MAT3040: Advanced Linear Algebra

Table Of Contents

Chapter 1

Week1

1.1. Monday for MAT3040

1.1.1. Introduction to Advanced Linear Algebra

Advanced Linear Algebra is one of the most important course in MATH major, with pre-request MAT2040. This course will offer the really linear algebra knowledge.

What the content will be covered?.

- In MAT2040 we have studied the space $\mathbb{R}^n$; while in MAT3040 we will study the general vector space V.
- In MAT2040 we have studied the *linear transformation* between Euclidean spaces, i.e., $T : \mathbb{R}^n \to \mathbb{R}^m$; while in MAT3040 we will study the linear transformation from vector spaces to vector spaces: $T : V \to W$
- In MAT2040 we have studied the eigenvalues of $n \times n$ matrix $\boldsymbol{A}$; while in MAT3040 we will study the eigenvalues of a **linear operator** $T : V \to V$.
- In MAT2040 we have studied the dot product $\boldsymbol{x} \cdot \boldsymbol{y} = \sum_{i=1}^{n} x_i y_i$; while in MAT3040 we will study the **inner product** $\langle \boldsymbol{v}_1, \boldsymbol{v}_2 \rangle$.

Why do we do the generalization?. We are studying many other spaces, e.g., $C(\mathbb{R})$ is called the space of all functions on $\mathbb{R}$, $C^\infty(\mathbb{R})$ is called the space of all infinitely differentiable functions on $\mathbb{R}$, $\mathbb{R}[x]$ is the space of polynomials of one-variable.

■ **Example 1.1**

1. Consider the Laplace equation $\Delta f = 0$ with linear operator Δ:

$$\Delta : C^\infty(\mathbb{R}^3) \to C^\infty(\mathbb{R}^3) \quad f \mapsto \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right) f$$

The solution to the PDE $\Delta f = 0$ is the 0-eigenspace of Δ.

2. Consider the Schrödinger equation $\hat{H} f = E f$ with the linear operator

$$\hat{H} : C^\infty(\mathbb{R}^3) \to C^\infty(\mathbb{R}^3), \quad f \to \left[\frac{-\hbar^2}{2\mu}\nabla^2 + V(x, y, z)\right] f$$

Solving the equation $\hat{H} f = E f$ is equivalent to finding the eigenvectors of $\hat{H}$. In fact, the eigenvalues of $\hat{H}$ are **discrete**.

■

1.1.2. Vector Spaces

Definition 1.1 [Vector Space] A **vector space** over a field $\mathbb{F}$ (in particular, $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$) is a set of objects V equipped with vector addiction and scalar multiplication such that

1. the vector addiction + is closed with the rules:
   (a) **Commutativity**: $\forall \boldsymbol{v}_1, \boldsymbol{v}_2 \in V$, $\boldsymbol{v}_1 + \boldsymbol{v}_2 = \boldsymbol{v}_2 + \boldsymbol{v}_1$.
   (b) **Associativity**: $\boldsymbol{v}_1 + (\boldsymbol{v}_2 + \boldsymbol{v}_3) = (\boldsymbol{v}_1 + \boldsymbol{v}_2) + \boldsymbol{v}_3$.
   (c) **Addictive Identity**: $\exists \mathbf{0} \in V$ such that $\mathbf{0} + \boldsymbol{v} = \boldsymbol{v}$, $\forall \boldsymbol{v} \in V$.
2. the **scalar multiplication** is closed with the rules:
   (a) **Distributive**: $\alpha(\boldsymbol{v}_1 + \boldsymbol{v}_2) = \alpha\boldsymbol{v}_1 + \alpha\boldsymbol{v}_2, \forall \alpha \in \mathbb{F}$ and $\boldsymbol{v}_1, \boldsymbol{v}_2 \in V$
   (b) **Distributive**: $(\alpha_1 + \alpha_2)\boldsymbol{v} = \alpha_1\boldsymbol{v} + \alpha_2\boldsymbol{v}$
   (c) **Compatibility**: $a(b\boldsymbol{v}) = (ab)\boldsymbol{v}$ for $\forall a, b \in \mathbb{F}$ and $\boldsymbol{b} \in V$.
   (d) $0\boldsymbol{v} = \mathbf{0}$, $1\boldsymbol{v} = \boldsymbol{v}$.

■

Here we study several examples of vector spaces:

■ **Example 1.2** For $V = \mathbb{F}^n$, we can define

1. Addictive Identity:
$$\mathbf{0} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$$
2. Scalar Multiplication:
$$\alpha \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \alpha x_1 \\ \vdots \\ \alpha x_n \end{pmatrix}$$
3. Vector Addiction:
$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{pmatrix}$$

■

■ **Example 1.3** 1. It is clear that the set $V = M_{n\times n}(\mathbb{F})$ (the set of all $m \times n$ matrices) is a vector space as well.

2. The set $V = C(\mathbb{R})$ is a vector space:
   (a) Vector Addiction:
$$(f+g)(x) = f(x) + g(x), \forall f, g \in V$$
   (b) Scalar Multiplication:
$$(\alpha f)(x) = \alpha f(x), \forall \alpha \in \mathbb{R}, f \in V$$
   (c) Addictive Identity is a zero function, i.e., $\mathbf{0}(x) = 0$ for all $x \in \mathbb{R}$.

■

Definition 1.2 A sub-collection $W \subseteq V$ of a vector space V is called a **vector subspace of** V if W itself forms a vector space, denoted by $W \leq V$.

■ **Example 1.4**

1. For $V = \mathbb{R}^3$, we claim that $W = \{(x, y, 0) \mid x, y \in \mathbb{R}\} \leq V$
2. $W = \{(x, y, 1) \mid x, y \in \mathbb{R}\}$ is not the vector subspace of V.

Proposition 1.1 $W \subseteq V$ is a **vector subspace** of V iff for $\forall \boldsymbol{w}_1, \boldsymbol{w}_2 \in W$, we have $\alpha \boldsymbol{w}_1 + \beta \boldsymbol{w}_2 \in W$, for $\forall \alpha, \beta \in \mathbb{F}$.

■ **Example 1.5**

1. For $V = M_{n \times n}(\mathbb{F})$, the subspace $W = \{A \in V \mid \boldsymbol{A}^{\mathrm{T}} = \boldsymbol{A}\} \leq V$
2. For $V = C^{\infty}(\mathbb{R})$, define $W = \{f \in V \mid \frac{\mathrm{d}^2}{\mathrm{d}x^2} f + f = 0\} \leq V$. For $f, g \in W$, we have

$$(\alpha f + \beta g)'' = \alpha f'' + \beta g'' = \alpha(-f) + \beta(-g) = -(\alpha f + \beta g),$$

   which implies $(\alpha f + \beta g)'' + (\alpha f + \beta g) = 0$.

1.4. Wednesday for MAT3040

1.4.1. Review

1. Vector Space: e.g., $\mathbb{R}, M_{n\times n}(\mathbb{R}), C(\mathbb{R}^n), \mathbb{R}[x]$.
2. Vector Subspace: $W \le V$, e.g.,
   (a) $V = \mathbb{R}^2$, the set $W := \mathbb{R}^2_+$ is not a vector subspace since W is not closed under scalar multiplication;
   (b) the set $W = \mathbb{R}^2_+ \bigcup \mathbb{R}^2_-$ is not a vector subspace since it is not closed under addition.
   (c) For $V = \mathbb{M}_{3\times 3}(\mathbb{R})$, the set of invertible 3×3 matrices is not a vector subspace, since we cannot define zero vector inside.
   (d) Exercise: How about the set of all singular matrices? Answer: it is not a vector subspace since the vector addition does not necessarily hold.

1.4.2. Spanning Set

Definition 1.11 [Span] Let V be a vector space over $\mathbb{F}$:

1. A linear combination of a subset S in V is of the form

$$\sum_{i=1}^{n} \alpha_i \boldsymbol{s}_i, \quad \alpha_i \in \mathbb{F}, \boldsymbol{s}_i \in S$$

   Note that the summation should be finite.
2. The **span** of a subset $S \subseteq V$ is

$$\operatorname{span}(S) = \left\{ \sum_{i=1}^{n} \alpha_i \boldsymbol{s}_i \,\middle|\, \alpha_i \in \mathbb{F}, \boldsymbol{s}_i \in S \right\}$$

3. S is a spanning set of V, or say S spans V, if

$$\operatorname{span}(S) = V.$$

■ **Example 1.12** For $V = \mathbb{R}[x]$, define the set

$$S = \{1, x^2, x^4, \ldots, x^6\},$$

then $2 + x^4 + \pi x^{106} \in \text{span}(S)$, while the series $1 + x^2 + x^4 + \cdots \notin \text{span}(S)$.

It is clear that $\text{span}(S) \neq V$, but S is the spanning set of $W = \{p \in V \mid p(x) = p(-x)\}$. ■

■ **Example 1.13** For $V = M_{3\times 3}(\mathbb{R})$, let $W_1 = \{\boldsymbol{A} \in V \mid \boldsymbol{A}^{\mathrm{T}} = \boldsymbol{A}\}$ and $W_2 = \{\boldsymbol{B} \in V \mid \boldsymbol{B}^{\mathrm{T}} = -\boldsymbol{B}\}$ (the set of skew-symmetric matrices) be two vector subspaces. Define the set

$$\boldsymbol{S} := W_1 \bigcup W_2$$

Exercise: $\boldsymbol{S}$ spans V. ■

Proposition 1.7 Let S be a subset in a vector space V.

1. $S \subseteq \text{span}(S)$
2. $\text{span}(S) = \text{span}(\text{span}(S))$
3. If $\boldsymbol{w} \in \text{span}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\} \setminus \text{span}\{\boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$, then

$$\boldsymbol{v}_1 \in \text{span}\{\boldsymbol{w}, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\} \setminus \text{span}\{\boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$$

Proof. 1. For each $\boldsymbol{s} \in S$, we have

$$\boldsymbol{s} = 1 \cdot \boldsymbol{s} \in \text{span}(S)$$

2. From (1), it's clear that $\text{span}(S) \subseteq \text{span}(\text{span}(S))$, and therefore suffices to show $\text{span}(\text{span}(S)) \subseteq \text{span}(S)$:

Pick $\boldsymbol{v} = \sum_{i=1}^{n} \alpha_i \boldsymbol{v}_i \in \text{span}(\text{span}(S))$, where $\boldsymbol{v}_i \in \text{span}(S)$. Rewrite

$$\boldsymbol{v}_i = \sum_{j=1}^{n_i} \beta_{ij} \boldsymbol{s}_j, \quad \boldsymbol{s}_j \in S,$$

which implies

$$\begin{aligned}\boldsymbol{v} &= \sum_{i=1}^{n} \alpha_i \sum_{j=1}^{n_i} \beta_{ij} \boldsymbol{s}_j \\ &= \sum_{i=1}^{n} \sum_{j=1}^{n_i} (\alpha_i \beta_{ij}) \boldsymbol{s}_j,\end{aligned}$$

i.e., $\boldsymbol{v}$ is the finite combination of elements in S, whcih implies $\boldsymbol{v} \in \text{span}(S)$.

3. By hypothesis, $\boldsymbol{w} = \alpha_1 \boldsymbol{v}_1 + \cdots + \alpha_n \boldsymbol{v}_n$ with $\alpha_1 \neq 0$, which implies

$$\boldsymbol{v}_1 = -\frac{\alpha_2}{\alpha_1} \boldsymbol{v}_2 + \cdots + \left(-\frac{1}{\alpha_1} \boldsymbol{w}\right)$$

which implies $\boldsymbol{v}_1 \in \text{span}\{\boldsymbol{w}, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$. It suffices to show $\boldsymbol{v}_1 \notin \text{span}\{\boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$. Suppose on the contrary that $\boldsymbol{v}_1 \in \text{span}\{\boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$. It's clear that $\text{span}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\} = \text{span}\{\boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$. (left as exercise). Therefore,

$$\emptyset = \text{span}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\} \setminus \text{span}\{\boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\},$$

which is a contradiction.

■

1.4.3. Linear Independence and Basis

Definition 1.12 [Linear Independence] Let S be a (not necessarily finite) subset of V. Then S is **linearly independent** (l.i.) on V if for any finite subset $\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_k\}$ in S,

$$\sum_{i=1}^{k} \alpha_i \boldsymbol{s}_i = 0 \Longleftrightarrow \alpha_i = 0, \forall i$$

■ **Example 1.14** For $V = C(\mathbb{R})$,

1. let $S_1 = \{\sin x, \cos x\}$, which is l.i., since

$$\alpha \sin x + \beta \cos x = \mathbf{0}(\text{means zero function})$$

   Taking $x = 0$ both sides leads to $\beta = 0$; taking $x = \frac{\pi}{2}$ both sides leads to $\alpha = 0$.

2. let $S_2 = \{\sin^2 x, \cos^2 x, 1\}$, which is linearly dependent, since

$$1 \cdot \sin^2 x + 1 \cdot \cos^2 x + (-1) \cdot 1 = 0, \forall x$$

3. Exercise: For $V = \mathbb{R}[x]$, let $S = \{1, x, x^2, x^3, \dots, \}$, which is l.i.:

   Pick $x^{k_1}, \dots, x^{k_n} \in S$ with $k_1 < \cdots < k_n$. Consider that the euqation

$$\alpha_1 x^{k_1} + \cdots + \alpha_n x^{k_n} = \mathbf{0}$$

   holds for all x, and try to solve for $\alpha_1, \dots, \alpha_n$ (one way is differentation.)

Definition 1.13 [Basis] A subset S is a **basis** of V if

(a) S spans V;

(b) S is l.i.

Example 1.15

1. For $V = \mathbb{R}^n$, $S = \{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$ is a basis of V
2. For $V = \mathbb{R}[x]$, $S = \{1, x, x^2, \dots\}$ is a basis of V
3. For $V = M_{2\times 2}(\mathbb{R})$,

$$S = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \right\}$$

   is a basis of V

R Note that there can be many basis for a vector space V.

Proposition 1.8 Let $V = \text{span}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m\}$, then there exists a subset of $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m\}$, which is a basis of V.

Proof. If $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m\}$ is l.i., the proof is complete.

Suppose not, then $\alpha_1 \boldsymbol{v}_1 + \cdots + \alpha_m \boldsymbol{v}_m = \boldsymbol{0}$ has a non-trivial solution. w.l.o.g., $\alpha_1 \neq 0$, which implies

$$\boldsymbol{v}_1 = -\frac{\alpha_2}{\alpha_1}\boldsymbol{v}_2 + \cdots + \left(\frac{\alpha_m}{\alpha_1}\right)\boldsymbol{v}_m \implies \boldsymbol{v}_1 \in \text{span}\{\boldsymbol{v}_2, \ldots, \boldsymbol{v}_m\}$$

By the proof in (c), Proposition (1.7),

$$\text{span}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m\} = \text{span}\{\boldsymbol{v}_2, \ldots, \boldsymbol{v}_m\},$$

which implies $V = \text{span}\{\boldsymbol{v}_2, \ldots, \boldsymbol{v}_m\}$.

Continuse this argument finitely many times to guarantee that $\{\boldsymbol{v}_i, \boldsymbol{v}_{i+1}, \ldots, \boldsymbol{v}_m\}$ is l.i., and spans V. The proof is complete. ∎

Corollary 1.1 If $V = \text{span}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m\}$ (i.e., V is finitely generated), then V has a basis. (The same holds for non-finitely generated V).

Proposition 1.9 If $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ is a basis of V, then every $\boldsymbol{v} \in V$ can be expressed uniquely as

$$\boldsymbol{v} = \alpha_1 \boldsymbol{v}_1 + \cdots + \alpha_n \boldsymbol{v}_n$$

Proof. Since $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ spans V, so $\boldsymbol{v} \in V$ can be written as

$$\boldsymbol{v} = \alpha_1 \boldsymbol{v}_1 + \cdots + \alpha_n \boldsymbol{v}_n \tag{1.1}$$

Suppose further that

$$\boldsymbol{v} = \beta_1 \boldsymbol{v}_1 + \cdots + \beta_n \boldsymbol{v}_n, \tag{1.2}$$

it suffices to show that $\alpha_i = \beta_i$ for $\forall i$:

Subtracting (1.1) into (1.2) leads to

$$(\alpha_1 - \beta_1)\boldsymbol{v}_1 + \cdots + (\alpha_n - \beta_n)\boldsymbol{v}_n = 0.$$

By the hypothesis of linear independence, we have $\alpha_i - \beta_i = 0$ for $\forall i$, i.e., $\alpha_i = \beta_i$. $\blacksquare$

Chapter 2

Week2

2.1. Monday for MAT3040

Reviewing.

1. Linear Combination and Span
2. Linear Independence
3. Basis: a set of vectors $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_k\}$ is called a **basis** for V if $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_k\}$ is linearly independent, and $V = \operatorname{span}\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_k\}$.

   Lemma: Given $V = \operatorname{span}\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_k\}$, we can find a basis for this set. Here V is said to be **finitely generated**.
4. Lemma: The vector $\boldsymbol{w} \in \operatorname{span}\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_n\} \setminus \operatorname{span}\{\boldsymbol{v}_2, \dots, \boldsymbol{v}_n\}$ implies that

$$\boldsymbol{v}_1 \in \operatorname{span}\{\boldsymbol{w}, \boldsymbol{v}_2, \dots, \boldsymbol{v}_n\} \setminus \operatorname{span}\{\boldsymbol{v}_2, \dots, \boldsymbol{v}_n\}$$

2.1.1. Basis and Dimension

Theorem 2.1 Let V be a finitely generated vector space. Suppose $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_m\}$ and $\{\boldsymbol{w}_1, \dots, \boldsymbol{w}_n\}$ are two basis of V. Then $m = n$. (where m is called the **dimension**)

Proof. Suppose on the contrary that $m \neq n$. Without loss of generality (w.l.o.g.), assume that $m < n$. Let $\boldsymbol{v}_1 = \alpha_1 \boldsymbol{w}_1 + \cdots + \alpha_n \boldsymbol{w}_n$, with some $\alpha_i \neq 0$. w.l.o.g., assume $\alpha_1 \neq 0$. Therefore,

$$\boldsymbol{v}_1 \in \operatorname{span}\{\boldsymbol{w}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_n\} \setminus \operatorname{span}\{\boldsymbol{w}_2, \dots, \boldsymbol{w}_n\} \tag{2.1}$$

which implies that $\boldsymbol{w}_1 \in \operatorname{span}\{\boldsymbol{v}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_n\} \setminus \operatorname{span}\{\boldsymbol{w}_2, \dots, \boldsymbol{w}_n\}$.

Then we claim that $\{\boldsymbol{v}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_n\}$ is a basis of V:

1. Note that $\{\boldsymbol{v}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_n\}$ is a spannning set:

$$\begin{aligned}\boldsymbol{w}_1 \in \operatorname{span}\{\boldsymbol{v}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_n\} &\implies \{\boldsymbol{w}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_n\} \subseteq \operatorname{span}\{\boldsymbol{v}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_n\} \\ &\implies \operatorname{span}\{\boldsymbol{w}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_n\} \subseteq \operatorname{span}\{\operatorname{span}\{\boldsymbol{v}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_n\}\} \subseteq \operatorname{span}\{\boldsymbol{v}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_n\}\end{aligned}$$

   Since $V = \operatorname{span}\{\boldsymbol{w}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_n\}$, we have $\operatorname{span}\{\boldsymbol{v}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_n\} = V$.

2. Then we show the linear independence of $\{\boldsymbol{v}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_n\}$. Consider the equation

$$\beta_1 \boldsymbol{v}_1 + \beta_2 \boldsymbol{v}_2 + \dots + \beta_n \boldsymbol{w}_n = \boldsymbol{0}$$

   (a) When $\beta_1 \neq 0$, we imply

$$\boldsymbol{v}_1 = \left(-\frac{\beta_2}{\beta_1}\right)\boldsymbol{w}_2 + \dots + \left(-\frac{\beta_n}{\beta_1}\right)\boldsymbol{w}_n \in \operatorname{span}\{\boldsymbol{w}_2, \dots, \boldsymbol{w}_n\},$$

   which contradicts (2.1).

   (b) When $\beta_1 = 0$, then $\beta_2 \boldsymbol{w}_2 + \dots + \beta_n \boldsymbol{w}_n = \boldsymbol{0}$, which implies $\beta_2 = \dots = \beta_n = 0$, due to the independence of $\{\boldsymbol{w}_2, \dots, \boldsymbol{w}_n\}$.

Therefore, $\boldsymbol{v}_2 \in \operatorname{span}\{\boldsymbol{v}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_n\}$, i.e.,

$$\boldsymbol{v}_2 = \gamma_1 \boldsymbol{v}_1 + \dots + \gamma_n \boldsymbol{v}_n,$$

where $\gamma_2, \dots, \gamma_n$ cannot be all zeros, since otherwise $\{\boldsymbol{v}_1, \boldsymbol{v}_2\}$ are linearly dependent, i.e., $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_m\}$ cannot form a basis. w.l.o.g., assume $\gamma_2 \neq 0$, which implies

$$\boldsymbol{w}_2 \in \operatorname{span}\{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{w}_3, \dots, \boldsymbol{w}_n\} \setminus \operatorname{span}\{\boldsymbol{v}_1, \boldsymbol{w}_3, \dots, \boldsymbol{w}_n\}.$$

Following the simlar argument above, $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{w}_3, \dots, \boldsymbol{w}_n\}$ forms a basis of V.

Continuing the argument above, we imply $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_m, \boldsymbol{w}_{m+1}, \dots, \boldsymbol{w}_n\}$ is a basis of V.

Since $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_m\}$ is a basis as well, we imply

$$\boldsymbol{w}_{m+1} = \delta_1 \boldsymbol{v}_1 + \cdots + \delta_m \boldsymbol{v}_m$$

for some $\delta_i \in \mathbb{F}$, i.e., $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_m, \boldsymbol{w}_{m+1}\}$ is linearly dependent, which is a contradction. ■

Example 2.1 A vector space may have more than one basis.

Suppose $V = \mathbb{F}^n$, it is clear that $\dim(V) = n$, and

$$\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\} \text{ is a basis of } V, \text{ where } \boldsymbol{e}_i \text{ denotes a unit vector.}$$

There could be other basis of V, such as

$$\left\{ \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \dots, \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}, \right\}$$

Actually, the columns of any invertible $n \times n$ matrix forms a basis of V. ■

Example 2.2 Suppose $V = M_{m\times n}(\mathbb{R})$, we claim that $\dim(V) = mn$:

$$\left\{ E_{ij} \middle| \begin{matrix} 1 \le i \le m \\ 1 \le j \le n \end{matrix} \right\} \text{ is a basis of } V,$$

where E_{ij} is $m \times n$ matrix with 1 at (i,j)-th entry, and 0s at the remaining entries. ■

Example 2.3 Suppose $V = \{\text{all polynomials of degree} \le \text{n}\}$, then $\dim(V) = n+1$. ■

Example 2.4 Supppose $V = \{\boldsymbol{A} \in M_{n\times n}(\mathbb{R}) \mid \boldsymbol{A}^{\mathrm{T}} = \boldsymbol{A}\}$, then $\dim(V) = \frac{n(n+1)}{2}$. ■

Example 2.5 Let $W = \{\boldsymbol{B} \in M_{n\times n}(\mathbb{R}) \mid \boldsymbol{B}^{\mathrm{T}} = -\boldsymbol{B}\}$, then $\dim(V) = \frac{n(n-1)}{2}$. ■

R Sometimes it should be classified the field $\mathbb{F}$ for the scalar multiplication to define a vector space. Conside the example below:

1. Let $V = \mathbb{C}$, then $\dim(\mathbb{C}) = 1$ for the scalar multiplication defined under the field $\mathbb{C}$.
2. Let $V = \text{span}\{1, i\} = \mathbb{C}$, then $\dim(\mathbb{C}) = 2$ for the scalar multiplication defined under the field $\mathbb{R}$, since all $z \in V$ can be written as $z = a + bi$, $\forall a, b \in \mathbb{R}$.
3. Therefore, to aviod confusion, it is safe to write

$$\dim_{\mathbb{C}}(\mathbb{C}) = 1, \quad \dim_{\mathbb{R}}(\mathbb{C}) = 2.$$

2.1.2. Operations on a vector space

Note that the basis for a vector space is characterized as the **maximal linearly independent set**.

Theorem 2.2 — Basis Extension. Let V be a finite dimensional vector space, and $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_k\}$ be a linearly independent set on V, Then we can extend it to the basis $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_k, \boldsymbol{v}_{k+1}, \dots, \boldsymbol{v}_n\}$ of V.

Proof.

- Suppose $\dim(V) = n > k$, and $\{\boldsymbol{w}_1, \dots, \boldsymbol{w}_n\}$ is a basis of V. Consider the set $\{\boldsymbol{w}_1, \dots, \boldsymbol{w}_n\} \bigcup \{\boldsymbol{v}_1, \dots, \boldsymbol{v}_k\}$, which is linearly dependent, i.e.,

$$\alpha_1 \boldsymbol{w}_1 + \cdots + \alpha_n \boldsymbol{w}_n + \beta_1 \boldsymbol{v}_1 + \cdots + \beta_k \boldsymbol{v}_k = \boldsymbol{0},$$

with some $\alpha_i \neq 0$, since otherwise this equation will only have trivial solution. w.l.o.g., assume $\alpha_1 \neq 0$.

- Therefore, consider the set $\{\boldsymbol{w}_2, \dots, \boldsymbol{w}_n\} \bigcup \{\boldsymbol{v}_1, \dots, \boldsymbol{v}_k\}$. We keep removing elements from $\{\boldsymbol{w}_2, \dots, \boldsymbol{w}_n\}$ until we first get the set

$$S \bigcup \{\boldsymbol{v}_1, \dots, \boldsymbol{v}_k\},$$

with $S \subseteq \{\boldsymbol{w}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_n\}$ and $S \bigcup \{\boldsymbol{v}_1, \dots, \boldsymbol{v}_k\}$ is linearly independent, i.e., S is a maximal subset of $\{\boldsymbol{w}_1, \dots, \boldsymbol{w}_n\}$ such that $S \bigcup \{\boldsymbol{v}_1, \dots, \boldsymbol{v}_k\}$ is linearly independent.

- Rewrite $S = \{\boldsymbol{v}_{k+1}, \dots, \boldsymbol{v}_m\}$ and therefore $S' = \{\boldsymbol{v}_1, \dots, \boldsymbol{v}_k, \boldsymbol{v}_{k+1}, \dots, \boldsymbol{v}_m\}$ are linearly independent. It suffices to show S' spans V.
  - Indeed, for all $\boldsymbol{w}_i \in \{\boldsymbol{w}_1, \dots, \boldsymbol{w}_n\}$, $\boldsymbol{w}_i \in \text{span}(S')$, since otherwise the equation

$$\alpha \boldsymbol{w}_i + \beta_1 \boldsymbol{v}_1 + \cdots + \beta_m \boldsymbol{v}_m = \boldsymbol{0} \implies \alpha = 0,$$

    which implies that $\beta_1 \boldsymbol{v}_1 + \cdots + \beta_m \boldsymbol{v}_m = \boldsymbol{0}$ admits only trivial solution, i.e.,

$$\{\boldsymbol{w}_i\} \bigcup S' = \{\boldsymbol{w}_i\} \bigcup S \bigcup \{\boldsymbol{v}_1, \dots, \boldsymbol{v}_k\} \text{ is linearly independent,}$$

    which violetes the maximality of S.

  Therefore, all $\{\boldsymbol{w}_1, \dots, \boldsymbol{w}_n\} \subseteq \text{span}(S')$, which implies $\text{span}(S') = V$.

  Therefore, S' is a basis of V.

■

R Start with a spanning set, we keep removing something to form a basis; start with independent set, we keep adding something to form a basis.

In other words, the basis is both the minimal spanning set, and the maximal linearly independent set.

Definition 2.1 [Direct Sum] Let W_1, W_2 be two vector subspaces of V, then

1. $W_1 \bigcap W_2 := \{\boldsymbol{w} \in V \mid \boldsymbol{w} \in W_1, \text{ and } \boldsymbol{w} \in W_2\}$
2. $W_1 + W_2 := \{\boldsymbol{w}_1 + \boldsymbol{w}_2 \mid \boldsymbol{w}_i \in W_i\}$
3. If furthermore that $W_1 \bigcap W_2 = \{\boldsymbol{0}\}$, then $W_1 + W_2$ is denoted as $W_1 \oplus W_2$, which is called **direct sum**.

Proposition 2.1 $W_1 \bigcap W_2$ and $W_1 + W_2$ are vector subspaces of V.

2.4. Wednesday for MAT3040

Reviewing.

- Basis, Dimension
- Basis Extension
- $W_1 \bigcap W_2 = \emptyset$ implies $W_1 \oplus W_2 = W_1 + W_2$ (Direct Sum).

2.4.1. Remark on Direct Sum

Proposition 2.13 The set $W_1 + W_2 = W_1 \oplus W_2$ iff any $\boldsymbol{w} \in W_1 + W_2$ can be uniquely expressed as

$$\boldsymbol{w} = \boldsymbol{w}_1 + \boldsymbol{w}_2,$$

where $\boldsymbol{w}_i \in W_i$ for $i = 1, 2$.

R We can also define addiction among finite set of vector spaces $\{W_1, \dots, W_k\}$. If $\boldsymbol{w}_1 + \cdots + \boldsymbol{w}_k = \boldsymbol{0}$ implies $\boldsymbol{w}_i = 0, \forall i$, then we can write $W_1 + \cdots + W_k$ as

$$W_1 \oplus \cdots \oplus W_k$$

Proposition 2.14 — **Complementation.** Let $W \le V$ be a vector subspace of a fintie dimension vector space V. Then there exists $W' \le V$ such that

$$W \oplus W' = V.$$

Proof. It's clear that $\dim(W) := k \le n := \dim(V)$. Suppose $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_k\}$ is a basis of W.

By the basis extension proposition, we can extend it into $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_k, \boldsymbol{v}_{k+1}, \dots, \boldsymbol{v}_n\}$, which is a basis of V.

Therefore, we take $W' = \text{span}\{\boldsymbol{v}_{k+1}, \dots, \boldsymbol{v}_n\}$, which follows that

1. $W + W' = V$: $\forall \mathbf{v} \in V$ has the form

$$\mathbf{v} = (\alpha_1 \mathbf{v}_1 + \cdots + \alpha_k \mathbf{v}_k) + (\alpha_{k+1} \mathbf{v}_{k+1} + \cdots + \alpha_n \mathbf{v}_n),$$

   where $\alpha_1 \mathbf{v}_1 + \cdots + \alpha_k \mathbf{v}_k \in W$ and $\alpha_{k+1} \mathbf{v}_{k+1} + \cdots + \alpha_n \mathbf{v}_n \in W'$.

2. $W \bigcap W' = \{\mathbf{0}\}$: Suppose $\mathbf{v} \in W \bigcap W'$, i.e.,

$$\begin{aligned}\mathbf{v} &= (\beta_1 \mathbf{v}_1 + \cdots + \beta_k \mathbf{v}_k) + (0\mathbf{v}_{k+1} + \cdots + 0\mathbf{v}_n) \in W \\ &= (0\mathbf{v}_1 + \cdots + 0\mathbf{v}_k) + (\beta_{k+1} \mathbf{v}_{k+1} + \cdots + \beta_n \mathbf{v}_n) \in W'.\end{aligned}$$

   By the uniqueness of coordinates, we imply $\beta_1 = \cdots = \beta_n = 0$, i.e., $\mathbf{v} = \mathbf{0}$.

Therefore, we conclude that $W \oplus W' = V$. ■

2.4.2. Linear Transformation

Definition 2.7 [Linear Transformation] Let V, W be vector spaces. Then $T : V \to W$ is a **linear transformation** if

$$T(\alpha \mathbf{v}_1 + \beta \mathbf{v}_2) = \alpha T(\mathbf{v}_1) + \beta T(\mathbf{v}_2),$$

for $\forall \alpha, \beta \in \mathbb{F}$ and $\mathbf{v}_1, \mathbf{v}_2 \in V$.

Proposition 2.15

1. Suppose that $S : V \to W$ and $T : W \to U$ are linear transformations, then so is $T \circ S : V \to U$.
2. For any linear transformation $T : V \to W$, we have

$$T(\mathbf{0}_V) = \mathbf{0}_W$$

Proof. Simply apply the definition of the linear transformation. ■

Example 2.12

1. The transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ defined as $\mathbf{x} \mapsto \mathbf{A}\mathbf{x}$ (where $\mathbf{A} \in \mathbb{R}^{m \times n}$) is a linear transformation.

2. The transformation $T : \mathbb{R}[x] \to \mathbb{R}[x]$ defined as

$$p(x) \mapsto T(p(x)) = p'(x), \quad p(x) \mapsto T(p(x)) = \int_0^x p(t)\,\mathrm{d}t$$

is a linear transformation

3. The transformation $T : M_{n\times n}(\mathbb{R}) \to \mathbb{R}$ defined as

$$\boldsymbol{A} \mapsto \operatorname{trace}(\boldsymbol{A}) := \sum_{i=1}^{n} a_{ii}$$

is a linear transformation.

However, the transformation

$$\boldsymbol{A} \mapsto \det(\boldsymbol{A})$$

is not a linear transformation.

■

Definition 2.8 [Kernel/Image] Let $T : V \to W$ be a linear transfomation.

1. The **kernel** of T is

$$\ker(T) = T^{-1}(\mathbf{0}) = \{\boldsymbol{v} \in V \mid T(\boldsymbol{v}) = \mathbf{0}\}$$

2. The **image** (or range) of T is

$$\operatorname{Im}(T) = T(\boldsymbol{v}) = \{T(\boldsymbol{v}) \in W \mid \boldsymbol{v} \in V\}$$

■

■ **Example 2.13** 1. Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation with $T(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}$, then

$$\ker(T) = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{A}\boldsymbol{x} = \mathbf{0}\} = \operatorname{Null}(\boldsymbol{A}) \qquad \text{Null Space}$$

and

$$\text{Im}(T) = \{\boldsymbol{A}\boldsymbol{x} \mid \boldsymbol{x} \in \mathbb{R}^n\} = \text{Col}(\boldsymbol{A}) = \text{span}\{\text{columns of } \boldsymbol{A}\} \qquad \text{Column Space}$$

2. For $T(p(x)) = p'(x)$, $\ker(T) = \{\text{constant polynomials}\}$ and $\text{Im}(T) = \mathbb{R}[x]$.

Proposition 2.16 The kernel or image for a linear transformation $T : V \to W$ also forms a vector subspace:

$$\ker(T) \le V, \quad \text{Im}(T) \le W$$

Proof. For $\boldsymbol{v}_1, \boldsymbol{v}_2 \in \ker(T)$, we imply

$$T(\alpha \boldsymbol{v}_1 + \beta \boldsymbol{v}_2) = \boldsymbol{0},$$

which implies $\alpha \boldsymbol{v}_1 + \beta \boldsymbol{v}_2 \in \ker(T)$.

The remaining proof follows similarly. ■

Definition 2.9 [Rank/Nullity] Let V, W be finite dimensional vector spaces and $T : V \to W$ a linear transformation. Then we define

$$\text{rank}(T) = \dim(\text{im}(T))$$
$$\text{nullity}(T) = \dim(\ker(T))$$

R Let

$$\text{Hom}_{\mathbb{F}}(V, W) = \{\text{all linear transformations } T : V \to W\},$$

and we can define the addiction and scalar multiplication to make it a vector space:

1. For $T, S \in \text{Hom}_{\mathbb{F}}(V, W)$, define

$$(T + S)(\boldsymbol{v}) = T(\boldsymbol{v}) + S(\boldsymbol{v}),$$

which implies $T+S \in \mathrm{Hom}_{\mathbb{F}}(V,W)$.

2. Also, define

$$(\gamma T)(\boldsymbol{v}) = \gamma T(\boldsymbol{v}), \qquad \text{for } \forall \gamma \in \mathbb{F},$$

which implies $\gamma T \in \mathrm{Hom}_{\mathbb{F}}(V,W)$.

In particular, if $V = \mathbb{R}^n, W = \mathbb{R}^m$, then

$$\mathrm{Hom}_{\mathbb{F}}(V,W) = M_{m\times n}(\mathbb{R}).$$

Proposition 2.17 If $\dim(V) = n, \dim(W) = m$, then $\dim(\mathrm{Hom}_{\mathbb{F}}(V,W)) = mn$.

