

# Linear Algebra I

Kon Yi

November 6, 2025

### **Abstract**

The lecture note of Linear Algebra I by professor 余正道.

# Contents

<b>1</b>	<b>Vector Space</b>	<b>2</b>
1.1	Introduction to vector and vector space . . . . .	2
1.2	Formal definition of vector spaces . . . . .	3
1.3	Vector Space over general field . . . . .	4
1.4	Subspaces . . . . .	5
1.5	Linear Combination . . . . .	5
1.6	Linearly independent . . . . .	5
1.7	Basis . . . . .	6
1.8	More on subspaces . . . . .	9
1.9	Space of linear maps . . . . .	15
1.10	Map/matrix correspondence . . . . .	17
<b>2</b>	<b>Dual space</b>	<b>24</b>
2.1	Dual of Dual space/Evaluation . . . . .	27
<b>3</b>	<b>Eigenvalue and Eigenvector</b>	<b>28</b>
3.1	Diagonalization . . . . .	36
3.2	Minimal polynomial . . . . .	39
3.3	Invariant subspaces . . . . .	40

# Chapter 1

## Vector Space

### Lecture 1

#### 1.1 Introduction to vector and vector space

3 Sep. 10:20

In high school, our vectors are in  $\mathbb{R}^2$  and  $\mathbb{R}^3$ , and we have define the addition and scalar multiplication of vectors.

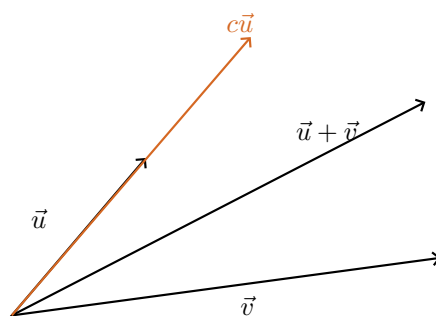


Figure 1.1: Vectors in  $\mathbb{R}^2$

**Example 1.1.1.**  $\mathbb{R}^n = \{(a_1, a_2, \dots, a_n \mid a_i \in \mathbb{R})\}$

With this type of space, we can define addition and multiplication as

$$(a_1, a_2, \dots, a_n) + (b_1, b_2, \dots, b_n) = \{a_1 + b_1, a_2 + b_2, \dots, a_n + b_n\}$$
$$\alpha \cdot (a_1, a_2, \dots, a_n) = (\alpha a_1, \alpha a_2, \dots, \alpha a_n)$$

Also, if we define a space:

**Example 1.1.2.**  $V = \{\text{function } f : (a, b) \rightarrow \mathbb{R}\}$ , where  $(a, b)$  is an open interval.

then this can also be a vector space after defining addition and multiplication.

**Note 1.1.1.** In a vector space, we have to make sure the existence of 0-element, which means  $0(x) = 0$ .

Now we give a more abstract example:

**Example 1.1.3.** Suppose  $S$  is any set, then define  $V = \{\text{all functions from } S \text{ to } \mathbb{R}\}$

If we define  $(f + g)(s) = f(s) + g(s)$  and  $(\alpha \cdot f)(s) = \alpha \cdot f(s)$ , and  $0(s) = 0$ , then this is also a vector space.

### Put some linear conditions

**Example 1.1.4.** In  $\mathbb{R}^n$ , fix  $\vec{a} = (a_1, a_2, \dots, a_n) \in \mathbb{R}^n$ , if we define

$$W = \{(x_1, x_2, \dots, x_n) \in \mathbb{R}^n \mid a_1x_1 + a_2x_2 + \dots + a_nx_n = 0\},$$

then this is also a vector space.

However, if we have

$$W' = \{(x_1, \dots, x_n) \in \mathbb{R}^n \mid a_1x_1 + \dots + a_nx_n = 1\},$$

then this is not a vector space because it is not close.

**Example 1.1.5.** In  $V = \{(a, b) \rightarrow \mathbb{R}\}$  or  $W_1 = \{\text{polynomial defined on } (a, b)\}$ , these are both vector space.

**Remark 1.1.1.** In the later course, we will learn that  $W_1$  is a subspace of  $V$ .

**Example 1.1.6.** If we furtherly defined  $W_1^{(k)} = \{\text{polynomial degree } \leq k\}$ , then this is also a vector space.

**Remark 1.1.2.**  $W_1^{(k)}$  is actually isomorphic to  $\mathbb{R}^{k+1}$  since

$$a_0 + a_1x + a_2x^2 + \dots + a_kx^k \leftrightarrow (a_0, a_1, a_2, \dots, a_n).$$

**Example 1.1.7.**  $W_2 = \{\text{continuous function on } (a, b)\}$  and  $W_3 = \{\text{differentiable functions}\}$  are also both vector spaces.

**Example 1.1.8.**  $W_4 = \left\{\frac{d^2f}{dx^2} = 0\right\}$  and  $W_5 = \left\{\frac{d^2f}{dx^2} = -f\right\}$  are both vector spaces.

**Proof.**

$$\begin{aligned} W_4 &= \{a_0 + a_1x\} \\ W_5 &= \{a_1 \cos x + a_2 \sin x\} \end{aligned}$$

⊛

## 1.2 Formal definition of vector spaces

### 1.2.1 Vector Spaces Over $\mathbb{R}$

**Definition 1.2.1.** Suppose  $V$  is a non-empty set equipped with

- addition:  $V \times V \rightarrow V$ , that is, given  $u, v \in V$ , defining  $u + v \in V$
- scalare multiplication:  $\mathbb{R} \times V \rightarrow V$ , that is, given  $\alpha \in \mathbb{R}$  and  $v \in V$ , we need to have  $\alpha v \in V$

Also, we need some good properties or conditions

- For addition,
  - $u + v = v + u$
  - $(u + v) + w = u + (v + w)$
- There exists  $0 \in V$  such that  $u + 0 = u = 0 + u$

- Given  $v \in V$ , there exists  $-v \in V$  such that  $v + (-v) = 0 = (-v) + v$
- For scalar multiplication,
  - $1 \cdot v = v$  for all  $v \in V$
  - $(\alpha\beta)v = \alpha \cdot (\beta v)$  for all  $\alpha, \beta \in \mathbb{R}$  and  $v \in V$ .
- For addition and multiplication,
  - $\alpha(u + v) = \alpha u + \alpha v$
  - $(\alpha + \beta)u = \alpha u + \beta u$

## Lecture 2

### 1.3 Vector Space over general field

5 Sep. 10:20

Now we introduce the concept of field.

**Definition 1.3.1 (Field).** A set  $F$  with  $+$  and  $\cdot$  is called a **field** if

- $\alpha + \beta = \beta + \alpha$  and  $(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$ .
- There exists  $0 \in F$  such that  $\alpha + 0 = 0 + \alpha = \alpha$ .
- For  $\alpha \in F$ , there exists  $-\alpha$  such that  $\alpha + (-\alpha) = 0$ .
- $\alpha\beta = \beta\alpha$  and  $(\alpha\beta)\gamma = \alpha(\beta\gamma)$
- $\exists 1$  such that  $1 \neq 0$  and  $1 \cdot \alpha = \alpha$ .
- For  $\alpha \neq 0$ ,  $\exists \alpha^{-1} \in F$  such that  $\alpha\alpha^{-1} = 1$ .
- $\alpha(\beta + \gamma) = \alpha\beta + \alpha\gamma$

**Example 1.3.1.**  $\mathbb{Q} \subseteq \mathbb{R} \subseteq \mathbb{C}$  are all fields but  $\mathbb{Z}$  is not.

**Example 1.3.2.**  $\{0, 1\}$  is also a field.

Now we know the concept of field, so we can make a vector space over a field.

**Theorem 1.3.1 (Cancellation law).** Suppose  $v_1, v_2, w \in V$ , a vector space, then if  $v_1 + w = v_2 + w$ , then  $v_1 = v_2$ .

**Proof.**

$$v_1 = v_1 + (w + (-w)) = (v_1 + w) + (-w) = (v_2 + w) + (-w) = v_2 + (w + (-w)) = v_2.$$

■

**Theorem 1.3.2.** The zero vector  $0$  is unique.

**Proof.** Suppose we have  $0, 0'$  both zero vector, then for some  $0 = 0 + 0' = 0'$ .

■

**Theorem 1.3.3.** For any  $v \in V$ ,  $0 \cdot u = 0$ .

**Proof.**  $0 \cdot u = (0 + 0) \cdot u = 0 \cdot u + 0 \cdot u$ , so  $0 = 0 \cdot u$  by [cancellation law](#).

■

**Theorem 1.3.4.**  $(-1) \cdot u = -u$ .

**Theorem 1.3.5.** Given any  $u \in V$  is unique,  $-u$  is unique.

## 1.4 Subspaces

**Definition 1.4.1 (subspace).** Let  $V$  be a vector space. A non-empty subset  $W \subseteq V$  is called a subspace of  $V$  if  $W$  is itself a vector space under  $+$  and  $\cdot$  on  $V$ .

**Example 1.4.1.**  $M_n(F) = \{n \times n \text{ matrix with entries in } F\}$  is a vector space, and

$$U_n(F) = \left\{ \begin{pmatrix} a_{11} & & & \\ 0 & a_{22} & & \\ \vdots & & \ddots & \\ 0 & \cdots & 0 & a_{nn} \end{pmatrix} \right\}$$

is a subspace of  $M_n(F)$ .

**Proposition 1.4.1.** Suppose  $V$  is a vector space, and  $W \subseteq V$  is non-empty, then

$W$  is a subspace  $\Leftrightarrow$  For  $u, v \in W, \alpha \in F$ , we have  $u + v \in W$  and  $\alpha \cdot u \in W$ .

