1 + ... + \alpha_m w_m) w_i \\
&= \sum \alpha_i w_i \\
&= F(v)
\end{aligned}$$

♠

## 5.2 Tensor Algebra

**Definition 5.2.1.** Let $F$ be a field, an ***F-Algebra*** is a vector space $A$ over $F$ equipped with a multiplication map such that, for all $a, b, c \in A$, $\alpha \in F$, we have

1. $a(bc) = (ab)c$
2. $a(b+c) = ab + ac$
3. $(b+c)a = ba = ca$
4. $\alpha(ab) = (\alpha a)b = a(\alpha b)$

Let $V$ be an $F$ vector space. For $K \in \mathbb{N}$, we define $T^k(V) = \bigotimes_{i=1}^k V = V \otimes V \otimes ... \otimes V$. The elements in $T^k(V)$ is called a $k-$tensor. We also define $T^0(V) = F$.

Aside, let $V$ be $F$ vector space, and $\{W_i\}$ be countable many subspaces of $V$. The direct product

$$\prod_{i=1}^{\infty} W_i = \{(a_1, a_2, ...) : a_i \in W_i\}$$

The direct sum

$$\bigoplus_{i=1}^{\infty} W_i = \{(a_1, a_2, ...) : a_i \in W_i, \text{ where } a_i = 0 \text{ for all but infinite many } i\}$$

We then define the ***tensor algebra*** of $V$ by $T(V) = \bigoplus_{i=0}^{\infty} T^i(V)$.

**Example 5.2.2.** Let $x, y \in V$, $F = \mathbb{R}$, then an element in $T(V)$ would like

$$3 + 2(x \otimes y) - \frac{1}{7}(x \otimes x \otimes y) + 87(x \otimes x \otimes x \otimes y \otimes x) \in T(V)$$

Here, in $T(V)$, multiplication is given by $(v_1 \otimes ... \otimes v_k)(u_1 \otimes ... \otimes u_l) = v_1 \otimes ... \otimes v_k \otimes u_1 \otimes ... \otimes u_l$ and extend by distributivity.

## 5.3 Exterior ALgebra

**Definition 5.3.1.** Let $V$ be $F$ vector space, let $A(V)$ be the ideal of $T(V)$ generated by elements of the form $V \otimes V$. We define the ***exterior algebra*** of $V$ by $\Lambda(V) = T(V)/A(V)$.

Then, we have $\overline{x} + \overline{y} = \overline{x + y}$, $\alpha\overline{x} = \overline{\alpha x}$ and $\overline{x} \cdot \overline{y} = \overline{xy}$.

Then, we write $\overline{v_1 \otimes ... \otimes v_k} \in \Lambda(V)$ as $v_1 \wedge v_2 \wedge ... \wedge v_k$.

**Example 5.3.2.** Note $(x+y) \wedge (x+y) = x \wedge x + x \wedge y + y \wedge x + y \wedge y = x \wedge y + y \wedge x$. However, $(x + y) \wedge (x + y) = 0$ and so $x \wedge y = -y \wedge x$.

# Chapter 6

# Intro to functional analysis

**Remark 6.0.1.** Recall that $(V, \|\cdot\|)$ is normed vector space where $\|\cdot\| : V \to [0, \infty)$ such that $\|v\| = 0 \Leftrightarrow v = 0$ and $\|\alpha v\| = |\alpha| \cdot \|v\|$ and $\|u + v\| \leq \|u\| + \|v\|$.

**Definition 6.0.2.** Let $(V, \|\cdot\|)$,

1. A sequence $(x_n)$ in $V$ converges to $x$, we write $x_n \to x$, if $\forall \epsilon > 0$, $\exists N \in \mathbb{N}$,

$$(n \geq \mathbb{N}) \Rightarrow \|x_n - x\| < \epsilon$$

2. A sequence $(x_n)$ in $V$ is Cauchy if

$$\forall \epsilon > 0, \exists (n, m \geq N) \Rightarrow \|x_n - x_m\| < \epsilon$$

3. We say $V$ is complete if every Cauchy sequence converges in $V$.
4. If a normed vector space is complete, we call it a ***Banach space***. If $\|v\| = \sqrt{\langle v, v \rangle}$ for some inner product on $V$, and $(V, \|\cdot\|)$ is complete, we call $V$ a ***Hilbert space***.