Proposition 2.18 There are anternative characterizations for the injectivity and surjectivity of lienar transformation T:

1. The linear transformation T is injective if and only if

$$\ker(T) = 0, \Longleftrightarrow \mathrm{nullity}(T) = 0.$$

2. The linear transformation T is surjective if and only if

$$\mathrm{im}(T) = W, \Longleftrightarrow \mathrm{rank}(T) = \dim(W).$$

3. If T is bijective, then T^{-1} is a linear transformation.

Proof. 1. (a) For the forward direction of (1),

$$\boldsymbol{x} \in \ker(T) \implies T(\boldsymbol{x}) = 0 = T(\mathbf{0}) \implies \boldsymbol{x} = \mathbf{0}$$

(b) For the reverse direction of (1),

$$T(\boldsymbol{x}) = T(\boldsymbol{y}) \implies T(\boldsymbol{x}-\boldsymbol{y}) = \mathbf{0} \implies \boldsymbol{x}-\boldsymbol{y} \in \ker(T) = \mathbf{0} \implies \boldsymbol{x} = \boldsymbol{y}$$

2. The proof follows similar idea in (1).

3. Let $T^{-1}: W \to V$. For all $\boldsymbol{w}_1, \boldsymbol{w}_2 \in W$, there exists $\boldsymbol{v}_1, \boldsymbol{v}_2 \in V$ such that $T(\boldsymbol{v}_i) = \boldsymbol{w}_i$, i.e.,

$T^{-1}(\mathbf{w}_i) = \mathbf{v}_i \; i = 1,2.$

Consider the mapping

$$T(\alpha\mathbf{v}_1 + \beta\mathbf{v}_2) = \alpha T(\mathbf{v}_1) + \beta T(\mathbf{v}_2)$$
$$= \alpha\mathbf{w}_1 + \beta\mathbf{w}_2,$$

which implies $\alpha\mathbf{v}_1 + \beta\mathbf{v}_2 = T^{-1}(\alpha\mathbf{w}_1 + \beta\mathbf{w}_2)$, i.e.,

$$\alpha T^{-1}(\mathbf{w}_1) + \beta T^{-1}(\mathbf{w}_2) = T^{-1}(\alpha\mathbf{w}_1 + \beta\mathbf{w}_2).$$

■

Definition 2.10 [isomorphism] We say that the vector subspaces V and W are isomorphic if there exists a bijective linear transfomation $T : V \to W$. $(V \cong W)$

This mapping T is called an **isomorphism** from V to W.

R If $\dim(V) = \dim(W) = n < \infty$, then $V \cong W$:

Take $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}, \{\mathbf{w}_1, \dots, \mathbf{w}_n\}$ as basis of V and W, respectively. Then one can construct $T : V \to W$ satisfying $T(\mathbf{v}_i) = \mathbf{w}_i$ for $\forall i$ as follows:

$$T(\alpha_1\mathbf{v}_1 + \dots + \alpha_n\mathbf{v}_n) = \alpha_n\mathbf{w}_1 + \dots + \alpha_n\mathbf{w}_n \; \forall \alpha_i \in \mathbb{F}$$

It's clear that our constructed T is a linear transformation.

R $V \cong W$ doesn't imply any linear transformations $T : V \to W$ is an isomorphism. e.g., $T(\mathbf{v}) = \mathbf{0}$ is not an isomorphic if $W \neq \{\mathbf{0}\}$.

Theorem 2.3 — Rank-Nullity Theorem. Let $T : V \to W$ be a linear transformation with $\dim(V) < \infty$. Then

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

Proof. Since $\ker(T) \leq V$, by proposition (2.14), there exists $V_1 \leq V$ such that

$$V = \ker(T) \oplus V_1.$$

1. Consider the transformation $T|_{V_1} : V_1 \to T(V_1)$, which is an isomorphism, since:

   - Surjectivity is immediate
   - For $\boldsymbol{v} \in \ker(T|_{V_1})$,
   $$T(\boldsymbol{v}) = \mathbf{0} \implies \boldsymbol{v} \in \ker(T),$$
   which implies $\boldsymbol{v} = \mathbf{0}$ since $\boldsymbol{v} \in \ker(T) \cap V_1 = 0$, i.e., the injectivity follows.

   Therefore, $\dim(V_1) = \dim(T(V_1))$.
2. Secondly, given an isomorphism T from X to Y with $\dim(X) < \infty$, then $\dim(X) = \dim(T(X))$. The reason follows from assignment 1 questions (8-9):
$$\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_k\} \text{ is a basis of } X \implies \{T(\boldsymbol{v}_1), \dots, T(\boldsymbol{v}_k)\} \text{ is a basis of } Y$$
3. Note that $T(V_1) = T(V) = \operatorname{im}(T)$, since:

   - for $\forall \boldsymbol{v} \in V$, $\boldsymbol{v} = \boldsymbol{v}_k + \boldsymbol{v}_1$, where $\boldsymbol{v}_k \in \ker(T), \boldsymbol{v}_1 \in V_1$, which implies
   $$T(\boldsymbol{v}) = T(\boldsymbol{v}_k) + T(\boldsymbol{v}_1) = \mathbf{0} + T(\boldsymbol{v}_1),$$
   i.e., $T(V) \subseteq T(V_1) \subseteq T(V)$, i.e., $T(V) = T(V_1)$.
4. We can show that $\dim(V) = \dim(\ker(T)) + \dim(V_1)$: Let $\{v_1, \dots, v_k\}$ be a basis of $\ker(T)$, and $\{v_{k+1}, \dots, v_n\}$ be a basis of V_1, then by the proof of complementation proposition (2.14), we imply $\{v_1, \dots, v_n\}$ is a basis of V, i.e., $\dim(V) = n = k + (n - k) = \dim(\ker(T)) + \dim(V_1)$.

Therefore, we imply

$$\begin{aligned}\dim(V) &= \dim(\ker(T)) + \dim(V_1)\\ &= \text{nullity}(T) + \dim(T(V_1))\\ &= \text{nullity}(T) + \dim(T(V))\\ &= \text{nullity}(T) + \dim(\text{im}(T))\\ &= \text{nullity}(T) + \text{rank}(T).\end{aligned}$$

■

Chapter 3

Week3

3.1. Monday for MAT3040

Reviewing.

1. Complementation. Suppose $\dim(V) = n < \infty$, then $W \le V$ implies that there exists W' such that
$$W \oplus W' = V.$$
2. Given the linear transformation $T : V \to W$, define the set $\ker(T)$ and $\operatorname{Im}(T)$.
3. Isomorphism of vector spaces: $T : V \cong W$
4. Rank-Nullity Theorem

3.1.1. Remarks on Isomorphism

Proposition 3.1 If $T : V \to W$ is an isomorphism, then

1. the set $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_k\}$ is linearly independent in V if and only if $\{T\boldsymbol{v}_1, \dots, T\boldsymbol{v}_k\}$ is linearly independent.
2. The same goes if we replace the linearly independence by spans.
3. If $\dim(V) = n$, then $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_n\}$ forms a basis of V if and only if $\{T\boldsymbol{v}_1, \dots, T\boldsymbol{v}_n\}$ forms a basis of W. In particular, $\dim(V) = \dim(W)$.
4. Two vector spaces with finite dimensions are isomorphic if and only if they have the same dimension:

Proof. It suffices to show the reverse direction. Let $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_n\}$ and $\{\boldsymbol{w}_1, \dots, \boldsymbol{w}_n\}$ be two basis of V, W, respectively. Define the linear transformation $T : V \to W$ by

$$T(a_1\boldsymbol{v}_1 + \cdots + a_n\boldsymbol{v}_n) = a_1\boldsymbol{w}_1 + \cdots + a_n\boldsymbol{w}_n$$

Then T is surjective since $\{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n\}$ spans W; T is injective since $\{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n\}$ is linearly independent. ■

3.1.2. Change of Basis and Matrix Representation

Definition 3.1 [Coordinate Vector] Let V be a finite dimensional vector space and $B = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ an **ordered** basis of V. Any vector $\boldsymbol{v} \in V$ can be uniquely written as

$$\boldsymbol{v} = \alpha_1\boldsymbol{v}_1 + \cdots + \alpha_n\boldsymbol{v}_n,$$

Therefore we define the map $[\cdot]_{\mathcal{B}} : V \to \mathbb{F}^n$, which maps any vector in $\boldsymbol{v}$ into its **coordinate vector**:

$$[\boldsymbol{v}]_{\mathcal{B}} = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}$$

R Note that $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$ and $\{\boldsymbol{v}_2, \boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ are distinct ordered basis.

■ **Example 3.1** Given $V = M_{2\times 2}(\mathbb{F})$ and the ordered basis

$$\mathcal{B} = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \right\}$$

Any matrix has the coordinate vector w.r.t. $\mathcal{B}$, i.e.,

$$\left[\begin{pmatrix} 1 & 4 \\ 2 & 3 \end{pmatrix}\right]_{\mathcal{B}} = \begin{pmatrix} 1 \\ 4 \\ 2 \\ 3 \end{pmatrix}$$

However, if given another ordered basis

$$\mathcal{B}_1 = \left\{\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \right\},$$

the matrix may have the different coordinate vector w.r.t. $\mathcal{B}_1$:

$$\left[\begin{pmatrix} 1 & 4 \\ 2 & 3 \end{pmatrix}\right]_{\mathcal{B}_1} = \begin{pmatrix} 4 \\ 1 \\ 2 \\ 3 \end{pmatrix}$$

Theorem 3.1 The mapping $[\cdot]_{\mathcal{B}} : V \to \mathbb{F}^n$ is an isomorphism.

Proof. 1. First show the operator $[\cdot]_{\mathcal{B}}$ is well-defined, i.e., the same input gives the same output. Suppose that

$$[\boldsymbol{v}]_{\mathcal{B}} = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} \quad [\boldsymbol{v}]_{\mathcal{B}} = \begin{pmatrix} \alpha_1' \\ \vdots \\ \alpha_n' \end{pmatrix},$$

then we imply

$$\begin{aligned} \boldsymbol{v} &= \alpha_1 \boldsymbol{v}_1 + \cdots + \alpha_n \boldsymbol{v}_n \\ &= \alpha_1' \boldsymbol{v}_1 + \cdots + \alpha_n' \boldsymbol{v}_n. \end{aligned}$$

By the uniqueness of coordinates, we imply $\alpha_i = \alpha_i'$ for $i = 1, \dots, n$.

2. It's clear that the operator $[\cdot]_{\mathcal{B}}$ is a linear transformation, i.e.,

$$[p\boldsymbol{v} + q\boldsymbol{w}]_{\mathcal{B}} = p[\boldsymbol{v}]_{\mathcal{B}} + q[\boldsymbol{w}]_{\mathcal{B}} \quad \forall p, q \in \mathbb{F}$$

3. The operator $[\cdot]_B$ is surjective:

$$[\boldsymbol{v}]_{\mathcal{B}} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} \implies \boldsymbol{v} = 0\boldsymbol{v}_1 + \cdots + 0\boldsymbol{v}_n = \mathbf{0}.$$

4. The injective is clear, i.e., $[\boldsymbol{v}]_{\mathcal{B}} = [\boldsymbol{w}]_{\mathcal{B}}$ implies $\boldsymbol{v} = \boldsymbol{w}$.

Therefore, $[\cdot]_B$ is an isomorphism. ■

We can use the Theorem (3.1) to simplify computations in vector spaces:

▪ **Example 3.2** Given a vector sapce $V = P_3[x]$ and its basis $B = \{1, x, x^2, x^3\}$.

To check if the set $\{1 + x^2, 3 - x^3, x - x^3\}$ is linearly independent, by part (1) in Proposition (3.1) and Theorem (3.1), it suffices to check whether the corresponding coordinate vectors

$$\left\{ \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 3 \\ 0 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \\ -1 \end{pmatrix} \right\}$$

is linearly independent, i.e., do Gaussian Elimination and check the number of pivots. ▪

Here gives rise to the question: if $\mathcal{B}_1, \mathcal{B}_2$ form two basis of V, then how are $[\boldsymbol{v}]_{\mathcal{B}_1}, [\boldsymbol{v}]_{\mathcal{B}_2}$ related to each other?

Here we consider an easy example first:

■ **Example 3.3** Consider $V = \mathbb{R}^n$ and its basis $\mathcal{B}_1 = \{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$. For any $\boldsymbol{v} \in V$,

$$\boldsymbol{v} = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} = \alpha_n \boldsymbol{e}_1 + \cdots + \alpha_n \boldsymbol{e}_n \implies [\boldsymbol{v}]_{\mathcal{B}_1} = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}$$

Also, we can construct a different basis of V:

$$\mathcal{B}_2 = \left\{ \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \dots, \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \right\},$$

which gives a different coordinate vector of $\boldsymbol{v}$:

$$[\boldsymbol{v}]_{\mathcal{B}_2} = \begin{pmatrix} \alpha_1 - \alpha_2 \\ \alpha_2 - \alpha_3 \\ \vdots \\ \alpha_{n-1} - \alpha_n \\ \alpha_n \end{pmatrix}$$

■

Proposition 3.2 — Change of Basis. Let $\mathcal{A} = \{\boldsymbol{v}_1, \dots, \boldsymbol{v}_n\}$ and $\mathcal{A}' = \{\boldsymbol{w}_1, \dots, \boldsymbol{w}_n\}$ be two ordered basis of a vector space V. Define the **change of basis** matrix from $\mathcal{A}$ to $\mathcal{A}'$, say $C_{\mathcal{A}',\mathcal{A}} := [\alpha_{ij}]$, where

$$\boldsymbol{v}_j = \sum_{i=1}^{m} \alpha_{ij} \boldsymbol{w}_i$$

Then for any vector $\boldsymbol{v} \in V$, the *change of basis amounts to left-multiplying the change of basis matrix*:

$$C_{\mathcal{A}',\mathcal{A}}[\boldsymbol{v}]_A = [\boldsymbol{v}]_{A'} \tag{3.1}$$

Define matrix $C_{\mathcal{A},\mathcal{A}'} := [\beta_{ij}]$, where

$$\boldsymbol{w}_j = \sum_{i=1}^{n} \beta_{ij} \boldsymbol{v}_i$$

Then we imply that

$$(C_{\mathcal{A},\mathcal{A}'})^{-1} = C_{\mathcal{A}',\mathcal{A}}$$

Proof. 1. First show (3.1) holds for $\boldsymbol{v} = \boldsymbol{v}_j$, $j = 1, \dots, n$:

$$\text{LHS of (3.1)} = [\alpha_{ij}]\boldsymbol{e}_j = \begin{pmatrix} \alpha_{1j} \\ \vdots \\ \alpha_{nj} \end{pmatrix}$$

$$\text{RHS of (3.1)} = [\boldsymbol{v}_j]_{\mathcal{A}'} = \left[\sum_{i=1}^{n} \alpha_i \boldsymbol{w}_i \right]_{\mathcal{A}'} = \begin{pmatrix} \alpha_{1j} \\ \vdots \\ \alpha_{nj} \end{pmatrix}$$

Therefore,

$$C_{\mathcal{A}',\mathcal{A}}[\boldsymbol{v}_j]_{\mathcal{A}} = [\boldsymbol{v}_j]_{\mathcal{A}'}, \quad \forall j = 1, \dots, n. \tag{3.2}$$

2. Then for any $\boldsymbol{v} \in V$, we imply $\boldsymbol{v} = r_1\boldsymbol{v}_1 + \cdots + r_n\boldsymbol{v}_n$, which implies that

$$\begin{aligned}
C_{\mathcal{A}',\mathcal{A}}[\boldsymbol{v}]_{\mathcal{A}} &= C_{\mathcal{A}',\mathcal{A}}[r_1\boldsymbol{v}_1 + \cdots + r_n\boldsymbol{v}_n]_{\mathcal{A}} && \text{(3.3a)} \\
&= C_{\mathcal{A}',\mathcal{A}}(r_1[\boldsymbol{v}_1]_A + \cdots + r_n[\boldsymbol{v}_n]_{\mathcal{A}}) && \text{(3.3b)} \\
&= \sum_{j=1}^{n} r_j C_{\mathcal{A}',\mathcal{A}}[\boldsymbol{v}_j]_{\mathcal{A}} && \text{(3.3c)} \\
&= \sum_{j=1}^{n} r_j [\boldsymbol{v}_j]_{\mathcal{A}'} && \text{(3.3d)} \\
&= \left[\sum_{j=1}^{n} r_j \boldsymbol{v}_j \right]_{\mathcal{A}'} && \text{(3.3e)} \\
&= [\boldsymbol{v}]_{\mathcal{A}'} && \text{(3.3f)}
\end{aligned}$$

where (3.3a) and (3.3e) is by applying the lineaity of $[\cdot]_{\mathcal{A}}$ and $[\cdot]_{\mathcal{A}'}$; (3.3d) is by applying the result (3.12). Therefore (3.1) is shown for $\forall \boldsymbol{v} \in V$.

3. Now we show that $(C_{\mathcal{A}\mathcal{A}'}C_{\mathcal{A}'\mathcal{A}}) = \boldsymbol{I}_n$. Note that

$$\begin{aligned}\boldsymbol{v}_j &= \sum_{i=1}^n \alpha_{ij}\boldsymbol{w}_i \\ &= \sum_{i=1}^n \alpha_{ij} \sum_{k=1}^n \beta_{ki}\boldsymbol{v}_k \\ &= \sum_{k=1}^n \left(\sum_{i=1}^n \beta_{ki}\alpha_{ij}\right)\boldsymbol{v}_i\end{aligned}$$

By the uniqueness of coordinates, we imply

$$\left(\sum_{i=1}^n \beta_{ki}\alpha_{ij}\right) = \delta_{jk} := \begin{cases}1, & j = k \\ 0, & j \neq k\end{cases}$$

By the matrix multiplication, the (k,j)-th entry for $C_{\mathcal{A}\mathcal{A}'}C_{\mathcal{A}'\mathcal{A}}$ is

$$[C_{\mathcal{A}\mathcal{A}'}C_{\mathcal{A}'\mathcal{A}}]_{kj} = \left(\sum_{i=1}^n \beta_{ki}\alpha_{ij}\right) = \delta_{jk} \implies (C_{\mathcal{A}\mathcal{A}'}C_{\mathcal{A}'\mathcal{A}}) = \boldsymbol{I}_n$$

Noew, suppose

$$\begin{aligned}\boldsymbol{v}_j &= \sum_{i=1}^n \alpha_{ij}\boldsymbol{w}_i \\ &= \sum_{i=1}^n \alpha_{ij} \sum_{k=1}^n \beta_{ki}\boldsymbol{v}_k \\ &= \sum_{k=1}^n \left(\sum_{i=1}^n \beta_{ki}\alpha_{ij}\right)\boldsymbol{v}_i\end{aligned}$$

By the uniqueness of coordinates, we imply

$$\left(\sum_{i=1}^n \beta_{ki}\alpha_{ij}\right) = \begin{cases}1, & j = k \\ 0, & j \neq k\end{cases}$$

where

$$\left(\sum_{i=1}^n \beta_{ki}\alpha_{ij}\right) = (C_{AA'}C_{A'A}).$$

Therefore, $(C_{AA'}C_{A'A}) = \boldsymbol{I}_n$. ■

■ **Example 3.4** Back to Example (3.3), write $\mathcal{B}_1, \mathcal{B}_2$ as

$$\mathcal{B}_1 = \{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}, \quad \mathcal{B}_2 = \{\boldsymbol{w}_1, \dots, \boldsymbol{w}_n\}$$

and therefore $\boldsymbol{w}_i = \boldsymbol{e}_1 + \cdots + \boldsymbol{e}_i$. The change of basis matrix is given by

$$C_{\mathcal{B}_1, \mathcal{B}_2} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 0 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

which implies that for $\boldsymbol{v}$ in the example,

$$C_{\mathcal{B}_1, \mathcal{B}_2}[\boldsymbol{v}]_{\mathcal{B}_2} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 0 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \begin{pmatrix} \alpha_1 - \alpha_2 \\ \vdots \\ \alpha_{n-1} - \alpha_n \\ \alpha_n \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} = [\boldsymbol{v}]_{\mathcal{B}_1}$$

■

Definition 3.2 Let $T : V \to W$ be a linear transformation, and

$$\mathcal{A} = \{\boldsymbol{v}_1, \dots, \boldsymbol{v}_m\}, \quad \mathcal{B} = \{\boldsymbol{w}_1, \dots, \boldsymbol{w}_m\}$$

be basis of V and W, respectively. The **matrix representation** of T with respect to (w.r.t.) $\mathcal{A}$ and $\mathcal{B}$ is defined as $(T)_{\mathcal{B}\mathcal{A}} := (\alpha_{ij}) \in M_{m\times m}(\mathbb{F})$, where

$$T(\boldsymbol{v}_j) = \sum_{i=1}^{m} \alpha_{ij} \boldsymbol{w}_i$$

■

3.4. Wednesday for MAT3040

3.4.1. Remarks for the Change of Basis

Reviewing.

- $[\cdot]_{\mathcal{A}} : V \to \mathbb{F}^n$ denotes coordinate vector mapping
- Change of Basis matrix: $C_{\mathcal{A}',\mathcal{A}}$
- $T : V \to W$, $\mathcal{A} = \{\boldsymbol{v}_1,\dots,\boldsymbol{v}_n\}$ and $\boldsymbol{B} = \{\boldsymbol{w}_1,\dots,\boldsymbol{w}_m\}$.

  $\mathrm{Hom}_{\mathbb{F}}(V,W) \to M_{m\times n}(\mathbb{F})$

▪ **Example 3.10** Let $V = \mathbb{P}_3[x]$ and $\mathcal{A} = \{1,x,x^2,x^3\}$.

Let $T : V \to V$ defined as $p(x) \mapsto p'(x)$:

$$\begin{cases} T(1) = 0\cdot 1 + 0\cdot x + 0\cdot x^2 + 0\cdot x^3 \\ T(x) = 1\cdot 1 + 0\cdot x + 0\cdot x^2 + 0\cdot x^3 \\ T(x^2) = 0\cdot 1 + 2\cdot x + 0\cdot x^2 + 0\cdot x^3 \\ T(x^3) = 0\cdot 1 + 0\cdot x + 3\cdot x^2 + 0\cdot x^3 \end{cases}$$

We can define the change of basis matrix for a linear transformation T as well, w.r.t. $\mathcal{A}$ and $\mathcal{A}$:

$$C_{\mathcal{A},\mathcal{A}} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Also, we can define a different basis $\mathcal{A}' = \{x^3,x^2,x,1\}$ for the output space for T, say $T : V_{\mathcal{A}} \to V_{\mathcal{A}'}$:

$$(T)_{\mathcal{A},\mathcal{A}'} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 2 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

Our observation is that the corresponding coordinate vectors before and after linear transformation admits a matrix multiplication:

$$(2x^2 + 4x^3) \xrightarrow{T} ((4x + 12x^2))$$

$$(2x^2 + 4x^3)_{\mathcal{A}} = \begin{pmatrix} 0 \\ 0 \\ 2 \\ 4 \end{pmatrix} \qquad (4x + 12x^2)_{\mathcal{A}} = \begin{pmatrix} 0 \\ 4 \\ 12 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 2 \\ 4 \end{pmatrix} = \begin{pmatrix} 0 \\ 4 \\ 12 \\ 0 \end{pmatrix}$$

$$C_{\mathcal{A}\mathcal{A}} \cdot (2x^2 + 4x^3)_{\mathcal{A}} = (4x + 12x^2)_{\mathcal{A}}$$

Theorem 3.3 — Matrix Representation. Let $T : V \to W$ be a linear transformation of finite dimensional vector sapces. Let $\mathcal{A}, \mathcal{B}$ the ordered basis of V, W, respectively. Then the following diagram holds:

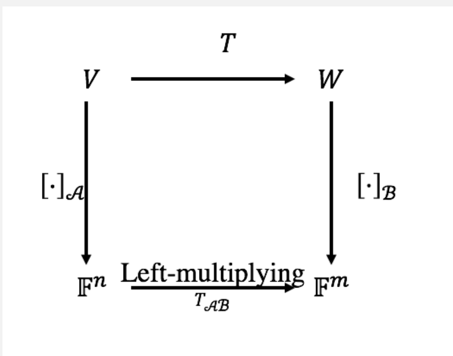


Figure 3.2: Diagram for the matrix reprentation, where $n := \dim(V)$ and $m := \dim(W)$

namely, for any $\boldsymbol{v} \in V$,

$$(T)_{\mathcal{B},\mathcal{A}}(\boldsymbol{v})_{\mathcal{A}} = (T\boldsymbol{v})_{\mathcal{B}}$$

Therefore, we can compute $T\boldsymbol{v}$ by matrix multiplication.

Therefore, linear transformation corresponds to coordinate matrix multiplication.

Proof. Suppose $\mathcal{A} = \{\boldsymbol{v}_1, \dots, \boldsymbol{v}_n\}$ and $\mathcal{B} = \{\boldsymbol{w}_1, \dots, \boldsymbol{w}_n\}$. The proof of this theorem follows the same procedure of that in Theorem (3.1)

1. We show this result for $\boldsymbol{v} = \boldsymbol{v}_j$ first:

$$\text{LHS} = [\alpha_{ij}]\boldsymbol{e}_j = \begin{pmatrix} \alpha_{1j} \\ \vdots \\ \alpha_{nj} \end{pmatrix}$$

$$\text{RHS} = (T\boldsymbol{v}_j)_{\mathcal{B}} = \left(\sum_{i=1}^{m} \alpha_{ij}\boldsymbol{w}_i\right)_{\mathcal{B}} = \begin{pmatrix} \alpha_{1j} \\ \vdots \\ \alpha_{nj} \end{pmatrix}$$

2. Then we show the theorem holds for any $\boldsymbol{v} := \sum_{j=1}^{n} r_j \boldsymbol{v}_j$ in V:

$$\begin{aligned}
(T)_{\mathcal{B}\mathcal{A}}(\boldsymbol{v})_{\mathcal{A}} &= (T)_{\mathcal{B}\mathcal{A}}\left(\sum_{j=1}^{n} r_j \boldsymbol{v}_j\right)_{\mathcal{A}} && (3.8a)\\
&= (T)_{\mathcal{B}\mathcal{A}}\left(\sum_{j=1}^{n} r_j (\boldsymbol{v}_j)_{\mathcal{A}}\right) && (3.8b)\\
&= \sum_{j=1}^{n} r_j (T)_{\mathcal{B}\mathcal{A}}(\boldsymbol{v}_j)_{\mathcal{A}} && (3.8c)\\
&= \sum_{j=1}^{n} r_j (T\boldsymbol{v}_j)_{\mathcal{B}} && (3.8d)\\
&= \left(\sum_{j=1}^{n} r_j (T\boldsymbol{v}_j)\right)_{\mathcal{B}} && (3.8e)\\
&= \left[T(\sum_{j=1}^{n} r_j \boldsymbol{v}_j)\right]_{\mathcal{B}} && (3.8f)\\
&= (T\boldsymbol{v})_{\mathcal{B}} && (3.8g)
\end{aligned}$$

The justification for (3.8a) is similar to that shown in Theorem (3.1). The proof is complete.

■

R Consider a special case for Theorem (3.3), i.e., $T = \text{id}$ and $\mathcal{A}, \mathcal{A}'$ are two ordered basis for the input and output space, respectively. Then the result in Theorem (3.3) implies

$$C_{\mathcal{A}',\mathcal{A}}(\boldsymbol{v})_{\mathcal{A}} = (\boldsymbol{v})_{\mathcal{A}'}$$

i.e., the matrix representation theorem (3.3) is a general case for the change of basis theorem (3.1)

Proposition 3.6 — Functoriality. Suppose V, W, U are finite dimensional vector spaces, and let $\mathcal{A}, \mathcal{B}, C$ be the ordered basis for V, W, U, respectively. Suppose that

$$T : V \to W, \quad S : W \to U$$

are given two linear transformations, then

$$(S \circ T)_{C,\mathcal{A}} = (S)_{C,\mathcal{B}}(T)_{\mathcal{B},\mathcal{A}}$$

Composition of linear transformation corresponds to the multiplication of change of basis matrices.

Proof. Suppose the ordered basis $\mathcal{A} = \{\boldsymbol{v}_1, \dots, \boldsymbol{v}_n\}$, $\mathcal{B} = \{\boldsymbol{w}_1, \dots, \boldsymbol{w}_m\}$, $C = \{\boldsymbol{u}_1, \dots, \boldsymbol{u}_p\}$. By defintion of change of basis matrices,

$$T(\boldsymbol{v}_j) = \sum_i (T_{\mathcal{B},\mathcal{A}})_{ij} \boldsymbol{w}_i$$

$$S(\boldsymbol{w}_i) = \sum_k (S_{C,\mathcal{B}})_{ki} \boldsymbol{u}_k$$

We start from the j-th column of $(S \circ T)_{C,\mathcal{A}}$ for $j = 1, \dots, n$, namely

$$
\begin{aligned}
(S \circ T)_{C,\mathcal{A}}(\boldsymbol{v}_j)_{\mathcal{A}} &= (S \circ T(\boldsymbol{v}_j))_C &\text{(3.9a)}\\
&= \left[S \circ \left(\sum_i (T_{\mathcal{B},\mathcal{A}})_{ij}\boldsymbol{w}_i\right)\right]_C &\text{(3.9b)}\\
&= \sum_i (T_{\mathcal{B},\mathcal{A}})_{ij}\,(S(\boldsymbol{w}_i))_C &\text{(3.9c)}\\
&= \sum_i (T_{\mathcal{B},\mathcal{A}})_{ij}\left(\sum_k (S_{C,\mathcal{B}})_{ki}\boldsymbol{u}_k\right)_C &\text{(3.9d)}\\
&= \sum_k \sum_i (S_{C,\mathcal{B}})_{ki}(T_{\mathcal{B},\mathcal{A}})_{ij}(\boldsymbol{u}_k)_C &\text{(3.9e)}\\
&= \sum_k (S_{C,\mathcal{B}}T_{\mathcal{B},\mathcal{A}})_{kj}(\boldsymbol{u}_k)_C &\text{(3.9f)}\\
&= \sum_k (S_{C,\mathcal{B}}T_{\mathcal{B},\mathcal{A}})_{kj}\boldsymbol{e}_k &\text{(3.9g)}\\
&= j\text{-th column of } [S_{C\mathcal{B}}T_{\mathcal{B},\mathcal{A}}] &\text{(3.9h)}
\end{aligned}
$$

where (3.9a) is by the result in theorem (3.3); (3.9b) and (3.9d) follows from definitions of $T(\boldsymbol{v}_j)$ and $S(\boldsymbol{w}_i)$; (3.9c) and (3.9e) follows from the linearity of C; (3.9f) follows from the matrix multiplication definition; (3.9g) is because $(\boldsymbol{u}_k)_C = \boldsymbol{e}_k$.

Therefore, $(S \circ T)_{C\mathcal{A}}$ and $(S_{C,\mathcal{B}})(T_{\mathcal{B},\mathcal{A}})$ share the same j-th column, and thus equal to each other. ∎

Corollary 3.2 Suppose that S and T are two identity mappings $V \to V$, and consider $(S)_{\mathcal{A}'\mathcal{A}}$ and $(T)_{\mathcal{A},\mathcal{A}'}$ in proposition (3.6), then

$$(S \circ T)_{\mathcal{A}',\mathcal{A}'} = (S)_{\mathcal{A}'\mathcal{A}}(T)_{\mathcal{A},\mathcal{A}'}$$

Therefore,

$$\text{Identity matrix} = C_{\mathcal{A}',\mathcal{A}}C_{\mathcal{A},\mathcal{A}'}$$

Proposition 3.7 Let $T : V \to W$ with $\dim(V) = n, \dim(W) = m$, and let

- $\mathcal{A}, \mathcal{A}'$ be ordered basis of V

- $\mathcal{B}, \mathcal{B}'$ be ordered basis of W

then the change of basis matrices admit the relation

$$(T)_{\mathcal{B}',\mathcal{A}'} = C_{\mathcal{B}',\mathcal{B}}(T)_{\mathcal{B},\mathcal{A}}C_{\mathcal{A}\mathcal{A}'} \tag{3.10}$$

Here note that $(T)_{\mathcal{B}',\mathcal{A}'}, (T)_{\mathcal{B},\mathcal{A}} \in \mathbb{F}^{m\times n}$; $C_{\mathcal{B}',\mathcal{B}} \in \mathbb{F}^{m\times m}$; and $C_{\mathcal{A}\mathcal{A}'} \in \mathbb{F}^{n\times n}$.

Proof. Let $\mathcal{A} = \{\boldsymbol{v}_1, \dots, \boldsymbol{v}_n\}, \mathcal{A}' = \{\boldsymbol{v}'_1, \dots, \boldsymbol{v}'_n\}$. Consider simplifying the j-th column for the LHS and RHS of (3.10) and showing they are equal:

$$\begin{aligned}
\text{LHS} &= (T)_{\mathcal{B}',\mathcal{A}'}\boldsymbol{e}_j \\
&= (T)_{\mathcal{B}',\mathcal{A}'}(\boldsymbol{v}'_j)_{\mathcal{A}'} \\
&= (T\boldsymbol{v}'_j)_{\mathcal{B}'}
\end{aligned}$$

$$\begin{aligned}
\text{RHS} &= C_{\mathcal{B}',\mathcal{B}}(T)_{\mathcal{B},\mathcal{A}}C_{\mathcal{A}\mathcal{A}'}\boldsymbol{e}_j \\
&= C_{\mathcal{B}',\mathcal{B}}(T)_{\mathcal{B},\mathcal{A}}C_{\mathcal{A}\mathcal{A}'}(\boldsymbol{v}'_j)_{\mathcal{A}'} \\
&= C_{\mathcal{B}',\mathcal{B}}(T)_{\mathcal{B},\mathcal{A}}(\boldsymbol{v}'_j)_{\mathcal{A}} \\
&= C_{\mathcal{B}',\mathcal{B}}(T\boldsymbol{v}'_j)_{\mathcal{B}} \\
&= (T\boldsymbol{v}'_j)_{\mathcal{B}'}
\end{aligned}$$

∎

R Let $T : V \to V$ be a linear operator with $\mathcal{A}, \mathcal{A}'$ being two ordered basisof V, then

$$(T)_{\mathcal{A}'\mathcal{A}'} = C_{\mathcal{A}',\mathcal{A}}(T)_{\mathcal{A}\mathcal{A}}C_{\mathcal{A},\mathcal{A}'} = (C_{\mathcal{A},\mathcal{A}'})^{-1}(T)_{\mathcal{A}\mathcal{A}}C_{\mathcal{A},\mathcal{A}'}$$

Therefore, the change of basis matrices $(T)_{\mathcal{A}'\mathcal{A}'}$ and $(T)_{\mathcal{A}\mathcal{A}}$ are similar to each other, which means they share the same eigenvalues, determinant, trace.

Therefore, two similar matrices cooresponds to same linear transformation using different basis.

Chapter 4

Week4

4.1. Monday for MAT3040

4.1.1. Quotient Spaces

Now we aim to divide a big **vector space** into many pieces of slices.

- For example, the Cartesian plane can be expressed as union of set of vertical lines as follows:

$$\mathbb{R}^2 = \bigcup_{m\in\mathbb{R}} \left\{ \begin{pmatrix} m \\ 0 \end{pmatrix} + \operatorname{span}\{(0,1)\} \right\}$$

- Another example is that the set of integers can be expressed as union of three sets:

$$\mathbb{Z} = Z_1 \cup Z_2 \cup Z_3,$$

where Z_i is the set of integers z such that $z \bmod 3 = i$.

Definition 4.1 [Coset] Let V be a vector space and $W \leq V$. For any element $\boldsymbol{v} \in V$, the **(right) coset** determined by $\boldsymbol{v}$ is the set

$$\boldsymbol{v} + W := \{\boldsymbol{v} + \boldsymbol{w} \mid \boldsymbol{w} \in W\}$$

For example, consider $V = \mathbb{R}^3$ and $W = \operatorname{span}\{(1,2,0)\}$. Then the coset determined by

$\boldsymbol{v} = (5,6,-3)$ can be written as

$$\boldsymbol{v} + W = \{(5+t, 6+2t, -3) \mid t \in \mathbb{R}\}$$

It's interesting that the coset determined by $\boldsymbol{v}' = \{(4,4,-3)\}$ is exactly the same as the coset shown above:

$$\boldsymbol{v}' + W = \{(4+t, 4+2t, -3) \mid t \in \mathbb{R}\} = \boldsymbol{v} + W.$$

Therefore, write the exact expression of $\boldsymbol{v} + W$ may sometimes become tedious and hard to check the equivalence. We say $\boldsymbol{v}$ is a **representative** of a coset $\boldsymbol{v} + W$.