**proof of  $\Rightarrow$ .** Clear. ■

**proof of  $\Leftarrow$ .** First, we would want to check  $0 \in W$ , and we can pick any  $u \in W$ , and pick  $\alpha = -1$ , so we know  $-u \in W$ , and thus  $0 = u + (-u) \in W$ . ■

**Corollary 1.4.1.** If we want to check  $W$  is a subspace, we just need to check for  $u, v \in W, \alpha \in F$ ,  $u + \alpha v \in W$  or not.

## 1.5 Linear Combination

**Definition 1.5.1 (Linear combination).** Given  $v_1, v_2, \dots, v_n \in V$ , a linear combination of them is a vector of the form

$$\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n.$$

**Proposition 1.5.1.** Given  $v_1, v_2, \dots, v_n \in V$ ,

1.  $W = \{\text{all linear combinations of } v_1, \dots, v_n\}$  is a subspace.
2. This subspace is the smallest subspace containing  $v_1, \dots, v_n$ . That is, if  $W' \subseteq V$  is a subspace containing  $v_1, \dots, v_n$ , then  $W \subseteq W'$ .

**Notation.**  $\text{span}\{v_1, v_2, \dots, v_n\} = \{\text{all linear combinations of } v_1, v_2, \dots, v_n\}$

## 1.6 Linearly independent

**Definition.** Now we talk about the linear dependence and linear independence.

**Definition 1.6.1 (Linearly dependent).**  $v_1, v_2, \dots, v_n$  are linearly dependent if

$$\alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_n v_n = 0$$

for some  $\alpha_1, \alpha_2, \dots, \alpha_n$  not all zeros.

**Definition 1.6.2 (Linearly independent).**  $v_1, v_2, \dots, v_n$  are called linearly independent if they are not linearly dependent.

**Corollary 1.6.1.** Say  $\alpha_i \neq 0$ , then  $v_i \in \text{span}\{\hat{v}_1, \hat{v}_2, \dots, \hat{v}_k\}$  suppose the corresponding  $\alpha_i$  of  $\hat{v}_1, \dots, \hat{v}_k$  are not zeros.

**Corollary 1.6.2.** Linearly independent means if  $\alpha_1 v_1 + \dots + \alpha_n v_n = 0$ , then  $\alpha_1 = \alpha_2 = \dots = \alpha_n = 0$ .

**Corollary 1.6.3.** Linearly independent means if  $\sum \alpha_i v_i = \sum \beta_i v_i$ , then  $\alpha_i = \beta_i$  for all  $i$ .

**Example 1.6.1.**

- $v \in V$  is linearly independent iff  $v \neq 0$ .
- $v, w \in V$  are linearly independent iff  $v$  is not a scalar of  $w$  and  $w$  is not a scalar of  $v$ .

**Lemma 1.6.1.**  $v_1, \dots, v_n$  are linearly independent iff  $v_i \notin \text{span}\{v_1, \dots, v_{i-1}, v_{i+1}, \dots, v_n\}$ .

## 1.7 Basis

**Definition.** We now talking about basis

**Definition 1.7.1 (Basis).**  $B = \{v_1, v_2, \dots, v_n\}$  is called a basis of  $V$  if  $B$  spans  $V$  and  $B$  is linearly independent.

**Definition 1.7.2 (Dimension).** In this case,  $n$  is called the dimension of  $V$ , and denoted by  $\dim V$ .

**Notation.**  $\text{span}\{v_1, v_2, \dots, v_n\} = \langle v_1, v_2, \dots, v_n \rangle$

**Notation.**  $\text{span}(S) = \langle S \rangle$

**Theorem 1.7.1.** For any  $v \in V$ , it has a unique expression  $v = \sum_{i=1}^n \alpha_i v_i$ .

## Lecture 3

**As previously seen.** A basis of a vector space  $V$  is a set  $\{v_1, v_2, \dots, v_n\}$  that is linearly independent and simultaneously spans  $V$ . That is, suppose we have  $\sum a_i v_i = 0$  for some scalars  $a_i$ , then  $a_i = 0$  for all  $i$ . Also, we call the number  $n$ , the dimension of  $V$ .

10 Sep 10:20



**Example 1.7.1.** Suppose we have  $V = F^n = \{(\alpha_1, \alpha_2, \dots, \alpha_n) \mid \alpha_i \in F\}$ , then we have a **standard basis**, which is

$$\begin{aligned} e_1 &= (1, 0, \dots, 0) \\ e_2 &= (0, 1, \dots, 0) \\ &\vdots \\ e_n &= (0, 0, \dots, 1) \end{aligned}$$

since  $\{e_i\}_{i=1}^n$  is linearly independent and for every  $\vec{a} = (a_1, \dots, a_n)$ , we know

$$\vec{a} = \sum_{i=1}^n a_i e_i.$$

**Example 1.7.2.** Suppose

$$V = M_{n \times n}(F) = \left\{ \begin{pmatrix} \alpha_{11} & \alpha_{12} & \dots & \alpha_{1n} \\ \alpha_{21} & \ddots & & \alpha_{2n} \\ \vdots & & & \\ \alpha_{n1} & \dots & & \alpha_{nn} \end{pmatrix} \right\},$$

then we know

$$\{e_{ij}\}_{1 \leq i, j \leq n} = \begin{pmatrix} 0 & 0 & & \\ 0 & & & \\ & & 1 & \\ 0 & & & 0 \\ 0 & & & 0 \end{pmatrix},$$

where the 1 is in the  $i$ -th row and  $j$ -th column.

**Theorem 1.7.2.** Suppose  $V$  is a vector space, and  $V = \langle v_1, v_2, \dots, v_n \rangle$  and  $\{w_1, w_2, \dots, w_m\}$  is linearly independent, then  $m \leq n$ . Furthermore, one can make

$$\langle w_1, w_2, \dots, w_m, v_{m+1}, \dots, v_n \rangle = V$$

after rearrangement of  $v_1, \dots, v_n$ .

**Proof.** We can do induction on  $m$ . It is trivial that  $m = 0$  is true. Suppose the statement holds for a fixed  $m$  with  $m \leq n$ . Let  $w_1, w_2, \dots, w_{m+1}$  be linearly independent. In particular,  $w_1, w_2, \dots, w_m$  is linearly independent.

**Claim 1.7.1.**  $m + 1 \leq n$ .

**Proof.** Otherwise, if  $m + 1 > n$ , then since  $m \leq n$ , so  $m = n$ . Hence, by induction hypothesis, we know  $\langle w_1, w_2, \dots, w_m \rangle = V$ . However, by [Lemma 1.7.1](#) and the note following it, we know

$$\{w_1, w_2, \dots, w_m\} \cup \{w_{m+1}\}$$

can not be linearly independent since  $w_{m+1} \in V = \langle w_1, \dots, w_m \rangle$ . ⊗

Now we know  $m + 1 \leq n$ . By induction hypothesis, we know

$$\langle w_1, w_2, \dots, w_m, v_{m+1}, \dots, v_n \rangle = V$$

**Claim 1.7.2.** One of  $v_{m+1}, \dots, v_n$  can be replaced by  $w_{m+1}$ .

**Proof.** Since

$$w_{m+1} = \sum_{i=1}^m \alpha_i w_i + \sum_{j=m+1}^n \beta_j v_j.$$

Trivially, one of  $\beta_j \neq 0$ , say  $\beta_{m+1} \neq 0$ . Check

$$\langle w_1, \dots, w_m, w_{m+1}, v_{m+2}, \dots, v_n \rangle = V.$$

⊗

■

**Corollary 1.7.1.** If  $\{v_1, v_2, \dots, v_n\}$  and  $\{w_1, w_2, \dots, w_m\}$  are bases of  $V$ , then  $n = m$ .

**Remark 1.7.1.** Corollary 1.7.1 tells us  $\dim V$  is well-defined, which means the size of the bases of a vector space is unique.

**Corollary 1.7.2.** Suppose  $\dim V = n$ , then if  $\langle v_1, v_2, \dots, v_m \rangle = V$ , then  $m \geq n$ . If  $\{w_1, w_2, \dots, w_m\}$  is linearly independent, then  $m \leq n$ . Also, any  $\{v_i\}_{i=1}^m$  with  $m > n$  is linearly dependent.

**Lemma 1.7.1.** Suppose  $v_1, v_2, \dots, v_n$  is linearly independent. If  $w \notin \langle v_1, v_2, \dots, v_n \rangle$ , then

$$\{v_1, v_2, \dots, v_n, w\}$$

is linearly independent.

**Proof.** Suppose  $\sum_{i=1}^n \alpha_i v_i + \alpha_{n+1} w = 0$ , then if  $\alpha_{n+1} = 0$ , we know  $\alpha_1 = \alpha_2 = \dots = \alpha_n = 0$  since  $\{v_i\}_{i=1}^n$  is linearly independent. If  $\alpha_{n+1} \neq 0$ , then  $w = \frac{1}{\alpha_{n+1}} \sum_{i=1}^n \alpha_i v_i \in \langle v_1, v_2, \dots, v_n \rangle$ , which is a contradiction. ■

**Note 1.7.1.** The reverse of Lemma 1.7.1 is still correct and is trivial. That is, if  $w \notin \{v_1, \dots, v_n\}$  and  $\{v_1, v_2, \dots, v_n, w\}$  is linearly independent, then  $\{v_1, \dots, v_n\}$  is linearly independent.