**Example 6.0.3.**

1. $(\mathbb{R}^n, \|\cdot\|), (\mathbb{C}^n, \|\cdot\|)$ are Hilbert space.
2. $C_{00} = \{(x_n)_{n=1}^{\infty} : x_n \in \mathbb{R}, x_n = 0$ for all but finitely many $n \in \mathbb{N}\}$. Let $\|(x_n)_{n=1}^{\infty}\| = max_{n \in \mathbb{N}}\{|x_n|\}$ be the norm. Let $x_n = (1, ..., \frac{1}{n}, 0, ...)$, then $(x_n)_{n=1}^{\infty}$ is Cauchy. Indeed, let $\epsilon > 0$ be given, then choose $N \in \mathbb{N}$ so that $\frac{1}{N} < \epsilon$. Suppose $n, m \geq N$, and without lose of generality, suppose $n < m$, we have $\|x_n - x_m\| = \frac{1}{1+n} < \frac{1}{n} \leq \frac{1}{N} < \epsilon$. Thus $(x_n)$ is Cauchy. However, we have that $x_n \to x$ where $x = (1, \frac{1}{2}, \frac{1}{3}, ...) \notin C_{00}$.
3. $C_0 = \{(x_n) : x_n \in \mathbb{R}, \lim\limits_{n \to \infty} x_n = 0\}$ with the same norm as $C_{00}$ is a Banach space.

**Example 6.0.4.** Show $l^{\infty}$ is Banach space, where $l^{\infty} = \{(a_n) : a_n \in \mathbb{R}, \sup_{n \in \mathbb{N}} |a_n| < \infty\}$ with the norm $\|(a_n)\| = \sup_{n \in \mathbb{N}}\{|a_n|\}$.

*Solution.* Let $(x_n)$ be a Cauchy sequence. We write $x_n = (x_n^1, x_n^2, x_n^3, ...)$. Let $\epsilon > 0$ be given. Thus, there exists $N_1 \in \mathbb{N}$ such that $\|x_n - x_m\| < \frac{\epsilon}{2}$ for all $n, m \geq N_1$.

For every $i \in \mathbb{N}$, for $n, m \geq N_1$, we have $|x_n^i - x_m^i| \leq \|x_n - x_m\| < \frac{\epsilon}{2} < \epsilon$ and thus $(x_n^i)_{n=1}^{\infty}$ converges in $\mathbb{R}$ for each $i$ as it is Cauchy. Say $x_n^i \to a_i \in \mathbb{R}$ for each $i$. We claim $\lim\limits_{n\to\infty} x_n = (a_i)_{i=1}^{\infty}$.

First, let $x = (a_1, a_2, a_3, ...)$, we need to show $x \in l^{\infty}$. For $\epsilon = 1$, there exists $N_2 \in \mathbb{N}$ such that $\|x_n - x_m\| < 1$ for $n, m \geq N_2$. Then, for $n, m \geq N_2$, we have $|x_n^i - x_m^i| < 1$. Now, consider $\lim\limits_{m\to\infty} |x_n^i - x_m^i| = |x_n^i - a_i|$ as $|x|$ is continuous. Thus, we have $|x_n^i - a_i| \leq 1$, therefore, for $n \geq N_2$,

$$\sup_i |a_i| = \sup_i |a_i - x_n^i + x_n^i| \leq \sup_i |a_i - x_n^i| + |x_n^i| \leq 1 + \|x_n\| < \infty$$

Thus $x \in l^{\infty}$. Next, we show $\lim\limits_{n\to\infty} x_n = x$. For $i \in \mathbb{N}$, $n, m \geq \mathbb{N}$,

$$\begin{aligned}
&|x_m^i - x_n^i| \leq \|x_m - x_n\| < \epsilon/2 \\
&\Rightarrow \lim_{n\to\infty} |x_m^i - x_n^i| = |x_m^i - a_i| \leq \epsilon/2 \\
&\text{For } m \geq N_1, \|x_m - x_n\| = \sup_i |x_m^i - a_i| \leq \epsilon/2 < \epsilon
\end{aligned}$$

Thus, $(l^{\infty}, \|\cdot\|)$ is a Banach space. ♠

**Remark 6.0.5.** A closed subset is Banach space is a Banach space. Thus, $C_0$ is a Banach space.

**Example 6.0.6.** Let $1 \leq p \leq \infty$, let $(a_n)_{n=1}^{\infty}$ be a sequence in $\mathbb{R}$. Define $\|(a_n)\|_p = (\sum_{n=1}^{\infty} |a_n|^p)^{\frac{1}{p}}$. Let $l^p = \{(a_n)_n : \|(a_n)\|_p < \infty\}$. Then, $l^p$ with $\|\cdot\|_p$ is a Banach space.

**Remark 6.0.7.** When $p = 2$, $l^2$ is a Hilbert space with the inner product

$$\langle (a_n), (b_n) \rangle = \sum_{i=1}^{\infty} a_i b_i$$

**Remark 6.0.8.** Let $V$ be inner product space, note $\|x+y\|^2 + \|x-y\|^2 = 2(\|x\|^2 + \|y\|^2)$. A norm is induced by inner product if and only if the parallelogram law holds.