Proposition 4.1 Two cosets are the same iff the subtraction for the corresponding representatives is in W, i.e.,

$$\boldsymbol{v}_1 + W = \boldsymbol{v}_2 + W \iff \boldsymbol{v}_1 - \boldsymbol{v}_2 \in W$$

Proof. *Necessity.* Suppose that $\boldsymbol{v}_1 + W = \boldsymbol{v}_2 + W$, then $\boldsymbol{v}_1 + \boldsymbol{w}_1 = \boldsymbol{v}_2 + \boldsymbol{w}_2$ for some $\boldsymbol{w}_1, \boldsymbol{w}_2 \in W$, which implies

$$\boldsymbol{v}_1 - \boldsymbol{v}_2 = \boldsymbol{w}_2 - \boldsymbol{w}_1 \in W$$

Sufficiency. Suppose that $\boldsymbol{v}_1 - \boldsymbol{v}_2 = \boldsymbol{w} \in W$. It suffices to show $\boldsymbol{v}_1 + W \subseteq \boldsymbol{v}_2 + W$. For any $\boldsymbol{v}_1 + \boldsymbol{w}' \in \boldsymbol{v}_1 + W$, this element can be expressed as

$$\boldsymbol{v}_1 + \boldsymbol{w}' = (\boldsymbol{v}_2 + \boldsymbol{w}) + \boldsymbol{w}' = \boldsymbol{v}_2 + \underbrace{(\boldsymbol{w} + \boldsymbol{w}')}_{\text{belong to } W} \in \boldsymbol{v}_2 + W.$$

Therefore, $\boldsymbol{v}_1 + W \subseteq \boldsymbol{v}_2 + W$. Similarly we can show that $\boldsymbol{v}_2 + W \subseteq \boldsymbol{v}_1 + W$. ∎

Exercise: Two cosets with representatives $\boldsymbol{v}_1, \boldsymbol{v}_2$ have no intersection iff $\boldsymbol{v}_1 - \boldsymbol{v}_2 \notin W$.

Definition 4.2 [Quotient Space] The **quotient space** of V by the subspace W, is the collection of all cosets $\boldsymbol{v} + W$, denoted by V/W.

To make the quotient space a vector space structure, we define the addition and scalar multiplication on V/W by:

$$(\boldsymbol{v}_1 + W) + (\boldsymbol{v}_2 + W) := (\boldsymbol{v}_1 + \boldsymbol{v}_2) + W$$
$$\alpha \cdot (\boldsymbol{v} + W) := (\alpha \cdot \boldsymbol{v}) + W$$

For example, consider $V = \mathbb{R}^2$ and $W = \operatorname{span}\{(0,1)\}$. Then note that

$$\left(\begin{pmatrix}1\\0\end{pmatrix} + W\right) + \left(\begin{pmatrix}2\\0\end{pmatrix} + W\right) = \left(\begin{pmatrix}3\\0\end{pmatrix} + W\right)$$
$$\pi \cdot \left(\begin{pmatrix}1\\0\end{pmatrix} + W\right) = \left(\begin{pmatrix}\pi\\0\end{pmatrix} + W\right)$$

Proposition 4.2 The addition and scalar multiplication is well-defined.

Proof.

1. Suppose that
$$\begin{cases} \boldsymbol{v}_1 + W = \boldsymbol{v}_1' + W \\ \boldsymbol{v}_2 + W = \boldsymbol{v}_2' + W \end{cases}, \tag{4.1}$$
and we need to show that $(\boldsymbol{v}_1 + \boldsymbol{v}_2) + W = (\boldsymbol{v}_1' + \boldsymbol{v}_2') + W$.
From (4.1) and proposition (4.1), we imply
$$\boldsymbol{v}_1 - \boldsymbol{v}_1' \in W, \quad \boldsymbol{v}_2 - \boldsymbol{v}_2' \in W$$
which implies
$$(\boldsymbol{v}_1 - \boldsymbol{v}_1') + (\boldsymbol{v}_2 - \boldsymbol{v}_2') = (\boldsymbol{v}_1 + \boldsymbol{v}_2) - (\boldsymbol{v}_1' + \boldsymbol{v}_2') \in W$$
By proposition (4.1) again we imply $(\boldsymbol{v}_1 + \boldsymbol{v}_2) + W = (\boldsymbol{v}_1' + \boldsymbol{v}_2') + W$
2. For scalar multiplication, similarly, we can show that $\boldsymbol{v}_1 + W = \boldsymbol{v}_1' + W$ implies $\alpha\boldsymbol{v}_1 + W = \alpha\boldsymbol{v}_1' + W$ for all $\alpha \in \mathbb{F}$.

■

Proposition 4.3 The canonical projection mapping

$$\begin{aligned} \pi_W : V &\to V/W, \\ \boldsymbol{v} &\mapsto \boldsymbol{v} + W, \end{aligned}$$

is a **surjective linear transformation** with $\ker(\pi_W) = W$.

Proof. 1. First we show that $\ker(\pi_W) = W$:

$$\pi_W(\boldsymbol{v}) = 0 \implies \boldsymbol{v} + W = \mathbf{0}_{V/W} \implies \boldsymbol{v} + W = \mathbf{0} + W \implies \boldsymbol{v} = (\boldsymbol{v} - \mathbf{0}) \in W$$

Here note that the zero element in the quotient space V/W is the coset with representative $\mathbf{0}$.

2. For any $\boldsymbol{v}_0 + W \in V/W$, we can construct $\boldsymbol{v}_0 \in V$ such that $\pi_W(\boldsymbol{v}_0) = \boldsymbol{v}_0 + W$. Therefore the mapping π_W is surjective.
3. To show the mapping π_W is a linear transformation, note that

$$\begin{aligned} \pi_W(\alpha \boldsymbol{v}_1 + \beta \boldsymbol{v}_2) &= (\alpha \boldsymbol{v}_1 + \beta \boldsymbol{v}_2) + W \\ &= (\alpha \boldsymbol{v}_1 + W) + (\beta \boldsymbol{v}_2 + W) \\ &= \alpha(\boldsymbol{v}_1 + W) + \beta(\boldsymbol{v}_2 + W) \\ &= \alpha \pi_W(\boldsymbol{v}_1) + \beta \pi_W(\boldsymbol{v}_2) \end{aligned}$$

■

4.1.2. First Isomorphism Theorem

The key of linear algebra is to solve the linear system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ with $\boldsymbol{A} \in \mathbb{R}^{m \times n}$. The general step for solving this linear system is as follows:

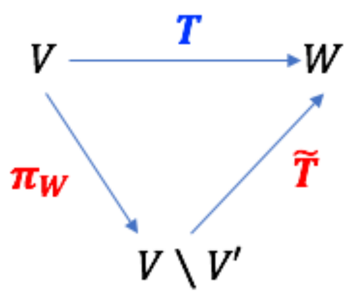
1. Find the solution set for $\boldsymbol{A}\boldsymbol{x} = \mathbf{0}$, i.e., the set $\ker(\boldsymbol{A})$
2. Find a particular solution $\boldsymbol{x}_0$ such that $\boldsymbol{A}\boldsymbol{x}_0 = \boldsymbol{b}$.

Then the general solution set to this linear system is $\boldsymbol{x}_0 + \ker(\boldsymbol{A})$, which is a coset in the space $\mathbb{R}^n/\ker(\boldsymbol{A})$. Therefore, to solve the linear system $\boldsymbol{Ax} = \boldsymbol{b}$ suffices to study the quotient space $\mathbb{R}^n/\ker(\boldsymbol{A})$:

Proposition 4.4 — Universal Property I. Suppose that $T : V \to W$ is a linear transformation, and that $V' \leq \ker(T)$. Then the mapping

$$\begin{aligned} \tilde{T} : V/V' &\to W \\ \boldsymbol{v} + V' &\mapsto T(\boldsymbol{v}) \end{aligned}$$

is a well-defined linear transformation. As a result, the diagram below commutes:



In other words, we have $T = \tilde{T} \circ \pi_W$.

Proof. First we show the well-definedness. Suppose that $\boldsymbol{v}_1 + V' = \boldsymbol{v}_2 + V'$ and suffices to show $\tilde{T}(\boldsymbol{v}_1 + V') = \tilde{T}(\boldsymbol{v}_2 + V')$, i.e., $T(\boldsymbol{v}_1) = T(\boldsymbol{v}_2)$. By proposition (4.1), we imply

$$\boldsymbol{v}_1 - \boldsymbol{v}_2 \in V' \leq \ker(T) \implies T(\boldsymbol{v}_1 - \boldsymbol{v}_2) = \boldsymbol{0} \implies T(\boldsymbol{v}_1) - T(\boldsymbol{v}_2) = \boldsymbol{0}.$$

Then we show $\tilde{(T)}$ is a linear transformation:

$$\begin{aligned} \tilde{T}(\alpha(\boldsymbol{v}_1 + V') + \beta(\boldsymbol{v}_2 + V')) &= \tilde{T}((\alpha\boldsymbol{v}_1 + \beta\boldsymbol{v}_2) + V') \\ &= T(\alpha\boldsymbol{v}_1 + \beta\boldsymbol{v}_2) \\ &= \alpha T(\boldsymbol{v}_1) + \beta T(\boldsymbol{v}_2) \\ &= \alpha\tilde{T}(\boldsymbol{v}_1 + V') + \beta\tilde{T}(\boldsymbol{v}_2 + V') \end{aligned}$$

∎

Actually, if we let $V' = \ker(T)$, the mapping $\tilde{T} : V/V' \to T(V)$ forms an isomorphism, In particular, if further T is surjective, then $T(V) = W$, i.e., the mapping $\tilde{T} : V/V' \to W$ forms an isomorphism.

Theorem 4.1 — **First Isomorphism Theorem.** Let $T : V \to W$ be a surjective linear transformation. Then the mapping

$$\begin{aligned} \tilde{T} : V/\ker(T) &\to W \\ \boldsymbol{v} + \ker(T) &\mapsto T(\boldsymbol{v}) \end{aligned}$$

is an isomorphism.

Proof. *Injectivity.* Suppose that $\tilde{T}(\boldsymbol{v}_1 + \ker(T)) = \tilde{T}(\boldsymbol{v}_2 + \ker(T))$, then we imply

$$T(\boldsymbol{v}_1) = T(\boldsymbol{v}_2) \implies T(\boldsymbol{v}_1 - \boldsymbol{v}_2) = \mathbf{0}_W \implies \boldsymbol{v}_1 - \boldsymbol{v}_2 \in \ker(T),$$

i.e., $\boldsymbol{v}_1 + \ker(T) = \boldsymbol{v}_2 + \ker(T)$.

Surjectivity. For $\boldsymbol{w} \in W$, due to the surjectivity of T, we can find a $\boldsymbol{v}_0$ such that $T(\boldsymbol{v}_0) = \boldsymbol{w}$. Therefore, we can construct a set $\boldsymbol{v}_0 + \ker(T)$ such that

$$\tilde{T}(\boldsymbol{v}_0 + \ker(T)) = \boldsymbol{w}.$$

∎

4.4. Wednesday for MAT3040

Reviewing.

- Quotient Space:

$$V/W = \{\boldsymbol{v} + W \mid \boldsymbol{v} \in V\}$$

  The elements in V/W are cosets. Note that V/W does not mean a subset of V.

- Define the canonical projection mapping

$$\begin{aligned} &\pi_W : V \to V/W, \\ \text{with} \quad &\boldsymbol{v} \mapsto \boldsymbol{v} + W, \end{aligned}$$

  then we imply π_W is a surjective linear transformation with $\ker(\pi_W) = W$.

  If $\dim(V) < \infty$, then by Rank-Nullity Theorem (2.3), we imply that

$$\dim(V) = \dim(W) + \dim(V/W),$$

  i.e., $\dim(V/W) = \dim(V) - \dim(W)$.

- **(Universal Property I)** Every linear transformation $T : V \to W$ with $V' \le \ker(T)$ can be descended to the composition of the canonical projection mapping $\pi_{V'}$ and the mapping

$$\begin{aligned} &\tilde{T} : V/V' \to W \\ \text{with} \quad &\boldsymbol{v} + V' \mapsto T(\boldsymbol{v}). \end{aligned}$$

  In other words, the diagram (2.1) commutes:

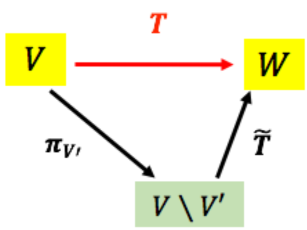


Diagram (2.1)

In other words, the mapping starting from either the black or red line gives the same result, i.e., $T(\boldsymbol{v}) = \tilde{T} \circ \pi_{V'}(\boldsymbol{v}) = \tilde{T}(\boldsymbol{v} + V')$ for any $\boldsymbol{v} \in V$.

- **(First Isomorphism Theorem)** Under the setting of Universal Property I (UPI), if T is a surjective linear transformation with $V' = \ker(T)$, then the $\tilde{T}$ is an isomorphism.

■ **Example 4.2** Suppose that $U, W \le V$ with $U \cap W = \{\mathbf{0}\}$, then define the mapping

$$\begin{aligned} \phi &: U \oplus W \to U \\ \text{with} \quad & \phi(\boldsymbol{u} + \boldsymbol{w}) = \boldsymbol{u} \end{aligned}$$

R Exercise: if $U, W \le V$ but $U \cap W \ne \{\mathbf{0}\}$, then the mapping

$$\begin{aligned} \phi &: U + W \to U \\ \text{with} \quad & \boldsymbol{u} + \boldsymbol{w} \mapsto \boldsymbol{u} \end{aligned} \qquad \text{is not well-defined:}$$

Suppose that $\mathbf{0} \ne \boldsymbol{v} \in U \cap W$ and for any $\boldsymbol{u} \in U, \boldsymbol{w} \in W$, we construct

$$\boldsymbol{u}' = \boldsymbol{u} - \boldsymbol{v} \in U, \quad \boldsymbol{w}' = \boldsymbol{w} + \boldsymbol{v} \in V \quad \Longrightarrow \phi(\boldsymbol{u}' + \boldsymbol{w}') = \boldsymbol{u} - \boldsymbol{v}$$

Therefore we get $\boldsymbol{u} + \boldsymbol{w} = \boldsymbol{u}' + \boldsymbol{w}'$ but $\phi(\boldsymbol{u} + \boldsymbol{w}) \ne \phi(\boldsymbol{u}' + \boldsymbol{w}')$.

Back to the situation $U \cap W = \{\mathbf{0}\}$, then it's clear that $\phi : U \oplus W \to U$ is surjective linear transformation with $\ker(\phi) = W$. Therefore, construct the new mapping

$$\begin{aligned} \tilde{\phi} &: U \oplus W / W \to U \\ \text{with} \quad & \boldsymbol{u} + \boldsymbol{w} + W \mapsto \phi(\boldsymbol{u} + \boldsymbol{w}) \end{aligned}$$

We imply $\tilde{\phi}$ is an isomorphism by First Isomorphism Theorem. ■

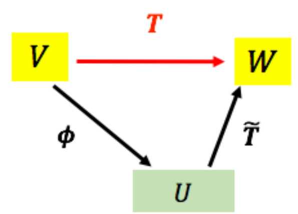
Now we study the generalized quotients, which is defined to satisfy the generalized version of universal property I.

Definition 4.7 [Universal Property for Quotients] Let V be a vector space and $V' \leq V$. Consider the collection of linear transformations

$$\text{Obj} = \left\{ T : V \to W \middle| \begin{array}{l} T \text{ is a linear transformation} \\ V' \leq \ker(T) \end{array} \right\}$$

(For example, $\pi_{V'} : V \to V/V'$ is an element from the set Obj.)

An element $(\phi : V \to U) \in \text{Obj}$ is said to satisfy the **universal property** if it satisfies the following:

> Given any element $(T : V \to W) \in \text{Obj}$, we can extend the transformation ϕ with a **uniquely existing** $\tilde{T} : U \to W$ so that the diagram (2.2) commutes:
>
> 
>
>
> Diagram (2.2)
>
> Or equivalently, for given $(T : V \to W) \in \text{Obj}$, there exists the **unique** mapping $\tilde{T} : U \to W$ such that $T = \tilde{T} \circ \phi$.

Theorem 4.3 — **Universal Property II.**

1. The mapping $(\pi_{V'} : V \to V/V') \in \text{Obj}$ is a universal object, i.e., it satisfies the universal property.
2. If $(\phi : V \to U)$ is a universal object, then $U \cong V/V'$, i.e., there is intrinsically "one" element in the set of universal objects.

Proof.

1. Consider any linear transformation $T : V \to W$ such that $V' \leq \ker(T)$, then define (construct) the same $\tilde{T} : V/V' \to W$ as that in UPI. Therefore, for given T, applying the result of UPI, we imply $T = \tilde{T} \circ \pi_{V'}$, i.e., $\pi_{V'}$ satisfies the diagram (2.2).

To show the uniqueness of $\tilde{T}$, suppose there exists $\tilde{S} : V/V' \to W$ such that the diagram (2.3) commutes.

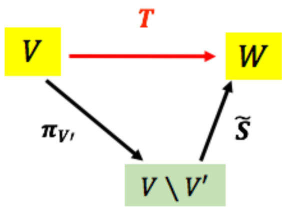


Diagram (2.3)

It suffices to show the mapping $\tilde{S} = \tilde{T}$: for any $\boldsymbol{v} + V' \in V/V'$, we have

$$\tilde{S}(\boldsymbol{v} + V') := \tilde{S} \circ \pi_{V'}(\boldsymbol{v}) = T(\boldsymbol{v}),$$

where the first equality is due to the surjectivity of $\pi_{V'}$. By the result of UPI, $T(\boldsymbol{v}) = \tilde{T}(\boldsymbol{v} + V')$. Therefore $\tilde{T}(\boldsymbol{v} + V') = \tilde{S}(\boldsymbol{v} + V')$ for all $\boldsymbol{v} + V' \in V/V'$. The proof is complete.

2. Suppose that $(\phi : V \to U)$ satisfies the universal property. In particular, the following two diagrams hold:

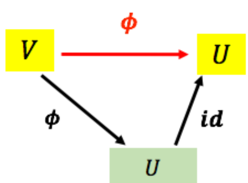


Diagram (2.4)

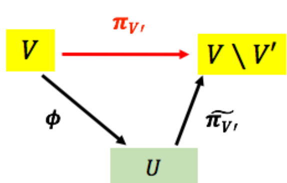


Diagram (2.5)

Since $(\pi_{V'})$ satisfies the universal property, in particular, the following two diagrams hold:

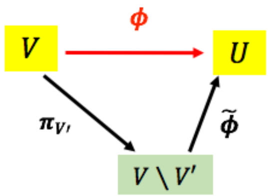


Diagram (2.6)

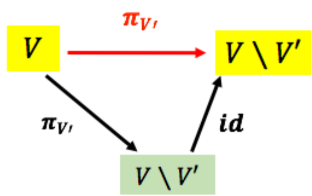


Diagram (2.7)

Then we claim that: Combining Diagram (2.5) and (2.6), we imply the diagram (2.8):

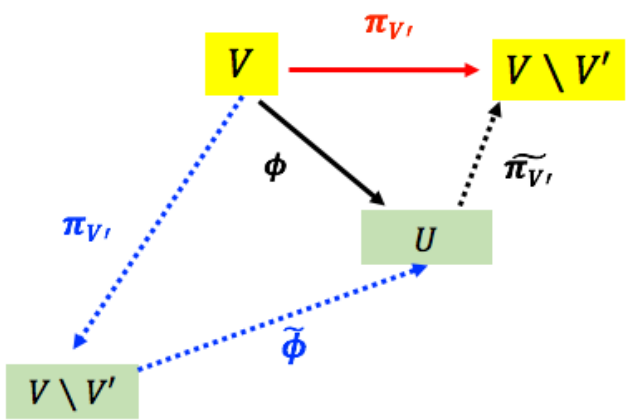


Diagram (2.8)

Graph Description: Note that this diagram commutes, i.e., the mapping starting from either the red line or the dash line gives the same result, i.e., $\pi_{V'} = \tilde{\pi}_{V'} \circ \tilde{\phi} \circ \pi_{V'}$. Comparing Diagram (2.7) and Diagram (2.8), we have $\tilde{\pi}_{V'} \circ \tilde{\phi} = id$, by the **uniqueness** of the universal object.

Therefore, $\tilde{\pi}_{V'} \circ \tilde{\phi} = id$ implies $\tilde{\pi}_{V'}$ is surjective and $\tilde{\phi}$ is injective.

Also, combining Diagram (2.6) and (2.5), we imply diagram (2.9):

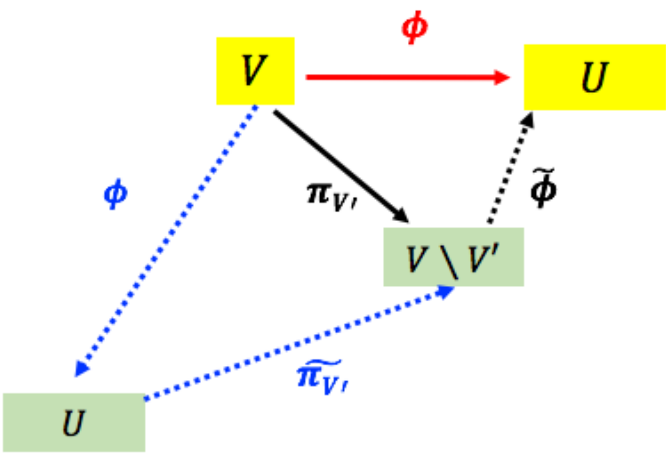


Diagram (2.9)

Graph Description: Note that this diagram commutes, i.e., the mapping starting from either the red line or the dash line gives the same result, i.e., $\phi = \tilde{\phi} \circ \tilde{\pi}_{V'} \circ \phi$. Comparing Diagram (2.9) and Diagram (2.4), we have $\tilde{\phi} \circ \tilde{\pi}_{V'} = id$, by the **uniqueness** of the universal object

Therefore, $\tilde{\phi} \circ \tilde{\pi}_{V'} = id$ implies $\tilde{\phi}$ is surjective and $\tilde{\pi}_{V'}$ is injective.

Therefore, both $\tilde{\phi} : U \to V/V'$ and $\tilde{\pi}_{V'} : V/V' \to U$ are bijective, i.e., $U \cong V/V'$. The proof is complete.

■

4.4.1. Dual Space

Definition 4.8 Let V be a vector space over a field $\mathbb{F}$. The **dual vector space** V^* is defined as

$$\begin{aligned} V^* &= \mathrm{Hom}_{\mathbb{F}}(V, \mathbb{F}) \\ &= \{ f : V \to \mathbb{F} \mid f \text{ is a linear transformation} \} \end{aligned}$$

■ **Example 4.3** 1. Consider $V = \mathbb{R}^n$ and define $\phi_i : V \to \mathbb{R}$ as the i-th component of input:

$$\phi_i \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = x_i,$$

Then we imply $\phi_i \in V^*$. On the contrary, $\phi_i^2 \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = x_i^2$ is not in V^*

2. Consider $V = \mathbb{F}[x]$ and define $\phi : V \to \mathbb{F}$ as:

$$\phi(p(x)) = p(1),$$

It's clear that $\phi \in V^*$:

$$\begin{aligned} \phi(ap(x) + bq(x)) &= ap(1) + bq(1) \\ &= a\phi(p(x)) + b\phi(q(x)) \end{aligned}$$

3. Also, $\psi : V \to \mathbb{F}$ by $\psi(p(x)) = \int_0^1 p(x)\,\mathrm{d}x$ is in V^*.
4. Also, for $V = M_{n\times n}(\mathbb{F})$, the mapping $\mathrm{tr} : V \to \mathbb{F}$ by $\mathrm{tr}(M) = \sum_{i=1}^n M_{ii}$ is in V^*. However, the $\det : V \to \mathbb{F}$ is not in V^*

■

Definition 4.9 Let V be a vector space, with basis $B = \{v_i \mid i \in I\}$ (I can be finite or countable, or uncountable). Define

$$B^* = \{f_i : V \to \mathbb{F} \mid i \in I\},$$

where f_i's are defined on the basis B:

$$f_i(v_j) = \delta_{ij} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases}$$

Then we extend f_i's linearly, i.e., for $\sum_{j=1}^{N} \alpha_j v_j \in V$,

$$f_i(\sum_{j=1}^{N} \alpha_j v_j) = \sum_{i=1}^{N} \alpha_j f_i(v_j).$$

It's clear that $f_i \in V^*$ is well-defined.

Our question is that whether the B^* can be the basis of V^*?

Chapter 5

Week5

5.1. Monday for MAT3040

Reviewing.

- Dual space: the set of linear transformations from V to $\mathbb{F}$, denoted as $\mathrm{Hom}(V,\mathbb{F})$.
- Suppose $B = \{\boldsymbol{v}_i \mid i \in I\}$ is the basis of V, define $B^* = \{f_i \mid i \in I\}$ by

$$f_i(\boldsymbol{v}_j) = \delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$$

  Actually, the above recipe uniquely defines a linear transformation $f_i : V \to \mathbb{F}$: For any $\boldsymbol{v} \in V$, it can be written as $\boldsymbol{v} = \sum_{i\in I} \alpha_i \boldsymbol{v}_i$, and therefore

$$f_i(\boldsymbol{v}) = f_i(\sum_{i\in I} \alpha_i \boldsymbol{v}_i) = \sum_{i\in I} \alpha_i f_i(\boldsymbol{v}_i).$$

■ **Example 5.1** Consider $V = \mathbb{R}^n, B = \{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$. Then we imply $B^* = \{\phi_i\}_{i=1}^n$, where ϕ_i is the mapping $V \to \mathbb{R}$ defined by

$$\phi_i \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \phi(x_1\boldsymbol{e}_1 + \cdots + x_n\boldsymbol{e}_n) = \sum_{j=1}^n x_j \phi_i(\boldsymbol{e}_j) = x_i$$

■

5.1.1. Remarks on Dual Space

Proposition 5.1

1. B^* is always lienarly independent, i.e., any finite subset of B^* is linearly independent.
2. If V has finite dimension, then B^* is a basis of V^*.

Proof.

1. Suppose that
$$\alpha_1 f_{i_1} + \alpha_2 f_{i_2} + \cdots + \alpha_k f_{i_k} = \mathbf{0}_{V^*}.$$
In particular, let the input of these linear transformations be $\boldsymbol{v}_{i_1}$, we imply
$$\begin{aligned}\alpha_1 f_{i_1}(\boldsymbol{v}_{i_1}) + \alpha_2 f_{i_2}(\boldsymbol{v}_{i_1}) + \cdots + \alpha_k f_{i_k}(\boldsymbol{v}_{i_1}) = \mathbf{0}(\boldsymbol{v}_{i_1}) \equiv \mathbf{0} \\ = \alpha_1 \cdot 1 + \cdots + 0 \\ = \alpha_1\end{aligned}$$
Applying the same trick, one can show that $\alpha_2 = \cdots = \alpha_k = 0$. Therefore, $\{f_{i_1}, \ldots, f_{i_k}\}$ is linearly independent.
2. Suppose that $B = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ and $B^* = \{f_1, \ldots, f_n\}$. For any $f \in V^*$, construct the linear transformation
$$g := \sum_{i=1}^{n} f(\boldsymbol{v}_i) \cdot f_i \in \text{span}\{B^*\}.$$
It follows that for $j = 1, 2, \ldots, n$,
$$g(\boldsymbol{v}_j) = \sum_{i=1}^{n} f(\boldsymbol{v}_i) \cdot f_i(\boldsymbol{v}_j) = f(\boldsymbol{v}_j).$$
It's clear that $g(\boldsymbol{v}) = f(\boldsymbol{v})$ for all $\boldsymbol{v} \in V$, i.e., $f \equiv g \in \text{span}(B^*)$. Therefore B^* spans V^*, i.e., forms a basis of V^*.

■

Corollary 5.1 If $\dim(V) = n$, then $\dim(V^*) = n$.

Proof. It's eay to show the mapping defined as

$$\begin{aligned} & V \to V^* \\ \text{with} \quad & \boldsymbol{v}_i \mapsto f_i \end{aligned}$$

is an isomorphism from $V \to V^*$. Note that this constructed isomorphism depends on **the choice of basis** B in V. (We say this is not a **natural isomorphism**.) ■

(R) The part 2 for proposition (5.1) does not hold for V with infinite dimension. The reason is that the spanning set is defined with **finite** linear combinations. Check the example below for a counter-example.

■ **Example 5.2** Suppose that $V = \mathbb{F}[x]$, and $B^* = \{1, x, x^2, \dots,\}$ forms a basis of V. We imply that $B^* = \{\phi_0, \phi_1, \phi_2, \dots,\}$, where ϕ_i is the mapping defined as

$$\phi_i(x^j) = \begin{cases} 1, & i = j \\ 0, & \text{otherwise} \end{cases}$$

Consider a special element $\phi \in V^*$ with $f(p(x)) = p(1)$:

$$\phi(1) = 1, \quad \phi(x) = 1, \quad \phi(x^2) = 1, \quad \cdots \quad \phi(x^n) = 1, \quad \forall n \in \mathbb{N}.$$

If following the proof in proposition (5.1), we expect that

$$g := \sum_{n=0}^{\infty} \phi(x^n)\phi_n = \sum_{n=0}^{\infty} \phi_n \in \text{span}\{B^*\},$$

which is a contradiction, since $\text{span}\{B^*\}$ consists of finite sum of ϕ_i's only. ■

(R) Therefore, if V is not finite-dimensional, we can say the cardinality of V is strictly less than the cardinality of V^*.

Any subspace of a given vector space has some gap. Now we want to describe this gap formally from the perspective of the dual space.

5.1.2. Annihilators

Definition 5.1 Let V be a vector space, $S \subseteq V$ be a subset. The **annihilator** of S is defined as

$$\text{Ann}(S) = \{f \in V^* \mid f(s) = 0, \forall s \in S\}$$

■

■ **Example 5.3** Consider $V = \mathbb{R}^4$, $B = \{\boldsymbol{e}_1, \dots, \boldsymbol{e}_4\}$. Let $B^* = \{f_1, \dots, f_4\}$, $S = \{\boldsymbol{e}_3, \boldsymbol{e}_4\}$.

- Then $f_1 \in \text{Ann}(S)$, since

$$f_1(\boldsymbol{e}_3) = 0, \quad f_1(\boldsymbol{e}_4) = 0$$

Indeed, any $a \cdot f_1 + b \cdot f_2 \in V^*$ is in $\text{Ann}(S)$. ■

Proposition 5.2 1. The set $\text{Ann}(S)$ is a vector subspace of V^*

2. The mapping $\text{Ann}(\cdot)$ is **inclusion-reversing**, i.e., if $W_1 \subseteq W_2 \subseteq V$, then

$$\text{Ann}(W_1) \supseteq \text{Ann}(W_2)$$

3. The mapping $\text{Ann}(\cdot)$ is **idempotent**, i.e., $\text{Ann}(S) = \text{Ann}(\text{span}(S))$.
4. If V has finite dimension, and $W \leq V$, then $\text{Ann}(W)$ fills in the gap, i.e.,

$$\dim(W) + \dim(\text{Ann}(W)) = \dim(V)$$

Proof. 1. Suppose that $f, g \in \text{Ann}(S)$, i.e., $f(s) = g(s) = 0, \forall s \in S$. It's clear that $(af + bg) \in \text{Ann}(S)$.

2. Suppose that $f \in \text{Ann}(W_2)$, we imply $f(\boldsymbol{w}) = 0$ for any $\boldsymbol{w} \in W_2$. Therefore, $f(\boldsymbol{w}_1) = 0$ for any $\boldsymbol{w}_1 \in W_1 \subseteq W_2$, i.e., $f \in \text{Ann}(W_1)$.
3. Note that $S \subseteq \text{span}(S)$. Therefore we imply $\text{Ann}(S) \supseteq \text{Ann}(\text{span}(S))$ from part (b). It suffices to show $\text{Ann}(S) \subseteq \text{Ann}(\text{span}(S))$:

For any $f \in \text{Ann}(S)$ and any $\sum_{i=1}^{n} k_i \boldsymbol{s}_i \in \text{span}(S)$, we imply

$$\begin{aligned} f\left(\sum_{i=1}^{n} k_i \boldsymbol{s}_i\right) &= \sum_{i=1}^{n} k_i f(\boldsymbol{s}_i) \\ &= \sum_{i=1}^{n} k_i \cdot 0 \\ &= 0, \end{aligned}$$

i.e., $f \in \text{Ann}(\text{span}(S))$.

4. Let $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_k\}$ be a basis of W. By basis extension, we construct a basis of V:

$$B = \{\boldsymbol{v}_1, \dots, \boldsymbol{v}_k, \boldsymbol{v}_{k+1}, \dots, \boldsymbol{v}_n\}.$$

Let $B^* = \{f_1, \dots, f_k, f_{k+1}, \dots, f_n\}$ be a basis of V^*. We claim that $\{f_{k+1}, \dots, f_n\}$ is a basis of $\text{Ann}(W)$:

- Firstly, f_j's are the elements in $\text{Ann}(W)$ for $j = k+1, \dots, n$, since for any $\boldsymbol{w} = \sum_{i=1}^{k} \alpha_i(\boldsymbol{v}_i) \in W$, we have

$$\begin{aligned} f_j(\boldsymbol{w}) &= \sum_{i=1}^{k} \alpha_i f_j(\boldsymbol{v}_i) \\ &= \sum_{i=1}^{k} \alpha_i \cdot 0 \\ &= 0, \quad j = k+1, k+2, \dots, n \end{aligned}$$

- Secondly, the set $\{f_{k+1}, \dots, f_n\}$ is linearly independent, since the set $B^* = \{f_1, \dots, f_n\}$ is linearly independent.

- Thirdly, $\{f_{k+1}, \dots, f_n\}$ spans $\text{Ann}(W)$: for any $g \in \text{Ann}(W) \subseteq V^*$, it can be expressed as $g = \sum_{i=1}^{n} \beta_i f_i$. It follows that

$$g(\mathbf{v}_1) = \sum_{i=1}^{n} \beta_i f_i(\mathbf{v}_1) = 0 \implies \beta_1 = 0$$
$$\vdots$$
$$g(\mathbf{v}_k) = \sum_{i=1}^{n} \beta_i f_i(\mathbf{v}_k) = 0 \implies \beta_k = 0$$

Substituting $\beta_1 = \cdots = \beta_k = 0$ into $g = \sum_{i=1}^{n} \beta_i f_i$, we imply

$$g = \beta_{k+1} f_{k+1} + \cdots + \beta_n f_n \in \operatorname{span}\{f_{k+1}, \dots, f_n\}.$$

Therefore, $\{f_{k+1}, \dots, f_n\}$ forms a basis for $\operatorname{Ann}(W)$, i.e., $\dim(\operatorname{Ann}(W)) = n - k$.

■

R Let $W \le V$, where V has finite dimension, recall that we have obtained two relations below:

$$\dim(\operatorname{Ann}(W)) = \dim(V) - \dim(W)$$
$$\dim((V/W)^*) = \dim(V/W) = \dim(V) - \dim(W)$$

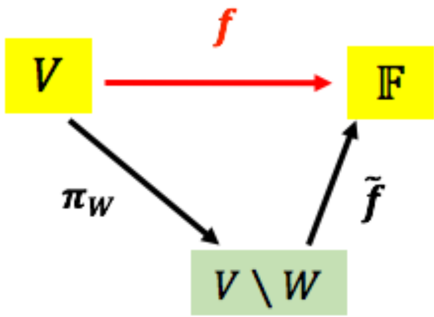
Therefore, $\dim((V/W)^*) = \dim(\operatorname{Ann}(W))$, i.e.,

$$(V/W)^* \cong \operatorname{Ann}(W).$$

The question is that can we construct an isomorphism explicitly? We claim that the mapping defined below is an isomorphism:

$$\begin{aligned} &\operatorname{Ann}(W) \to (V/W)^* \\ \text{with} \quad &f \mapsto \tilde{f}, \end{aligned}$$

where $\tilde{f} : V/W \to \mathbb{F}$ is constructed from the **universal property I**, i.e., given the mapping $f \in \mathrm{Ann}(W)$, since $W \leq \ker(f)$, there exists $\tilde{f} : V/W \to \mathbb{F}$ such that the diagram below commutes:



i.e., $\tilde{f}(\boldsymbol{v} + W) = f(\boldsymbol{v})$.

5.4. Wednesday for MAT3040

There will be a quiz on next Monday.

$$\text{Scope}: \quad \text{From Week 1 up to (including) the definition of } B^*.$$

Reviewing.

1. If V is finite dimensional, and B a basis of V, then B^* is a basis of the dual space V^*.
2. Define the Annihilator $\text{Ann}(S) \le V^*$:
$$\text{Ann}(S) = \{f \in V^* \mid f(s) = 0, \forall s \in S\}$$
3. If V is finite dimensional, and $W \le V$, then $\text{Ann}(W)$ fills the gap, i.e.,
$$\dim(\text{Ann}(W)) = \dim(V) - \dim(W)$$
4. Define a map
$$\begin{aligned} \Phi: \quad & \text{Ann}(W) \to (V/W)^* \\ & f \mapsto \tilde{f} \end{aligned}$$
where $\tilde{f}$ is defined such that the diagram (5.1) below commutes

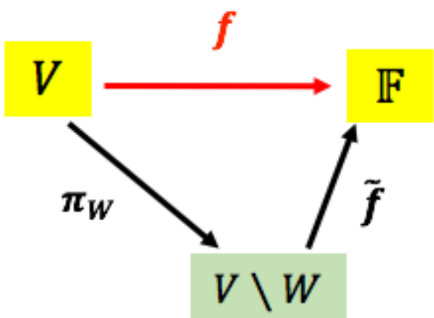


Figure 5.1: Construction of $\tilde{f}$

Or equivalently, $\tilde{f} : V/W \to \mathbb{F}$ is such that $\tilde{f}(\boldsymbol{v} + W) = f(\boldsymbol{v})$.