**Corollary 1.7.3.** If  $W \subseteq V$  is a subspace of  $V$ , then  $\dim W \leq \dim V$ .

**Proof.** If  $\dim V = n$ , and  $\{w_i\}_{i=1}^m$  is a basis of  $W$ , then this basis is linearly independent in  $V$ , which means  $m \leq n$  by Theorem 1.7.2. ■

**Corollary 1.7.4.** If  $v_1, v_2, \dots, v_m$  is linearly independent, then  $\{v_1, v_2, \dots, v_m\}$  forms a basis after adding some  $v_{m+1}, \dots, v_n$  to it.

**Theorem 1.7.3 (Dual version).** If  $\langle v_1, v_2, \dots, v_n \rangle = V$ , then  $\{v_1, v_2, \dots, v_m\}$  forms a basis after rearrangement, where  $m \leq n$ .

**Remark 1.7.2.** Most of the time, we consider finite-dimensional vector spaces.

**Remark 1.7.3 (Examples of  $\infty$ -dim vector space).**

•

$$V = \{\text{all polynomials over } F\} = F[x] = \{a_0 + a_1x + \dots + a_nx^n \text{ for some } n \text{ where } a_i \in F\}.$$

•

$$W = \{(a_0, a_1, \dots) \mid a_i \in \mathbb{R}\}.$$

Notice that

$$W' = \{\text{convergent sequence}\} \subseteq W.$$

and

$$W'' = l^2 = \left\{ (a_i) \mid \sum_{i=0}^{\infty} a_i^2 \text{ finite} \right\} \subseteq W'$$

**Remark 1.7.4.** We define  $\dim \{0\} = 0$ , which is the only vector space with dimension 0, and we define  $\langle \emptyset \rangle = \{0\}$ , which means  $\emptyset$  is the basis of  $\{0\}$ .

**Note 1.7.2.** We call a subspace  $W \subsetneq V$  is proper.

## 1.8 More on subspaces

**Theorem 1.8.1.** If  $W_1$  and  $W_2$  are subspace of  $V$ , then  $W_1 \cap W_2$  is a subspace.

**Theorem 1.8.2.** If  $W_1, W_2$  are subspaces of  $V$ , then  $W_1 + W_2$  is still a subspace of  $V$ .

**Remark 1.8.1.** If  $W_1, W_2$  are subspaces of  $V$ , then  $W_1 \cup W_2$  may not be a subspace. (See HW1).

**Remark 1.8.2.** In fact,  $W_1 \cap W_2$  is the largest subspaces contained in  $W_1$  and  $W_2$ .

**Remark 1.8.3.** In fact,  $W_1 + W_2$  is the smallest subspace containing both  $W_1$  and  $W_2$ .

**Corollary 1.8.1.** Suppose  $S$  is the index set, and for all  $i \in S$ ,  $W_i$  is a subspace of  $V$ , then

$$\bigcap_{i \in S} W_i = \{v \in V \mid v \in W_i \forall i\}$$

is also a subspace of  $V$ .

**Corollary 1.8.2.** Suppose  $S$  is the index set, and for all  $i \in S$ ,  $W_i$  is a subspace of  $V$ , then

$$\sum_{i \in S} W_i = \{w_{i_1} + w_{i_2} + \dots + w_{i_n} \text{ for some } i_j \in S\}$$

is also a subspace of  $V$ .

**Proposition 1.8.1 (Dimension theorem).** Suppose  $W_1, W_2 \subseteq V$  are subspaces of  $V$ , then

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

## Lecture 4

In calculus,  $f : \mathbb{R} \rightarrow \mathbb{R}$  is called continuous if  $f(\lim_{x \rightarrow a} x) = \lim_{x \rightarrow a} f(x)$ .

12 Sep 10:20

**Definition 1.8.1** (Linear transformation). Suppose  $V, W$  are vector spaces over  $F$ . A function

$$\begin{aligned} T : V &\rightarrow W \\ v &\mapsto T(v) \end{aligned}$$

is called a linear transformation or a linear map if

$$T(u + v) = T(u) + T(v) \quad T(\alpha v) = \alpha T(v),$$

or equivalently,

$$T(\alpha u + v) = \alpha T(u) + T(v).$$

**Corollary 1.8.3.** Suppose  $T$  is a linear transformation, then

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

**Example 1.8.1.** Suppose  $V = \{\text{functions from } (-1, 1) \text{ to } \mathbb{R}\}$ , and define  $T_a(f) = f(a)$ , then  $T_a$  is a linear transformation.

**Example 1.8.2.** Consider the space of column vectors,

$$F^n = \left\{ \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix} \mid \alpha_i \in F \right\},$$

and define  $A = (a_{ij}) \in M_{n \times n}(F)$  by

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix},$$

then if we have  $T_A : F^n \rightarrow F^n$  where

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \mapsto A \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix},$$

then  $T_A$  is a linear map.

**Note 1.8.1.**

$$\begin{pmatrix} \vdots \\ \alpha_{i1} & \cdots & \alpha_{in} \\ \vdots \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \vdots \\ \sum_{j=1}^n \alpha_{ij} x_j \\ \vdots \end{pmatrix}$$

**Example 1.8.3.** Consider row of vector space,

$$F^m = \{(\alpha_1, \dots, \alpha_m) \mid \alpha_i \in F\},$$

and  $A \in M_{m \times n}(F)$ , then if  $T_A : F^m \rightarrow F^n$  where

$$T_A : u = (u_1, \dots, u_m) \mapsto (u_1, \dots, u_m) \cdot A$$

is a linear map.

Observe that a linear map  $T : V \rightarrow W$  is determined by  $T(v_i)$ , where  $\{v_1, \dots, v_n\}$  is a basis of  $V$ .

**Proposition 1.8.2.** Suppose  $\{v_1, v_2, \dots, v_n\}$  is a basis of  $V$ , then pick any  $w_1, \dots, w_n \in W$ . Then there is a unique linear map  $T : V \rightarrow W$  satisfying  $T(v_i) = w_i$ .

**Proof.** Since any  $v \in V$  has a unique representation  $v = \sum_{i=1}^n \alpha_i v_i$ . Hence, for a linear map  $T : V \rightarrow W$ , and for any  $v \in V$ , we know

$$T(v) = T\left(\sum_{i=1}^n \alpha_i v_i\right) = \sum_{i=1}^n \alpha_i T(v_i) = \sum_{i=1}^n \alpha_i w_i.$$

Hence, if such map exists, then it must be unique. Now we have to show the existence of this map. Now if we define a map

$$T\left(\sum_{i=1}^n \alpha_i v_i\right) = \sum_{i=1}^n \alpha_i w_i,$$

then we can check this is a linear map. ■

**Example 1.8.4.** Suppose  $F^n$  is the span of column vectors, and  $A \in M_{m \times n}(F)$ , and define  $T_A(v) = Av$ , then we can check  $T_A(e_i) = c_i$ , where  $c_i$  is the  $i$ -th column of  $A$ . This is the linear map that sends  $e_i$  to  $c_i \in F^m$ . If we pick  $c_1, c_2, \dots, c_n \in F^m$ , then there is a unique map sending  $e_i$  to  $c_i$ . In fact, this map is

$$T_A : v \mapsto Av$$

, where the  $i$ -th column of  $A$  is  $c_i$ .

**Definition.** Given  $T : V \rightarrow W$ , where  $T$  is linear.

**Definition 1.8.2 (Kernel).** The kernel/nullspace of  $T$  is defined as

$$\ker(T) = \{v \in V \mid T(v) = 0\} \subseteq V.$$

**Definition 1.8.3 (Image).** The image/range of  $T$  is defined as

$$\text{Im}(T) = \{T(v) \mid v \in V\} \subseteq W.$$

**Remark 1.8.4.** Kernel and Image are subspaces.

## Lecture 5

**As previously seen.** Given such a linear map  $T : V \rightarrow W$ , we define

17 Sep. 10:20

$$\ker T = T^{-1}(0) \quad \text{kernel/null space of } T$$

$$\text{Im } T = T(V) \quad \text{image/range of } T,$$

and  $\ker T$  is a subspace of  $V$ , and  $\text{Im } T$  is a subspace of  $W$ .

**Definition.** Now we define the nullity and rank of a linear map.

**Definition 1.8.4 (nullity).** The nullity of  $T$  is the number

$$\nu(T) = \dim \ker T.$$

**Definition 1.8.5 (rank).** The rank of  $T$  is the number  $\text{rank } T = \dim \text{Im } T$ .

**Example 1.8.5.** Suppose  $T : F^n \rightarrow F^m$ , where  $F^n$  is the column space of dimension  $n$ , then  $T = T_A$  for a matrix  $A \in M_{m \times n}(F)$  and  $T_A(v) = Av$ .