**Example 6.0.9.** Consider $l^{\infty}$ with $\|\cdot\|_{\infty}$. Let $x = (1, 0, ...)$ and $y = (0, 1, ...)$, then $\|x+y\|^2 + \|x-y\|^2 = 1 + 1 = 2$, but $2(\|x\|^2 + \|y\|^2) = 4$. Thus $l^{\infty}$ is not Hilbert space.

Moreover, consider $l^p$ with $\|\cdot\|_p$. We have $\|x+y\|^2 + \|x-y\|^2 = 2 \cdot 2^{\frac{2}{p}}$ and $2(\|x\|^2 + \|y\|^2) = 4$. Thus, $l^p$ is Hilbert space if and only if $p = 2$.

**Definition 6.0.10.** Let $V, W$ be normed vector spaces, $T : V \to W$ be linear. Then $T$ is continuous at $v \in V$ if for all $\epsilon > 0$, there exists $\delta > 0$ so that $\forall x \in V$, $(\|x - v\| < \delta) \Rightarrow (\|T(x) - T(v)\| < \epsilon)$. Then, $T$ is continuous if $T$ is continuous at every $v \in V$. Moreover, $T$ be bounded if there exists $c \geq 0$ so $\|T(x)\| \leq C\|x\|$ for all $x \in V$.

**Theorem 6.0.11.** *Let $V, W$ be normed vector spaces, let $T : V \to W$ be linear. Then, the following are equivalent:*

*a) $T$ is continuous*
*b) $T$ is continuous at $0$*
*c) $T$ is bounded*
*d) $n_1 = \sup\{\|T(x)\| : \|x\| \leq 1\} < \infty$*
*e) $n_2 = \sup\{\|T(x)\| : \|x\| = 1\} < \infty$*

*Proof.* ($a \Rightarrow b$) Trivial.

($b \Rightarrow c$) Suppose $T$ is continuous at 0. For $\epsilon = 1$, there exists $\delta > 0$ such that if $\|x\| < \delta$ then $\|T(x)\| < 1$. For $0 \neq x \in V$, we have $\left\|T(\frac{\delta}{2\|x\|})x\right\| < 1$. Thus, $\|T(x)\| < \frac{2\|x\|}{\delta}$.

($c \Rightarrow d$) Suppose $T$ is bounded, say $\|T(x)\| \leq C\,\|x\|$, $C > 0$. Then, for $x \in V$ with $\|x\| \leq 1$, then $\|T(x)\| \leq C\,\|x\| \leq C$ which imply $n_1 \leq C$.

($d \Rightarrow e$) Trivial.

($e \Rightarrow a$) Suppose $n_2 < \infty$. Let $v \in V$, let $\epsilon > 0$. Choose $\delta = \frac{\epsilon}{n_2+1}$. Suppose $x \in V$ with $\|x - v\| < \delta$. If $x = v$ then we are done. Assume $x \neq v$. Then,

$$\begin{aligned}
\|T(x) - T(v)\| &= \|T(x - v)\| \\
&= \|x - v\| \cdot \left\|T(\frac{x - v}{\|x - v\|})\right\| \\
&\leq n_2\delta < \epsilon
\end{aligned}$$

♡

**Remark 6.0.12.** Let $T : V \to W$ be continuous and $n_1, n_2 < \infty$. Clearly $n_2 \leq n_1$. If $x \in V$, $0 < \|x\| \leq 1$, then $\left\|T(\frac{x}{\|x\|})\right\| = \frac{1}{\|x\|}\|T(x)\| \leq n_2$. Thus $\|T(x)\| \leq n_2\,\|x\| \leq n_2$. Thus, $n_1 \leq n_2$ and so $n_1 = n_2$.

**Definition 6.0.13.** Let $\|T\| = \sup_{\|x\|=1}\{\|T(x)\|\}$ be the operator norm.

# Chapter 7

# Final

**Definition 7.0.1.** PL- proof from lecture.

PA- proof from assignments.

C- computational.

**Remark 7.0.2.** 6 questions, 10 mark each.

1. JCF, five parts, worth 2 marks each. Similar to midterm.
2. Part a, PL, 7 marks. Part b, C, 3 marks.
3. Part a, PL, 3 marks. Part b, PL, 3 marks. Part c, C, 4 marks.
4. Part a, PA, 4 marks. Part b, PL, 4 marks. Part c, PL, 2 marks.
5. Part a, PL, 4 marks. Part b, PL, 3 marks. Part c, new proofs, 3 marks.
6. Part a, similarly to assignment, 7 marks. Part b, new proofs, 3 marks.