5.4.1. Adjoint Map

The natural question is that whether Φ is the isomorphism between $\text{Ann}(W)$ and $(V/W)^*$:

Proposition 5.4 Φ is a linear transformation, i.e.,

$$\Phi(af+bg) = a\cdot\Phi(f) + b\cdot\Phi(g).$$

Proof. Itt suffices to show that

$$\overline{af+bg} = a\overline{f} + b\overline{g}$$

■

Therefore, we need to answer whether Φ a bijective map. We will show this conjucture at the end of this lecture. The definition of Φ is **natural**, i.e., we do not need to specify any basis to define this Φ . However, as studied in Monday, the constructed isomorphism $V \to V^*$ with $\boldsymbol{v}_i \mapsto f_i$ is not natural.

Definition 5.3 [Adjoint Map] Let $T: V \to W$ be a linear transformation. Define the **adjoint** of T by

$$T^*: \quad W^* \to V^*$$

such that for any $f \in W^*$,

$$[T^*(f)](\boldsymbol{v}) := f(T(\boldsymbol{v})), \ \forall \boldsymbol{v} \in V.$$

■

R

1. In other words, $T^*(f) = f \circ T$, i.e., a linear transformation from V to $\mathbb{F}$, i.e., belongs to V^*.
2. Moreover, the mapping T^* itself is a linear transformation: For $f, g \in W^*$,

and $\forall \boldsymbol{v} \in V$,

$$
\begin{aligned}
[T^*(af+bg)](\boldsymbol{v}) &= (af+bg)[T(\boldsymbol{v})] \\
&= af(T(\boldsymbol{v})) + bg(T(\boldsymbol{v})) && \text{definition of } W^* \text{ as a vector space} \\
&= a[T^*(f)](\boldsymbol{v}) + b[T^*(g)](\boldsymbol{v}) \\
&= [aT^*(f) + bT^*(g)](\boldsymbol{v}) && \text{definition of } V^* \text{ as a vector space}
\end{aligned}
$$

Proposition 5.5 Let $T : V \to W$ be a linear transformation.

1. If T is **injective**, then T^* is **surjective**.
2. If T is **surjetive**, then T^* is **injective**.

This statement is quite intuitive, since T^* reverses the dual of output into the dual of input:

$$T : V \to W$$

$$T^* : W^* \to V^*$$

Proof. We only give a proof of (2), i.e., suffices to show $\ker(T) = \{\mathbf{0}\}$.

Consider any $g \in W^*$ such that $T^*(g) = \mathbf{0}_{V^*}$. It follows that

$$[T^*(g)](\boldsymbol{v}) = \mathbf{0}_{V^*}(\boldsymbol{v}), \quad \forall \boldsymbol{v} \in V. \Longleftrightarrow g(T(\boldsymbol{v})) = \mathbf{0}, \quad \forall \boldsymbol{v} \in V. \tag{5.4}$$

To show $g = \mathbf{0}_{W^*}$, it suffices to show $g(\boldsymbol{w}) = \mathbf{0}$ for $\forall \boldsymbol{w} \in W$. For all $\boldsymbol{w} \in W$, by the surjectivity of T, there exists $\boldsymbol{v}' \in V$ such that

$$\boldsymbol{w} = T(\boldsymbol{v}').$$

By substituting $\boldsymbol{w}$ with $T(\boldsymbol{v}')$ and (5.4), we imply

$$g(\boldsymbol{w}) = g(T(\boldsymbol{v}')) = \mathbf{0}.$$

The proof is complete. ■

Proposition 5.6 Let $T : V \to W$ be a linear transformation, and $\mathcal{A} = \{\boldsymbol{v}_1, \dots, \boldsymbol{v}_n\}, \mathcal{B} = \{\boldsymbol{w}_1, \dots, \boldsymbol{w}_m\}$ be the bases of V and W, respectively. Let $\mathcal{A}^* = \{f_1, \dots, f_n\}, \mathcal{B}^* = \{g_1, \dots, g_m\}$

be bases of dual spaces V^* and W^*, respectively. Then $T^* : W^* \to V^*$ admits a matrix representation

$$(T^*)_{\mathcal{A}^*\mathcal{B}^*} = \text{transpose}((T)_{\mathcal{B}\mathcal{A}})$$

where $(T^*)_{\mathcal{A}^*\mathcal{B}^*} \in \mathbb{F}^{n\times m}$ and $(T)_{\mathcal{B}\mathcal{A}} \in \mathbb{F}^{m\times n}$

Proof. Let $(T)_{\mathcal{B}\mathcal{A}} = (\alpha_{ij})$ and $(T^*)_{\mathcal{A}^*\mathcal{B}^*} = (\beta_{ij})$. By definition of matrix representation,

$$T(\boldsymbol{v}_j) = \sum_{i=1}^{m} \alpha_{ij}\boldsymbol{w}_i, \qquad T^*(g_i) = \sum_{k=1}^{n} \beta_{ki} f_k \in V^*$$

As a result,

$$\begin{aligned}
[T^*(g_i)](\boldsymbol{v}_j) &= g_i(T(\boldsymbol{v}_j)) \\
&= g_i\left(\sum_{\ell=1}^{m} \alpha_{\ell j}\boldsymbol{w}_\ell\right) \\
&= \sum_{\ell=1}^{m} \alpha_{\ell j} g_i(\boldsymbol{w}_\ell) \\
&= \alpha_{ij}
\end{aligned}$$

and

$$\begin{aligned}
[T^*(g_i)](\boldsymbol{v}_j) &= \left(\sum_{k=1}^{n} \beta_{ki} f_k\right)(\boldsymbol{v}_j) \\
&= \sum_{k=1}^{n} \beta_{ki} f_k(\boldsymbol{v}_j) \\
&= \beta_{ji}
\end{aligned}$$

Therefore, $\beta_{ji} = \alpha_{ij}$. The proof is complete. ∎

5.4.2. Relationship between Annihilator and dual of quotient spaces

■ **Example 5.5** Consider the canonical projection mapping $\pi_W : V \to V/W$ with its **adjoint** mapping:

$$(\pi_W)^* : (V/W)^* \to V^*$$

The understanding of $(\pi_W)^*$ is as follows:

1. Take $h \in (V/W)^*$ and study $(\pi_W)^*(h) \in V^*$
2. Take $\boldsymbol{v} \in V$ and understand

$$[(\pi_W)^*(h)](\boldsymbol{v}) = h(\pi_W(\boldsymbol{v})) = h(\boldsymbol{v} + W)$$

(a) In particular, for all $\boldsymbol{w} \in W \le V$, we have

$$[(\pi_W)^*(h)](\boldsymbol{w}) = h(\boldsymbol{w} + W) = h(\mathbf{0}_{V/W}) = \mathbf{0}_{\mathbb{F}}$$

Therefore,

$$(\pi_W)^*(h) \in \mathrm{Ann}(W).$$

i.e., $(\pi_W)^*$ is a mapping from $(V/W)^*$ to $\mathrm{Ann}(W)$.

(b) By proposition (5.5), π_W is surjective implies $(\pi_W)^*$ is injective.

Combining (a) and (b), it's clear that (i.e., left as homework problem)

$$\Phi \circ \pi_W^* = \mathrm{id}_{(V/W)^*} \text{ and } \pi_W^* \circ \Phi = \mathrm{id}_{\mathrm{Ann}(W)}$$

This relationship implies Φ is an isomorphism. ■

Chapter 6

Week6

6.1. Monday for MAT3040

6.1.1. Polynomials

We recall some useful properties of polynomial before studying eigenvalues/eigenvectors.

Definition 6.1 [Polynomial]

1. A polynomial over $\mathbb{F}$ has the form

$$p(z) = a_m z^m + \cdots + a_1 z + a_0, \quad (a_m \neq 0).$$

   Here $a_m z^m$ is called the **leading term** of $p(z)$; m is called the degree; a_m is called the **leading coefficient**; $a_m, \cdots, a_0$ are called the coefficients of this polynomial.
2. A polynomial over $\mathbb{F}$ is monic if its leading coefficient is $1_{\mathbb{F}}$.
3. A polynomial $p(z) \in \mathbb{F}[z]$ is **irreducible** if for any $a(z), b(z) \in \mathbb{F}[z]$,

$$p(z) = a(z)b(z) \implies \text{either } a(z) \text{ or } b(z) \text{ is a constant polynomial.}$$

   Otherwise $p(z)$ is **reducible**.

■ **Example 6.1** For example, the polynomial $p(x) = x^2 + 1$ is irreducible over $\mathbb{R}$; but $p(x) = (x - i)(x + i)$ is **reducible** over $\mathbb{C}$. ■

Theorem 6.1 — **Division Theorem.** For all $p, q \in \mathbb{F}[z]$ such that $p \neq 0$, there exists unique $s, r \in \mathbb{F}[x]$ satisfying $\deg(r) < \deg(f)$, such that

$$p(z) = s(z) \cdot q(z) + r(z).$$

Here $r(z)$ is called the **remainder**.

■ **Example 6.2** Given $p(x) = x^4 + 1$ and $q(x) = x^2 + 1$, the junior school knowledge tells us that uniquely

$$x^4 + 1 = (x^2 - 1)(x^2 + 1) + 2.$$

■

Theorem 6.2 — **Root Theorem.** For $p(x) \in \mathbb{F}[x]$, and $\lambda \in \mathbb{F}$, $x - \lambda$ divides p if and only if $p(\lambda) = 0$.

Proof.

1. If $(x - \lambda)$ divides p, then $p = (x - \lambda)q$ for some $q \in \mathbb{F}[x]$. Thus clearly $p(\lambda) = 0$.
2. For the other direction, suppose that $p(\lambda) = 0$. By division theorem, there exists $s, r \in \mathbb{F}[x]$ such that

$$p = (x - \lambda)s + r \quad \text{with } \deg(r) < \deg(x - \lambda) = 1. \tag{6.1}$$

Therefore, the polynomial r must be constant.

Substituting λ into x both sides in (6.1), we have

$$0 = p(\lambda) = 0 \cdot s + r \implies r = 0.$$

Therefore, $p = (x - \lambda) \cdot s$, i.e., $(x - \lambda)$ divides p.

■

6.4. Wednesday for MAT3040

Reviewing: Root Theorem: $p(\lambda) = 0$ iff $(x - \lambda)$ divdes $p(x)$.

Corollary 6.2 A polynomial with degree n has at most n roots counting multiplicity.

For example, the polynomial $(x-3)^2$ has one root $x = 3$ with multiplicity 2. When counting multiplicity, we say the polynomial $(x-3)^2$ has two roots.

Definition 6.5 [Algebraically Closed] A field $\mathbb{F}$ is called **algebraically closed** if every non-constant polynomial $p(x) \in \mathbb{F}[x]$ has a root $\lambda \in \mathbb{F}$.

Theorem 6.5 — **Fundamental Theroem of Algebra.** The set of complex numbers $\mathbb{C}$ is algebraically closed.

Proof. One way is by complex analysis; Another way is by the topology on $\mathbb{C} \setminus \{0\}$. ■

(R) By induction, we can show that every polynomial with degree n on algebraically closed field $\mathbb{F}$ has **exactly** n roots, counting multiplicity. Therefore, for any $p(x)$ on algebraically closed field $\mathbb{F}$,

$$p(x) = c(x - \lambda_1) \cdots (x - \lambda_n) \tag{6.3}$$

for $c, \lambda_1, \dots, \lambda_n \in \mathbb{F}$.

The polynomials on general field $\mathbb{F}$ may not necessarily be factorized as in (6.3) , but still admit unique factorization property:

Theorem 6.6 — **Unique Factorization.** Every $f(x) = a_n x^n + \cdots + a_0$ in $\mathbb{F}[x]$ can be factorized as

$$f(x) = a_n [p_1(x)]^{e_1} \cdots [p_k(x)]^{e_k}$$

where p_i's are **monic, irreducible,distinct**. Furthermore, this expression is unique up to the permutation of factors.

Definition 6.6 [Factor] If $p(x) = q(x)s(x)$ with $p, q, s \in \mathbb{F}[x]$, then we say

- $p(x)$ is **divisible** by $s(x)$;
- $s(x)$ is a **factor** of $p(x)$;
- $s(x) | p(x)$
- $s(x)$ **divides** $p(x)$
- $p(x)$ is **multiple** of $s(x)$

Definition 6.7 [Common Factor]

1. The polynomial $g(x)$ is said to be a **common factor** of $f_1, \dots, f_k \in \mathbb{F}[x]$ if

$$g | f_i, i = 1, \dots, k$$

2. The polynomial $g(x)$ is said to be a **greatest common divisor** of $f_1, \dots, f_k$ if
   - g is **monic**.
   - g is common factor of $f_1, \dots, f_k$
   - g is of largest possible (maximal) degree.

R

- $\gcd(f_1, \dots, f_k) = \gcd(\gcd(f_1, f_2), f_3, \dots, f_k) = \gcd(\gcd(f_1, f_2, f_3), \dots, f_k)$
- $\gcd(f_1, \dots, f_k)$ is unique.
- If $\gcd(f_1, \dots, f_k) = 1$, we say $f_1, \dots, f_k$ is **relatively prime**
- Polynomials $f_1, \dots, f_k$ are relatively prime does not necessarily mean $\gcd(f_i, f_j) = 1$ for any $i \neq j$.

Counter-example: Let $a_1, \ldots, a_n$ distinct irreducible polynomials, and

$$f_i(x) = a_1(x) \cdots \hat{a}_i(x) \cdots a_n(x) := a_1 \cdots a_{i-1} a_{i+1} \cdots a_n,$$

then $\gcd(f_1, \ldots, f_n) = 1$, but $\gcd(f_i, f_j) = a_1 \cdots \hat{a}_i \cdots \hat{a}_j \cdots a_n$, which does not necessarily equal to 1.

■ **Example 6.6** The $\gcd(f_1, f_2)$ is easy to compute for factorized polynomials. For example, let $f_1(x) = (x^2 + x + 1)^3 (x - 3)^2 x^4$ and $f_2(x) = (x^2 + 1)(x - 3)^4 x^2$ in $\mathbb{R}[x]$, then

$$\gcd(f_1, f_2) = (x - 3)^2 x^2$$

■

The question is how to find $\gcd(f_1, f_2)$ for given un-factorized polynomials?

Theorem 6.7 — Bezout. Let $g = \gcd(f_1, f_2)$, then there exists $r_1, r_2 \in \mathbb{F}[x]$ such that

$$g(x) = r_1(x) f_1(x) + r_2(x) f_2(x)$$

More generally, $g = \gcd(f_1, \ldots, f_k)$ implies there exists $r_1, \ldots, r_k$ such that

$$g = r_1 f_1 + \cdots + r_k f_k$$

The derivation of r_i's is by applying **Euclidean algorithm**. For example, given $x^3 + 6x + 7$ and $x^2 + 3x + 2$, we imply

$$x^3 + 6x + 7 - (x - 3)(x^2 + 3x + 2) = 13x + 13$$

and

$$x^2 + 3x + 2 - \frac{x + 2}{13}(13x + 13) = 0$$

Therefore, $\gcd(x^3 + 6x + 7, x^2 + 3x + 2) = \gcd(x^2 + 3x + 2, 13x + 13) = x + 2$.

6.4.1. Eigenvalues & Eigenvectors

Definition 6.8 [Eigenvalues] Let $T : V \to V$ be a linear operator.

1. We say $\boldsymbol{v} \in V \setminus \{\mathbf{0}\}$ is an eigenvector of T with eigenvalue λ if $T(\boldsymbol{v}) = \lambda \boldsymbol{v}$;
2. Or equivalently, $\boldsymbol{v} \in \ker(T - \lambda I)$, the λ-eigenspace of T. Here the mapping $I : V \to V$ denotes identity map, i.e., $I(\boldsymbol{v}) = \boldsymbol{v}, \forall \boldsymbol{v} \in V$.

Definition 6.9 A vector $\boldsymbol{v} \in V \setminus \{\mathbf{0}\}$ is a **generalized eigenvector** of T with **generalized eigenvalue** λ if $\boldsymbol{v} \in \ker((T - \lambda I)^k)$ for some $k \in \mathbb{N}^+$.

Note that an eigenvector is a generalized eigenvector of T; while the converse does not necessarily hold.

Example 6.7 Consider the linear transformation $A : \mathbb{R}^2 \to \mathbb{R}^2$ with

$$
\begin{aligned}
A: &\quad \mathbb{R}^2 \to \mathbb{R}^2 \\
\text{with} &\quad \boldsymbol{x} \to \boldsymbol{A}\boldsymbol{x} \\
\text{where} &\quad \boldsymbol{A} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}
\end{aligned}
$$

1. Note that $[1,0]^{\mathrm{T}}$ is an eigenvector with eigenvalue 1, since

$$A\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

2. However, $[0,1]^{\mathrm{T}}$ is not an eigenvector, since

$$A\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Note that

$$(A-I)^2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad (A-I)^3 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

and therefore

$$\begin{pmatrix} 0 \\ 1 \end{pmatrix} \in \ker(A-I)^2,$$

i.e., a generalized eigenvector with generalized eigenvalue 1.

■

■ **Example 6.8** Consider $V = C^\infty(\mathbb{R})$, which is a set of all infinitely differentiable functions. Define the linear operator $T : V \to V$ as $T(f) = f''$. Then the (-1)-eigenspace of T has $f \in V$ satisfying

$$f'' = -f$$

From ODE course, we imply $\{\sin x, \cos x\}$ forms a basis of (-1)-eigenspace. ■

Assumption. From now on, we assume V has finite dimension by default.

Definition 6.10 [Determinant] Let $T : V \to V$ be a linear operator. The **determinant** of T is given by

$$\det(T) = \det((T)_{\mathcal{A},\mathcal{A}})$$

where $\mathcal{A}$ is some basis of V. ■

R Assume we have complete knowledge about $\det(M)$ for matrices for now. The determinant is well-defined, i.e., independent of the choice of basis $\mathcal{A}$. For another basis $\mathcal{B}$, we imply

$$\det(T_{\mathcal{B},\mathcal{B}}) = \det(C_{\mathcal{B},\mathcal{A}} T_{\mathcal{A},\mathcal{A}} C_{\mathcal{A},\mathcal{B}}) = \det(C_{\mathcal{B},\mathcal{A}})\det(T_{\mathcal{A},\mathcal{A}})\det(C_{\mathcal{A},\mathcal{B}}) = \det(T_{\mathcal{A},\mathcal{A}})$$

Definition 6.11 [characteristic polynomial] The **characteristic polynomial** $\mathcal{X}_T(x)$ of $T : V \to V$ is defined as

$$\mathcal{X}_T(x) = \det((T)_{\mathcal{A},\mathcal{A}} - xI)$$

for any basis $\mathcal{A}$

In the next few lectures, we will study

- Cayley-Hamilton Theorem
- Jordan Canonical Form

These theorems can be stated using matrices, and they both hold up to change of basis. We have a unified statement of these theorem using vecotor space rather than $\mathbb{R}^n$.

Chapter 7

Week7

7.1. Monday for MAT3040

Reviewing. Define the characteristic polynomial for an linear operator T:

$$\mathcal{X}_T(x) = \det((T)_{\mathcal{A},\mathcal{A}} - x\boldsymbol{I})$$

We will use the notation "$I/\boldsymbol{I}$" in two different occasions:

1. I denotes the identity transformation from V to V with $I(\boldsymbol{v}) = \boldsymbol{v}, \forall \boldsymbol{v} \in V$
2. $\boldsymbol{I}$ denotes the identity matrix $(I)_{\mathcal{A},\mathcal{A}}$, defined based on any basis $\mathcal{A}$.

7.1.1. Minimal Polynomial

Definition 7.1 [Linear Operator Induced From Polynomial] Let $f(x) := a_m x^m + \cdots + a_0$ be a polynomial in $\mathbb{F}[x]$, and $T : V \to V$ be a linear operator. Then the mapping

$$f(T) = a_m T^m + \cdots + a_1 T + a_0 I : \quad V \to V,$$

is called a linear operator induced from the polynomial $f(x)$. ■

R

1. The composition of linear operators is not abelian, e.g., in general $S \circ T = T \circ S$ does not hold. The reason follows similarly from the fact that square-matrix multiplication is not abelian in general.

2. However, we always have $f(T)T = Tf(T)$, where $f(T)$ is a linear operator induced from the polynomial $f(x)$:

   Proof. We can show that $T^nT = TT^n, \forall n$ by induction. Suppose that $f(x) = \sum_i a_i x^i$, which follows that

   $$f(T)T = \sum_i a_i T^i T = \sum_i a_i T T^i = T \sum_i a_i T^i = Tf(T).$$

   ■

3. We can generalize the statement in (2) into the fact that the composition of linear operators induced from polynomials is abelian, i.e.,

   $$f(T)g(T) = g(T)f(T)$$

   for any polynomials $f(x), g(x)$.

Definition 7.2 [Minimal Polynomial] Let $T : V \to V$ be a linear operator. The **minimal polynomial** $m_T(x)$ is a **nonzero monic polynomial** of least (minimal) degree such that

$$m_T(T) = \mathbf{0}_{V\to V}.$$

where $\mathbf{0}_{V\to V}$ denotes the zero vector in $\text{Hom}(V,V)$.

■ **Example 7.1** 1. Let $\boldsymbol{A} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, then $\boldsymbol{A}$ defines a linear operator:

$$\begin{aligned} A: \quad & \mathbb{F}^2 \to \mathbb{F}^2 \\ \text{with} \quad & \boldsymbol{x} \mapsto \boldsymbol{A}\boldsymbol{x} \end{aligned}$$

Here $\mathcal{X}_A(x) = (x-1)^2$ and $\boldsymbol{A} - \boldsymbol{I} = \mathbf{0}$, which gives $m_A(x) = x - 1$.

2. Let $\boldsymbol{B} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, which implies

$$\mathcal{X}_B(x) = (x-1)^2,$$

The question is that can we get the minimal polynomial with degree 1?

The answer is no, since $\boldsymbol{B} - k\boldsymbol{I} = \begin{pmatrix} 1-k & 1 \\ 0 & 1-k \end{pmatrix} \neq \boldsymbol{0}$.

In fact, $m_B(x) = (x-1)^2$, since

$$(\boldsymbol{B}-\boldsymbol{I})^2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}^2 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

■

Two questions naturally arises:

1. Does $m_T(x)$ exist? If exists, is it unique?
2. What's the relationship between $m_T(x)$ and $\mathcal{X}_T(x)$?

Regarding to the first question, the minimal polynomial $m_T(x)$ may not exist, if V has infinite dimension:

■ **Example 7.2** Consider $V = \mathbb{R}[x]$ and the mapping

$$\begin{aligned} T: \quad & V \to V \\ & p(x) \mapsto \int_0^x p(t)\,\mathrm{d}t \end{aligned}$$

In particular, $T(x^n) = \frac{1}{n+1}x^{n+1}$. Suppose $m_T(x)$ is with degree n, i.e.,

$$m_T(x) = x^n + \cdots + a_1 x + a_0,$$

then

$$m_T(T) = T^n + \cdots + a_0 I \text{ is a zero linear transformation}$$

It follows that

$$[m_T(T)](x) = \frac{1}{n!}x^n + a_{n-1}\frac{1}{(n-1)!}x^{n-1} + \cdots + a_1 x + a_0 = 0_{\mathbb{F}},$$

which is a contradiction since the coefficients of x^k is nonzero on LHS for $k = 1, \dots, n$, but zero on the RHS.

Proposition 7.1 The minimal polynomial $m_T(x)$ always exists for $\dim(V) = n < \infty$.

Proof. It's clear that $\{I, T, \dots, T^n, T^{n+1}, \cdots, T^{n^2}\} \subseteq \text{Hom}(V, V)$. Since $\dim(\text{Hom}(V, V)) = n^2$, we imply $\{I, T, \dots, T^n, T^{n+1}, \cdots, T^{n^2}\}$ is linearly dependent, i.e., there exists a_i's that are not all zero such that

$$a_0 I + a_1 T + \cdots + a_{n^2} T^{n^2} = 0$$

i.e., there is a polynomial $g(x)$ of degree less than n^2 such that $g(T) = 0$.

The proof is complete. ■

Proposition 7.2 The minimal polynomial $m_T(x)$, if exists, then it exists uniquely.

Proof. Suppose f_1, f_2 are two distinct minimal polynomials with $\deg(f_1) = \deg(f_2)$. It follows that

- $\deg(f_1 - f_2) < \deg(f_1)$.
- $f_1 - f_2 \neq 0$
- $(f_1 - f_2)(T) = f_1(T) - f_2(T) = 0_{V \to V}$

By scaling $f_1 - f_2$, there is a monic polynomial g with lower degree satisfying $g(T) = 0$, which contradicts the definition for minimal polynomial. ■

Proposition 7.3 Suppose $f(x) \in \mathbb{F}[x]$ satisfying $f(T) = \mathbf{0}$, then

$$m_T(x) \mid f(x).$$

Proof. It's clear that $\deg(f) \geq \deg(m_T)$. The division algorithm gives

$$f(x) = q(x) m_T(x) + r(x).$$

Therefore, for any $\boldsymbol{v} \in V$

$$[r(T)](\boldsymbol{v}) = [f(T)](\boldsymbol{v}) - [q(T)m_T(T)](\boldsymbol{v}) = \mathbf{0}_V - q(T)\mathbf{0}_V = \mathbf{0}_V - \mathbf{0}_V = \mathbf{0}_V$$

Therefore, $r(T) = \mathbf{0}_{V \to V}$. By definition of minimal polynomial, we imply $r(x) \equiv 0$. ■

Proposition 7.4 If $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{F}^{n \times n}$ are similar to each other, then $m_A(x) = m_B(x)$.

Proof. Suppose that $\boldsymbol{B} = \boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$, and that

$$m_A(x) = x^k + \cdots + a_1 x + a_0, \quad m_B(x) = x^\ell + \cdots + b_0.$$

It follows that

$$\begin{aligned} m_A(\boldsymbol{B}) &= \boldsymbol{B}^k + \cdots + a_0 I \\ &= \boldsymbol{P}^{-1}\boldsymbol{A}^k\boldsymbol{P} + \cdots + a_0\boldsymbol{P}^{-1}\boldsymbol{P} \\ &= \boldsymbol{P}^{-1}(\boldsymbol{A}^k + \cdots + a_0\boldsymbol{I})\boldsymbol{P} \\ &= \boldsymbol{P}^{-1}(m_A(\boldsymbol{A}))\boldsymbol{P} \end{aligned}$$

Therefore, $m_A(\boldsymbol{B}) = \mathbf{0}$ since $m_A(\boldsymbol{A}) = \mathbf{0}$. By proposition (7.3), we imply $m_B(x) \mid m_A(x)$. Similarly, $m_A(x) \mid m_B(x)$. Since $m_A(x)$ and $m_B(x)$ are monic, we imply $m_A(x) = m_B(x)$. ■

R Proposition (7.4) claims that the minimal polynomial is **similarity-invariant**; actually, the characteristic polynomial is **similarity-invariant** as well.

Assumption. We will asssume V has finite dimension from now on. Now we study the vanishing of a single vector $\boldsymbol{v} \in V$.

Notation. The $m_T(x)$ is a nonzero monic poylnomial of least degree such that

$$m_T(T) = \mathbf{0}_{V \to V}.$$

7.1.2. Minimal Polynomial of a vector

Definition 7.3 [Minimal Polynomial of a vector] Similar to the minimal polynomial, we define the **minimal polynomial of a vector $\boldsymbol{v}$ relative to** T, say $m_{T,\boldsymbol{v}}(x)$, as the monic polynomial of least degree such that

$$m_{T,\boldsymbol{v}}(T)(\boldsymbol{v}) = 0$$

The existence of minimal polynomial of a vector is due to the existence of minimal polynomial; the uniqueness follows similarly as in proposition (7.2).

Proposition 7.5 Let $T : V \to V$ be a linear operator and $\boldsymbol{v} \in V$. The degree of the minimal polynomial of a vector is upper bounded by:

$$\deg(m_{T,\boldsymbol{v}}(x)) \le \dim(V).$$

Proof. It's clear that $\{\boldsymbol{v}, T\boldsymbol{v}, \dots, T^n\boldsymbol{v}\} \subseteq V$ and the proof follows similarly as in proposition (7.1). ■

Similar to the division property in proposition (7.3), we have the division proprty for minimal polynomial of a vector:

Proposition 7.6 Suppose $f(x) \in \mathbb{F}[x]$ satisfying $f(T)(\boldsymbol{v}) = \mathbf{0}_V$, then

$$m_{T,\boldsymbol{v}}(x) \mid f(x).$$

In particular, $m_{T,\boldsymbol{v}} \mid m_T(x)$.

Proof. The proof follows similarly as in proposition (7.3). ■

Proposition 7.7 Suppose that $m_{T,\boldsymbol{v}}(x) = f_1(x)f_2(x)$, where f_1, f_2 are both monic. Let $\boldsymbol{w} = f_1(T)\boldsymbol{v}$, then

$$m_{T,\boldsymbol{w}}(x) = f_2(x)$$

Proof.

1. $$f_2(T)\boldsymbol{w} = f_2(T)f_1(T)\boldsymbol{v} = m_{T,\boldsymbol{v}}(T)\boldsymbol{v} = \mathbf{0}$$

   By the proposition (7.3), we imply $m_{T,\boldsymbol{w}}|f_2$.

2. On the other hand,

   $$\mathbf{0} = m_{T,\boldsymbol{w}}(T)(\boldsymbol{w}) = m_{T,\boldsymbol{w}}(T)f_1(T)\boldsymbol{v} = f_1(T)m_{T,\boldsymbol{w}}(T)\boldsymbol{v},$$

   which implies that $m_{T,\boldsymbol{v}}(x) \mid f_1(x)m_{T,\boldsymbol{w}}(x),$, i.e.,

   $$f_1 \cdot f_2 \mid f_1 \cdot m_{T,\boldsymbol{w}} \implies f_2 \mid m_{T,\boldsymbol{w}}.$$

   The proof is complete.

■

7.4. Wednesday for MAT3040

Reviewing.

- Given the polynomial $f(x) \in \mathbb{F}[x]$, we extend it into the linear operator $f(T): V \to V$.
- The minimal polynomial $m_T(x)$ is defined to be the polynomial with least degree such that
$$m_T(T) = \mathbf{0}_{V \to V},$$
i.e., $[m_T(T)]\boldsymbol{v} = 0_{\boldsymbol{V}}, \forall \boldsymbol{v} \in V$.
- The minimial polynomial of a vector $\boldsymbol{v}$ relative to T is defined to be the polynomial $m_{T,\boldsymbol{v}}(x)$ with the least degree such that
$$m_{T,\boldsymbol{v}}(T)(\boldsymbol{v}) = 0$$
- If $f(T) = \mathbf{0}_{V \to V}$, then we imply $m_T(x) \mid f(x)$. If $[g(T)](\boldsymbol{w}) = 0_V$, following the similar argument, we imply $m_{T,\boldsymbol{w}}(x) \mid g(x)$.
- In particular, $m_T(T)\boldsymbol{w} = \mathbf{0}$, which implies $m_{T,\boldsymbol{w}}(x) \mid m_T(x)$.

7.4.1. Cayley-Hamiton Theorem

Let's raise an motivative example first:

■ **Example 7.8** Consider the matrix and its induced mapping $\boldsymbol{A} = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$. It has the characteristic polynomial
$$\mathcal{X}_A = (x-1)(x-2).$$

- Note that $m_A(x)$ cannot be with degree one, since otherwise $m_A(x) = x - k$ with some k, and

$$m_A(\boldsymbol{A}) = \boldsymbol{A} - k\boldsymbol{I} = \begin{pmatrix} 1-k & 0 \\ 0 & 2-k \end{pmatrix} \ne \boldsymbol{0}, \quad \forall k,$$

which is a contradiction.

- However, one can verify that the $m_A(x)$ is with degree 2:

$$m_A(x) = (x-1)(x-2).$$

- The minimal polynomial with eigenvectors can be with degree 1:

$$\boldsymbol{w} = [0,1]^{\mathrm{T}} \implies (A-2I)\boldsymbol{w} = \boldsymbol{0} \implies m_{A,\boldsymbol{w}}(x) = x-2$$

R More generally, given an eigen-pair $(\lambda, \boldsymbol{v})$, the minimal polynomial of an $\boldsymbol{v}$ has the explicit form

$$m_{T,\boldsymbol{v}}(x) = (x-\lambda) \implies (x-\lambda) \mid m_T(x)$$

Now we want to relate the characterstic polynomial $m_T(x)$ with $\mathcal{X}_T(x)$. Suppose that

$$\mathcal{X}_T(x) = (x-\lambda_1)^{e_1} \cdots (x-\lambda_k)^{e_k} \in \mathbb{F}[x]. \tag{7.1}$$

Then we imply

- λ_i is an eigenvalue of T;
- $(x-\lambda_i) \mid m_T(x)$;

which implies that $(x-\lambda_1)\cdots(x-\lambda_k) \mid m_T(x)$.

Furthermore, (a). does $m_T(x)$ possess other factors, e.g., does there exist $\mu \ne \lambda_i, i = 1, \dots, k$ such that $(x-\mu) \mid m_T(x)$? (b). does $(x-\lambda_i)^{f_i} \mid m_T(x)$ when $f_i > e_i$?

The answer is no for both question (a) and (b).

Theorem 7.1 — **Cayley-Hamilton.** $m_T(x) \mid \mathcal{X}_T(x)$. In particular, $\mathcal{X}_T(T) = \mathbf{0}$.

The nice equality in (7.1) does not necessarily hold. Sometimes $\mathcal{X}_T(x)$ cannot be factorized into linear factors in $\mathbb{F}[x]$, e.g., $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ in $\mathbb{R}$.

However, for every $f(x) \in \mathbb{F}[x]$, we can extend $\mathbb{F}$ into the algebraically closed set $\overline{F} \supseteq \mathbb{F}$ such that

$$f(x) = (x - \lambda_1)^{e_1} \cdots (x - \lambda_k)^{e_k}$$

where $\lambda_i \in \overline{\mathbb{F}}$.

For example, for $f(x) = x^2 + 1 \in \mathbb{R}[x]$, we can extend $\mathbb{R}$ into $\mathbb{C}$ to obtain

$$f(x) = (x + i)(x - i).$$

Therefore, the general proof outline for the Cayley-Hamilton Theorem is as follows:

- Consider the case where $m_T(x), \mathcal{X}_T(x)$ are both in $\overline{F}[x]$
- Show that $m_T(x) \mid \mathcal{X}_T(x)$ under $\overline{F}[x]$.

Before the proof, let's study the invariant subspaces, which leads to the decomposition of charactersitc polynomial:

Assumption. From now on, we assume that V is finite dimensional by default.

Definition 7.12 [Invariant Subspace] An **invariant subspace** of a linear operator $T : V \to V$ is a subspace $W \leq V$ that is preserved by T, i.e., $T(W) \subseteq W$. We also call W as T-invariant.

R If $W \leq V$ is T-invariant, then the restriction of the linear operator $T : V \to V$ induces the linear operator

$$T\mid_W : W \to W.$$

Example 7.9 1. V itself is T-invariant.

2. For the eigenvalue λ, the associated λ-eigenspace $U = \ker(T - \lambda I)$ is T-invariant.

3. More generally, $U = \ker(g(T))$ is T-invariant for any polynomial g:

   If $\boldsymbol{v} \in \ker(g(T))$, i.e., $g(T)\boldsymbol{v} = \boldsymbol{0}$, it suffices to show $T(\boldsymbol{v}) \in \ker(g(T))$:

$$\begin{aligned} g(T)[T(\boldsymbol{v})] &= (a_m T^m + \cdots + a_0 I)[T(\boldsymbol{v})] \\ &= (a_m T \circ T^m + \cdots + a_1 T \circ T + a_0 T \circ I)(\boldsymbol{v}) \\ &= T[g(T)\boldsymbol{v}] = T(\boldsymbol{0}) = \boldsymbol{0} \end{aligned}$$

4. For $\boldsymbol{v} \in \ker(T - \lambda I)$, $U = \mathrm{span}\{\boldsymbol{v}\}$ is T-invariant.

Proposition 7.10 Suppose that $T : V \to V$ is a linear transformation and $W \leq V$ is T-invariant, then we construct the subspace mapping and the recipe mapping

$$\begin{aligned} T\,|_W : &\quad W \to W \\ \text{with} &\quad \boldsymbol{w} \mapsto T(\boldsymbol{w}) \end{aligned} \tag{7.2a}$$

$$\begin{aligned} \tilde{T} : &\quad V/W \to V/W \\ \text{with} &\quad \boldsymbol{v} + W \mapsto T(\boldsymbol{v}) + W \end{aligned} \tag{7.2b}$$

(Here the well-definedness of the recipe mapping $\tilde{T}$ is shown in Hw2, Exercise 4),

which leads to the decomposition of the charactersitic polynomial:

$$\mathcal{X}_T(x) = \mathcal{X}_{T|_W}(x)\mathcal{X}_{\tilde{T}}(x).$$

Proof. Suppose $C = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$ is a basis of W, and extend it into the basis of V, denoted as

$$\mathcal{B} = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k, \boldsymbol{v}_{k+1}, \ldots, \boldsymbol{v}_n\}$$

Therefore, $\overline{\mathcal{B}} = \{\boldsymbol{v}_{k+1} + W, \ldots, \boldsymbol{v}_n + W\}$ is a basis of V/W. By Homework 2, Question 5, the representation $(T)_{\mathcal{B},\mathcal{B}}$ can be written as the block matrix

$$(T)_{\mathcal{B},\mathcal{B}} = \begin{pmatrix} (T|_W)_{C,C} & \times \\ \mathbf{0} & (\tilde{T})_{\overline{\mathcal{B}},\overline{\mathcal{B}}} \end{pmatrix}_{(k+(n-k))\times(k+(n-k))}$$

Therefore, the characteristic polynomial of T can be calculated as:

$$\begin{aligned} \mathcal{X}_T(x) &= \det((T)_{\mathcal{B},\mathcal{B}} - xI) \\ &= \det((T|_U)_{C,C} - xI) \cdot \det((\tilde{T})_{\overline{\mathcal{B}},\overline{\mathcal{B}}} - xI) \end{aligned}$$

■

Proposition 7.11 Suppose that

$$\mathcal{X}_T(x) = (x - \lambda_1)\cdots(x - \lambda_n)$$

where λ_i's are not necessarily distinct. Then there exists a basis of V, say $\mathcal{A}$, such that

$$(T)_{\mathcal{A},\mathcal{A}} = \begin{pmatrix} \lambda_1 & \times & \times & \times \\ 0 & \lambda_2 & \cdots & \times \\ 0 & \cdots & \ddots & \times \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

Proof. The proof is by induction on n, i.e., suppose the results hold for all vector spaces with dimension no more than $n-1$, and we aim to show this result holds for dimension n.