**Proof.** Suppose  $A = (c_1, c_2, \dots, c_n)$ , where  $c_i$  is the  $i$ -th column vector of  $A$ . Consider the standard basis  $\{e_1, e_2, \dots, e_n\}$  of  $F^n$ , where  $e_i$  is the column vector with  $i$ -th position 1 and the other entries are all 0's. Then,  $T_A(e_i) = c_i \in F^m$ . Explicitly,

$$T_A \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = (c_1 \quad \dots \quad c_n) \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = x_1 c_1 + \dots + x_n c_n$$

since we know

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \sum_i x_i e_i.$$

and  $T_A(e_i) = c_i$ . In this case,

$$\begin{aligned} \ker T_A &= \text{all linear relations among } c_1, \dots, c_n \subseteq F^n \\ \text{Im } T_A &= \text{span } \{c_1, \dots, c_n\} \subseteq F^m. \end{aligned}$$

If we want to solve  $\ker T_A$ , then we need to solve

$$0 = x_1 \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix} + x_2 \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix} + \dots + x_n \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix}.$$

Hence, we have to solve

$$\begin{cases} a_{11}x_1 + \dots + a_{1n}x_n = 0 \\ a_{21}x_1 + \dots + a_{2n}x_n = 0 \\ \vdots \\ a_{m1}x_1 + \dots + a_{mn}x_n = 0. \end{cases}$$

Given  $A = (c_1, \dots, c_n)_{m \times n}$ , then the column rank is  $\dim \langle c_1, \dots, c_n \rangle$ . If we rewrite  $A = (r_1, \dots, r_m)^t$ , where  $r_i$  is the  $i$ -th row of  $A$ , then the row rank is  $\dim \langle r_1, r_2, \dots, r_m \rangle$ . Since we can define  $S_A : F^m \rightarrow F^n$ , where

$$v = (x_1, \dots, x_m) \mapsto vA.$$

**Remark 1.8.5.** In fact, column rank is equal to row rank in a matrix, and we will prove it later.

⊛

**Theorem 1.8.3** (rank and nullity theorem). Suppose  $T : V \rightarrow W$  is a linear map, then

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

**Proof.** Since  $\ker T \subseteq V$ , so take a basis  $\{v_1, \dots, v_\nu\}$  of  $\ker T$ , and  $\text{Im } T \subseteq W$ , so take a basis  $\{w_1, \dots, w_r\}$  of  $\text{Im } T$ . Take  $u_j$  s.t.  $T(u_j) = w_j$ .

**Claim 1.8.1.**  $S = \{v_1, \dots, v_\nu, u_1, \dots, u_r\}$  forms a basis of  $V$ .

**Proof.** We first show that  $S$  is linearly independent. Suppose  $\sum \alpha_i v_i + \sum \beta_j u_j = 0$ . Apply  $T$  on it, we get

$$0 = \sum \alpha_i T(v_i) + \sum \beta_j T(u_j) = \sum \alpha_i T(v_i) + \sum \beta_j w_j = \sum \beta_j w_j.$$

However,  $\{w_j\}$  is linearly independent, so  $\beta_j = 0$  for all  $j$ . Now we know  $\sum \alpha_i v_i = 0$ , which means  $\alpha_i = 0$  for all  $i$ , so  $S$  is linearly independent. Now we want to show  $\langle S \rangle = V$ . Given  $v \in V$ , we know  $T(v) \in \text{Im } T$ , and thus we can represent it as  $T(v) = \sum \beta_j w_j$ . We want to show

$$v = \sum \alpha_i v_i + \sum \beta_j u_j.$$

Thus, we want to show  $v - \sum \beta_j u_j \in \ker T$ , but note that

$$T\left(v - \sum \beta_j u_j\right) = T(v) - \sum \beta_j w_j = \sum \beta_j w_j - \sum \beta_j w_j = 0,$$

so we're done, and thus we have

$$v - \sum \beta_j u_j = \sum \alpha_i v_i$$

for some  $\alpha_i$ 's, and we're done. ⊗

Hence,  $\dim V = |S| = \nu T + \text{rank } T$ . ■

**Remark 1.8.6.** If  $\dim V > \dim W$ , then  $\nu(T) > 0$ . Since,  $\text{rank } T \leq \dim W$ , so if  $\dim V > \dim W$ , then we have  $\nu(T) = \dim V - \text{rank } T \geq \dim V - \dim W > 0$ .

**As previously seen.** A map  $f : X \rightarrow Y$  is called one-to-one or 1-1 or injective if  $f(x_1) = f(x_2)$  implies  $x_1 = x_2$ .  $f$  is called onto, surjective if  $f(X) = Y$ .  $f$  is called bijective if it is both 1-1 and onto. In this case, there is the inverse map  $f^{-1} : Y \rightarrow X$  with  $y \mapsto x$  if  $f(x) = y$ .

**Proposition 1.8.3.** Let  $T : V \rightarrow W$  be linear, then  $T$  is injective iff  $\ker T = \{0\}$ .

**Proof.**

( $\Rightarrow$ ) If  $v \in \ker T$ , then since  $T(0) = 0$ , so  $v = 0$ .

( $\Leftarrow$ ) If  $T(v_1) = T(v_2)$ , then  $T(v_1 - v_2) = 0$ , which means  $v_1 - v_2 \in \ker T = \{0\}$ , so  $v_1 = v_2$ , which means  $T$  is linear. ■

**Proposition 1.8.4.** If  $T : V \rightarrow W$  is a linear map, and if  $b$  is a basis of  $V$ , then  $T$  is injective if and only if  $T(b)$  is linearly independent.

**Proof.**

( $\Rightarrow$ ) Suppose  $v_1, v_2, \dots, v_n$  is a basis of  $V$  and we want to show  $T(v_1), \dots, T(v_n)$  is linearly inde-

pendent. Suppose  $\sum \alpha_i T(v_i) = 0$ , then  $T(\sum \alpha_i v_i) = 0$ , so  $\sum \alpha_i v_i = 0$ , and thus  $\alpha_i = 0$  for all  $i$ .

( $\Leftarrow$ )  $T$  sends one particular basis  $v_1, \dots, v_n$  to a linearly independent set. We want to show  $\ker T = \{0\}$ . Suppose  $v \in \ker T$ , then if  $v = \sum \alpha_i v_i$ , we have

$$0 = T\left(\sum \alpha_i v_i\right) = \sum \alpha_i T(v_i),$$

but since  $\{T(v_i)\}$  is linearly independent, so  $\alpha_i = 0$  for all  $i$ , which means  $v = 0$ . ■

**Proposition 1.8.5.** If  $T : V \rightarrow W$  is a linear map, then TFAE

- (a)  $T$  is surjective
- (b)  $T$  sends any basis to a generating set.
- (c)  $T$  sends one basis to a generating set.

**Theorem 1.8.4 (isomorphism).** Suppose  $T : V \rightarrow W$  is linear and bijective, then there is the inverse map  $T^{-1} : W \rightarrow V$ , and  $T^{-1}$  is also linear. In this case,  $T : V \rightarrow W$  is called an isomorphism.

**Definition 1.8.6.** If  $T$  is both injective and surjective, then  $T$  is an isomorphism.

**Remark 1.8.7.** If there is an isomorphism from  $V$  to  $W$ , we say  $V$  is isomorphic to  $W$ , or  $V$  and  $W$  are isomorphic.

**Example 1.8.6 (Coordinates).** If  $\dim V = n$ , then  $V$  is isomorphic to  $F^n$ , we write  $V \simeq F^n$ .

**Proof.** In fact, given an order basis  $B = \{v_1, \dots, v_n\}$  of  $V$ , then we know  $v = \sum_{i=1}^n \alpha_i v_i$ , where

$$v = \sum_{i=1}^n \alpha_i v_i \mapsto [v]_B = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix},$$

and this is a bijection. Note that this map is well-defined since any  $v$  has unique coordinate under  $B$ . Hence, we have  $v_i \mapsto [v_i]_B = e_i$ . ⊗

Hence, if  $T : V \rightarrow W$ , and we know  $V \simeq F^n$  and  $W \simeq F^m$ , and we know there is a matrix sends  $F^n$  to  $F^m$ , called  $[T]_{B'}^B$ , and we can use it to represent the transformation from  $V$  to  $W$ , which is  $T$ .

**Exercise 1.8.1.**  $T^{-1}(w_1 + w_2) = T^{-1}(w_1) + T^{-1}(w_2)$ .

**Proof.** Suppose  $T(v_3) = w_1 + w_2$ , we want to show  $v_3 = v_1 + v_2$ . Hence, we need to check

$$w_1 + w_2 = T(T^{-1}(w_1) + T^{-1}(w_2)) = T(T^{-1}(w_1)) + T(T^{-1}(w_2)) = w_1 + w_2,$$

which is true. ■

## Lecture 6

**As previously seen.**  $T$  is called an isomorphism if  $T$  is both injective and surjective.

19 Sep. 10:20



**Proposition 1.8.6.** Suppose  $\dim V = \dim W = n$ , then TFAE

- (i)  $T$  is an isomorphism.
- (ii)  $T$  is injective.
- (iii)  $T$  is surjective.
- (iv)  $T$  sends any basis of  $V$  to a basis of  $W$ .
- (v)  $T$  sends one basis to a basis.

**Example 1.8.7.** Suppose  $A \in M_{m \times n}(F)$ , say  $A = (c_1, c_2, \dots, c_n)$ , then  $T_A$  is injective if and only if  $\{c_1, \dots, c_n\}$  is linearly independent. (which means  $n \leq m$ ).

**Proof.** Since  $T_A(e_i) = c_i$  and  $\{e_i\}_{i=1}^n$  forms a basis. \*

**Example 1.8.8.** Following the last example,  $T_A$  is surjective if and only if  $\{c_1, c_2, \dots, c_n\}$  spans  $W$ . (which means  $n \geq m$ ).

## 1.9 Space of linear maps

Consider

$$\{f : V \rightarrow W\},$$

and then we can define addition and multiplication by

$$(f + g)(v) = f(v) + g(v) \quad (\alpha \cdot f)(v) = \alpha f(v).$$

Hence, we know it is a vector space. Now if we collect all linear maps, say

$$\mathcal{L}(V, W) = \{\text{linear } T : V \rightarrow W\}.$$

Observe that  $\mathcal{L}(V, W)$  is a vector space since we can similarly define the addition and multiplication.