1. **Step 1**: Argue that there exists the associated eigenvector $\boldsymbol{v}$ of λ_1 under the linear operator T.

   Consider any basis $\mathcal{M}$, by MAT2040, there exists associated eigenvector of λ_1, say $\boldsymbol{y} \in \mathbb{C}^n$ such that

   $$(T)_{\mathcal{M},\mathcal{M}} \cdot \boldsymbol{y} = \lambda_1 \boldsymbol{y}$$

   Since the operator $(\cdot)_{\mathcal{M}} : V \to \mathbb{C}^n$ is an isomorphism, there exists $\boldsymbol{v} \in V \setminus \{\mathbf{0}\}$ such that $(\boldsymbol{v})_{\mathcal{M}} = \boldsymbol{y}$. It follows that

$$(T)_{\mathcal{M},\mathcal{M}}(\boldsymbol{v})_{\mathcal{M}} = \lambda_1(\boldsymbol{v})_{\mathcal{M}} \implies (T\boldsymbol{v})_{\mathcal{M}} = (\lambda_1\boldsymbol{v})_{\mathcal{M}} \implies T\boldsymbol{v} = \lambda_1\boldsymbol{v}$$

2. **Step 2**: Dimensionality reduction of $\mathcal{X}_T(x)$: Construct $W = \text{span}\{\boldsymbol{v}\}$, which is T-invariant. By the proof of proposition (7.11), we have $\tilde{T} : V/W \to V/W$ admits the characteristic polynomial

$$\mathcal{X}_{\tilde{T}}(x) = (x - \lambda_2)\cdots(x - \lambda_n)$$

3. **Step 3:** Applying the induction, there exists basis $\overline{C}$ of V/W, i.e.,

$$\overline{C} = \{\boldsymbol{w}_2 + W, \dots, \boldsymbol{w}_n + W\}$$

such that

$$(\tilde{T})_{\overline{C},\overline{C}} = \begin{pmatrix} \lambda_2 & \times & \times & \times \\ 0 & \lambda_3 & \cdots & \times \\ 0 & \cdots & \ddots & \times \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

4. **Step 4:** Therefore, we construct the set $\mathcal{A} := \{\boldsymbol{v}, \boldsymbol{w}_2, \dots, \boldsymbol{w}_n\}$. We claim that
   - $\mathcal{A}$ is a basis of V (left as exercise in Hw2, Question 2)
   - $$(T)_{\mathcal{A},\mathcal{A}} = \begin{pmatrix} \lambda_1 & \times \\ \boldsymbol{0} & (\tilde{T})_{\overline{C},\overline{C}} \end{pmatrix} = \begin{pmatrix} \lambda_1 & \times & \times & \times \\ 0 & \lambda_2 & \cdots & \times \\ 0 & \cdots & \ddots & \times \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

     (This statement is also left as exercise in Hw2, Question 5.)

∎

Proposition 7.12 Suppose that $\mathcal{X}_T(x) = (x - \lambda_1)\cdots(x - \lambda_n)$, then $\mathcal{X}_T(T) = \boldsymbol{0}$.

R One special case is that $\boldsymbol{A} = \text{diag}(\lambda_1, \dots, \lambda_n)$. The results for proposition (7.12) gives

$$(A - \lambda_1 \boldsymbol{I}) \cdots (A - \lambda_n \boldsymbol{I}) \text{ is a zero matrix}$$

Chapter 8

Week8

8.1. Monday for MAT3040

Reviewing.

- If $\mathcal{X}_T(x) = (x-\lambda_1)\cdots(x-\lambda_n)$, then

$$(T)_{\mathcal{A},\mathcal{A}} = \begin{pmatrix} \lambda_1 & \times & \times & \times \\ 0 & \lambda_2 & \cdots & \times \\ 0 & \cdots & \ddots & \times \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

for some basis $\mathcal{A}$. In other words, T is **triangularizable** with the diagonal entries $\lambda_1,\dots,\lambda_n$.

R I hope you appreciate this result. Consider the example below: In linear algebra we have studied that the matrix $\boldsymbol{A} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ is not diagonalizable, and the characteristic polynomial is given by

$$\mathcal{X}_A(x) = (x-1)^2.$$

However, the theorem above claims that $\boldsymbol{A}$ is *triangularizable*, with diagonal entries 1 and 1. The diagonalization of $\boldsymbol{A}$ only uses the eigenvector of $\boldsymbol{A}$, but the 1-eigenspace has only 1 dimension. Fortunately, the triangularization gives a rescue such that we can make use of the generalized eigenvector

$(0,1)^{\mathrm{T}}$ (but not an eigenvector) of $\boldsymbol{A}$ by considering the mapping below:

$$
\begin{aligned}
U &= \operatorname{span}\left\{\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right\} \\
\bar{A}: \quad & V/U \to V/U
\end{aligned}
$$

Here $(0,1)^{\mathrm{T}} + U$ is an eigenvector of $\bar{A}$, with eigenvalue 1.

Theorem 8.1 The linear operator T is triangularizable with diagonal entries $(\lambda_1, \ldots, \lambda_n)$ if and only if

$$\mathcal{X}_T = (x - \lambda_1) \cdots (x - \lambda_n)$$

Proof. It suffices to show only the sufficiency. Suppose that there exists basis $\mathcal{A}$ such that

$$(T)_{\mathcal{A},\mathcal{A}} = \begin{pmatrix} \lambda_1 & \times & \times & \times \\ 0 & \lambda_2 & \cdots & \times \\ 0 & \cdots & \ddots & \times \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

Then we compute the characteristic polynomial directly:

$$
\begin{aligned}
\mathcal{X}_T(x) &= \det[(xI - T)_{\mathcal{A},\mathcal{A}}] \\
&= \det \begin{pmatrix} x - \lambda_1 & \times & \times & \times \\ 0 & x - \lambda_2 & \cdots & \times \\ 0 & \cdots & \ddots & \times \\ 0 & 0 & \cdots & x - \lambda_n \end{pmatrix} \\
&= (x - \lambda_1) \cdots (x - \lambda_n)
\end{aligned}
$$

■

8.1.1. Cayley-Hamiton Theorem

Proposition 8.1 — A Useful Lemma. Suppose that $\mathcal{X}_T(x) = (x - \lambda_1)\cdots(x - \lambda_n)$, then $\mathcal{X}_T(T) = 0$.

Proof. Since $\mathcal{X}_T(x) = (x - \lambda_1)\cdots(x - \lambda_n)$, we imply T is triangularizable under some basis $\mathcal{A}$. Note that

- $T \mapsto (T)_{\mathcal{A},\mathcal{A}}$ is an isomorphism between $\text{Hom}(V, V)$ and $M_{n\times n}(\mathbb{F})$,
- $(\underbrace{T \circ T \circ \cdots \circ T}_{m \text{ times}})_{\mathcal{A},\mathcal{A}} = [(T)_{\mathcal{A},\mathcal{A}}]^m$, for any m,

It suffices to show $\mathcal{X}_T((T)_{\mathcal{A},\mathcal{A}})$ is the zero matrix (why?):

$$\mathcal{X}_T((T)_{\mathcal{A},\mathcal{A}}) = ((T)_{\mathcal{A},\mathcal{A}} - \lambda_1 \boldsymbol{I})\cdots((T)_{\mathcal{A},\mathcal{A}} - \lambda_n \boldsymbol{I}).$$

Observe the matrix multiplication

$$((T)_{\mathcal{A},\mathcal{A}} - \lambda_i \boldsymbol{I})\begin{pmatrix} x_1 \\ \vdots \\ x_i \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} \lambda_1 - \lambda_i & \times & \times & \times \\ 0 & \lambda_2 - \lambda_i & \cdots & \times \\ 0 & \cdots & \ddots & \times \\ 0 & 0 & \cdots & \lambda_n - \lambda_i \end{pmatrix}\begin{pmatrix} x_1 \\ \vdots \\ x_i \\ 0 \\ \vdots \\ 0 \end{pmatrix} \in \text{span}\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_{i-1}\}$$

Therefore, for any $\boldsymbol{v} \in V$,

$$((T)_{\mathcal{A},\mathcal{A}} - \lambda_n \boldsymbol{I})\boldsymbol{v} \in \text{span}\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_{n-1}\}.$$

Applying the same trick, we conclude that

$$((T)_{\mathcal{A},\mathcal{A}} - \lambda_1 \boldsymbol{I})\cdots((T)_{\mathcal{A},\mathcal{A}} - \lambda_n \boldsymbol{I})\boldsymbol{v} = \boldsymbol{0}, \quad \forall \boldsymbol{v} \in V,$$

i.e., $\mathcal{X}_T((T)_{\mathcal{A},\mathcal{A}}) = ((T)_{\mathcal{A},\mathcal{A}} - \lambda_1 \boldsymbol{I})\cdots((T)_{\mathcal{A},\mathcal{A}} - \lambda_n \boldsymbol{I})$ is a zero matrix. ∎

Now we are ready to give a proof for the Cayley-Hamiton Theorem:

Proof. Suppose that $\mathcal{X}_T(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_0 \in \mathbb{F}[x]$. By considering algebrically closed field $\overline{\mathbb{F}} \supseteq \mathbb{F}$, we imply

$$\mathcal{X}_T(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_0 \tag{8.1a}$$

$$= (x - \lambda_1) \cdots (x - \lambda_n), \quad \lambda_i \in \overline{\mathbb{F}} \tag{8.1b}$$

By applying proposition (8.1), we imply $\mathcal{X}_T(T) = 0$, where the coefficients in the formula $\mathcal{X}_T(T) = 0$ w.r.t. T are in $\overline{\mathbb{F}}$.

Then we argue that these coefficients are essentially in $\mathbb{F}$. Expand the whole map of $\mathcal{X}_T(T)$:

$$\mathcal{X}_T(T) = (T - \lambda_1 I) \cdots (T - \lambda_n I) \tag{8.2a}$$

$$= T^n - (\lambda_1 + \cdots + \lambda_n)T^{n-1} + \cdots + (-1)^n \lambda_1 \cdots \lambda_n I \tag{8.2b}$$

$$= T^n + a_{n-1}T^{n-1} + \cdots + a_0 I \tag{8.2c}$$

where the derivation of (8.2c) is because that the polynomial coefficients for (8.1a) and (8.1b) are all identical.

Therefore, we conclude that $\mathcal{X}_T(T) = 0$, under the field $\mathbb{F}$. ∎

Corollary 8.1 $m_T(x) \mid \mathcal{X}_T(x)$. More precisely, if

$$\mathcal{X}_T(x) = [p_1(x)]^{e_1} \cdots [p_k(x)]^{e_k}, \ e_i > 0, \forall i$$

where p_i's are distinct, monic, and irreducible polynomials. Then

$$m_T(x) = [p_1(x)]^{f_1} \cdots [p_k(x)]^{f_k}, \text{ for some } 0 < f_i \le e_i, \forall i$$

Proof. The statement $m_T(x) \mid \mathcal{X}_T(x)$ is from Cayley-Hamiton Theorem. Therefore, $0 \le f_i \le e_i, \forall i$. Suppose on the contrary that $f_i = 0$ for some i. w.l.o.g., $i = 1$.

It's clear that $\gcd(p_1, p_j) = 1$ for $\forall j \neq 1$, which implies

$$a(x)p_1(x) + b(x)p_j(x) = 1, \text{ for some } a(x), b(x) \in \mathbb{F}[x].$$

Considering the field extension $\overline{\mathbb{F}} \supseteq \mathbb{F}$, we have $p_1(x) = (x - \mu_1)\cdots(x - \mu_\ell)$. For any root μ_m of p_1, $m = 1, \ldots, \ell$, we have

$$a(\mu_m)p_1(\mu_m) + b(\mu_m)p_j(\mu_m) = 1 \implies b(\mu_m)p_j(\mu_m) = 1 \implies p_j(\mu_m) \neq 0,$$

i.e., μ_m is not a root of p_j, $\forall j \neq 1$.

Therefore, μ_m is a root of $\mathcal{X}_T(x)$, but not a root of $m_T(x)$. Then μ_m is an eigenvalue of T, e.g., $T\boldsymbol{v} = \mu_m \boldsymbol{v}$ for some $\boldsymbol{v} \neq \boldsymbol{0}$. Recall that $m_{T,\boldsymbol{v}} = x - \mu_m$, we imply $m_{T,\boldsymbol{v}} = x - \mu_m \mid m_T(x)$, which is a contradiction. ■

▪ **Example 8.1** We can use Corollary (8.1), a stronger version of Cayley-Hamiltion Theorem to determine the minimal polynomials:

1. For matrix $\boldsymbol{A} = \begin{pmatrix} 0 & -1 \\ 1 & 1 \end{pmatrix}$, we imply $\mathcal{X}_A(x) = (x^2 + x + 1)^1$. Since $x^2 + x + 1$ is irreducible in $\mathbb{R}$, we have $m_A(x) = x^2 + x + 1$.
2. For matrix
$$\boldsymbol{A} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix},$$
we imply $\mathcal{X}_A(x) = (x-1)^2(x-2)^2$.

   By Corollary (8.1), we imply both $(x-1)$ and $(x-2)$ should be roots of $m_T(x)$, i.e., $m_A(x)$ may have the four options:
$$\begin{aligned} (x-1)^2(x-2)^2, &\text{ or} \\ (x-1)(x-2)^2, &\text{ or} \\ (x-1)^2(x-2), &\text{ or} \\ (x-1)(x-2). \end{aligned}$$

   By trial and error, one sees that $m_A(x) = (x-1)^2(x-2)$.

8.1.2. Primary Decomposition Theorem

We know that not every linear operator is diagonalizable, but diagonalization has some nice properties:

Definition 8.1 [diagonalizable] The linear operator $T : V \to V$ is diagonalizable over $\mathbb{F}$ if and only if there exists a basis $\mathcal{A}$ of V such that

$$(T)_{\mathcal{A},\mathcal{A}} = \operatorname{diag}(\lambda_1, \ldots, \lambda_n),$$

where λ_i's are not necessarily distinct.

Proposition 8.2 If the linear operator $T : V \to V$ is diagonalizable, then

$$m_T(x) = (x - \mu_1) \cdots (x - \mu_k),$$

where μ_i's are **distinct**.

Proof. Suppose T is diagonalizable, then there exists a basis $\mathcal{A}$ of V such that

$$(T)_{\mathcal{A},\mathcal{A}} = \operatorname{diag}(\mu_1, \ldots, \mu_1, \mu_2, \ldots, \mu_2, \ldots, \mu_k, \ldots, \mu_k)$$

It's clear that $((T)_{\mathcal{A},\mathcal{A}} - \mu_1 \boldsymbol{I}) \cdots ((T)_{\mathcal{A},\mathcal{A}} - \mu_k \boldsymbol{I}) = \mathbf{0}$, i.e., $m_T(x) \mid (x - \mu_1) \cdots (x - \mu_k)$.

Then we show the minimality of $(x - \mu_1) \cdots (x - \mu_k)$. In particular, if $(x - \mu_i)$ is omitted for any $1 \le i \le k$, then it's easy to show

$$(T_{\mathcal{A},\mathcal{A}} - \mu_1 \boldsymbol{I}) \cdots (T_{\mathcal{A},\mathcal{A}} - \mu_{i-1} \boldsymbol{I})(T_{\mathcal{A},\mathcal{A}} - \mu_{i+1} \boldsymbol{I}) \cdots (T_{\mathcal{A},\mathcal{A}} - \mu_k \boldsymbol{I}) \ne \mathbf{0},$$

since all μ_i's are distinct. Therefore, $m_T(x)$ will not divide $(x - \mu_1) \cdots (x - \mu_{i-1})(x - \mu_{i+1}) \cdots (x - \mu_k)$ for any i, i.e.,

$$m_T(x) = (x - \mu_1) \cdots (x - \mu_k)$$

■

R The converse of proposition (8.2) is also true, which is a special case for the Primary Decomposition Theorem.

Theorem 8.2 — Primary Decomposition Theorem. Let $T : V \to V$ be a linear operator with

$$m_T(x) = [p_1(x)]^{e_1} \cdots [p_k(x)]^{e_k},$$

where p_i's are distinct, monic, and irreducible polynomials. Let $V_i = \ker([p_i(x)]^{e_i}) \le V, i = 1, \dots, k$, then

1. Each V_i is T-invariant ($T(V_i) \le V_i$)
2. $V = V_1 \oplus V_2 \oplus \cdots \oplus V_k$
3. Consider $T\,|_{V_i} : V_i \to V_i$, then

$$m_{T|V_i}(x) = [p_i(x)]^{e_i}$$

Chapter 9

Week9

9.1. Monday for MAT3040

Reviewing.

- $\mathcal{X}_T(x) = (x-\lambda_1)\cdots(x-\lambda_n)$ over $\mathbb{F}$ if and only if T is triangularizable over $\mathbb{F}$.
- $m_T(x) = (x-\mu_1)\cdots(x-\mu_k)$, where μ_i's are distinct over $\mathbb{F}$ if and only if T is diagonalizable over $\mathbb{F}$.

  The converse for this statement is the proposition (8.2). Let's focus on the proof for the forward direction.

9.1.1. Remarks on Primary Decomposition Theorem

Theorem 9.1 — **Primary Decomposition Theorem.** Let $T : V \to V$ be a linear operator with $\dim(V) < \infty$, and

$$m_T(x) = [p_1(x)]^{e_1}\cdots[p_k(x)]^{e_k}$$

where p_i's are distinct, monic, irreducible polynomials. Let $V_i = \ker(p_i(T)^{e_i})$, then

1. each V_i is T-invariant (*i.e.*,$T(V_i) \le V_i$)
2. $V = V_1 \oplus \cdots \oplus V_k$
3. $T\mid_{V_i}$ has the minimal polynomial $p_i(x)^{e_i}$.

Proof. 1. (1) follows from part (2) for example (**??**).

2. Let $q_i(x) = [p_1(x)]^{e_1} \cdots \widehat{[p_i(x)]^{e_i}} \cdots [p_k(x)]^{e_k} := m_T(x)/[p_i(x)]^{e_i}$, then it is clear that

   (a) $\gcd(q_1, \ldots, q_k) = 1$

   (b) $\gcd(q_i, p_i^{e_i}) = 1$

   (c) $q_i \cdot p_i^{e_i} = m_T$

   (d) If $i \neq j$, then $m_T(x) \mid q_i(x)q_j(x)$

   - By (a) and Bezout's Theorem (6.7), there exists polynomials $a_1, \ldots, a_k$ such that

$$a_1(x)q_1(x) + \cdots + a_k(x)q_k(x) = 1,$$

     which implies

$$\underbrace{a_1(T)q_1(T)\boldsymbol{v}}_{\boldsymbol{v}_1} + \cdots + \underbrace{a_k(T)q_k(T)\boldsymbol{v}}_{\boldsymbol{v}_k} = \boldsymbol{v}$$

     Therefore, $\boldsymbol{v} = \boldsymbol{v}_1 + \cdots + \boldsymbol{v}_k$ for our constructed $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$.

   - Note that

$$p_i(T)^{e_i}\boldsymbol{v}_i = p_i(T)^{e_i}a_i(T)q_i(T)\boldsymbol{v} = a_i(T)[q_i(T)p_i(T)^{e_i}]\boldsymbol{v} = a_i(T)m_T(T)\boldsymbol{v} = \mathbf{0},$$

     whcih implies $\boldsymbol{v}_i \in \ker([p_i(T)]^{e_i}) := V_i$, and therefore

$$V = V_1 + \cdots + V_k \tag{9.1}$$

   - To show that the summation in (9.3) is essentially the direct sum, consider

$$\mathbf{0} = \boldsymbol{v}_1' + \cdots + \boldsymbol{v}_k', \quad \forall \boldsymbol{v}_i' \in V_i. \tag{9.2}$$

     By (a) and Bezout's Theorem (6.7), there exists $b_i(x), c_i(x)$ such that

$$b_i(x)q_i(x) + c_i(x)p_i(x)^{e_i} = 1 \implies b_i(T)q_i(T) + c_i(T)p_i(T)^{e_i} = I,$$

     i.e.,

$$b_i(T)q_i(T)\boldsymbol{v}_i' + c_i(T)p_i(T)^{e_i}\boldsymbol{v}_i' = b_i(T)q_i(T)\boldsymbol{v}_i' = \boldsymbol{v}_i'.$$

Appying the mapping $b_i(T)q_i(T)$ into equality (9.4) both sides, $i = 1,\dots,k$, we obtain

$$\mathbf{0} = b_i(T)q_i(T)\mathbf{0} = b_i(T)q_i(T)\boldsymbol{v}_1' + \cdots + b_i(T)q_i(T)\boldsymbol{v}_k'$$

Note that all terms on RHS vanish except for $b_i(T)q_i(T)\boldsymbol{v}_i' = \boldsymbol{v}_i'$, since $q_i(x) = [p_1(x)]^{e_1}\cdots\widehat{[p_i(x)]^{e_i}}\cdots[p_k(x)]^{e_k}$ and $\boldsymbol{v}_j' \in \ker([p_j(x)]^{e_j})$. Therefore, $\boldsymbol{v}_i' = 0$ for $i = 1,\dots,k$, i.e., $V = V_1 \oplus \cdots \oplus V_k$.

3. For any $\boldsymbol{v}_i \in V_i$, we have $p_i(T)^{e_i}\boldsymbol{v}_i = \mathbf{0}$, which implies $m_{T|V_i}(x) \mid p_i(x)^{e_i}$. Together with Corollary (8.1), $m_{T|v_i}(x) = p_i(x)^{f_i}$ for some $1 \le f_i \le e_i$.
Suppose on the contrary that there exists $f_i < e_i$ for some i, consider any $\boldsymbol{v} := \boldsymbol{v}_1 + \cdots + \boldsymbol{v}_k \in V$, and

$$p_1(T)^{f_1}\cdots p_k(T)^{f_k}\boldsymbol{v} = p_1(T)^{f_1}\cdots p_k(T)^{f_k}(\boldsymbol{v}_1 + \cdots + \boldsymbol{v}_k)$$

The term on the RHS vanishes since $p_j(T)^{f_j}\boldsymbol{v}_j = \mathbf{0}$, which implies

$$m_T \mid p_1^{f_1}\cdots p_k^{f_k},$$

but there exists i such that $e_i > f_i$, which is a contradiction.

■

Corollary 9.1 If $m_i(x) = (x-\mu_1)\cdots(x-\mu_k)$ over $\mathbb{F}$, where μ_i's are distinct, then T is diagonalizable over $\mathbb{F}$. (the converse actually also holds, see proposition (8.2))

Proof. By primary decomposition theorem,

$$V = \underbrace{\ker(T-\mu_1 I)}_{V_1}\oplus\cdots\oplus\underbrace{\ker(T-\mu_k I)}_{V_k}$$

Take B_i as a basis of V_i, an μ_i-eigenspace of T. Then $B := \cup_{i=1}^k B_i$ is a basis consisting of eigenvectors of T.

It's clear that $(T|_{V_i})_{\mathcal{B},\mathcal{B}} = \operatorname{diag}(\mu_i, \dots, \mu_i)$, and T is diagonalizable with

$$(T)_{\mathcal{B},\mathcal{B}} = \operatorname{diag}((T|_{V_1})_{\mathcal{B},\mathcal{B}}, \cdots, (T|_{V_k})_{\mathcal{B},\mathcal{B}}).$$

■

Corollary 9.2 [Spectral Decomposition] Suppose $T : V \to V$ is diagonalizable, then there exists a linear operator $p_i : V \to V$ for $1 \le i \le k$ such that

- $p_i^2 = p_i$ (idempotent)
- $p_i p_j = 0, \forall i \neq j$
- $\sum_{i=1}^k p_i = I$
- $p_i T = T p_i, \forall i$

and scalars $\mu_1, \dots, \mu_k$ such that

$$T = \mu_1 p_1 + \cdots + \mu_k p_k$$

Proof. Diagonlization of T is equivalent to say that $m_T(x) = (x - \mu_1) \cdots (x - \mu_k)$, where μ_i's are distinct. Construct

- $V_i := \ker(T - \mu_i I)$
- $p_i : V \to V$ given by $p_i = a_i(T)q_i(T)$ as in the proof of primary decomposition theorem

Then:

- $p_i T = T p_i$ is obvious
- $\sum_{i=1}^k p_i = \sum_{i=1}^k a_i(T)q_i(T) = I$
- $p_i p_j = a_i(T)a_j(T)q_i(T)q_j(T) := a_i(T)a_j(T)s(T)m_T(T) = \mathbf{0}$
- $p_i^2 = p_i(p_1 + \cdots + p_k) = p_i \cdot I = p_i$

For the last part, note that

- $p_i V \le V_i, \forall i$: for $\forall \boldsymbol{v} \in V$,

$$(T - \mu_i I)p_i \boldsymbol{v} = (T - \mu_i I)a_i(T)q_i(T)\boldsymbol{v} = a_i(T)m_T(x)\boldsymbol{v} = \boldsymbol{0}$$

  Therefore, $p_i V \le \ker(T - \mu_i I) = V_i$

- Now, for all $\boldsymbol{w} \in V$,

$$\begin{aligned}
T\boldsymbol{w} &= T(p_1 + \cdots + p_k)\boldsymbol{w} \\
&= Tp_1\boldsymbol{w} + \cdots + Tp_k\boldsymbol{w} \\
&= (\mu_1 p_1)\boldsymbol{w} + \cdots + (\mu_k p_k)\boldsymbol{w}
\end{aligned}$$

  and therefore $T = \mu_1 p_1 + \cdots + \mu_k p_k$

■

Organization of future two weeks. We are interested in under which condition does the T is diagonalizable. One special case is $T = A$, where $\boldsymbol{A}$ is a symmetric matrix. We will study normal operators, which includes the case for symmetric matrices.

Question: what happens if $m_T(x)$ contains repeated linear factors? We will spend the next whole class to show the Jordan Normal Form:

Theorem 9.2 — Jordan Normal Form. Let $\mathbb{F}$ be algebraically closed field such that every linear operator $T : V \to V$ has the form

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

where λ_i's are distinct.

Then there exists basis $\mathcal{A}$ of V such that

$$(T)_{\mathcal{A},\mathcal{A}} = \operatorname{diag}(\boldsymbol{J}_1, \ldots, \boldsymbol{J}_k)$$

where

$$\boldsymbol{J}_i = \begin{pmatrix} \mu & 1 & 0 & 0 \\ 0 & \mu & 1 & 0 \\ 0 & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mu \end{pmatrix}$$

for some $\mu \in \{\lambda_1, \ldots, \lambda_k\}$

9.4. Wednesday for MAT3040

9.4.1. Jordan Normal Form

Theorem 9.3 — Jordan Normal Form. Suppose that $T : V \to V$ has minimial polynomial

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

then there exists a basis $\mathcal{A}$ such that

$$(T)_{\mathcal{A},\mathcal{A}} = \text{diag}(J_1, \dots, J_\ell),$$

where each block J_i is a square matrix of the form

$$J_i = \begin{bmatrix} \mu_i & 1 & & \\ & \mu_i & \ddots & \\ & & \ddots & 1 \\ & & & \mu_i \end{bmatrix}.$$

R By primary decomposition theorem,

$$V = V_1 \oplus \cdots \oplus V_k, \quad \text{where } V_i = \ker((T - \lambda_i I)^{e_i}),\ i = 1, \dots, k,$$

and each V_i is T-invariant.

We pick basis $\mathcal{B}_i$ for each subspace V_i, then $\mathcal{B} := \cup_{i=1}^{k} \mathcal{B}_i$ is a basis of V, and

$$(T)_{\mathcal{B},\mathcal{B}} = \begin{pmatrix} (T\mid_{V_1})_{\mathcal{B}_1,\mathcal{B}_1} & 0 & \cdots & 0 \\ 0 & (T\mid_{V_2})_{\mathcal{B}_2,\mathcal{B}_2} & \vdots & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & \vdots & (T\mid_{V_k})_{\mathcal{B}_k,\mathcal{B}_k} \end{pmatrix}$$

with $m_{T|_{V_i}}(x) = (x - \lambda_i)^{e_i}$.

Therefore, it suffices to show the Jordan normal form holds for the linear operator

T with minimal polynomial $m_T(x) = (x-\lambda)^e$.

Firstly, we consider the case where the minimal polynomial has the form x^m:

Proposition 9.6 Suppose $T : V \to V$ is such that $m_T(x) = x^m$, then the theorem (9.3) holds, i.e., there exists a basis $\mathcal{A}$ such that

$$(T)_{\mathcal{A},\mathcal{A}} = \text{diag}(\boldsymbol{J}_1, \dots, \boldsymbol{J}_\ell),$$

where each block J_i is a square matrix of the form

$$J_i = \begin{bmatrix} 0 & 1 & & \\ & 0 & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{bmatrix}.$$

Proof.

- Suppose that $m_T(x) = x^m$, then it is clear that

$$\{0\} := \ker(T^0) \le \ker(T) \le \ker(T^2) \le \cdots \le \ker(T^m) := V$$

  Furthermore, we have $\ker(T^{i-1}) \subsetneqq \ker(T^i)$ for $i = 1, \dots, m$: Note that $\ker(T^{m-1}) \subsetneqq \ker(T^m) := V$ due to the minimality of $m_T(x)$; and $\ker(T^{m-2}) \subsetneqq \ker(T^{m-1})$ since otherwise for any $\boldsymbol{x} \in \ker(T^m)$,

$$T^{m-1}(T\boldsymbol{x}) = \boldsymbol{0} \implies T\boldsymbol{x} \in \ker(T^{m-1}) = \ker(T^{m-2}) \implies T^{m-2}(T\boldsymbol{x}) = T^{m-1}(\boldsymbol{x}) = \boldsymbol{0},$$

  i.e., $\boldsymbol{x} \in \ker(T^{m-1})$, which contradicts to the fact that $\ker(T^{m-1}) \subsetneqq \ker(T^m)$. Proceeding this trick sequentially for $i = m, m-1, \dots, 1$, we proved the disired result.

- Then construct the quotient space $W_i = \ker(T^i)/\ker(T^{i-1})$ and define $\mathcal{B}_i'$ to be a basis of W_i:

$$\mathcal{B}_i' = \{a_1^i + \ker(T^{i-1}), \dots, a_{\ell_i}^i + \ker(T^{i-1})\}$$

  Construct $\mathcal{B}_i = \{a_1^i, \dots, a_{\ell_i}^i\}$, then we claim that $B := \cup_{i=1}^m \mathcal{B}_i$ forms a basis of V:

- First proof the case $m = 2$ first: let $U \le V$ ($\dim(V) < \infty$), and $\mathcal{B}_1 = \{a_1^1, \dots, a_{k_1}^1\}$ be a basis of U, and

$$\mathcal{B}_2' = \{a_1^2 + U, \dots, a_{k_2}^2 + U\}$$

be a basis of V/U. Then to show the statement suffices to show that

$$\bigcup_{i=1}^{2} \{a_1^i, \dots, a_{k_i}^i\} \text{ forms a basis of } V.$$

It's clear that $\cup_{i=1}^{2}\{a_1^i, \dots, a_{k_i}^i\}$ spans V. Furthermore, $\dim(V) = \dim(U) + \dim(V/U) = k_1 + k_2$, i.e., $\cup_{i=1}^{2}\{a_1^i, \dots, a_{k_i}^i\}$ contains correct amount of vectors. The proof is complete.

- This result can be extended from 2 to general m, thus the claim is shown.

- For $i < m$, consider the set $S_i = \{T(\boldsymbol{w}_j) + \ker(T^{i-1}) \mid \boldsymbol{w}_j \in B_{i+1}\}$. Note that
  - Since $T^{i+1}(\boldsymbol{w}_j) = \mathbf{0}$, $T^i(T(\boldsymbol{w}_j)) = \mathbf{0}$, we imply $T(\boldsymbol{w}_j) \in \ker(T^i)$, i.e., $S_i \subseteq W_i$.
  - The set S_i is linearly independent: consider the equation

$$\sum_j k_j(T(\boldsymbol{w}_j) + \ker(T^{i-1})) = \mathbf{0}_{W_i} \iff T\left(\sum_j k_j \boldsymbol{w}_j\right) + \ker(T^{i-1}) = \mathbf{0}_{W_i}$$

i.e.,

$$T\left(\sum_j k_j \boldsymbol{w}_j\right) \in \ker(T^{i-1}) \iff T^{i-1}(T(\sum_j k_j \boldsymbol{w}_j)) = \mathbf{0}_V,$$

i.e., $\sum_j k_j \boldsymbol{w}_j \in \ker(T^i)$, i.e.,

$$\sum_j k_j \boldsymbol{w}_j + \ker(T^i) = \mathbf{0}_{W_{i+1}} \iff \sum_j k_j(\boldsymbol{w}_j + \ker(T^i)) = \mathbf{0}_{W_{i+1}}.$$

Since $\{\boldsymbol{w}_j + \ker(T^i), \ \forall j\}$ fomrs a basis of W_{i+1}, we imply $k_j = 0, \forall j$.

From $\mathcal{B}_{i+1}$ we construct S_i, which is linearly independent in W_i. Therefore, we imply $|T(\mathcal{B}_{i+1})| \le |\mathcal{B}_i|$ for $\forall i < m$ (why?).

- Now we start to construct a basis $\mathcal{A}$ of V:
  - Start with $\mathcal{B}_m' := \{u_1^m + \ker(T^{m-1}), \dots, u_{\ell_m}^m + \ker(T^{m-1})\}$, and $\mathcal{B}_m = \{u_1^m, \dots, u_{\ell_m}^m\}$.