Now if we have  $U, V, W$ , three vector spaces, and  $f : U \rightarrow V$  is a linear map, then if we define a map

$$\begin{aligned} R_f : \mathcal{L}(V, W) &\rightarrow \mathcal{L}(U, W) \\ T &\mapsto T \circ f, \end{aligned}$$

then this map is linear. Similarly,

$$\begin{aligned} L_f : \mathcal{L}(W, U) &\rightarrow \mathcal{L}(W, V) \\ T &\mapsto f \circ T, \end{aligned}$$

then this is also a linear map.

**Note 1.9.1.** We just need to check something like

$$R_f(T + S) = R_f(T) + R_f(S) \quad R_f(\alpha T) = \alpha R_f(T).$$

Now if we consider

$$\begin{aligned} \mathcal{L}(V, W) \times \mathcal{L}(U, V) &\rightarrow \mathcal{L}(U, W) \\ (T, S) &\mapsto T \circ S, \end{aligned}$$

then this is also a linear map.

**Example 1.9.1.**  $\mathcal{L}(F^n, F^m) = M_{m \times n}(F)$ .

**Proof.** Check that

$$T_A + T_B = T_{A+B}.$$

**Note 1.9.2.** More precisely, they are isomorphic, that is,  $\mathcal{L}(F^n, F^m) \cong M_{m \times n}(F)$ .

⊛

**Example 1.9.2.** Consider

$$\mathcal{L}(F^n, F^m) \times \mathcal{L}(F^p, F^n) \rightarrow \mathcal{L}(F^p, F^m),$$

we know this is a linear map, and by [Example 1.9.1](#), we know

$$M_{m \times n}(F) \times M_{n \times p}(F) \rightarrow M_{m \times p}(F)$$

is a linear map.

**Proof.** Check

$$(T_A \circ T_B)(v) = T_{AB}(v) \Leftrightarrow A(Bv) = (AB)(v).$$

⊛

**Definition 1.9.1.** We call

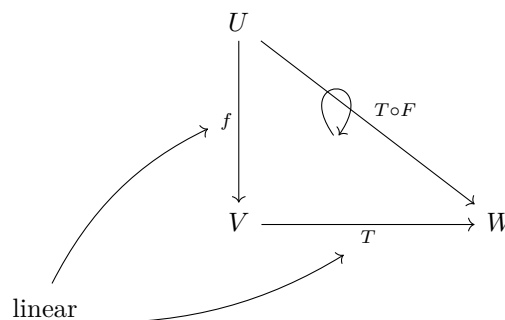
$$V \cong F^n$$

a basic isomorphisms if  $\dim V = n$ .

**Corollary 1.9.1.**  $\mathcal{L}(F^n, F^m) \cong M_{m \times n}(F)$ .

**Remark 1.9.1.** If you change  $F^n$  to  $V$  and  $F^m$  to  $W$ , then this is also correct since  $F^n \cong V$  and  $F^m \cong W$ . (We suppose  $\dim V = n$  and  $\dim W = m$ .)

## Lecture 7



24 Sep. 10:20

There is a special case,

$$\mathcal{L}(V, V) := \mathcal{L}(V) = \{\text{linear } T : V \rightarrow V\},$$

which is the space of linear operators on  $V$ .

Now consider linear  $T_A : F^n \rightarrow F^m, T_B : F^p \rightarrow F^m$ , then we can define a map  $T_{AB} = T_A \circ T_B$ , and it will be a linear map.

$$\begin{array}{ccc}
 F^p & & \\
 \downarrow T_B & \searrow T_A \circ T_B = T_{AB} & \\
 F^n & \xrightarrow{T_A} & F^m
 \end{array}$$

Also, note that  $T_A, T_B$  corresponds to two matrices  $A, B$ , respectively, and it turns out that  $T_{AB}$  corresponds to the matrix  $AB$ . (Check)

Hence,  $\mathcal{L}(F^n) = M_n(F)$ .

A matrix  $P$  is called invertible if  $T_P$  is bijective. In this case,

$$\begin{array}{ccc}
 F^n & \xrightarrow{T_P} & F^m \\
 & \xleftarrow{T_Q} &
 \end{array}$$

Hence, there exists  $Q \in M_n(F)$  s.t.  $QP = PQ = I_n$  since we know  $T_P \circ T_Q = T_Q \circ T_P = I$ .

Thus, we have

$$P = (c_1, c_2, \dots, c_n) \text{ invertible} \Leftrightarrow \{c_1, \dots, c_n\} \text{ is a basis.}$$

by [Proposition 1.8.6](#).

## 1.10 Map/matrix correspondence

$$\begin{array}{ccc}
 V & \xrightarrow{T} & W \\
 \downarrow [\cdot]_B & \circlearrowleft & \downarrow [\cdot]_{B'} \\
 F^n & \xrightarrow{\text{What is this?}} & F^m
 \end{array}$$

Take an ordered basis  $B = \{v_1, v_2, \dots, v_n\}$  and  $B' = \{w_1, \dots, w_m\}$ , and says

$$T(v_j) = \sum_{i=1}^m \alpha_{ij} w_i \mapsto \begin{pmatrix} \alpha_{1j} \\ \vdots \\ \alpha_{mj} \end{pmatrix}.$$

Now consider the matrix

$$A = (\alpha_{ij}) = ([T(v_1)]_{B'}, [T(v_2)]_{B'}, \dots),$$

and then we called  $A$  the matrix of  $T$  relative to  $B$  and  $B'$ . (matrix representative of  $T$ ), and we denote this by  $[T]_{B'}^B$ .

### Theorem 1.10.1.

$$[T(v)]_{B'} = [T]_{B'}^B [v]_B.$$

**Theorem 1.10.2.** We have  $[\cdot]_{B'}^B : \mathcal{L}(V, W) \rightarrow M_{m \times n}(F)$ , and this matrix representative  $[\cdot]_{B'}^B$  is an isomorphism, which means

- $[T + S]_{B'}^B = [T]_{B'}^B + [S]_{B'}^B$ .
- It is bijective.

**Corollary 1.10.1.** if  $\dim V = n$  and  $\dim W = m$ , then

$$\dim(\mathcal{L}(V, W)) = \dim V \cdot \dim W.$$

**Theorem 1.10.3.**

$$[T]_{B'}^B [S]_B^{B''} = [T \circ S]_{B'}^{B''}.$$

$$\begin{array}{ccccc}
 & & V & \xrightarrow{\quad} & W \\
 & & \downarrow & & \downarrow \\
 v_j & & F^n & \xrightarrow{\quad} & F^m \\
 \uparrow & & & & \\
 e_j & \xrightarrow{\quad} & c_j = (\alpha_{1j}, \dots, \alpha_{mj})^t & & \sum_{i=1}^n \alpha_{ij} w_i
 \end{array}$$

Special case:

$$\mathcal{L}(V) \rightarrow M_n(F).$$

Take an ordered basis  $B = \{v_1, \dots, v_n\}$ . If  $T \in \mathcal{L}(V)$ , then we can define  $[T]_B = [T]_B^B$ .

**Corollary 1.10.2.** Given  $T : V \rightarrow W$ . There are  $B = \{v_1, \dots, v_n\}$  and  $B' = \{w_1, \dots, w_m\}$  where  $B$  is a basis of  $V$  and  $B'$  is a basis of  $W$  and

$$[T]_{B'}^B = \begin{pmatrix} I_p & 0 \\ 0 & 0 \end{pmatrix},$$

where  $p = \text{rank}(T)$ .

**Proof.** We can let  $B = \{v_1, \dots, v_r, v_{r+1}, \dots, v_n\}$ , where  $\{v_{r+1}, \dots, v_n\}$  is a basis of  $\ker T$  and  $T(v_1), \dots, T(v_r)$  is a basis of  $\text{Im}(T)$ , (Recall the proof in [Theorem 1.8.3](#)), then we can let  $B' = \{T(v_1), \dots, T(v_r), \dots\}$ . ■

**Example 1.10.1.** Suppose  $V = \{\text{polynomials with degree } \leq k\}$  and  $W$  is the space of polynomials with degree  $\leq k+1$ , then if  $T : V \rightarrow W$  and  $p(x) \mapsto \int_0^x p(t) dt$ , then we know an ordered basis  $B = \{1, x, x^2, \dots, x^k\}$  and  $B' = \{1, x, x^2, \dots, x^{k+1}\}$ , and then

$$[T]_{B'}^B = \begin{pmatrix} 0 & 0 & & & \\ 1 & 0 & & & \\ 0 & \frac{1}{2} & & & \\ \vdots & 0 & \ddots & & 0 \\ 0 & 0 & & \frac{1}{k+1} & \end{pmatrix}.$$

**Example 1.10.2.** Suppose  $V$  is the space of polynomials of degree  $\leq k$ , and  $B = \{1, x, x^2, \dots, x^k\}$ , and  $B' = \{1, y, y^2, \dots, y^k\}$  with  $y = x - 1$ . Then, if  $T$  is the identity transformation, note that

$$x^j = (y+1)^j = 1 + j \cdot y + \binom{j}{2} y^2 + \dots + \binom{j}{j} y^j.$$

Hence, we have

$$[T]_{B'}^B = \begin{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} & \begin{pmatrix} 1 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} & \begin{pmatrix} 2 \\ 2 \\ 1 \\ \vdots \\ 0 \end{pmatrix} & \ddots \end{pmatrix}$$

**Question.** Given  $V$ , and  $B, B'$  are ordered basis, then what is the relation between  $[v]_B$  and  $[v]_{B'}$ ?