– By the previous result,

$$\{T(u_1^m) + \ker(T^{m-2}), \ldots, T(u_{\ell_m}^m) + \ker(T^{m-2})\}$$

is linear independent in W_{m-1}. By basis extension, we get a basis $\mathcal{B}'_{m-1}$ of W_{m-1}, and let

$$\mathcal{B}_{m-1} = \{T(u_1^m), \ldots, T(u_{\ell_m}^m)\} \cup \xi_{m-1}$$

where $\xi_{m-1} := \{u_1^{m-1}, \ldots, u_{\ell_{m-1}}^{m-1}\}$

– Continue the process above to obtain $\mathcal{B}_{m-2}, \ldots, \mathcal{B}_1$, and $\cup_{i=1}^m \mathcal{B}_i$ forms a basis of V:

$\mathcal{B}_1$	$\mathcal{B}_2$	$\cdots$	$\mathcal{B}_{m-1}$	$\mathcal{B}_m$
$\{T^{m-1}(u_1^m), \ldots, T^{m-1}(u_{\ell_m}^m)\}$ $\{T^{m-2}(u_1^{m-1}), \ldots, T^{m-2}(u_{\ell_{m-1}}^{m-1})\}$ $\vdots$ $\{T(u_1^2), \ldots, T(u_{\ell_2}^2)\}$ $\{u_1^1, \ldots, u_{\ell_1}^1)\}$	$\{T^{m-2}(u_1^m), \ldots, T^{m-2}(u_{\ell_m}^m)\}$ $\{T^{m-3}(u_1^{m-1}), \ldots, T^{m-3}(u_{\ell_{m-1}}^{m-1})\}$ $\vdots$ $\{u_1^2, \ldots, u_{\ell_2}^2)\}$	$\cdots$ $\cdots$ $\cdots$	$\{T(u_1^m), \ldots, T(u_{\ell_m}^m)\}$ $\{u_1^{m-1}, \ldots, u_{\ell_{m-1}}^{m-1}\}$	$\{u_1^m, \ldots, u_{\ell_m}^m\}$

– Now construct the ordered basis $\mathcal{A}$:

$$\mathcal{A} = \left\{ \begin{array}{ccccc}
T^{m-1}(u_1^m) & \cdots & T^2(u_1^m) & T(u_1^m) & u_1^m \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
T^{m-1}(u_{\ell_m}^m) & \cdots & T^2(u_{\ell_m}^m) & T(u_{\ell_m}^m) & u_{\ell_m}^m \\
& T^{m-2}(u_1^{m-1}) & \cdots & T(u_1^{m-1}) & u_1^{m-1} \\
& \vdots & \ddots & \vdots & \vdots \\
& T^{m-2}(u_{\ell_{m-1}}^{m-1}) & \cdots & T(u_{\ell_{m-1}}^{m-1}) & u_{\ell_{m-1}}^{m-1} \\
& & \vdots & \ddots & \vdots \\
& & & & u_1^1 \\
& & & & \vdots \\
& & & & u_{\ell_1}^1
\end{array} \right\}$$

- Then the diagonal entries of $(T)_{\mathcal{A},\mathcal{A}}$ should be all zero, since

$$T(T^{i-1}(u_j^i)) = T^i(u_j^i) = 0, \forall i = 1, \ldots, m, j = 1, \ldots, \ell_i,$$

and every entry on the superdiagonal is 1:

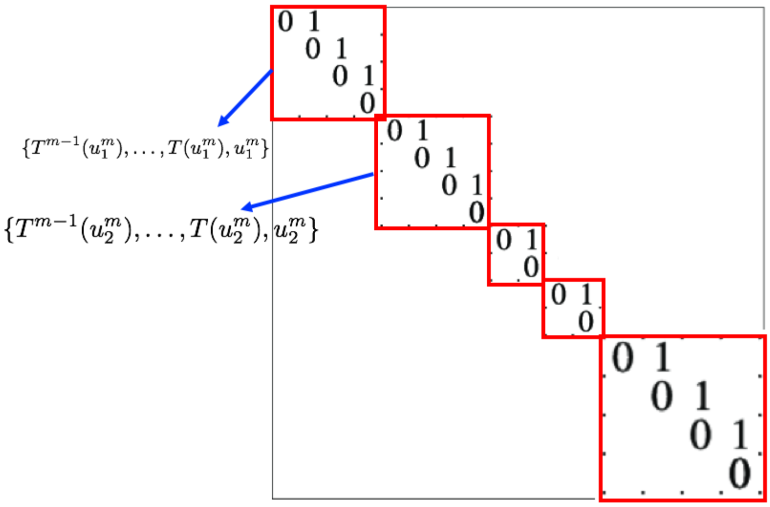


Figure 9.2: Illustration for $(T)_{\mathcal{A},\mathcal{A}}$

■

Then we consider the case where $m_T(x) = (x - \lambda)^e$:

Corollary 9.3 Suppose $T : V \to V$ is such that $m_T(x) = (x - \lambda)^e$, then the theorem (9.3) holds, i.e., there exists a basis $\mathcal{A}$ such that

$$(T)_{\mathcal{A},\mathcal{A}} = \operatorname{diag}(J_1, \ldots, J_\ell),$$

where each block J_i is a square matrix of the form

$$J_i = \begin{bmatrix} \lambda & 1 & & \\ & \lambda & \ddots & \\ & & \ddots & 1 \\ & & & \lambda \end{bmatrix}.$$

Proof. Suppose that $m_T(x) = (x - \lambda)^e$. Consider the operator $U := T - \lambda I$, then $m_U(x) = x^e$.

By applying proposition (9.6),

$$(U)_{\mathcal{A},\mathcal{A}} = \mathrm{diag}(\boldsymbol{J}_1, \dots, \boldsymbol{J}_\ell),$$

where

$$J_i = \begin{bmatrix} 0 & 1 & & \\ & 0 & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{bmatrix}.$$

Or equivalently,

$$(T)_{\mathcal{A},\mathcal{A}} - \lambda(I)_{\mathcal{A},\mathcal{A}} = \mathrm{diag}(\boldsymbol{J}_1, \dots, \boldsymbol{J}_\ell)$$

i.e.,

$$(T)_{\mathcal{A},\mathcal{A}} = \mathrm{diag}(\boldsymbol{K}_1, \dots, \boldsymbol{K}_\ell),$$

where

$$\boldsymbol{K}_i = \begin{bmatrix} \lambda & 1 & & \\ & \lambda & \ddots & \\ & & \ddots & 1 \\ & & & \lambda \end{bmatrix}$$

■

R The Jordan Normal Form Theorem (9.3) follows from our arguments using the primary decomposition.

Corollary 9.4 Any matrix $A \in M_{n\times n}(\mathbb{C})$ is similar to a matrix of the Jordan normal form

$$\mathrm{diag}(\boldsymbol{J}_1, \dots, \boldsymbol{J}_\ell).$$

9.4.2. Inner Product Spaces

Definition 9.8 [Bilinear] Let V be a vector space over $\mathbb{R}$. A bilinear form on V is a mapping

$$F : V \times V \to \mathbb{R}$$

satisfying

1. $F(\boldsymbol{u} + \boldsymbol{v}, \boldsymbol{w}) = F(\boldsymbol{u}, \boldsymbol{w}) + F(\boldsymbol{v}, \boldsymbol{w})$
2. $F(\boldsymbol{u}, \boldsymbol{v} + \boldsymbol{w}) = F(\boldsymbol{u}, \boldsymbol{v}) + F(\boldsymbol{u}, \boldsymbol{w})$
3. $F(\lambda \boldsymbol{u}, \boldsymbol{v}) = \lambda F(\boldsymbol{u}, \boldsymbol{v}) = F(\boldsymbol{u}, \lambda \boldsymbol{v})$

We say

- F is symmetric if $F(\boldsymbol{u}, \boldsymbol{v}) = F(\boldsymbol{v}, \boldsymbol{u})$
- F is non-degenerate if $F(\boldsymbol{u}, \boldsymbol{w}) = \mathbf{0}$ for $\forall \boldsymbol{u} \in V$ implies $\boldsymbol{w} = 0$
- F is positive definite if $F(\boldsymbol{v}, \boldsymbol{v}) > 0$ for $\forall \boldsymbol{v} \neq \mathbf{0}$

R If F is positive-definite, then F is non-degenerate: Suppose that $F(\boldsymbol{v}, \boldsymbol{v}) > 0, \forall \boldsymbol{v} \neq \mathbf{0}$. If we have $F(\boldsymbol{u}, \boldsymbol{w}) = 0$ for any $\boldsymbol{u} \in V$, then in particular, when $\boldsymbol{u} = \boldsymbol{w}$, we imply $F(\boldsymbol{w}, \boldsymbol{w}) = 0$. By positive-definiteness, $\boldsymbol{w} = \mathbf{0}$, i.e., F is non-degenerate.

Chapter 10

Week10

10.1. Monday for MAT3040

10.1.1. Inner Product Space

- Symmetric: $F(\boldsymbol{u},\boldsymbol{w}) = F(\boldsymbol{w},\boldsymbol{u}), \forall \boldsymbol{u},\boldsymbol{w}$
- Non-degenerate: $F(\boldsymbol{u},\boldsymbol{w}) = 0, \forall \boldsymbol{w}$ implies $\boldsymbol{u} = \boldsymbol{0}$
- Positive definite: $F(\boldsymbol{v},\boldsymbol{v}) > 0, \forall \boldsymbol{v} \neq \boldsymbol{0}$

Classification. When we say V be a vector space over $\mathbb{F}$, we treat $\alpha \in \mathbb{F}$ as a scalar.

Definition 10.1 [Sesqui-linear Form] Let V be a vector space over $\mathbb{C}$. A **sesquilinear form** on V is a function $F : V \times V \to \mathbb{C}$ such that

1. $F(\boldsymbol{u}+\boldsymbol{v},\boldsymbol{w}) = F(\boldsymbol{u},\boldsymbol{w}) + F(\boldsymbol{v},\boldsymbol{w})$
2. $F(\boldsymbol{u},\boldsymbol{v}+\boldsymbol{w}) = F(\boldsymbol{u},\boldsymbol{v}) + F(\boldsymbol{u},\boldsymbol{w})$
3. $F(\overline{\lambda}\boldsymbol{v},\boldsymbol{w}) = F(\boldsymbol{v},\lambda\boldsymbol{w}) = \lambda F(\boldsymbol{v},\boldsymbol{w}), \forall \lambda \in \mathbb{C}$

In this case, we say F is **conjugate symmetric** if

$$F(\boldsymbol{v},\boldsymbol{w}) = \overline{F(\boldsymbol{w},\boldsymbol{v})}, \quad \forall \boldsymbol{v},\boldsymbol{w} \in V.$$

The definition for non-degenerateness, and positve definiteness is the same as that in bilinear form.

R In the sesquilinear form, why there is a $\bar{\lambda}$ shown in condition (3)?

Partial Answer: We want our F to be positive definite in many cases:

- Suppose that $F(\boldsymbol{v},\boldsymbol{v}) > 0$ and we do not have $\bar{\lambda}$ in sesquilinear form F, it follows that
$$F(i\boldsymbol{v},i\boldsymbol{v}) = i^2F(\boldsymbol{v},\boldsymbol{v}) = -F(\boldsymbol{v},\boldsymbol{v}) < 0$$
As a result, there will be no positive bilinear form for vector space over $\mathbb{C}$.

Therefore, $\bar{\lambda}$ is essential to guarantee that we have a positive definite form on vector space over $\mathbb{C}$, i.e.,

$$F(i\boldsymbol{v},i\boldsymbol{v}) = \bar{i}iF(\boldsymbol{v},\boldsymbol{v}) = F(\boldsymbol{v},\boldsymbol{v})$$

Example 10.1 Consider $V = \mathbb{C}^n$, and a basic sesquilinear form is the Hermitian inner product:

$$F(\boldsymbol{v},\boldsymbol{u}) = \boldsymbol{v}^{\mathrm{H}}\boldsymbol{u} = \begin{pmatrix}\bar{v_1} & \cdots & \bar{v_n}\end{pmatrix}\begin{pmatrix}w_1\\ \vdots \\ w_n\end{pmatrix} = \sum_{i=1}^{n}\bar{v}_iw_i$$

In this case, we do not have symmetric property $F(\boldsymbol{v},\boldsymbol{w}) = F(\boldsymbol{w},\boldsymbol{v})$ any more, instead, we have the conjugate symmetric property $F(\boldsymbol{v},\boldsymbol{w}) = \overline{F(\boldsymbol{w},\boldsymbol{v})}$.

Definition 10.2 [Inner Product] A real (complex) vector space V with a bilinear (sesquilinear) form with symmetric (conjugate symmetric) and positive definite property is called an **inner product** on V. Any vector space equipped with inner product is called an **inner product space**.

Notation. We write $\langle\cdot,\cdot\rangle$ instead of $F(\cdot,\cdot)$ to denote inner product.

Definition 10.3 [Norm] The **norm** of a vector $\boldsymbol{v}$ is $\|\boldsymbol{v}\| = \sqrt{\langle\boldsymbol{v},\boldsymbol{v}\rangle}$.

R As a result, $\|\alpha\boldsymbol{v}\| = \sqrt{\langle\alpha\boldsymbol{v},\alpha\boldsymbol{v}\rangle} = \sqrt{\bar{\alpha}\alpha\langle\boldsymbol{v},\boldsymbol{v}\rangle} = \sqrt{|\alpha|^2\langle\boldsymbol{v},\boldsymbol{v}\rangle} = |\alpha|\|\boldsymbol{v}\|$.

The norm is well-defined since $\langle \boldsymbol{v}, \boldsymbol{v} \rangle \geq 0$ (positive definiteness of inner product).

Definition 10.4 [Orthogonal] We say a family of vectors $S = \{\boldsymbol{v}_i \mid i \in I\}$ is **orthogonal** if

$$\langle \boldsymbol{v}_i, \boldsymbol{v}_j \rangle = 0, \ \forall i \neq j$$

If furthermore $\langle \boldsymbol{v}_i, \boldsymbol{v}_i \rangle = 1, \forall i$, then we say S is an **orthonormal** set.

1. The Cauchy-Scharwz inequality holds for inner product space:

$$|\langle \boldsymbol{u}, \boldsymbol{v} \rangle| \leq \|\boldsymbol{u}\| \|\boldsymbol{v}\|, \ \forall \boldsymbol{u}, \boldsymbol{v} \in V.$$

Proof. The proof for $\langle \boldsymbol{u}, \boldsymbol{v} \rangle \in \mathbb{R}$ is the same as in MAT2040 course. Check Theorem (6.1) in the note

`https://walterbabyrudin.github.io/information/Notes/MAT2040.pdf`

However, for $\langle \boldsymbol{u}, \boldsymbol{v} \rangle \in \mathbb{C} \setminus \mathbb{R}$, we need the re-scaling technique: Let $\boldsymbol{w} = \frac{1}{\langle \boldsymbol{u}, \boldsymbol{v} \rangle} \boldsymbol{u}$, then $\langle \boldsymbol{w}, \boldsymbol{v} \rangle \in \mathbb{R}$:

$$\langle \boldsymbol{w}, \boldsymbol{v} \rangle = \langle \frac{1}{\overline{\langle \boldsymbol{u}, \boldsymbol{v} \rangle}} \boldsymbol{u}, \boldsymbol{v} \rangle = \overline{\left(\frac{1}{\overline{\langle \boldsymbol{u}, \boldsymbol{v} \rangle}} \right)} \langle \boldsymbol{u}, \boldsymbol{v} \rangle = \frac{1}{\langle \boldsymbol{u}, \boldsymbol{v} \rangle} \langle \boldsymbol{u}, \boldsymbol{v} \rangle = 1.$$

Applying the Cauchy-Scharwz inequality for $\langle \boldsymbol{w}, \boldsymbol{v} \rangle \in \mathbb{R}$ gives

$$\begin{aligned} \left| \langle \frac{1}{\overline{\langle \boldsymbol{u}, \boldsymbol{v} \rangle}} \boldsymbol{u}, \boldsymbol{v} \rangle \right| &= |\langle \boldsymbol{w}, \boldsymbol{v} \rangle| \\ &\leq \|\boldsymbol{w}\| \|\boldsymbol{v}\| = \left\| \frac{1}{\overline{\langle \boldsymbol{u}, \boldsymbol{v} \rangle}} \boldsymbol{u} \right\| \|\boldsymbol{v}\| \end{aligned}$$

Or equivalently,

$$\left| \frac{1}{\langle \boldsymbol{u}, \boldsymbol{v} \rangle} \right| |\langle \boldsymbol{u}, \boldsymbol{v} \rangle| \leq \left| \frac{1}{\overline{\langle \boldsymbol{u}, \boldsymbol{v} \rangle}} \right| \|\boldsymbol{u}\| \|\boldsymbol{v}\|$$

Since $\left|\frac{1}{\langle \boldsymbol{u},\boldsymbol{v}\rangle}\right| = \left|\frac{1}{\overline{\langle \boldsymbol{u},\boldsymbol{v}\rangle}}\right|$, we imply

$$|\langle \boldsymbol{u},\boldsymbol{v}\rangle| \le \|\boldsymbol{u}\|\|\boldsymbol{v}\|$$

■

2. The triangle inequality also holds for inner product process:

$$\|\boldsymbol{u}+\boldsymbol{v}\| \le \|\boldsymbol{u}\| + \|\boldsymbol{v}\|$$

3. The Gram-Schmidt process holds for finite set of vectors: let $S = \{\boldsymbol{v}_1,\dots,\boldsymbol{v}_n\}$ be (finite) linearly independent. Then we can construct an orthonormal set from S:

$$\boldsymbol{w}_1 = \boldsymbol{v}_1, \quad \boldsymbol{w}_{i+1} = \boldsymbol{v}_{i+1} - \frac{\langle \boldsymbol{v}_{i+1},\boldsymbol{w}_1\rangle}{\|\boldsymbol{w}_1\|^2} - \frac{\langle \boldsymbol{v}_{i+1},\boldsymbol{w}_2\rangle}{\|\boldsymbol{w}_2\|^2} - \cdots - \frac{\langle \boldsymbol{v}_{i+1},\boldsymbol{w}_i\rangle}{\|\boldsymbol{w}_i\|^2}, \; i = 1,\dots,n-1$$

Then after normalization, we obtain the constructed orthonormal set. Consequently, every finite dimensional inner product space has an orthonormal basis.

10.1.2. Dual spaces

Theorem 10.1 — Riesz Representation. Consider the mapping

$$\begin{aligned} \phi : \quad & V \to V^* \\ \text{with} \quad & \boldsymbol{v} \mapsto \phi_{\boldsymbol{v}} \\ \text{where} \quad & \phi_{\boldsymbol{v}}(w) = \langle \boldsymbol{v}, w\rangle, \ \forall w \in V \end{aligned}$$

Then the mapping ϕ is well-defined and it is an $\mathbb{R}$-linear transformation. Moreover, if V is finite dimensional, then ϕ is an isomorphism.

The $\mathbb{R}$-linear transformation $V \to V^*$ means that, when V, V^* are vector space over $\mathbb{R}$, the $\mathbb{R}$-linear transformation deduces into exactly the linear transformation.

R The $\mathbb{R}$-linear transformation $V \to V^*$ is **not** necessarily linear if V, V^* are vector spaces over $\mathbb{C}$.

However, we can transform a vector space over $\mathbb{C}$ into a vector space over $\mathbb{R}$:

- For example, suppose that $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_n\}$ is a basis of V over $\mathbb{C}$, i.e.,

$$\boldsymbol{v} = \sum_{j=1}^{n} \alpha_j \boldsymbol{v}_j$$

where $\alpha_j = p_j + iq_j, \forall p_j, q_j \in \mathbb{R}$, then

$$\boldsymbol{v} = \sum_j p_j \boldsymbol{v}_j + \sum_j q_j (i\boldsymbol{v}_j),\ p_j, q_j \in \mathbb{R}$$

Therefore, $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_n, i\boldsymbol{v}_1, \dots, i\boldsymbol{v}_n\}$ forms a basis of V over $\mathbb{R}$.

Note that $i\boldsymbol{v}_1$ cannot be considered as a linear combination of $\boldsymbol{v}_1$ over $\mathbb{R}$, but a linear combination of $\boldsymbol{v}_1$ over $\mathbb{C}$.

In particular, if $\phi : V \to V^*$ is a $\mathbb{R}$-linear transformation, then

$$\phi(i\boldsymbol{v}) \neq i\phi(\boldsymbol{v}), \text{ but } \phi(2\boldsymbol{v}) = 2\phi(\boldsymbol{v}).$$

Proof.

1. Well-definedness: We need to show $\phi_{\boldsymbol{v}} \in V^*$, i.e., for scalars a, b,

$$\phi_{\boldsymbol{v}}(a\boldsymbol{w}_1 + b\boldsymbol{w}_2) = \langle \boldsymbol{v}, a\boldsymbol{w}_1 + b\boldsymbol{w}_2 \rangle = a\langle \boldsymbol{v}, \boldsymbol{w}_1 \rangle + b\langle \boldsymbol{v}, \boldsymbol{w}_2 \rangle = a\phi_{\boldsymbol{v}}(\boldsymbol{w}_1) + b\phi_{\boldsymbol{v}}(\boldsymbol{w}_2)$$

   Therefore, $\phi_{\boldsymbol{v}} \in V^*$.

2. $\mathbb{R}$-linearity of ϕ: it suffices to show

$$\phi(c\boldsymbol{v}_1 + d\boldsymbol{v}_2) = c\phi(\boldsymbol{v}_1) + d\phi(\boldsymbol{v}_2), \quad \forall c, d \in \mathbb{R}, \boldsymbol{v}_1, \boldsymbol{v}_2 \in V.$$

   For all $\boldsymbol{w} \in V$, we have

$$\phi_{c\boldsymbol{v}_1 + d\boldsymbol{v}_2}(\boldsymbol{w}) = \langle c\boldsymbol{v}_1 + d\boldsymbol{v}_2, \boldsymbol{w} \rangle = c\langle \boldsymbol{v}_1, \boldsymbol{w} \rangle + d\langle \boldsymbol{v}_2, \boldsymbol{w} \rangle = c\phi_{\boldsymbol{v}_1}(\boldsymbol{w}) + d\phi_{\boldsymbol{v}_2}(\boldsymbol{w})$$

where the second equality holds because $c, d \in \mathbb{R}$.

Therefore,

$$\phi(c\boldsymbol{v}_1 + d\boldsymbol{v}_2) = c\phi(\boldsymbol{v}_1) + d\phi(\boldsymbol{v}_2).$$

∎

10.4. Wednesday for MAT3040

Reviewing. Consider the mapping

$$\begin{aligned} \phi: &\quad V \to V^* \\ \text{with} &\quad \phi(\boldsymbol{v}) = \phi_{\boldsymbol{v}} \\ \text{where} &\quad \phi_{\boldsymbol{v}}(\boldsymbol{w}) = \langle \boldsymbol{v}, \boldsymbol{w} \rangle \end{aligned}$$

The Riesz Representation Theorem claims that

1. ϕ is a $\mathbb{R}$-linear transformation.
2. ϕ is injective.
3. If $\dim(V) < \infty$, then ϕ is an isomorphism.

Proof for Claim (2). Consider the equality $\phi(\boldsymbol{v}) = \phi_{\boldsymbol{v}} = 0_{V^*}$, which implies

$$\phi_{\boldsymbol{v}}(\boldsymbol{w}) = \langle \boldsymbol{v}, \boldsymbol{w} \rangle = 0, \forall \boldsymbol{w} \in V$$

By the non-degenercy property, $\boldsymbol{v} = 0_{\boldsymbol{v}}$, i.e., ϕ is injective. ■

Proof for Claim (3). Since $\dim_{\mathbb{R}}(V) = \dim_{\mathbb{R}}(V^*)$, and ϕ is injective as a $\mathbb{R}$-linear transformation, we imply ϕ is an isomorphism from V to V^*, where V, V^* are treated as vector spaces over $\mathbb{R}$. ■

10.4.1. Orthogonal Complement

Definition 10.5 [Orthogonal Complement] Let $U \le V$ be a subspace of an inner product space. Then the **orthogonal complement** of U is

$$U^{\perp} = \{\boldsymbol{v} \in V \mid \langle \boldsymbol{v}, \boldsymbol{u} \rangle = 0, \forall \boldsymbol{u} \in U\}$$

The analysis for orthogonal complement for vector spaces over $\mathbb{C}$ is quite similar as what we have studied in MAT2040.

Proposition 10.7 1. $U^{\perp}$ is a subspace of V

2. $U \cap U^{\perp} = \{0\}$
3. $U_1 \subseteq U_2$ implies $U_2^{\perp} \leq U_1^{\perp}$.

Proof. 1. Suppose that $\boldsymbol{v}_1, \boldsymbol{v}_2 \in U^{\perp}$, where $a, b \in K$ ($K = \mathbb{C}$ or $\mathbb{R}$), then for all $\boldsymbol{u} \in U$,

$$\begin{aligned}\langle a\boldsymbol{v}_1 + b\boldsymbol{v}_2, \boldsymbol{u}\rangle &= \bar{a}\langle \boldsymbol{v}_1, \boldsymbol{u}\rangle + \bar{b}\langle \boldsymbol{v}_2, \boldsymbol{u}\rangle \\ &= \bar{a} \cdot 0 + \bar{b} \cdot 0 = 0\end{aligned}$$

Therefore, $a\boldsymbol{v}_1 + b\boldsymbol{v}_2 \in U^{\perp}$.

2. Suppose that $\boldsymbol{u} \in U \cap U^{\perp}$, then we imply $\langle \boldsymbol{u}, \boldsymbol{u}\rangle = 0$. By the positive-definiteness of inner product, $\boldsymbol{u} = \mathbf{0}$.
3. The statement (3) is easy.

■

Proposition 10.8 1. If $\dim(V) < \infty$ and $U \leq V$, then $V = U \oplus U^{\perp}$

2. If $U, W \leq V$, then

$$\begin{aligned}(U + W)^{\perp} &= U^{\perp} \cap W^{\perp} \\ (U \cap W)^{\perp} &\supseteq U^{\perp} + W^{\perp} \\ (U^{\perp})^{\perp} &\supseteq U\end{aligned}$$

Moreover, if $\dim(V) < \infty$, then these are equalities.

Proof. 1. Suppose that $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_k\}$ forms a basis for U, and by basis extension, we obtain $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_k, \boldsymbol{v}_{k+1}, \dots, \boldsymbol{v}_n\}$ is a basis for V.

By Gram-Schmidt Process, any finite basis induces an orthonormal basis.

Therefore, suppose that $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_k\}$ forms an orthonormal basis for U, and $\{\boldsymbol{e}_{k+1}, \dots, \boldsymbol{e}_n\}$ forms an orthonormal basis for $U^{\perp}$.

It's easy to show $V = U + U^{\perp}$ using orthonormal basis.

2. (a) The reverse part $(U + W)^{\perp} \supseteq U^{\perp} \cap W^{\perp}$ is trivial; for the forward part, suppose

$\boldsymbol{v} \in (U+W)^{\perp}$, then

$$\langle \boldsymbol{v}, \boldsymbol{u}+\boldsymbol{w}\rangle = 0,\ \forall \boldsymbol{u} \in U,\ \boldsymbol{w} \in W$$

Taking $\boldsymbol{u} \equiv \mathbf{0}$ in the equality above gives $\langle \boldsymbol{v}, \boldsymbol{w}\rangle = 0$, i.e., $\boldsymbol{v} \in U^{\perp}$. Similarly, $\boldsymbol{v} \in W^{\perp}$.

(b) Follow the similar argument as in (2a). If $\dim(V) < \infty$, then write down the orthonormal basis for $U^{\perp} + W^{\perp}$ and $(U \cap W)^{\perp}$.

(c) Follow the similar argument as in (2a). If $\dim(V) < \infty$, then

$$V = U^{\perp} \oplus (U^{\perp})^{\perp} = U \oplus U^{\perp}.$$

Therefore, $(U^{\perp})^{\perp} = U$.

■

Proposition 10.9 The mapping $\phi : V \to V^*$ maps $U^{\perp} \leq V$ injectively to $\mathrm{Ann}(U) \leq V^*$. If $\dim(V) < \infty$, then $U^{\perp} \cong \mathrm{Ann}(U)$ as $\mathbb{R}$-vector spaces

Proof. The injectivity of ϕ has been shown at the beginning of this lecture. For any $\boldsymbol{v} \in U^{\perp}$, we imply $\phi_{\boldsymbol{v}}(\boldsymbol{u}) = 0, \forall \boldsymbol{u} \in U$, i.e., $\phi_{\boldsymbol{v}} \in \mathrm{Ann}(U)$.

Therefore, $\phi(U^{\perp}) \leq \mathrm{Ann}(U)$.

Provided that $\dim(V) < \infty$, by (1) in proposition (10.8),

$$\dim(U) + \dim(U^{\perp}) = \dim(V)$$

Since $\dim(U) + \dim(\mathrm{Ann}(U)) = \dim(V)$, we imply $\dim(U^{\perp}) = \dim(\mathrm{Ann}(U))$.

Moreover,

$$\phi : U^{\perp} \to \mathrm{Ann}(U)$$

is an isomorphism between $\mathbb{R}$-vector spaces $U^{\perp}$ and $\mathrm{Ann}(U)$. ■

10.4.2. Adjoint Map

Motivation. Then we study the induced mapping based on a given linear operator T, denoted as T'. This induced mapping essentially plays the similar role as taking the Hermitian for a complex matrix.

Notation. Previously we have studied the **adjoint** of $T : V \to W$, denoted as $T^* : W^* \to V^*$. However, from now on, we use the same terminology but with different meaning. If $T : V \to V$ is a linear operator, then the **adjoint** of T is the linear operator $T' : V \to V$ defined as follows.

Definition 10.6 [Adjoint] Let $T : V \to V$ be a linear operator between inner product spaces. The **adjoint** of T is defined as $T' : V \to V$ satisfying

$$\langle T'(\boldsymbol{v}), \boldsymbol{w} \rangle = \langle \boldsymbol{v}, T(\boldsymbol{w}) \rangle, \ \forall \boldsymbol{w} \in V \tag{10.1}$$

■

Proposition 10.10 If $\dim(V) < \infty$, then T' exists, and it is unique. Moreove, T' is a linear map.

Proof. Fix any $\boldsymbol{v} \in V$. Consider the mapping

$$\alpha_{\boldsymbol{v}} : \boldsymbol{w} \xrightarrow{T} T(\boldsymbol{w}) \xrightarrow{\phi_{\boldsymbol{v}}} \langle \boldsymbol{v}, T(\boldsymbol{w}) \rangle$$

This is a linear transformation from V to $\mathbb{F}$, i.e., $\alpha_{\boldsymbol{v}} \in V^*$

By Riesz representation theorem, ϕ is an isomorphism from V to V^*. Therefore, for any $\alpha_{\boldsymbol{v}} \in V^*$, there exists a vector $T'(\boldsymbol{v}) \in V$ such that

$$\phi(T'(\boldsymbol{v})) = \alpha_{\boldsymbol{v}} \in V^*$$

Or equivalently, $\phi_{T'(\boldsymbol{v})}(\boldsymbol{w}) = \alpha_{\boldsymbol{v}}(\boldsymbol{w}), \forall \boldsymbol{w} \in V$, i.e., $\langle T'(\boldsymbol{v}), \boldsymbol{w} \rangle = \langle \boldsymbol{v}, T(\boldsymbol{w}) \rangle$.

Therefore, from $\boldsymbol{v}$ we have constructed $T'(\boldsymbol{v})$ satisfying (10.1). Now define $T' : V \to V$ by $\boldsymbol{v} \mapsto T'(\boldsymbol{v})$.

- Since the choice of $T'(\boldsymbol{v})$ is unique by the injectivity of ϕ, T' is well-defined.
- Now we show T' is a linear transformation: Let $\boldsymbol{v}_1, \boldsymbol{v}_2 \in V, a, b \in K$. For all $\boldsymbol{w} \in V$, we have

$$\begin{aligned}
\langle T'(a\boldsymbol{v}_1 + b\boldsymbol{v}_2), \boldsymbol{w}\rangle &= \langle a\boldsymbol{v}_1 + b\boldsymbol{v}_2, T(\boldsymbol{w})\rangle \\
&= \bar{a}\langle \boldsymbol{v}_1, T(\boldsymbol{w})\rangle + \bar{b}\langle \boldsymbol{v}_2, T(\boldsymbol{w})\rangle \\
&= \bar{a}\langle T'(\boldsymbol{v}_1), \boldsymbol{w}\rangle + \bar{b}\langle T'(\boldsymbol{v}_2), \boldsymbol{w}\rangle \\
&= \langle aT'(\boldsymbol{v}_1) + bT'(\boldsymbol{v}_2), \boldsymbol{w}\rangle
\end{aligned}$$

Therfore,

$$\langle T'(a\boldsymbol{v}_1 + b\boldsymbol{v}_2) - [aT'(\boldsymbol{v}_1) + bT'(\boldsymbol{v}_2)], \boldsymbol{w}\rangle = 0, \ \forall \boldsymbol{w} \in V$$

By the non-degeneracy of inner product,

$$T'(a\boldsymbol{v}_1 + b\boldsymbol{v}_2) - [aT'(\boldsymbol{v}_1) + bT'(\boldsymbol{v}_2)] = \mathbf{0},$$

i.e., $T'(a\boldsymbol{v}_1 + b\boldsymbol{v}_2) = aT'(\boldsymbol{v}_1) + bT'(\boldsymbol{v}_2)$

■

▪ **Example 10.2** Let $V = \mathbb{R}^n$, $\langle \cdot, \cdot \rangle$ as the usual inner product. Consider the matrix-multiplication mapping

$$\begin{aligned}
T: \quad & V \to V \\
& T(\boldsymbol{v}) = A\boldsymbol{v}
\end{aligned}$$

Then $\langle T'(\boldsymbol{v}), \boldsymbol{w}\rangle = \langle \boldsymbol{v}, T(\boldsymbol{w})\rangle$ implies

$$\begin{aligned}
(T'(\boldsymbol{v}))^{\mathrm{T}}\boldsymbol{w} &= \langle \boldsymbol{v}, \boldsymbol{A}\boldsymbol{w}\rangle \\
&= \boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{w} \\
&= (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{v})^{\mathrm{T}}\boldsymbol{w}
\end{aligned}$$

Therfore, $T'(\boldsymbol{v}) = A^{\mathrm{T}}\boldsymbol{v}$. ▪

Proposition 10.11 Let $T : V \to V$ be a linear transformation, V a inner product space. Suppose that $\mathcal{B} = \{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$ is an orthonormal basis of V, then

$$(T')_{\mathcal{B},\mathcal{B}} = \overline{((T)_{\mathcal{B},\mathcal{B}})^{\mathrm{T}}}$$

Proof. Suppose that $(T)_{\mathcal{B},\mathcal{B}} = (a_{ij})$, where $T(\boldsymbol{e}_j) = \sum_{k=1}^{n} a_{kj}\boldsymbol{e}_k$, then

$$\begin{aligned}\langle \boldsymbol{e}_i, T(\boldsymbol{e}_j)\rangle &= \langle \boldsymbol{e}_i, \sum_{k=1}^{n} a_{kj}\boldsymbol{e}_k\rangle \\ &= \sum_{k=1}^{n} a_{kj}\langle \boldsymbol{e}_i, \boldsymbol{e}_k\rangle \\ &= a_{ij}\end{aligned}$$

Also, suppose $(T')_{\mathcal{B},\mathcal{B}} = (b_{ij})$, we imply $T'(\boldsymbol{e}_j) = \sum_{k=1}^{n} b_{ij}\boldsymbol{e}_k$, which follows that

$$\langle \boldsymbol{e}_i, T'(\boldsymbol{e}_j)\rangle = b_{ij} \implies \overline{\langle T'(\boldsymbol{e}_j), \boldsymbol{e}_i\rangle} = b_{ij} \implies \overline{\langle \boldsymbol{e}_j, T(\boldsymbol{e}_i)\rangle} = b_{ij},$$

i.e., $\overline{a_{ji}} = b_{ij}$. ■

R Proposition (10.11) does not hold if $\mathcal{B}$ is not an orthonormal basis.

Chapter 11

Week11

11.1. Monday for MAT3040

Reviewing. Adjoint Operator: $\langle T'(\boldsymbol{v}), \boldsymbol{w}\rangle = \langle \boldsymbol{v}, T(\boldsymbol{w})\rangle$.

11.1.1. Self-Adjoint Operator

Definition 11.1 [Self-Adjoint] Let V be an inner product space and $T: V \to V$ be a linear operator. Then T is **self-adjoint** if $T' = T$.

Example 11.1 Let $V = \mathbb{C}^n$, and $\mathcal{B} = \{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$ be a orthonormal basis. Let $T: V \to V$ be given by

$$T(\boldsymbol{v}) = \boldsymbol{A}\boldsymbol{v}, \quad \text{where } A \in M_{n\times n}(\mathbb{C}).$$

Or equivalently, there exists basis $\mathcal{B}$ such that $(T)_{\mathcal{B},\mathcal{B}} = \boldsymbol{A}$.

In such case, T is self-adjoint if and only if $(T')_{\mathcal{B},\mathcal{B}} = (T)_{\mathcal{B},\mathcal{B}}$, i.e., $\overline{(T)^{\mathrm{T}}_{\mathcal{B},\mathcal{B}}} = (T)_{\mathcal{B},\mathcal{B}}$, i.e., $\boldsymbol{A}^{\mathrm{H}} = \boldsymbol{A}$.

Therefore, $T(\boldsymbol{v}) = \boldsymbol{A}\boldsymbol{v}$ is self-adjoint if and only if $\boldsymbol{A}^{\mathrm{H}} = \boldsymbol{A}$.

Moreover, if $\mathbb{C}$ is replaced by $\mathbb{R}$, then T is seld-adjoint if and only if $\boldsymbol{A}$ is symmetric.

(R) The notion of self-adjoint for linear operator is essentially the generalized notion of Hermitian for matrix that we have stuided in MAT2040.

We also have some nice properties for self-adjoint, and the proof for which are essentially the same for the proof in the case of Hermitian matrices.

Proposition 11.1 If λ is an eigenvalue of a self-adjoint operator T, then $\lambda \in \mathbb{R}$.

Proof. Suppose there is an eigen-pair $(\lambda, \boldsymbol{w})$ for $\boldsymbol{w} \neq \boldsymbol{0}$, then

$$\begin{aligned}
\lambda\langle \boldsymbol{w}, \boldsymbol{w}\rangle &= \langle \boldsymbol{w}, \lambda\boldsymbol{w}\rangle \\
&= \langle \boldsymbol{w}, T(\boldsymbol{w})\rangle = \langle T'(\boldsymbol{w}), \boldsymbol{w}\rangle \\
&= \langle T(\boldsymbol{w}), \boldsymbol{w}\rangle = \langle \lambda\boldsymbol{w}, \boldsymbol{w}\rangle \\
&= \bar{\lambda}\langle \boldsymbol{w}, \boldsymbol{w}\rangle
\end{aligned}$$

Since $\langle \boldsymbol{w}, \boldsymbol{w}\rangle \neq 0$ by non-degeneracy property, we have $\lambda = \bar{\lambda}$, i.e., $\lambda \in \mathbb{R}$. ■

Proposition 11.2 If $U \leq V$ is T-invariant over the self-adjoint operator T, then so is $U^\perp$.