**Answer.** Change of bases. ⊛

**Corollary 1.10.3.**

$$[id]_{B'}^B [v]_B = [v]_{B'}.$$

**Corollary 1.10.4.**

$$[id]_{B'}^B [id]_B^{B'} = [id]_{B'}^{B'}.$$

**Corollary 1.10.5.** Given any  $A \in M_{m \times n}(F)$ . There are invertible matrices  $P \in M_m(F)$  and  $Q \in M_n(F)$  s.t.

$$PAQ = \begin{pmatrix} I_p & 0 \\ 0 & 0 \end{pmatrix},$$

where  $p$  is the row rank of  $A$ .

**Proof.** Suppose  $A = [T]_B^{B'}$ , and by [Corollary 1.10.2](#), we know there exists  $b, b'$  s.t.  $[T]_b^{b'}$  is the matrix we want, then we can let  $Q = [id]_{b'}^{B'}$  and  $P = [id]_B^b$ , and we're done. ■

## Lecture 8

**Lemma 1.10.1.** Consider

$$V' \xrightarrow{f} V \xrightarrow{T} W \xrightarrow{g} W'$$

- Suppose  $g$  is injective, then  $\ker(g \circ T) = \ker T$ .
- Suppose  $f$  is surjective, then  $\text{Im}(T \circ f) = \text{Im } T$ .

26 Sep. 10:20

**Definition 1.10.1 (Matrix Equivalence).** Let  $A, B \in M_{m \times n}(\mathbb{F})$ . We say that  $A$  and  $B$  are *equivalent* if there exist invertible matrices  $P \in GL_m(\mathbb{F})$  and  $Q \in GL_n(\mathbb{F})$  such that

$$B = PAQ.$$

**Remark 1.10.1.** Matrix equivalence means that one can obtain  $B$  from  $A$  by a sequence of invertible row and column operations.

Equivalently, if  $A$  represents a linear map  $T : \mathbb{F}^n \rightarrow \mathbb{F}^m$ , then  $B$  represents the same linear map with respect to different bases of the domain and codomain.

**Theorem 1.10.4 (Row Rank Equals Column Rank).** Let  $A \in M_{m \times n}(\mathbb{F})$  be any matrix over a field  $\mathbb{F}$ . Then

$$\text{row rank}(A) = \text{column rank}(A).$$

**Proof.** We prove this using invertible row and column operations.

**Step 1: Reduce  $A$  to canonical form.**

It is a standard fact that any matrix  $A \in M_{m \times n}(\mathbb{F})$  can be transformed into a block matrix of the form

$$C = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}_{m \times n},$$

by multiplying on the left and right by invertible matrices  $P \in GL_m(\mathbb{F})$  and  $Q \in GL_n(\mathbb{F})$ :

$$C = PAQ.$$

Here  $r = \text{rank}(A)$  and  $I_r$  is the  $r \times r$  identity matrix. This uses Gaussian elimination (invertible row operations) and invertible column operations.

**Step 2: Row and column ranks of  $C$ .**

- The first  $r$  rows of  $C$  are linearly independent, and the remaining  $m - r$  rows are zero. So

$$\text{row rank}(C) = r.$$

- The first  $r$  columns of  $C$  are linearly independent, and the remaining  $n - r$  columns are zero. So

$$\text{column rank}(C) = r.$$

**Step 3: Equivalence preserves row and column ranks.**

We have  $C = PAQ$ .

1. *Left multiplication by  $P$  (row operations):* Multiplying  $A$  on the left by invertible  $P$  corresponds to invertible row operations. Row operations do not change the linear independence of the rows. Hence

$$\text{row rank}(PA) = \text{row rank}(A).$$

2. *Right multiplication by  $Q$  (column operations):* Each row of  $AQ$  is obtained by multiplying the corresponding row of  $A$  by  $Q$ :

$$\text{row}_i(AQ) = \text{row}_i(A) \cdot Q.$$

Since  $Q$  is invertible, this is an invertible linear transformation on  $\mathbb{F}^n$ , which preserves linear independence of the rows. Therefore

$$\text{row rank}(AQ) = \text{row rank}(A).$$

**Note 1.10.1.**

$$\sum_{i \in I} \alpha_i \text{row}_i(A) \cdot Q = 0 \Leftrightarrow \sum_{i \in I} \alpha_i \text{row}_i(A) = 0$$

since  $Q$  is invertible.

Combining the above, for  $C = PAQ$  we get

$$\text{row rank}(C) = \text{row rank}(A) = r,$$

and similarly

$$\text{column rank}(C) = \text{column rank}(A) = r.$$

**Step 4: Conclusion.**

From Step 2 and Step 3, we have

$$\text{row rank}(A) = \text{row rank}(C) = r = \text{column rank}(C) = \text{column rank}(A).$$

Hence, the row rank of  $A$  equals the column rank of  $A$ . ■

**Theorem 1.10.5.** Two matrices  $A$  and  $B$  of same sizes are equivalent if and only if  $\text{rank}(A) = \text{rank}(B)$ .

**Proof.** Suppose  $A, B$  equivalent, then  $A = PBQ$  for some invertible  $P, Q$ . By [Lemma 1.10.1](#), we know  $\text{Im}(BQ) = \text{Im } B$ , which gives  $\text{rank}(BQ) = \text{rank } B$ . Also, since  $\ker(P(BQ)) = \ker(BQ)$ , so  $\text{rank}(P(BQ)) = \text{rank}(BQ)$  by rank and nullity theorem. Hence, we have  $\text{rank } A = \text{rank}(PBQ) = \text{rank}(BQ) = \text{rank } B$ .

Now if  $\text{rank } A = \text{rank } B$ , then we know

$$PAQ = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} = P'BQ',$$

so  $A = P^{-1}P'BQ'Q^{-1}$ , which means  $A, B$  are equivalent. ■

**Theorem 1.10.6.** Let  $T : V \rightarrow W$  be a linear transformation between finite-dimensional vector spaces over a field  $\mathbb{F}$ . Let  $B = \{v_1, \dots, v_n\}$  be a basis for  $V$  and  $C = \{w_1, \dots, w_m\}$  be a basis for  $W$ . Let

$$A = [T]_{B,C} \in M_{m \times n}(\mathbb{F})$$

be the matrix of  $T$  with respect to the bases  $B$  and  $C$ . Then

$$\text{rank}(A) = \dim(\text{Im}(T)).$$

**Proof. Step 1: Express the image of  $T$  in terms of the basis.**

The matrix  $A$  is given by

$$A = [T(v_1)]_C [T(v_2)]_C \dots [T(v_n)]_C,$$

where  $[T(v_j)]_C$  denotes the coordinate vector of  $T(v_j)$  with respect to  $C$ .

Since  $B$  is a basis for  $V$ , any vector  $v \in V$  can be written as

$$v = c_1 v_1 + c_2 v_2 + \dots + c_n v_n$$

for some scalars  $c_1, \dots, c_n \in \mathbb{F}$ . By linearity of  $T$ ,

$$T(v) = c_1 T(v_1) + c_2 T(v_2) + \dots + c_n T(v_n).$$

Thus, every vector in  $\text{Im}(T)$  is a linear combination of

$$\{T(v_1), T(v_2), \dots, T(v_n)\},$$

and hence

$$\text{Im}(T) = \text{span}\{T(v_1), T(v_2), \dots, T(v_n)\}.$$

**Step 2: Relate  $\text{Im}(T)$  to the column space of  $A$ .**

The column space of  $A$ , denoted  $\text{Col}(A)$ , is

$$\text{Col}(A) = \text{span}\{[T(v_1)]_C, [T(v_2)]_C, \dots, [T(v_n)]_C\}.$$

The coordinate mapping  $[\cdot]_C : W \rightarrow \mathbb{F}^m$  is a linear isomorphism. In particular, it preserves linear independence and spanning sets. Therefore, the map

$$T(v_j) \longmapsto [T(v_j)]_C$$

establishes a linear isomorphism between  $\text{Im}(T)$  and  $\text{Col}(A)$ :

$$\text{Im}(T) \cong \text{Col}(A).$$

**Step 3: Compare dimensions.**

Since isomorphic vector spaces have the same dimension,

$$\dim(\text{Im}(T)) = \dim(\text{Col}(A)).$$

By definition, the rank of  $A$  is the dimension of its column space:

$$\text{rank}(A) = \dim(\text{Col}(A)).$$

Combining these equalities, we obtain

$$\text{rank}(A) = \dim(\text{Im}(T)),$$

as desired.

This shows that the rank of a matrix representing a linear transformation is independent of the choice of bases  $B$  and  $C$ , since  $\dim(\text{Im}(T))$  depends only on  $T$  itself. ■

## Lecture 9

Consider the system

1 Oct. 10:20

$$\begin{cases} a_{11}x_1 + \cdots + a_{1n}x_n = y_1 \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n = y_m. \end{cases}$$

We want to solve  $X$  s.t.  $AX = Y$ , where  $A = (a_{ij})_{m \times n}$  and  $Y = (y_i)_{i=1}^m$ . Suppose  $P \in M_{m \times m}(F)$  invertible, then if  $B = PA$ , we have  $BX = Z$ , which means doing row operations on the system. In this case, we call two systems are equivalent. We also call  $A, B$  are row equivalent.

Now we talk about the types of elementary row operations:

- (i) Replace  $i$ -th row with  $c \cdot r_i$  for some  $c \neq 0$ .
- (ii) Replace  $r_i$  with  $r_i + cr_j$  for some  $j \neq i$ .
- (iii) Interchange  $r_i$  and  $r_j$  for some  $i \neq j$ .

One can use (i) and (ii) in finite steps, making  $A$  into row reduced form (REF) of  $A$ , which means

- first entry of a non-zero row is 1, we called it leading 1
- entries below and above leading 1 are 0.

If allowing (iii), we can make  $A$  into RREF (row reduced echelon form), which means REF and all zero rows are at the bottom.

Note that  $AX = Y$  gives  $PAX = PY$ , so we can write  $P(A | Y) = (PA | PY)$ . Hence, we can do row operations on  $(X | Y)$  so that the  $X$  part becomes REF or RREF to solve the system. The system will be like

$$\begin{aligned} x_{k_1} + \cdots + 0 + \cdots &= z_1 \\ x_{k_2} + \cdots + 0 &= z_2 \\ &\vdots \end{aligned}$$

Suppose for the first  $n$  rows, there are  $r$  non-zero rows. If there is some  $z_i \neq 0$  for  $i > r$ , the system has no solution. If not, there is at least one solution, and there are  $n - r$  free variables.

**Note 1.10.2.** If  $n - r = 0$ , then the system has unique solution, and if  $n - r > 0$ , then it has infinitely many solutions.

In the homogeneous case (i.e.  $y_1 = y_2 = \cdots = y_m = 0$ ), we find  $\nu(A) = n - r$ . In this case, if  $n > m$ , then  $n - r > m - r \geq 0$ , so there are non-zero solutions to  $AX = 0$ .



**Some consequences:**

- If  $A \in M_n(F)$ , then TFAE
  - The system  $AX = 0$  has only trivial solution (injective).
  - For any  $Y$ ,  $AX = Y$  has a (unique) solution (surjective).
  - $A$  is invertible.

If  $P, Q$  are invertible, then  $(PQ)^{-1} = Q^{-1}P^{-1}$ . Also, by above mentioned things, we know every invertible matrix is a product of many elementary matrix, that is,  $A = (E_1)^{-1}(E_2)^{-1} \dots (E_m)^{-1}$  since we know

$$(E_m \dots E_2 E_1)A = I_m.$$

**Note 1.10.3.** If  $A$  is invertible, then  $AX = 0$  has only trivial solution, then its RREF is  $I$ , and thus  $A$  can be recovered to  $I$  by some row operations.

**As previously seen.** If  $\{v_1, \dots, v_n\}$  is linearly independent and  $\{w_1, \dots, w_m\}$  spans  $V$ , then  $n \leq m$ .

Suppose  $x_1 v_1 + \dots + x_n v_n = 0$ , where

$$v_i = a_{1i} w_1 + a_{2i} w_2 + \dots + a_{mi} w_m,$$

then we have

$$a_{i1} x_1 + \dots + a_{im} x_n = 0$$

for all  $1 \leq i \leq m$ . If  $n > m$ , then there exists a non-zero solution to this system, which contradicts to the fact that  $x_1 = x_2 = \dots = x_n = 0$ .

**Corollary 1.10.6.** For  $A \in M_{m \times n}(F)$ , we know there exists invertible  $P, Q$  s.t.

$$PAQ = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}.$$

**Corollary 1.10.7.** row rank is equal to col rank.

**Question.** How to show  $A$  invertible?

**Answer.** Check RREF of  $A$  is  $I_n$  or not. \*

**Question.** How to find  $A^{-1}$ ?

**Answer.** Calculate  $(A \mid I_n)$ . \*

# Chapter 2

## Dual space

Consider a vector space  $V$ , and  $V$  is over a field  $F$ , then we call

$$V^* = \mathcal{L}(V, F).$$

**Definition 2.0.1.** Suppose  $V$  is a vector space over  $F$  (with basis  $\{1\}$ ), then

- A linear functional  $f$  is a linear map  $f : V \rightarrow F$ .
- $V^* = \mathcal{L}(V, F)$  is called the dual space of  $V$ .

**Example 2.0.1.** Suppose  $V = F^n$ , then  $V^* = M_{1 \times n}(F)$ .

Note that Suppose  $f \in V^*$  corresponds to  $(a_1, a_2, \dots, a_n)$ , then  $f(e_i) = a_i$ .

**Example 2.0.2.** Suppose  $V = M_{n \times n}(F)$ , then the tract map

$$\text{tr} : M_{n \times n}(F) \rightarrow F \quad (a_{ij}) \mapsto \sum_{i=1}^n a_{ii}$$

is in  $V^*$ .

**Example 2.0.3.** We can define  $E_{pq}^* \in V^*$  by

$$E_{pq}^*((a_{ij})) = a_{pq},$$

then  $\{E_{ij}^*\}$  is a basis of  $V^*$ .

**Example 2.0.4.** Suppose

$$V = \{\text{continuous function } f : [p, q] \rightarrow \mathbb{R}\},$$

then we can define  $\text{ev}_s$ , the evaluation at  $s$ , by

$$\text{ev}_s(f) = f(s),$$

and we can define  $I : V \rightarrow \mathbb{R}$  with

$$I(f) = \int_p^q f(x) \, dx,$$

then  $\text{ev}_s$  and  $I$  are both elements of  $V^*$ .

## Lecture 10

3 Oct. 10:20

**Definition 2.0.2.**  $A, B \in M_n(F)$  are called similar or  $A \sim B$  iff  $B = P^{-1}AP$ .

**Notation.** We call  $\mathcal{L}(V, F)$

$$V^* \quad \text{or} \quad V^\vee \quad \text{or} \quad V^t.$$

**Theorem 2.0.1.**

$$\dim V = \dim V^*.$$

**Matrix relation proof.** Since  $V^* \simeq M_{1 \times n}(F)$ , where  $n = \dim V$ , so

$$\dim V^* = \dim M_{1 \times n}(F) = n = \dim V.$$

■

**Proof.** Suppose  $B = \{v_1, v_2, \dots, v_n\}$  is a basis of  $V$ , and define  $B^* = \{v_1^*, v_2^*, \dots, v_n^*\}$  as

$$v_i^*(v_j) = \delta_{ij}.$$

Note that  $v_i^* \in \mathcal{L}(V, F)$  for all  $i$ . Note that for all  $v = \sum_{i=1}^n \alpha_i v_i$ , we have

$$v_i^*(v) = \alpha_i.$$

Check  $B^*$  is linearly independent: Suppose  $f = \sum \alpha_i v_i^* = 0$ , then we know  $f(v_j) = \alpha_j = 0$  for all  $j$ . Also, note that  $B^*$  spans  $V^*$ . ■

**Remark 2.0.1.**

$$[v]_B = \begin{pmatrix} v_1^*(v) \\ \vdots \\ v_n^*(v) \end{pmatrix}$$

**Example 2.0.5.** Suppose  $V = F^2$  and  $B = \{e_1, e_2\}$ , then  $V^*$  is identified with

$$\mathcal{L}(F^2, F) = M_{1 \times 2}(F),$$

where  $B^* = \{e_1^*, e_2^*\}$  with

$$e_1^* = (1, 0) \quad e_2^* = (0, 1).$$

Now if we know  $T : V \rightarrow W$  is a linear map, then we can define  $T^* : W^* \rightarrow V^*$  by

$$T^* : f \mapsto f \circ T,$$

and we called it the transpose of  $T$ . We will show that if  $[T]_C^B = M$ , then  $[T^*]_{B^*}^{C^*} = N = M^t$ , which means if  $M = (m_{ij})_{m \times n}$  and  $N = (n_{ij})_{n \times m}$ , then  $n_{ij} = m_{ji}$  for all  $i, j$  with  $1 \leq i \leq n$  and  $1 \leq j \leq m$ .

**Proof.** Suppose  $T^*(w_j^*) = \sum_{p=1}^n n_{pj} v_p^*$ , then since

$$w_j^*(T(v_i)) = w_j^*\left(\sum_{q=1}^m m_{qi} v_q\right) = m_{ji},$$

so  $n_{ij} = m_{ji}$ . (See [Remark 2.0.1](#)) Note that the below one is the evaluation of the above equation at  $v_j$ . ■

## Lecture 11

8 Oct. 10:20

**Definition 2.0.3 (Annihilator).** Let  $S \subseteq V$  be a subset, then the annihilator  $S^0 \subseteq V^*$  is the subset defined by

$$\{f \in V^* \mid f(x) = 0 \quad \forall x \in S\}.$$

**Proposition 2.0.1.** For all  $S \subseteq V$ ,  $S^0$  is a subspace of  $V^*$ .

**Proof.** For all  $f, g \in S^0$ , we know

$$(cf + g)(x) = cf(x) + g(x) = 0 \quad \forall x \in S,$$

so  $cf + g \in S^0$ . ■

**Example 2.0.6.**  $\{0\}^0 = V^*$  and  $V^0 = \{0\}$ .

**Proposition 2.0.2.** If  $S_1 \subseteq S_2$ , then  $S_2^0 \subseteq S_1^0$ .

**Proof.** If  $f \in S_2^0$ , then  $f(x) = 0$  for all  $x \in S_2$ , so  $f(x) = 0$  for all  $x \in S_1$ , and thus  $f \in S_1^0$ , which means  $S_2^0 \subseteq S_1^0$ . ■

**Proposition 2.0.3.** If  $W = \langle S \rangle$ , then  $W^0 = S^0$ .

**Proof.** Since  $S \subseteq W$ , so we know  $W^0 \subseteq S^0$  by Proposition 2.0.2. Also, for all  $f \in S^0$ , we know for all  $x \in \langle S \rangle$ ,  $x = \sum \alpha_i x_i$  where  $x_i$ 's are elements of  $S$ , so

$$f(x) = f\left(\sum \alpha_i x_i\right) = \sum \alpha_i f(x_i) = 0,$$

which means  $S^0 \subseteq W^0$ . ■

**Example 2.0.7.** Suppose  $W_1 \subseteq W_2 \subseteq V$ , then  $W_1^0 \supseteq W_2^0 \supseteq V^0$ .

**Proposition 2.0.4.** Suppose  $V$  is finite dimensional and  $W \subseteq V$ , then  $\dim W + \dim W^0 = \dim V = \dim V^*$ .