Proof. It suffices to show $T(\boldsymbol{v}) \in U^\perp, \forall \boldsymbol{v} \in U^\perp$, i.e., for any $\boldsymbol{u} \in U$, check that

$$\langle \boldsymbol{u}, T(\boldsymbol{v})\rangle = \langle T'(\boldsymbol{u}), \boldsymbol{v}\rangle = \langle T(\boldsymbol{u}), \boldsymbol{v}\rangle = 0,$$

where the last equality is because that $T(\boldsymbol{u}) \in U$ and $\boldsymbol{v} \in U^\perp$. Therefore, $T(\boldsymbol{v}) \in U^\perp$. ■

Theorem 11.1 If $T : V \to V$ is self-adjoint, and $\dim(V) < \infty$, then there exists an orthonormal basis of eigenvectors of T, i.e., an orthonormal basis of V such that any element from this basis is an eigenvector of T.

Proof. We use the induction on $\dim(V)$:

- The result is trival for $\dim(V) = 1$.
- Suppose that this theorem holds for all vector spaces V with $\dim(V) \leq k$, then we want to show the theorem holds when $\dim(V) = k+1$:

  Suppose that $T : V \to V$ is self-adjoint with $\dim(V) = k+1$, then consider

  $$\mathcal{X}_T(x) = x^{k+1} + \cdots + a_1 x + a_0, \quad a_i \in \mathbb{K}, \text{ where } \mathbb{K} \text{ denotes } \mathbb{R} \text{ or } \mathbb{C}.$$

- If $\mathbb{K} = \mathbb{C}$, then $\mathcal{X}_T(x)$ can be decomposed as

$$\mathcal{X}_T(x) = (x - \lambda_1) \cdots (x - \lambda_{k+1})$$

  In paricular, we obtain the eigen-pair $(\lambda_1, \boldsymbol{v})$

- If $\mathbb{K} = \mathbb{R}$, i.e., we treat real number as scalars, then

$$\mathcal{X}_T(x) = (x - \lambda_1) \cdots (x - \lambda_{k+1}), \text{ where } \lambda_i \in \mathbb{C}.$$

  By proposition (11.1), we imply all λ_i's are in $\mathbb{R}$. Moreover, we also obtain the eigen-pair $(\lambda_1, \boldsymbol{v})$

Consider $U = \text{span}\{\boldsymbol{v}\}$, then

- U is T-invariant
- $V = U \oplus U^{\perp}$, since V is finite dimensional
- $U^{\perp}$ is T-invariant.

Consider $T\mid_{U^{\perp}}$, which is a self-adjoint operator on $U^{\perp}$, with $\dim(U^{\perp}) = k$.

By induction, there exists an orthonormal basis $\{\boldsymbol{e}_2, \dots, \boldsymbol{e}_{k+1}\}$ of eigenvectors of $T\mid_{U^{\perp}}$.

Consider the basis $\mathcal{B} = \{\boldsymbol{v}' = \boldsymbol{v}/\|\boldsymbol{v}\|, \boldsymbol{e}_2, \dots, \boldsymbol{e}_{k+1}\}$. As a result,

1. $\mathcal{B}$ forms a basis of V
2. All $\boldsymbol{v}', \boldsymbol{e}_i$ are of norm 1 eigenvectors of T.
3. $\mathcal{B}$ is an orthonormal set, e.g., $\langle \boldsymbol{v}', \boldsymbol{e}_i \rangle = 0$, where $\boldsymbol{v}' \in U$ and $\boldsymbol{e}_i \in U^{\perp}$.

Therefore, $\mathcal{B}$ is a basis of orthonormal eigenvectors of V.

∎

Corollary 11.1 If $\dim(V) < \infty$, and $T : V \to V$ is self-adjoint, then there exists orthonormal basis $\mathcal{B}$ such that

$$(T)_{\mathcal{B},\mathcal{B}} = \text{diag}(\lambda_1, \dots, \lambda_n)$$

In paticular, for all real symmtric matrix $\boldsymbol{A} \in \mathbb{S}^n$, there exists orthogonal matrix P ($P^{\mathrm{T}}P = \boldsymbol{I}_n$) such that

$$P^{-1}AP = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$$

Proof. 1. By applying theorem (11.1), there exists orthonormal basis of V, say $\mathcal{B} = \{\boldsymbol{v}_1, \dots, \boldsymbol{v}_n\}$ such that $T(\boldsymbol{v}_i) = \lambda_i \boldsymbol{v}_i$. Directly writing the basis representation gives

$$(T)_{\mathcal{B},\mathcal{B}} = \operatorname{diag}(\lambda_1, \dots, \lambda_n).$$

2. For the second part, consider $T : \mathbb{R}^n \to \mathbb{R}^n$ by $T(\boldsymbol{v}) = \boldsymbol{A}\boldsymbol{v}$. Since $\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{A}$, we imply T is self-adjoint. There exists orthonormal basis $\mathcal{B} = \{\boldsymbol{v}_1, \dots, \boldsymbol{v}_n\}$ such that

$$(T)_{\mathcal{B},\mathcal{B}} = \operatorname{diag}(\lambda_1, \dots, \lambda_n).$$

In particular, if $\mathcal{A} = \{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$, then $(T)_{\mathcal{A},\mathcal{A}} = \boldsymbol{A}$. We construct $P := C_{\mathcal{A},\mathcal{B}}$, which is the change of basis matrix from $\mathcal{B}$ to $\mathcal{A}$, then

$$P = \begin{pmatrix} \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_n \end{pmatrix}$$

and

$$P^{-1}(T)_{\mathcal{A},\mathcal{A}}P = (T)_{\mathcal{B},\mathcal{B}}$$

Or equivalently, $P^{-1}AP = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$, with

$$P^{\mathrm{T}}P = \begin{pmatrix} \boldsymbol{v}_1^{\mathrm{T}} \\ \vdots \\ \boldsymbol{v}_n^{\mathrm{T}} \end{pmatrix} \begin{pmatrix} \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_n \end{pmatrix} = \boldsymbol{I}$$

∎

11.1.2. Orthononal/Unitary Operators

Definition 11.2 A linear operator $T : V \to V$ over $\mathbb{K}$ with $\langle T(\boldsymbol{w}), T(\boldsymbol{v})\rangle = \langle \boldsymbol{w}, \boldsymbol{v}\rangle, \forall \boldsymbol{v}, \boldsymbol{w} \in V$, is called

1. **Orthogonal** if $\mathbb{K} = \mathbb{R}$
2. **Unitary** if $\mathbb{K} = \mathbb{C}$

Proposition 11.3 T is orthogonal / unitary if and only if $T' \circ T = I$

Proof. The reverse direction is by directly checking that

$$\langle T(\boldsymbol{w}), T(\boldsymbol{v})\rangle = \langle T' \circ T(\boldsymbol{w}), \boldsymbol{v}\rangle = \langle \boldsymbol{w}, \boldsymbol{v}\rangle$$

The forward direction is by checking $T' \circ T(\boldsymbol{w}) = \boldsymbol{w}, \forall \boldsymbol{w} \in V$:

$$\langle T' \circ T(\boldsymbol{w}), \boldsymbol{v}\rangle = \langle T(\boldsymbol{w}), T(\boldsymbol{v})\rangle = \langle \boldsymbol{w}, \boldsymbol{v}\rangle \implies \langle T' \circ T(\boldsymbol{w}) - \boldsymbol{w}, \boldsymbol{v}\rangle = 0, \forall \boldsymbol{v} \in V$$

By non-degeneracy, $T' \circ T(\boldsymbol{w}) - \boldsymbol{w} = 0$, i.e., $T' \circ T(\boldsymbol{w}) = \boldsymbol{w}, \forall \boldsymbol{w} \in V$. ∎

Example 11.2 Let $T : \mathbb{K}^n \to \mathbb{K}^n$ be given by $T(\boldsymbol{v}) = A\boldsymbol{v}$. Then T is orthogonal implies $(T')_{\mathcal{B},\mathcal{B}}(T)_{\mathcal{B},\mathcal{B}} = I$.

(Orthogonal) When $\mathbb{K} = \mathbb{R}$, then $A^{\mathrm{T}}A = I$

(Unitary) When $\mathbb{K} = \mathbb{C}$, then $A^{\mathrm{H}}A = I$.

Definition 11.3 [Orthogonal/Unitary Group]

$$\text{Orthognoal Group} : O(n, \mathbb{R}) = \{A \in M_{n\times n}(\mathbb{R}) \mid A^{\mathrm{T}}A = I\}$$

$$\text{Unitary Group} : U(n, \mathbb{C}) = \{A \in M_{n\times n}(\mathbb{C}) \mid A^{\mathrm{H}}A = I\}$$

11.4. Wednesday for MAT3040

Reviewing. Unitary Operators

$$\langle T\boldsymbol{v}, T\boldsymbol{w}\rangle = \langle \boldsymbol{v}, \boldsymbol{w}\rangle,\ \forall \boldsymbol{v}, \boldsymbol{w} \in V.$$

11.4.1. Unitary Operator

▪ **Example 11.8** Let $V = \mathbb{R}^n$ with usual inner product. For the linear operator $T(\boldsymbol{v}) = \boldsymbol{A}\boldsymbol{v}$, T is orthogonal if and only if $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{I}$.

Let $V = \mathbb{C}^n$ with usual inner product. For the linear operator $T(\boldsymbol{v}) = \boldsymbol{A}\boldsymbol{v}$, T is unitary if and only if $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A} = \boldsymbol{I}$. ▪

Proposition 11.8 Let $T: V \to V$ be a linear operator on a vector space over $\mathbb{K}$ satisfying $T'T = I$. Then for all eigenvalues λ of T, we have $|\lambda| = 1$.

Proof. Suppose we have the eigen-pair $(\lambda, \boldsymbol{v})$, then

$$\begin{aligned} \langle T\boldsymbol{v}, T\boldsymbol{v}\rangle &= \langle \boldsymbol{v}, \boldsymbol{v}\rangle \\ \iff \langle \lambda\boldsymbol{v}, \lambda\boldsymbol{v}\rangle &= \langle \boldsymbol{v}, \boldsymbol{v}\rangle \\ \iff \bar{\lambda}\lambda\langle \boldsymbol{v}, \boldsymbol{v}\rangle &= \langle \boldsymbol{v}, \boldsymbol{v}\rangle \end{aligned}$$

Since $\langle \boldsymbol{v}, \boldsymbol{v}\rangle \neq 0$ ($\boldsymbol{v} \neq \boldsymbol{0}$), we imply $|\lambda|^2 = 1$, i.e., $|\lambda| = 1$. ∎

Proposition 11.9 Let $T: V \to V$ be an operator on a finite dimension V over $\mathbb{K}$ satisfying $T'T = I$. If $U \le V$ is T-invariant, then U is also T^{-1}-invariant.

Proof. Since $T'T = I$, i.e., T is invertible, we imply 0 is not a root of $\mathcal{X}_T(x)$, i.e., 0 is not a root of $m_T(x)$. Since $m_T(0) \neq 0$, $m_T(x)$ has the form

$$m_T(x) = x^m + \cdots + a_1 x + a_0,\ a_0 \neq 0,$$

which follows that

$$m_T(T) = T^m + \cdots + a_0 I = 0 \implies T(T^{m-1} + \cdots + a_1 I) = -a_0 I$$

Or equivalently,

$$T\left(-\frac{1}{a_0}(T^{m-1} + \cdots + a_1 I)\right) = I$$

Therefore,

$$T^{-1} = -\frac{1}{a_0}T^{m-1} - \cdots - \frac{a_2}{a_0}T - \frac{a_1}{a_0}I,$$

i.e., the inverse T^{-1} can be expressed as a polynomial involving T only.

Sicne U is T-invariant, we imply U is T^m-invariant for $m \in \mathbb{N}$, and therefore U is T^{-1}-invariant since T^{-1} is a polynomial of T. ∎

Proposition 11.10 Let $T : V \to V$ satisfies $T'T = I$ ($\dim(V) < \infty$), then $U \leq V$ is T-invariant implies $U^\perp$ is T-invariant.

Proof. Let $v \in U^\perp$, it suffices to show $T(v) \in U^\perp$.

For all $u \in U$, we have

$$\langle u, T(v)\rangle = \langle T'(u), v\rangle = \langle T^{-1}(u), v\rangle$$

Since U is T^{-1}-invaraint, we imply $T^{-1}(u) \in U$, and therefore

$$\langle u, T(v)\rangle = \langle T^{-1}(u), v\rangle = 0 \implies T(v) \in U^\perp.$$

∎

Theorem 11.2 Let $T : V \to V$ be a unitary operator on finite dimension V (over $\mathbb{C}$), then there exists an orthonormal basis $\mathcal{A}$ such that

$$(T)_{\mathcal{A},\mathcal{A}} = \operatorname{diag}(\lambda_1, \ldots, \lambda_n), \ |\lambda_i| = 1, \ \forall i.$$

Proof Outline. Note that $\mathcal{X}_T(x)$ always admits a root in $\mathbb{C}$, so we can always find an eigenvector $\boldsymbol{v} \in V$ of T.

Then the theorem follows by the same argument before on seld-adjoint operators.

- Consider $U = \operatorname{span}\{\boldsymbol{v}\}$
- $V = U \oplus U^{\perp}$ and $U^{\perp}$ is T-invariant
- Use induction on the unitary operator $T\,|_{U^{\perp}} : U^{\perp} \to U^{\perp}$

■

R

- The argument fails for orthogonal operators

$$\begin{aligned} T \quad & : \mathbb{R} \to \mathbb{R}^2, \\ \text{with} \quad & T(\boldsymbol{v}) = \boldsymbol{A}\boldsymbol{v} \\ \text{where} \quad & \boldsymbol{A} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \end{aligned}$$

  The matrix $\boldsymbol{A}$ is not diagonalizable over $\mathbb{R}$. It has no real eigenvalues.

  However, if we treat $\boldsymbol{A}$ as $T : \mathbb{C}^2 \to \mathbb{C}^2$ with $T(\boldsymbol{v}) = \boldsymbol{A}\boldsymbol{v}$, then $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A} = \boldsymbol{I}$, and therefore T is unitary. Then $\boldsymbol{A}$ is diagonalizable over $\mathbb{C}$ with eigenvalues $e^{i\theta}, e^{-i\theta}$

- As a corollary of the theorem, for all $\boldsymbol{A} \in M_{n\times n}(\mathbb{C})$ satisfying $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A} = \boldsymbol{I}$, there exists $P \in M_{n\times n}(\mathbb{C})$ such that

$$P^{-1}AP = \operatorname{diag}(\lambda_1, \dots, \lambda_n), \quad |\lambda_i| = 1,$$

  where $P = (\boldsymbol{u}_1, \dots, \boldsymbol{u}_n)$, with $\{\boldsymbol{u}_1, \dots, \boldsymbol{u}_n\}$ forming orthonormal basis of $\mathbb{C}^n$.

  In fact,

$$P^{\mathrm{H}}P = \begin{pmatrix} \boldsymbol{u}_1^{\mathrm{H}} \\ \vdots \\ \boldsymbol{u}_n^{\mathrm{H}} \end{pmatrix} \begin{pmatrix} \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_n \end{pmatrix} = \begin{pmatrix} \langle \boldsymbol{u}_1, \boldsymbol{u}_1 \rangle & \cdots & \langle \boldsymbol{u}_1, \boldsymbol{u}_n \rangle \\ \vdots & \ddots & \vdots \\ \langle \boldsymbol{u}_n, \boldsymbol{u}_1 \rangle & \cdots & \langle \boldsymbol{u}_n, \boldsymbol{u}_n \rangle \end{pmatrix}$$

Conclusion: all matrices $\boldsymbol{A} \in M_{n\times n}(\mathbb{C})$ with $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A} = \boldsymbol{I}$ can be written as

$$\boldsymbol{A} = \boldsymbol{P}^{-1}\operatorname{diag}(\lambda_1,\dots,\lambda_n)\boldsymbol{P},$$

with some $\boldsymbol{P}$ satisfying $\boldsymbol{P}^{\mathrm{H}}\boldsymbol{P} = \boldsymbol{I}$.

Notation. Let $U(n) = \{\boldsymbol{A} \in M_{n\times n}(\mathbb{C}) \mid \boldsymbol{A}^{\mathrm{H}}\boldsymbol{A} = \boldsymbol{I})\}$ be the unitary group, then all $\boldsymbol{A} \in U(n)$ can be diagonalized by

$$A = P^{-1}\operatorname{diag}(\lambda_1,\dots,\lambda_n)P, \quad P \in U(n).$$

11.4.2. Normal Operators

Definition 11.10 [Normal] Let $T : V \to V$ be a linear operator over a $\mathbb{C}$ inner product vector space V. We say T is **normal**, if

$$T'T = TT'$$

■

■ **Example 11.9** • All self-adjoint operators are normal:

$$T = T' \implies TT' = T'T = T^2$$

• All (finite-dimensional) unitary operators are normal:

$$T'T = TT' = I$$

■

Proposition 11.11 Let T be a normal operator on V. Then

1. $\|T(\boldsymbol{v})\| = \|T'(\boldsymbol{v})\|, \forall \boldsymbol{v} \in V$.

   In particular, $T(\boldsymbol{v}) = 0$ if and only if $T'(\boldsymbol{v}) = 0$

2. $(T - \lambda I)$ is also a normal operator, for any $\lambda \in \mathbb{C}$
3. $T(\boldsymbol{v}) = \lambda \boldsymbol{v}$ if and only if $T'(\boldsymbol{v}) = \bar{\lambda}\boldsymbol{v}$.

Proof. 1.

$$\begin{aligned}
\langle T\boldsymbol{v}, T\boldsymbol{v}\rangle &= \langle T'T\boldsymbol{v}, \boldsymbol{v}\rangle \\
&= \langle TT'\boldsymbol{v}, \boldsymbol{v}\rangle \\
&= \overline{\langle \boldsymbol{v}, TT'\boldsymbol{v}\rangle} \\
&= \overline{\langle T'\boldsymbol{v}, T'\boldsymbol{v}\rangle} \\
&= \langle T'\boldsymbol{v}, T'\boldsymbol{v}\rangle
\end{aligned}$$

Therefore, $\|T(\boldsymbol{v})\|^2 = \|T'(\boldsymbol{v})\|^2$, i.e., $\|T(\boldsymbol{v})\| = \|T'(\boldsymbol{v})\|$.

2. By hw4, $(T - \lambda I)' = T' - \overline{\lambda} I$. It suffices to check

$$(T - \lambda I)'(T - \lambda I) = (T - \lambda I)(T - \lambda I)',$$

Expanding both sides out gives the desired result, i.e.,

$$(T - \lambda I)'(T - \lambda I) = (T' - \bar{\lambda} I)(T - \lambda I) = T'T - \bar{\lambda} T - \lambda T' + \lambda\bar{\lambda} I$$

and

$$(T - \lambda I)(T - \lambda I)' = (T - \lambda I)(T' - \bar{\lambda} I) = TT' - \bar{\lambda} T - \lambda T' + \lambda\bar{\lambda} I$$

3. The proof for (3) will be discussed in the next lecture.

■

Chapter 12

Week12

12.1. Monday for MAT3040

12.1.1. Remarks on Normal Operator

Proposition 12.1 If T is normal, then

1. $\|T(\boldsymbol{v})\| = \|T'(\boldsymbol{v})\|$ for any $\boldsymbol{v} \in V$
2. $(T - \lambda I)$ is normal for any $\lambda \in \mathbb{C}$
3. $T(\boldsymbol{v}) = \lambda \boldsymbol{v}$ if and only if $T'(\boldsymbol{v}) = \bar{\lambda}\boldsymbol{v}$
4. If $T(\boldsymbol{v}) = \lambda \boldsymbol{v}$ and $T(\boldsymbol{w}) = \mu \boldsymbol{w}$ with $\lambda \neq \mu$, then $\langle \boldsymbol{v}, \boldsymbol{w} \rangle = 0$.

Proof. (3)
- For the forward direction, if $(T - \lambda I)\boldsymbol{v} = 0$, then by part (2), $(T - \lambda I)$ is normal, which follows that

$$\|(T - \lambda I)'(\boldsymbol{v})\| = 0 \implies (T - \lambda I)'(\boldsymbol{v}) = 0 \implies T'\boldsymbol{v} = \bar{\lambda}\boldsymbol{v}.$$

- For the reverse direction, suppose that $(T' - \bar{\lambda} I)\boldsymbol{v} = 0$. Since T is normal, we imply T' is normal. Then by part (2), $(T' - \bar{\lambda} I)$ is normal. By applying the same trick,

$$(T' - \bar{\lambda} I)'\boldsymbol{v} = 0 \implies ((T')' - \bar{\bar{\lambda}} I)\boldsymbol{v} = 0.$$

  By hw4, $(T')' = T$. Therefore, $(T - \lambda I)\boldsymbol{v} = 0$.

(4) Observe that

$$\lambda\langle \boldsymbol{v}, \boldsymbol{w} \rangle = \langle \bar{\lambda}\boldsymbol{v}, \boldsymbol{w} \rangle \xrightarrow{\text{by (3)}} \lambda\langle \boldsymbol{v}, \boldsymbol{w} \rangle = \langle T'(\boldsymbol{v}), \boldsymbol{w} \rangle = \langle \boldsymbol{v}, T(\boldsymbol{w}) \rangle = \langle \boldsymbol{v}, \mu \boldsymbol{w} \rangle = \mu \langle \boldsymbol{v}, \boldsymbol{w} \rangle$$

Since $\lambda \neq \mu$, we imply $\langle \boldsymbol{v}, \boldsymbol{w} \rangle = 0$. The proof is complete.

■

Theorem 12.1 Let T be an operator on a finite dimensional ($\dim(V) = n$) $\mathbb{C}$-inner product vector space V satisfying $T'T = TT'$. Then there is an orthonormal basis of eigenvectors of V, i.e., an orthonormal basis of V such that any element from this basis is an eigenvector of T.

Proof. Since $\mathcal{X}_T(x)$ must have a root in $\mathbb{C}$, there must exist an eigen-pair $(\boldsymbol{v}, \lambda)$ of T.

- Construct $U = \text{span}\{\boldsymbol{v}\}$, and it follows that

$$T\boldsymbol{v} = \lambda \boldsymbol{v} \implies U \text{ is } T\text{-invariant.}$$

$$T'\boldsymbol{v} = \bar{\lambda} \boldsymbol{v} \implies U \text{ is } T'\text{-invariant.}$$

- Moreover, we claim that $U^{\perp}$ is T and T' invariant: let $\boldsymbol{w} \in U^{\perp}$, and for all $\boldsymbol{u} \in U$, we have

$$\langle \boldsymbol{u}, T(\boldsymbol{w}) \rangle = \langle T'(\boldsymbol{u}), \boldsymbol{w} \rangle = \langle \bar{\lambda} \boldsymbol{u}, \boldsymbol{w} \rangle = \lambda \langle \boldsymbol{u}, \boldsymbol{w} \rangle = 0,$$

  i.e., $U^{\perp}$ is T invariant.

$$\langle \boldsymbol{u}, T'(\boldsymbol{w}) \rangle = \langle T(\boldsymbol{u}), \boldsymbol{w} \rangle = \langle \lambda \boldsymbol{u}, \boldsymbol{w} \rangle = \bar{\lambda} \langle \boldsymbol{u}, \boldsymbol{w} \rangle = 0,$$

  which implies $U^{\perp}$ is T' invariant.

- Therefore, we construct the operator $T|_{U^{\perp}} : U^{\perp} \to U^{\perp}$, and

$$TT' = T'T \implies (T|_{U^{\perp}})(T'|_{U^{\perp}}) = (T'|_{U^{\perp}})(T|_{U^{\perp}}),$$

  i.e., $(T|_{U^{\perp}})$ is normal on $U^{\perp}$. Moreover, $\dim(U^{\perp}) = n - 1$.

- Applying the same trick as in Theorem (11.1), we imply there exists an orthonormal basis $\{\boldsymbol{e}_2,\dots,\boldsymbol{e}_n\}$ of eigenvectors of $(T\mid_{U^\perp})$. Then we can argue that

$$\mathcal{B}=\{\boldsymbol{v}'=\boldsymbol{v}/\|\boldsymbol{v}\|,\boldsymbol{e}_2,\dots,\boldsymbol{e}_{k+1}\}$$

is a basis of orthonormal eigenvectors of V.

■

Corollary 12.1 [Spectral Theorem for Normal Operator] Let $T:V\to V$ be a normal operator on a $\mathbb{C}$-inner product space with $\dim(V)<\infty$. Then there exists self-adjoint operators $P_1,\dots,P_k$ such that

$$P_i^2=P_i,\quad P_iP_j=0, i\ne j,\quad \sum_{i=1}^k P_i=I,$$

and $T=\sum_{i=1}^k\lambda_iP_i$, where λ_i's are the eigenvalues of T.

R These P_i's are the **orthogonal projections** from V to the λ_i-eigenspace $\ker(T-\lambda_iI)$ of T, i.e.,we have

$$\begin{aligned}&v=P_i(v)+(v-P_i(v)),\\ \text{where}\quad &P_i(v)\in\ker(T-\lambda_iI),\text{ and } v-P_i(v)\in(\ker(T-\lambda_iI))^\perp.\end{aligned}$$

You should know how to compute P_i's when $T(\boldsymbol{v})=\boldsymbol{A}\boldsymbol{v}$ in the course MAT2040.

Proof. Since T has a basis of eigenvectors, by definition, T is diagonalizable. By proposition (8.2),

$$m_T(x)=(x-\lambda_1)\cdots(x-\lambda_k),$$

where λ_i's are distinct. By spectral decomposition corollary (9.2), it suffices to show P_i's are self-disjoint.

- Recall that $P_i=a_i(T)q_i(T):=b_mT^m+\cdots+b_1T+b_0T$, i.e., a polynomial of T, and therefore

$$P_i'=\bar{b}_m(T')^m+\cdots+\bar{b}_1(T')+\bar{b}_0I.$$

We claim that P_i is normal: Since $T'T = TT'$, we imply

$$(T')^p T^q = T^q (T')^p, \forall p, q \in \mathbb{N}$$

which follows that

$$\begin{aligned}
P_i P_i' &= (b_m T^m + \cdots + b_0 I)(\bar{b}_m (T')^m + \cdots + \bar{b}_1 (T') + \bar{b}_0 I) \\
&= \sum_{1 \le x, y \le m} b_x \bar{b}_y (T)^x (T')^y \\
&= \sum_{1 \le x, y \le m} \bar{b}_y b_x (T')^y (T)^x \\
&= (\bar{b}_m (T')^m + \cdots + \bar{b}_1 (T') + \bar{b}_0 I)(b_m T^m + \cdots + b_0 I) \\
&= P_i' P_i
\end{aligned}$$

- In general, S is self-adjoint, which implies S is normal, but not vice versa. However, the converse holds if further all eigenvalues of S are real numbers:

  By Theorem (12.1), we imply S is orthonormally diagonalizable, and its diagonal representation is of the form

  $$(S)_{\mathcal{B},\mathcal{B}} = \mathrm{diag}(\lambda_1, \ldots, \lambda_k).$$

  Note that $\mathcal{B}$ is also a basis for S' and elements of $\mathcal{B}$ are eigenvalues of S', by part (3) in proposition (12.1). Therefore,

  $$(S')_{\mathcal{B},\mathcal{B}} = \mathrm{diag}(\lambda_1, \ldots, \lambda_k).$$

  Therefore, $S = S'$.

In particular, for $S = P_i$, we can easily show all eigenvalues of P_i are 0 or 1, which are real. Therefore, P_i's are self-adjoint. ■

Corollary 12.2 Let $T : V \to V$ be a linear operator on $\mathbb{C}$-inner product space with $\dim(V) < \infty$. Then T is normal if and only if $T' = f(T)$ for some polynomial $f(x) \in \mathbb{C}[x]$.

Proof.
- For the reverse direction, if $T' = f(T)$, then $T'T = f(T)T = Tf(T) = TT'$.
- For the forward direction, suppose that T is normal, then by corollary (12.1),

$$T = \sum_{i=1}^{k} \lambda_i P_i, \ P_i = f_i(T), \text{ where } P_i\text{'s are self-adjoint,}$$

which follows that

$$T' = \left(\sum_{i=1}^{k} \lambda_i P_i\right)' = \sum_{i=1}^{k} \bar{\lambda}_i P_i' = \sum_{i=1}^{k} \bar{\lambda}_i P_i = \sum_{i=1}^{k} \bar{\lambda}_i f_i(T)$$

■

(R) The normal operator is a generalization of Hermitian matrices, and it inherits many nice properties of Hermitian.

12.1.2. Tensor Product

Motivation. Let U, V, W be vector spaces. We want to study bilinear maps $f : U \times W \to U$, i.e.,

$$f(av_1 + bv_2, w) = af(v_1, w) + bf(v_2, w)$$

$$f(v, cw_1 + dw_2) = cf(v, w_1) + df(v, w_2)$$

Unfortunately, bilinear form usually is not a linear transformation!

Example 12.1
- Let $f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ be with $(u, v) \mapsto \langle u, v \rangle$.
- Let $f : M_{n\times n}(\mathbb{F}) \times M_{n\times n}(\mathbb{F}) \to M_{n\times n}(\mathbb{F})$ be with $f(A, B) = AB$.
- Let $f : \mathbb{F}[x] \times \mathbb{F}[x] \to \mathbb{F}$ be with $f(p(x), q(x)) = p(1)q(2)$

- Let $f : \mathbb{F}[x] \times \mathbb{F}[x] \to \mathbb{F}[x]$ be with $f(p(x), q(x)) = p(x)q(x)$.

12.4. Wednesday for MAT3040

Reviewing. Bilinear map: $f : V \times W \to U$, e.g.,

$$\begin{aligned} f: \quad & \mathbb{R}^3 \times \mathbb{R}^3 \\ \text{with} \quad & f(u,v) = u \times v \end{aligned}$$

Note that f is usually not a linear transformation, e.g.,

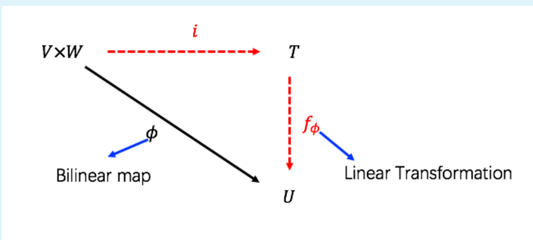
$$f(3(\boldsymbol{v},\boldsymbol{w})) = f(3\boldsymbol{v},3\boldsymbol{w}) = (3\boldsymbol{v}) \times (3\boldsymbol{w}) = 9\boldsymbol{v} \times \boldsymbol{w} \neq 3f(\boldsymbol{v},\boldsymbol{w}).$$

The vector space structure of $V \times W$ is not suited to study bilinear map, and the proper way is to study its induced linear transformation.

Definition 12.4 [Universal Property of Tensor Product] Let V, W be vector spaces. Consider the set

$$\text{Obj} := \{\phi : V \times W \to U \mid \phi \text{ is a bilinear map}\}$$

We say T, or $(i : V \times W \to T) \in \text{Obj}$ satisfies the **universal property** if for any $(\phi : V \times W \to T) \in \text{Obj}$, there exists an unique linear transformation $f_\phi : T \to U$ such that the diagram below commutes:



i.e., $\phi = f_\phi \circ i$.

Therefore, rather than studying bilinear map ϕ, it is better to study the linear transformation f_ϕ instead.

Question: does T exist?

Definition 12.5 [Spanning Set] Let V,W be vector spaces. Let $S = \{(\boldsymbol{v},\boldsymbol{w}) \mid \boldsymbol{v} \in V, \boldsymbol{w} \in W\}$, then we define

$$\mathfrak{X} = \mathrm{span}(S).$$

R

1. The spanning set $\mathfrak{X}$ is not addictive, e.g., $\mathfrak{x}_1 = 3(0,\boldsymbol{w}) \in \mathfrak{X}$ and $\mathfrak{x}_2 = 1(0,\boldsymbol{w}) + 1(0,2\boldsymbol{w}) \in \mathfrak{X}$, but $\mathfrak{x}_1 \neq \mathfrak{x}_2$.
2. Note that we assume no relations on the elements $(\boldsymbol{v},\boldsymbol{w}) \in \mathcal{S}$. More precisely, the set $S = \{(\boldsymbol{v},\boldsymbol{w}) \mid \boldsymbol{v} \in V, \boldsymbol{w} \in W\}$ is linearly independent in $\mathfrak{X}$. For example, $(0,\boldsymbol{w}) \perp (0,2\boldsymbol{w})$.
3. The only legitimate relationship is
$$2(\boldsymbol{v}_1,\boldsymbol{w}_1) + 3(\boldsymbol{v}_1,\boldsymbol{w}_1) = 5(\boldsymbol{v},\boldsymbol{w}),$$
which is not equal to $(5\boldsymbol{v},5\boldsymbol{w})$
4. $\mathcal{S}$ is a basis of $\mathfrak{X}$, and therefore $\mathcal{X}$ is of uncountable dimension.

Definition 12.6 [Special subspace of $\mathfrak{X}$] Let $y \leq \mathfrak{X}$ be a vector subspace spanned by vectors of the form

$$\{1(\boldsymbol{v}_1,\boldsymbol{v}_2,\boldsymbol{w}) - 1(\boldsymbol{v}_1,\boldsymbol{w}) - 1(\boldsymbol{v}_2,\boldsymbol{w})\}, \quad \text{and} \quad \{1(\boldsymbol{v},\boldsymbol{w}_1+\boldsymbol{w}_2) - 1(\boldsymbol{v},\boldsymbol{w}_1) - 1(\boldsymbol{v},\boldsymbol{w}_2)\}$$

and

$$\{1(k\boldsymbol{v},\boldsymbol{w}) - k(\boldsymbol{v},\boldsymbol{w}) \mid k \in \mathbb{F}\}$$

and

$$\{1(\boldsymbol{v},k\boldsymbol{w}) - k(\boldsymbol{v},\boldsymbol{w}) \mid k \in \mathbb{F}\}$$

Definition 12.7 [Tensor Product] We define the **tensor product** $V \otimes W$ by

$$V \otimes W = \mathcal{X}/y.$$

Therefore, $\boldsymbol{v} \otimes \boldsymbol{w} = (\boldsymbol{v}, \boldsymbol{w}) + y \in \mathcal{X}/y$

R

1. As a result, the tensor product is finitely addictive:

$$\begin{aligned}(\boldsymbol{v}_1 + \boldsymbol{v}_2) \otimes \boldsymbol{w} &= (\boldsymbol{v}_1 + \boldsymbol{v}_2, \boldsymbol{w}) + y \\ &= (\boldsymbol{v}_1 + \boldsymbol{v}_2, \boldsymbol{w}) - [(\boldsymbol{v}_1 + \boldsymbol{v}_2, \boldsymbol{w}) - (\boldsymbol{v}_1, \boldsymbol{w}) - (\boldsymbol{v}_2, \boldsymbol{w})] + y \\ &= 0(\boldsymbol{v}_1 + \boldsymbol{v}_2, \boldsymbol{w}) + (\boldsymbol{v}_1, \boldsymbol{w}) + (\boldsymbol{v}_2, \boldsymbol{w}) + y \\ &= [(\boldsymbol{v}_1, \boldsymbol{w}) + y] + [(\boldsymbol{v}_2, \boldsymbol{w}) + y] \\ &= \boldsymbol{v}_1 \otimes \boldsymbol{w} + \boldsymbol{v}_2 \otimes \boldsymbol{w}\end{aligned}$$

   Similarly,

$$\begin{aligned}\boldsymbol{v} \otimes (\boldsymbol{w}_1 + \boldsymbol{w}_2) &= (\boldsymbol{v} \otimes \boldsymbol{w}_1) + (\boldsymbol{v} \otimes \boldsymbol{w}_2) \\ (k\boldsymbol{v}) \otimes \boldsymbol{w} &= k(\boldsymbol{v} \otimes \boldsymbol{w}) \\ \boldsymbol{v} \otimes (k\boldsymbol{w}) &= k(\boldsymbol{v} \otimes \boldsymbol{w})\end{aligned}$$

2. The product space $V \times W$ is different from the tensor product space $V \otimes W$:

   (a) $(\boldsymbol{v}, \boldsymbol{0}) \neq \boldsymbol{0}_{V \times W}$ in $V \times W$; but $\boldsymbol{v} \otimes 0 \in 0_{V \otimes W}$:

$$\begin{aligned}V \otimes 0 &= V \otimes (0\boldsymbol{w}) \\ &= 0(V \otimes w) \\ &= 0_{V \otimes W}\end{aligned}$$

   Moreover, f is bilinear implies $f(\boldsymbol{v}, 0) = \boldsymbol{0}$.

(b) $(\boldsymbol{v}_1, \boldsymbol{w}_1) + (\boldsymbol{v}_2, \boldsymbol{w}_2) = (\boldsymbol{v}_1 + \boldsymbol{v}_2, \boldsymbol{w}_1 + \boldsymbol{w}_2)$; but $\boldsymbol{v}_1 \otimes \boldsymbol{w}_1 + \boldsymbol{v}_2 \otimes \boldsymbol{w}_2$ cannot be simplified further, unless $\boldsymbol{v}_1 = \boldsymbol{v}_2$:

$$\boldsymbol{v} \otimes \boldsymbol{w}_1 + \boldsymbol{v} \otimes \boldsymbol{w}_2 = \boldsymbol{v} \otimes (\boldsymbol{w}_1 + \boldsymbol{w}_2)$$

Theorem 12.3 The bilinear map

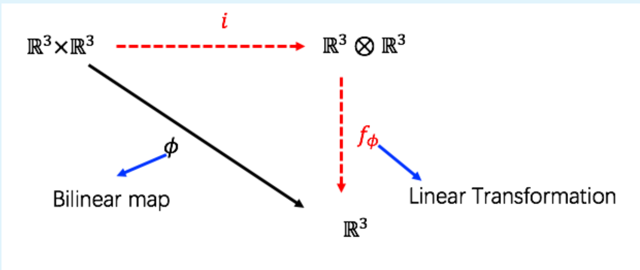
$$\begin{aligned} i: \quad & V \times W \to V \otimes W \quad (i \in \mathrm{Obj}) \\ \text{with} \quad & (\boldsymbol{v}, \boldsymbol{w}) \mapsto \boldsymbol{v} \otimes \boldsymbol{w} \end{aligned}$$

satisfies the universal property of tensor products.

■ **Example 12.5** Consider a common bilinear map

$$\begin{aligned} \phi: \quad & \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^3 \\ \text{with} \quad & (\boldsymbol{v}, \boldsymbol{w}) \mapsto \boldsymbol{v} \times \boldsymbol{w} \end{aligned}$$

By the universal property, there exists the linear transformation $f_\phi : \mathbb{R}^3 \otimes \mathbb{R}^3 \to \mathbb{R}^3$ such that the diagram below commutes:



■

Chapter 13

Week13

13.1. Monday for MAT3040

Reviewing.

1. Define $S = \{(\boldsymbol{v}, \boldsymbol{w}) \mid \boldsymbol{v} \in V, \boldsymbol{w} \in W\}$ and $\mathfrak{X} = \text{span}(S)$. In $\mathfrak{X}$, there are no relations between distinct elements of S, e.g.,

$$2(\boldsymbol{v}, 0) + 3(0, \boldsymbol{w}) \neq 1(2\boldsymbol{v}, 3\boldsymbol{w})$$

   General element in $\mathfrak{X}$:

$$a_1(\boldsymbol{v}_1, \boldsymbol{w}_1) + \cdots + a_n(\boldsymbol{v}_n, \boldsymbol{w}_n),$$

   where $(\boldsymbol{v}_i, \boldsymbol{w}_i)$ are distinct.

2. Define the space $V \otimes W = \mathfrak{X}/y$, with

$$\boldsymbol{v} \otimes \boldsymbol{w} = 1(\boldsymbol{v}, \boldsymbol{w}) + y \in V \otimes W.$$

   General element in $\mathfrak{X}/y := V \otimes W$:

$$\begin{aligned} a_1(\boldsymbol{v}_1, \boldsymbol{w}_1) + \cdots + a_n(\boldsymbol{v}_n, \boldsymbol{w}_n) + y &= a_1((\boldsymbol{v}_1, \boldsymbol{w}_1) + y) + \cdots + a_n((\boldsymbol{v}_n, \boldsymbol{w}_n) + y) \\ &= a_1(\boldsymbol{v}_1 \otimes \boldsymbol{w}_1) + \cdots + a_n(\boldsymbol{v}_n \otimes \boldsymbol{w}_n) \\ &= (a_1\boldsymbol{v}_1) \otimes \boldsymbol{w}_1 + \cdots + (a_n\boldsymbol{v}_n) \otimes \boldsymbol{w}_n \end{aligned}$$

Therefore, a general element in $V \otimes W$ is of the form

$$\boldsymbol{v}_1' \otimes \boldsymbol{w}_1 + \cdots + \boldsymbol{v}_n' \otimes \boldsymbol{w}_n, \ \boldsymbol{v}_i' \in V, \boldsymbol{w}_i \in W. \tag{13.1}$$

Note that $V \otimes W$ is different from $V \times W$, where all elements in $V \times W$ can be expressed as $(\boldsymbol{v}, \boldsymbol{w})$.

3. The tensor product mapping

$$\begin{aligned} i: &\quad V \times W \to V \otimes W \\ \text{with} &\quad (\boldsymbol{v}, \boldsymbol{w}) \mapsto \boldsymbol{v} \otimes \boldsymbol{w} \end{aligned}$$

satisfies the universal property.

Here we present an example for computing tensor product by making use of the rules below:

$$\begin{aligned} (\boldsymbol{v}_1 + \boldsymbol{v}_2) \otimes \boldsymbol{w} &= \boldsymbol{v}_1 \otimes \boldsymbol{w} + \boldsymbol{v}_2 \otimes \boldsymbol{w} \\ \boldsymbol{v} \otimes (\boldsymbol{w}_1 + \boldsymbol{w}_2) &= (\boldsymbol{v} \otimes \boldsymbol{w}_1) + (\boldsymbol{v} \otimes \boldsymbol{w}_2) \\ (k\boldsymbol{v}) \otimes \boldsymbol{w} &= k(\boldsymbol{v} \otimes \boldsymbol{w}) \\ \boldsymbol{v} \otimes (k\boldsymbol{w}) &= k(\boldsymbol{v} \otimes \boldsymbol{w}) \end{aligned}$$

▪ **Example 13.1** Let $V = W = \mathbb{R}^2$, with

$$\boldsymbol{e}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \boldsymbol{e}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Here we have

$$\begin{aligned}
\begin{pmatrix} 3 \\ 1 \end{pmatrix} \otimes \begin{pmatrix} -4 \\ 2 \end{pmatrix} &= (3\boldsymbol{e}_1 + 2\boldsymbol{e}_2) \otimes (-4\boldsymbol{e}_1 + 2\boldsymbol{e}_2) \\
&= (3\boldsymbol{e}_1) \otimes (-4\boldsymbol{e}_1 + 2\boldsymbol{e}_2) + (\boldsymbol{e}_2) \otimes (-4\boldsymbol{e}_1 + 2\boldsymbol{e}_2) \\
&= (3\boldsymbol{e}_1) \otimes (-4\boldsymbol{e}_1) + (3\boldsymbol{e}_1) \otimes (2\boldsymbol{e}_2) + (\boldsymbol{e}_2) \otimes (-4\boldsymbol{e}_1) + \boldsymbol{e}_2 \otimes (2\boldsymbol{e}_2) \\
&= -12(\boldsymbol{e}_1 \otimes \boldsymbol{e}_1) + 6(\boldsymbol{e}_1 \otimes \boldsymbol{e}_2) - 4(\boldsymbol{e}_2 \otimes \boldsymbol{e}_1) + 2(\boldsymbol{e}_2 \otimes \boldsymbol{e}_2)
\end{aligned}$$

Exercise: Check that $\boldsymbol{e}_1 \otimes \boldsymbol{e}_2 + \boldsymbol{e}_2 \otimes \boldsymbol{e}_1$ cannot be re-written as

$$(a\boldsymbol{e}_1 + b\boldsymbol{e}_2) \otimes (c\boldsymbol{e}_1 + d\boldsymbol{e}_2), \quad a, b, c, d \in \mathbb{R}.$$

13.1.1. Basis of $V \otimes W$

Motivation. Given that $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_n\}$ is a basis of V, and $\{\boldsymbol{w}_1, \dots, \boldsymbol{w}_m\}$ a basis of W, we aim to find a basis of $V \otimes W$ using $\boldsymbol{v}_i$'s and $\boldsymbol{w}_i$'s.

Proposition 13.1 The set $\{\boldsymbol{v}_i \otimes \boldsymbol{w}_j \mid 1 \le i \le n, 1 \le j \le m\}$ spans the tensor product space $V \otimes W$.

Proof. Consider any $\boldsymbol{v} \in V$ and $\boldsymbol{w} \in W$, and we want to express $\boldsymbol{v} \otimes \boldsymbol{w}$ in terms of $\boldsymbol{v}_i, \boldsymbol{w}_j$. Suppose that $\boldsymbol{v} = \alpha_1 \boldsymbol{v}_1 + \cdots + \alpha_n \boldsymbol{v}_n$ and $\boldsymbol{w} = \beta_1 \boldsymbol{w}_1 + \cdots + \beta_m \boldsymbol{w}_m$.

Substituting $\boldsymbol{v} = \alpha_1 \boldsymbol{v}_1 + \cdots + \alpha_n \boldsymbol{v}_n$ into the expression $\boldsymbol{v} \otimes \boldsymbol{w}$, we imply

$$\begin{aligned}
\boldsymbol{v} \otimes \boldsymbol{w} &= (\alpha_1 \boldsymbol{v}_1 + \cdots + \alpha_n \boldsymbol{v}_n) \otimes \boldsymbol{w} \\
&= (\alpha_1 \boldsymbol{v}_1) \otimes \boldsymbol{w}_1 + \cdots + (\alpha_n \boldsymbol{v}_n) \otimes \boldsymbol{w}_n \\
&= \alpha_1 (\boldsymbol{v}_1 \otimes \boldsymbol{w}) + \cdots + \alpha_n (\boldsymbol{v}_n \otimes \boldsymbol{w})
\end{aligned}$$

For each $\boldsymbol{v}_i \otimes \boldsymbol{w}$, $i = 1, \dots, n$, similarly,

$$\boldsymbol{v}_i \otimes \boldsymbol{w} = \beta_1 (\boldsymbol{v}_i \otimes \boldsymbol{w}_1) + \cdots + \beta_m (\boldsymbol{v}_i \otimes \boldsymbol{w}_m).$$

Therefore,

$$\boldsymbol{v} \otimes \boldsymbol{w} = \sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_i \beta_j (\boldsymbol{v}_i \otimes \boldsymbol{w}_j) \tag{13.2}$$

By (13.1), any vector in $V \otimes W$ is of the form

$$\boldsymbol{v}^{(1)} \otimes \boldsymbol{w}^{(1)} + \cdots + \boldsymbol{v}^{(\ell)} \otimes \boldsymbol{w}^{(\ell)}$$

By (13.2), each $\boldsymbol{v}^{(k)} \otimes \boldsymbol{w}^{(k)}, k = 1, \ldots, \ell$, can be expressed as

$$\boldsymbol{v}^{(k)} \otimes \boldsymbol{w}^{(k)} = \sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_i^{(k)} \beta_j^{(k)} (\boldsymbol{v}_i \otimes \boldsymbol{w}_j)$$

Therefore,

$$\boldsymbol{v}^{(1)} \otimes \boldsymbol{w}^{(1)} + \cdots + \boldsymbol{v}^{(\ell)} \otimes \boldsymbol{w}^{(\ell)} = \sum_{k=1}^{\ell} \sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_i^{(k)} \beta_j^{(k)} (\boldsymbol{v}_i \otimes \boldsymbol{w}_j)$$

In other words, $\{\boldsymbol{v}_i \otimes \boldsymbol{w}_j \mid 1 \le i \le n, 1 \le j \le m\}$ spans $V \otimes W$. ■

Theorem 13.1 A basis of $V \otimes W$ is $\{\boldsymbol{v}_i \otimes \boldsymbol{w}_j \mid 1 \le i \le n, 1 \le j \le m\}$

Proof. By proposition (13.1), it suffices to show that the set $\{\boldsymbol{v}_i \otimes \boldsymbol{w}_j \mid 1 \le i \le n, 1 \le j \le m\}$ is linear independent. Suppose that

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_{ij} (\boldsymbol{v}_i \otimes \boldsymbol{w}_j) = \mathbf{0} \tag{13.3}$$

Suppose that $\{\phi_1, \ldots, \phi_n\}$ is a dual basis of V^*, and $\{\psi_1, \ldots, \psi_m\}$ is a dual basis of W^*. Construct the mapping

$$\begin{aligned} \pi_{p,q} : &\quad V \times W \to \mathbb{F} \\ \text{with} &\quad \pi_{p,q} = \phi_p(\boldsymbol{v}) \psi_q(\boldsymbol{w}) \end{aligned}$$

- The mapping $\pi_{p,q}$ is actually bilinear: for instance,

$$\begin{aligned}\pi_{p,q}(a\boldsymbol{v}_1 + b\boldsymbol{v}_2, \boldsymbol{w}) &= \phi_p(a\boldsymbol{v}_1 + b\boldsymbol{v}_2)\psi_q(\boldsymbol{w}) \\ &= (a\phi_p(\boldsymbol{v}_1) + b\phi_p(\boldsymbol{v}_2))\psi_q(\boldsymbol{w}) \\ &= a\phi_p(\boldsymbol{v}_1)\psi_q(\boldsymbol{w}) + b\phi_p(\boldsymbol{v}_2)\psi_q(\boldsymbol{w}) \\ &= a\pi_{p,q}(\boldsymbol{v}_1, \boldsymbol{w}) + b\pi_{p,q}(\boldsymbol{v}_2, \boldsymbol{w}).\end{aligned}$$

  Following the similar ideas, we can check that $\pi_{p,q}(\boldsymbol{v}, a\boldsymbol{w}_1 + b\boldsymbol{w}_2) = a\pi_{p,q}(\boldsymbol{v}, \boldsymbol{w}_1) + b\pi_{p,q}(\boldsymbol{v}, \boldsymbol{w}_2)$.

- Therefore, $\pi_{p,q} \in \text{Obj}$. By the universal property of the tensor product, $\pi_{p,q}$ induces the unique linear transformation

$$\begin{aligned}&\prod_{p,q}: \quad V \otimes W \to \mathbb{F} \\ &\text{with} \quad \prod_{p,q}(\boldsymbol{v} \otimes \boldsymbol{w}) = \pi_{p,q}(\boldsymbol{v}, \boldsymbol{w})\end{aligned}$$

  In other words, $\prod_{p,q}(\boldsymbol{v} \otimes \boldsymbol{w}) = \phi_p(\boldsymbol{v})\psi_q(\boldsymbol{w})$.

- Applying the mapping $\Pi_{p,q}$ on both sides of (13.3), we imply

$$\Pi_{p,q}\left(\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_{ij}(\boldsymbol{v}_i \otimes \boldsymbol{w}_j)\right) = \Pi_{p,q}(\boldsymbol{0})$$

  Or equivalently,

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_{ij}\Pi_{p,q}(\boldsymbol{v}_i \otimes \boldsymbol{w}_j) = 0,$$

  i.e.,

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_{ij}\phi_p(\boldsymbol{v}_i)\psi_q(\boldsymbol{w}_j) = \alpha_{p,q} = 0$$

  Following this procedure, we can argue that $\alpha_{ij} = 0, \forall i, \forall j$.

■

Corollary 13.1 If $\dim(V), \dim(W) < \infty$, then $\dim(V \otimes W) = \dim(V)\dim(W)$

Proof. Check dimension of the basis of $V \otimes W$. ■

R The universal property can be very helpful. In particular, given a bilinear mapping, say $\phi: V \times W \to U$, we imply $\phi \in \text{Obj}$. By theorem (12.3), since i satisfies the universal property of tensor product, we can induce an unique linear transformation $\psi: V \otimes W \to U$.

Let's try another example for making use of the universal property:

Theorem 13.2 For finite dimension U and V,

$$V \otimes U \cong U \otimes V$$

Proof. Construct the mapping

$$\begin{aligned} \phi: \quad & V \times U \to U \otimes V \\ \text{with} \quad & \phi(\boldsymbol{v}, \boldsymbol{u}) = \boldsymbol{u} \otimes \boldsymbol{v} \end{aligned}$$

Indeed, ϕ is bilinear: for instance,

$$\begin{aligned} \phi(a\boldsymbol{v}_1 + b\boldsymbol{v}_2, \boldsymbol{u}) &= u \otimes (a\boldsymbol{v}_1 + b\boldsymbol{v}_2) \\ &= a(\boldsymbol{u} \otimes \boldsymbol{v}_1) + b(u \otimes \boldsymbol{v}_2) \\ &= a\phi(\boldsymbol{v}_1, \boldsymbol{u}) + b\phi(\boldsymbol{v}_2, \boldsymbol{u}) \end{aligned}$$

Therefore, $\phi \in \text{Obj}$. By the universal property of tensor product, we induce an unique linear transformation

$$\begin{aligned} \Phi: \quad & V \otimes U \to U \otimes V \\ \text{with} \quad & \Phi(\boldsymbol{v} \otimes \boldsymbol{u}) = \boldsymbol{u} \otimes \boldsymbol{v} \end{aligned}$$

Similarly, we may induce the linear transformation

$$\begin{aligned} \Psi: \quad & U \otimes V \to V \otimes U \\ \text{with} \quad & \Psi(\boldsymbol{u} \otimes \boldsymbol{v}) = \boldsymbol{v} \otimes \boldsymbol{u} \end{aligned}$$

Given any $\sum_i \boldsymbol{u}_i \otimes \boldsymbol{v}_i \in U \otimes V$, observe that

$$\begin{aligned}(\Phi \circ \Psi)\left(\sum_i \boldsymbol{u}_i \otimes \boldsymbol{v}_i\right) &= \Phi\left(\sum_i \Psi(\boldsymbol{u}_i \otimes \boldsymbol{v}_i)\right)\\ &= \Phi\left(\sum_i \boldsymbol{v}_i \otimes \boldsymbol{u}_i\right)\\ &= \sum_i \Phi(\boldsymbol{v}_i \otimes \boldsymbol{u}_i)\\ &= \sum_i \boldsymbol{u}_i \otimes \boldsymbol{v}_i\end{aligned}$$

Therefore, $\Phi \circ \Psi = \mathrm{id}_{U\otimes V}$. Similarly, $\Psi \circ \Phi = \mathrm{id}_{V\otimes U}$. Therefore,

$$U \otimes V \cong V \otimes U.$$

■

13.1.2. Tensor Product of Linear Transformation

Motivation. Given two linear transformations $T : V \to V'$ and $S : W \to W'$, we want to construct the tensor product

$$T \otimes S : V \otimes W \to V' \otimes W'$$

Question: is $T \otimes S$ a linear transformation?

Answer: Yes. Universal property plays a role!

13.4. Wednesday for MAT3040

13.4.1. Tensor Product for Linear Transformations

Proposition 13.5 Suppose that $T : V \to V'$ and $S : W \to W'$ are linear transformations, then there exists an unique linear transformation

$$\begin{aligned} T \otimes S : \quad & V \otimes W \to V' \otimes W' \\ \text{satisfying} \quad & (T \otimes S)(v \otimes w) = T(v) \otimes S(w) \end{aligned}$$

Proof. We construct the mapping

$$\begin{aligned} T \times S : \quad & V \times W \to V' \otimes W' \\ \text{with} \quad & (T \times S)(v, w) = T(v) \otimes S(w) \end{aligned}$$

This mapping is indeed bilinear: for instance, we can show that

$$(T \times S)(av_1 + bv_2, w) = a(T \times S)(v_1, w) + b(T \times S)(v_2, w)$$

Therefore, $T \times S \in \text{Obj}$. Since the tensor product satisfies the universal property, we imply there exists an unique linear transformation

$$\begin{aligned} T \otimes S \quad & V \otimes W \to V' \otimes W' \\ \text{satisfying} \quad & (T \otimes S)(v \otimes w) = T(v) \otimes S(w) \end{aligned}$$

■

Notation Warning. Does the notion $T \otimes S$ really form a tensor product, i.e., do we obtain the addictive rules for tensor product such as

$$(aT_1 + bT_2) \otimes S = a(T_1 \otimes S) + b(T_2 \otimes S)?$$

▪ **Example 13.2** Let $V = V' = \mathbb{F}^2$ and $W = W' = \mathbb{F}^3$. Define the matrix-multiply mappings:

$$\begin{cases} T: & V \to V \\ \text{with} & \boldsymbol{v} \mapsto \boldsymbol{A}\boldsymbol{v} \\ & \boldsymbol{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \end{cases} \qquad \begin{cases} S: & W \to W \\ \text{with} & \boldsymbol{w} \mapsto \boldsymbol{B}\boldsymbol{w} \\ & \boldsymbol{B} = \begin{pmatrix} p & q & r \\ s & t & u \\ v & w & x \end{pmatrix} \end{cases}$$

How does $T \otimes S : V \otimes W \to V \otimes W$ look like?

- Suppose $\{e_1, e_2\}, \{f_1, f_2, f_3\}$ are usual basis of V, W, respectively. Then the basis of $V \otimes W$ is given by:

$$C = \{e_1 \otimes f_1, e_1 \otimes f_2, e_1 \otimes f_3, e_2 \otimes f_1, e_2 \otimes f_2, e_2 \otimes f_3\}.$$

- As a result, we can compute $(T \otimes S)(e_i \otimes f_j)$ for $i = 1,2$ and $j = 1,2,3$. For instance,

$$\begin{aligned} (T \otimes S)(e_1 \otimes e_1) &= T(e_1) \otimes S(e_1) \\ &= (ae_1 + ce_2) \otimes (pe_1 + se_2 + ve_3) \\ &= (ap)e_1 \otimes e_1 + (as)e_1 \otimes e_2 + (av)e_1 \otimes e_3 + (cp)e_2 \otimes e_1 + (cs)e_2 \otimes e_2 + (cv)e_2 \otimes e_3 \end{aligned}$$

- Therefore, we obtain a matrix representation for the linear transformation $(T \otimes S)$:

We want a matrix representation for $(T \otimes S)$:

$$(T \otimes S)_{C,C} = \begin{pmatrix} aB & bB \\ cB & dB \end{pmatrix},$$

which is a large matrix formed by taking all possible products between the elements of $\boldsymbol{A}$ and those of $\boldsymbol{B}$. This operation is called the **Kronecker Tensor Product**, see the command *kron* in MATLAB for detail. ▪

Proposition 13.6 More generally, given the linear operator $T : V \to V$ and $S : W \to W$, let $\mathcal{A} = \{v_1, \dots, v_n\}, \mathcal{B} = \{w_1, \dots, w_m\}$ be a basis of V, W respectively, with

$$(T)_{\mathcal{A},\mathcal{A}} = (a_{ij}) \quad (S_{\mathcal{B},\mathcal{B}}) = (b_{ij}) := B$$

As a result, $(T \otimes S)_{C,C} = A \otimes B$, where $C = \{v_1 \otimes w_1, \dots, v_n \otimes w_m\}$, and $A \otimes B$ denotes the Kronecker tensor product, defined as the matrix

$$\begin{pmatrix} a_{1,1}B & \cdots & a_{1,n}B \\ \vdots & \ddots & \vdots \\ a_{n,1}B & \cdots & a_{n,n}B \end{pmatrix}.$$

Proof. Following the similar procedure as in Example (13.2) and applying the relation

$$\begin{aligned}(T \otimes S)(v_i \otimes w_j) &= T(v_i) \otimes S(w_j) \\ &= \left(\sum_{k=1}^{n} a_{ki} v_k\right) \otimes \left(\sum_{\ell=1}^{m} b_{\ell j} w_\ell\right) \\ &= \sum_{k=1}^{n} \sum_{\ell=1}^{m} (a_{ki} b_{\ell j}) v_k \otimes w_\ell\end{aligned}$$

■

Proposition 13.7 The operation $T \otimes S$ satisfies all the properties of tensor product. For example,

$$(aT_1 + bT_2) \otimes S = a(T_1 \otimes S) + b(T_2 \otimes S)$$

$$T \otimes (cS_1 + dS_2) = c(T \otimes S_1) + d(T \otimes S_2)$$

Therefore, the usage of the notion "$\otimes$" is justified for the definition of $T \otimes S$.

Proof using matrix multiplication. For instance, consider the operation $(T + T') \otimes S$, with $(T)_{\mathcal{A},\mathcal{A}} = (a_{ij})$, $(T')_{\mathcal{A},\mathcal{A}} = (c_{ij}), (S)_{\mathcal{B},\mathcal{B}} = B$.

We compute its matrix representation directly:

$$\begin{aligned}((T+T')\otimes S)_{\mathcal{C},\mathcal{C}} &= (T+T')_{\mathcal{A},\mathcal{A}}\otimes(S)_{\mathcal{B},\mathcal{B}}\\ &= [(T)_{\mathcal{A},\mathcal{A}}+(T')_{\mathcal{A},\mathcal{A}}]\otimes(S)_{\mathcal{B},\mathcal{B}}\\ &= (T)_{\mathcal{A},\mathcal{A}}\otimes(S)_{\mathcal{B},\mathcal{B}}+(T')_{\mathcal{A},\mathcal{A}}\otimes(S)_{\mathcal{B},\mathcal{B}}\end{aligned}$$

where the last equality is by the addictive rule for kronecker product for matrices. Therefore,

$$((T+T')\otimes S)_{\mathcal{C},\mathcal{C}} = (T\otimes S)_{\mathcal{C},\mathcal{C}}+(T'\otimes S)_{\mathcal{C},\mathcal{C}} \implies (T+T')\otimes S = T\otimes S+T'\otimes S$$

■

Proof using basis of $T\otimes S$. Another way of the proof is by computing

$$((T+T')\otimes S)(v_i\otimes w_j),$$

where $\{v_i\otimes w_j \mid 1\le i\le n, 1\le j\le m\}$ forms a basis of $(T+T')\otimes S$:

$$\begin{aligned}((T+T')\otimes S)(v_i\otimes w_j) &= (T+T')(v_i)\otimes S(w_j)\\ &= (T(v_i)+T'(v_i))\otimes S(w_j)\\ &= T(v_i)\otimes S(w_j)+T'(v_i)\otimes S(w_j)\\ &= (T\otimes S)(v_i\otimes w_j)+(T'\otimes S)(v_i\otimes w_j)\end{aligned}$$

Since $((T+T')\otimes S)(v_i\otimes w_j)$ coincides with $(T\otimes S+T'\otimes S)(v_i\otimes w_j)$ for all basis vectors $v_i\otimes w_j\in\mathcal{C}$, we imply

$$(T+T')\otimes S = T\otimes S+T'\otimes S$$

■

Proposition 13.8 Let A, C be linear operators from V to V, and B, D be linear operators from W to W, then

$$(A\otimes B)\circ(C\otimes D) = (AC)\otimes(BD)$$

Proposition 13.9 Define linear operators $A: V \to V$ and $B: W \to W$ with $\dim(V), \dim(W) < \infty$. Then

$$\det(A \otimes B) = (\det(A))^{\dim(W)} (\det(B))^{\dim(V)}$$

Corollary 13.3 There exists a linear transformation

$$\begin{aligned} \Phi: \quad & \text{Hom}(V,V) \otimes \text{Hom}(W,W) \to \text{Hom}(V \otimes W, V \otimes W) \\ \text{with} \quad & A \otimes B \mapsto A \otimes B \end{aligned}$$

where the input of Φ is the tensor product of linear transformations, and the output is the linear transformation.

Proof. Construct the mapping

$$\begin{aligned} \Phi \quad & : \text{Hom}(V,V) \times \text{Hom}(W,W) \to \text{Hom}(V \otimes W, V \otimes W) \\ \text{with} \quad & \Phi(A,B) = A \otimes B \end{aligned}$$

The Φ is indeed bilinear: for instance,

$$\begin{aligned} \Phi(pA + qC, B) &= (pA + qC) \otimes B \\ &= p(A \otimes B) + q(C \otimes B) \\ &= p\Phi(A,B) + q\Phi(C,B) \end{aligned}$$

This corollary follows from the universal property of tensor product. ∎

R If assuming that $\dim(V), \dim(W) < \infty$, we imply

$$\begin{aligned} \dim(\text{Input space of } \Phi) &= \dim(\text{Hom}(V,V)) \dim(\text{Hom}(W,W)) \\ &= [\dim(V)\dim(V)] \cdot [\dim(W)\dim(W)] = [\dim(V)\dim(W)]^2 \\ &= [\dim(V \otimes W)]^2 \\ &= \dim(\text{Hom}(V \otimes W, V \otimes W)) \\ &= \dim(\text{Output space of } \Phi) \end{aligned}$$

Therefore, is Φ is an isomorphism? If so, then every linear operator $\alpha : V \otimes W \to V \otimes W$ can be expressed as

$$\alpha = A_1 \otimes B_1 + \cdots + A_k \otimes B_k$$

where $A_i : V \to V$ and $B_j : W \to W$.

Chapter 14

Week14

14.1. Monday for MAT3040

14.1.1. Multilinear Tensor Product

Definition 14.1 [Tensor Product among More spaces] Let $V_1,\dots,V_p$ be vector spaces over $\mathbb{F}$. Let $S = \{(v_1,\dots,v_p) \mid v_i \in V_i\}$ (We assume no relations among distinct elements in S), and define $\mathfrak{X} = \text{span}(S)$.

1. Then define the tensor product space $V_1 \otimes \cdots \otimes V_p = \mathfrak{X}/y$, where y is the vector subspace of $\mathfrak{X}$ spanned by vectors of the form

$$(v_1,\dots,v_i+v_i',\dots,v_p)-(v_1,\dots,v_i,\dots,v_p)-(v_1,\dots,v_i',\dots,v_p),$$

   and

$$(v_1,\dots,\alpha v_i,\dots,v_p)-\alpha(v_1,\dots,v_i,\dots,v_p)$$

   where $i = 1,2,\dots,p$.

2. The tensor product for vectors is defined as

$$v_1\otimes\cdots\otimes v_p := \{(v_1,\dots,v_p)+y\} \in V_1\otimes\cdots\otimes V_p$$

R Similar as in tensor product among two space,

1. We have

$$v_1 \otimes \cdots \otimes (\alpha v_i + \beta v_i') \otimes \cdots \otimes v_p = \alpha(v_1 \otimes \cdots \otimes v_i \otimes \cdots \otimes v_p) + \beta(v_1 \otimes \cdots \otimes v_i' \otimes \cdots \otimes v_p)$$

2. A general vector in $V_1 \otimes \cdots \otimes V_p$ is

$$\sum_{i=1}^{n} (W_1^{(i)} \otimes \cdots \otimes W_p^{(i)}), \quad \text{where } W_j^{(i)} \in V_j, j = 1, \dots, p$$

3. Let $\mathcal{B}_i = \{v_i^{(1)}, \dots, v_i^{(\dim(V_i))}\}$ be a basis of $V_i, i = 1, \dots, p$, then

$$\mathcal{B} = \{V_1^{(\alpha_1)} \otimes \cdots \otimes V_p^{(\alpha_p)} \mid 1 \le \alpha_i \le \dim(V_i)\}$$

is a basis of $V_1 \otimes \cdots \otimes V_p$. As a result,

$$\dim(V_1 \otimes \cdots \otimes V_p) = (\dim(V_1)) \times \cdots \times (\dim(V_p))$$

Theorem 14.1 — **Universal Property of multi-linear tensor.** Let $\text{Obj} = \{\phi : V_1 \times \cdots \times V_p \to W \mid \phi \text{ is a } p\text{-linear map}\}$, i.e.,

$$\phi(v_1, \dots, \alpha v_i + \beta v_i', \dots, v_o) = \alpha\phi(v_1, \dots, v_i, \dots, v_p) + \beta\phi(v_1, \dots, v_i', \dots, v_p),$$
$$\forall v_i, v_i' \in V_i, i = 1, \dots, p, \forall \alpha, \beta \in \mathbb{F}.$$

For instance, the multiplication of p matrices is a p-linear map.

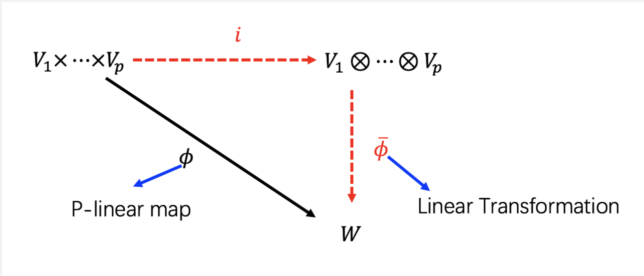
Then the mapping in the Obj,

$$\begin{aligned} i: &\quad V_1 \times V_p \to V_1 \otimes \cdots \otimes V_p \\ \text{with} &\quad (v_1, \dots, v_p) \mapsto v_1 \otimes \cdots \otimes v_p \end{aligned}$$

satisfies the universal property. In other words, for any $\phi : V_1 \times \cdots \times V_p \in \text{Obj}$, there exists the unqiue linear transformation

$$\bar{\phi} : V_1 \otimes \cdots \otimes V_p \to W$$

such that the diagram below commutes:



In other words, $\phi = \bar{\phi} \circ i$.

Corollary 14.1 Let $T_i : V_i \to V_i'$ be a linear transformation, $1 \le i \le p$. There is a unique linear transformation

$$\begin{aligned} (T_1 \otimes \cdots \otimes T_p) : \quad & V_1 \otimes \cdots \otimes V_p \to V_1' \otimes \cdots \otimes V_p' \\ \text{satisfying} \quad & (T_1 \otimes \cdots \otimes T_p)(v_1 \otimes \cdots \otimes v_p) = T_1(v_1) \otimes \cdots \otimes T_p(v_p) \end{aligned}$$

Proof. Construct the mapping

$$\begin{aligned} \phi : \quad & V_1 \times \cdots \times V_p \to V_1' \otimes \cdots \otimes V_p' \\ \text{with} \quad & (v_1, \ldots, v_p) \mapsto T_1(v_1) \otimes \cdots \otimes T_p(v_p) \end{aligned}$$

which is indeed p-linear.

By the universal property, we induce the unique linear transformation

$$\bar{\phi} : V_1 \otimes \cdots \otimes V_p \to V_1' \otimes \cdots \otimes V_p'$$

∎

Notation. To make life easier, from now on, we only consider $V_1 = \cdots = V_p = V$. Then for any linear transformation $T : V \to W$, we have

$$T^{\otimes p} : V \otimes \cdots \otimes V \to W \otimes \cdots \otimes W$$

We use the short-hand notation $V^{\otimes p}$ to denote $\underbrace{V \otimes \cdots \otimes V}_{p \text{ terms in total}}$

Final Exam Ends Here.

14.1.2. Exterior Power

Definition 14.2 A p-linear map $\phi : V \times \cdots \times V \to W$ is called **alternating** if

$$\phi(v_1, \dots, v_i, \dots, v_j, \dots, v_p) = \mathbf{0}_W, \quad \text{provided that there exists some } v_i = v_j \text{ for } i \neq j.$$

Also, we say ϕ is p-alternating

Example 14.1

1. The cross product mapping

$$\begin{aligned} \phi : \quad & \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^3 \\ \text{with} \quad & (\boldsymbol{v}, \boldsymbol{w}) \mapsto \boldsymbol{v} \times \boldsymbol{w} \end{aligned}$$

is alternating:

- ϕ is bilinear
- $\phi(\boldsymbol{v}, \boldsymbol{v}) = \boldsymbol{v} \times \boldsymbol{v} = \mathbf{0}$.

2. The determinant mapping

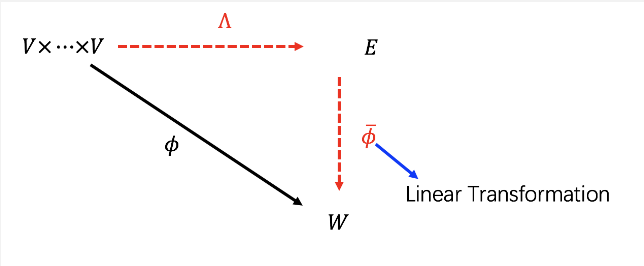
$$\begin{aligned} \phi : \quad & \underbrace{\mathbb{F}^n \times \cdots \times \mathbb{F}^n}_{n \text{ terms in total}} \to \mathbb{F} \\ \text{with} \quad & (\boldsymbol{v}_1, \dots, \boldsymbol{v}_n) \mapsto \det([\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n]) \end{aligned}$$

is alternating:

- ϕ is n-linear by MAT2040 knowledge
- ϕ is alternating by MAT2040 knowledge

Theorem 14.2 — **Universal Property for exterior power.** Let $\text{Obj} := \{\phi : \underbrace{V \times \cdots V}_{p \text{ terms}} \to W \mid \phi \text{ is } p\text{-alternating map}\}$. Then there exists $\{\Lambda : V \times \cdots \times V \to E\} \in \text{Obj}$ satisfying the following:

- For all $\phi : V \times \cdots \times V \to W \in \text{Obj}$, there exists unique linear transformation $\bar{\phi} : E \to W$ satisfying



  In other words, $\phi = \bar{\phi} \circ \Lambda$.