**Proof.** Let  $\dim W = m$  and  $\dim V = n$ , and take  $B = \{w_1, \dots, w_m\}$  a basis of  $W$  and  $C = \{w_1, \dots, w_m, v_{m+1}, \dots, v_n\}$  as a basis of  $V$ . If we take dual of  $C$ , suppose

$$C^* = \{w_1^*, w_2^*, \dots, w_m^*, v_{m+1}^*, \dots, v_n^*\},$$

and now we claim  $\{v_{m+1}^*, \dots, v_n^*\}$  is a basis of  $W^0$ . For all  $f \in V^*$ , we know  $f = \sum_{i=1}^m \alpha_i w_i^* + \sum_{j=m+1}^n \beta_j v_j^*$ . Now if  $f \in W^0$ , then we know  $f(w) = 0$  for all  $w \in W$ , so  $f(w_i) = 0$  for all  $w_i$ 's, and thus

$$f(w_i) = \sum_{i=1}^m \alpha_i w_i^*(w_i) + \sum_{j=m+1}^n \beta_j v_j^*(w_i) = \alpha_i = 0,$$

so we know  $f = \sum_{j=m+1}^n \beta_j v_j^*$ , which means  $f \in \langle v_{m+1}^*, \dots, v_n^* \rangle$ . Thus,  $W^0 \subseteq \langle v_{m+1}^*, \dots, v_n^* \rangle$ . Also,  $v_i^*(w) = 0$  for all  $w \in W$ , so we know  $\langle v_{m+1}^*, \dots, v_n^* \rangle \subseteq W^0$ , and we're done. ■

**Corollary 2.0.1.** If  $\dim V, \dim W < \infty$  and  $T : V \rightarrow W$  is linear, and we define  $T^* : W^* \rightarrow V^*$  as  $T$ 's transpose, then  $\text{rank } T = \text{rank } T^*$ .

**Proof.** First we show that  $\ker T^* = (\text{Im } T)^0$ . Suppose  $f \in \ker T^*$ , then

$$0 = T^*(f) = fT,$$

so  $fT(v) = 0$  for all  $v \in V$ , so  $f(w) = 0$  for all  $w \in \text{Im } T$ , so  $f \in (\text{Im } T)^0$ . Conversely, we can similarly show that  $(\text{Im } T)^0 \subseteq \ker T^*$ , and we're done. Note that

$$\dim W^* - \text{rank } T^* = \nu(T^*) = \dim (\text{Im}(T)^0) = \dim W - \dim(\text{Im } T) = \dim W - \text{rank } T,$$

and since  $\dim W = \dim W^*$ , so we know  $\text{rank } T = \text{rank } T^*$ . ■

**Corollary 2.0.2.** Suppose  $A$  is a matrix, then its row rank and column rank are same.

**Proof.** By regarding  $A$  as a linear map  $T$ 's corresponding matrix, then  $T^*$ 's corresponding matrix is  $A^t$ , and since we have shown that  $\text{rank } T = \text{rank } T^*$ , so  $A$ 's row rank is equal to  $A^t$ 's row rank, which is  $A$ 's column rank. ■

## 2.1 Dual of Dual space/Evaluation

We first define that  $V^{**} = (V^*)^*$ , and we can define a linear map

$$\text{ev} : V \rightarrow V^{**}, \quad x \mapsto \tilde{x},$$

where  $\tilde{x}$  is the functional

$$\tilde{x} : V^* \rightarrow F \quad f \mapsto f(x).$$

**Theorem 2.1.1.**  $\text{ev}$  is an isomorphism between  $V$  and  $V^{**}$ .

**Proof.** We can check  $\tilde{x}, \text{ev}$  are linear easily. DIY

**Lemma 2.1.1.** If  $v \in V$  is not zero, then there exists  $f \in V^*$  s.t.  $f(v) \neq 0$ .

**Proof.** Take  $B = \{v_1 = v, v_2, \dots, v_n\}$  as a basis of  $V$  and take dual  $B^*$ , then  $v_1^*(v) = 1$ . ■

**Claim 2.1.1.**  $\text{ev} : V \rightarrow V^{**}$  is injective.

**Proof.** Suppose  $v \in \ker \text{ev}$ , then  $\tilde{v} = 0$ , which means  $f(v) = 0$  for all  $f \in V^*$ , so  $v = 0$  by Lemma 2.1.1, and thus  $\text{ev}$  is injective. ■

Since  $\dim V = \dim V^* = \dim (V^*)^* = \dim V^{**}$ , so injectivity implies bijectivity. ■

**Corollary 2.1.1.** If  $T : V \rightarrow W$  is a linear map with inverse  $S : W \rightarrow V$ , then  $T^* : W^* \rightarrow V^*$ 's inverse is  $S^* : V^* \rightarrow W^*$ , where  $S^*$  is the transpose of  $S$ .

**Corollary 2.1.2 (Matrix ver).** Suppose  $A \in M_n(F)$  is invertible, then  $A^t$  is invertible, and

$$(A^t)^{-1} = (A^{-1})^t.$$

## Chapter 3

# Eigenvalue and Eigenvector

### Lecture 12

15 Oct. 10:20

**Question.** If  $V$  is a vector space and  $\dim V < \infty$ , if  $T : V \rightarrow V$  is a linear map, then is there a basis of  $V$ ,

$$B = \{v_1, v_2, \dots, v_n\}$$

s.t.  $T(v_i) = \lambda_i v_i$  for some  $\lambda_i \in F$  i.e.

$$[T]_B = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}.$$

Note that this question is equivalent to find some linearly independent  $\{v_i\}_{i=1}^n$  s.t.

$$A \underbrace{\begin{pmatrix} v_1 & v_2 & \cdots & v_n \end{pmatrix}}_P = \begin{pmatrix} \lambda_1 v_1 & \lambda_2 v_2 & \cdots & \lambda_n v_n \end{pmatrix} = \underbrace{\begin{pmatrix} v_1 & v_2 & \cdots & v_n \end{pmatrix}}_P \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix},$$

which means is there invertible  $P$  s.t.  $P^{-1}AP$ ?

**Question.** Why we want to diagonalize a matrix?

**Answer.** If we have  $A = PBP^{-1}$ , then  $A^k = PB^kP^{-1}$ , and if  $B$  is diagonal, say

$$B = \begin{pmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{pmatrix},$$

then

$$B^k = \begin{pmatrix} \lambda_1^k & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n^k \end{pmatrix},$$

and it is easy to compute. ⊗

One of the applications of diagonalization is about recurrence relation. If we have a sequence  $\{a_i\}_{i=0}^{\infty}$ , where

$$a_{k+2} = \alpha a_{k+1} + \beta a_k,$$

then suppose  $v_k = (a_k, a_{k+1})^t$ , then

$$v_k = \begin{pmatrix} 0 & 1 \\ \alpha & \beta \end{pmatrix} \begin{pmatrix} a_{k-1} \\ a_k \end{pmatrix} = Av_{k-1},$$

so we have  $v_k = A^k v_0$ , and thus if we know diagonalization, then we can compute  $A^k$  quickly.

Now we talk about how to find  $\lambda, v$  s.t.  $T(v) = \lambda v$ . If  $v = 0$ , then it is trivial, so we suppose  $v \neq 0$ , and thus it is equivalent to find  $\lambda, v$  s.t.

$$(T - \lambda I)(v) = 0.$$

**Definition 3.0.1 (Singular).** A matrix or linear operator is singular if it is not invertible.

Thus, we want to find  $\lambda$  s.t.  $T - \lambda I$  is singular since if  $T - \lambda I$  is invertible, then  $v = 0$ .

**Definition 3.0.2 (Adjoint of a matrix).** If  $A \in M_n(F)$ , then we define the adjoint of  $A$  to be  $\text{adj}(A) \in M_n(F)$  where

$$(\text{adj}(A))_{ij} = (-1)^{i+j} \det(A(j | i)),$$

where  $A(j | i)$  is  $A$  deleting its  $j$ -th row and  $i$ -th column.

**Note 3.0.1.** If we look at  $M_2(F)$  and  $M_3(F)$ , we can find that

$$A \cdot \text{adj}(A) = \det(A)I.$$

In fact, this is true for square matrices of all sizes.

**Remark 3.0.1.**  $A$  is invertible iff  $\det(A) \neq 0$ .

**Proof.** We will later show the proof. ■

We first introduce some good properties:

- (1) Multilinear.
- (2) Alternating.
- (3)  $\det(I_n) = 1$ .

**Definition 3.0.3 (Multilinear).** Consider a function  $D$  of  $n$  row vectors in  $F^n$  as its input, and the output is  $D(v_1, v_2, \dots, v_n) \in F$ , then  $D$  is called multilinear or  $n$ -linear if

$$\begin{aligned} D(u + \alpha w, v_2, \dots, v_n) &= D(u, v_2, \dots, v_n) + \alpha D(w, v_2, \dots, v_n) \\ &\vdots \\ D(v_1, v_2, \dots, u + \alpha w) &= D(v_1, v_2, \dots, u) + \alpha D(v_1, v_2, \dots, w). \end{aligned}$$

**Example 3.0.1.** If we suppose  $A \in M_n(F)$ , and  $r_i$  is the  $i$ -th row of  $A$ , where  $r_i = (a_{i1}, a_{i2}, \dots, a_{in})$ , then if we define  $D(A) = a_{1k_1} a_{2k_2} \dots a_{nk_n}$ , then in fact  $D$  is multilinear if we regard  $D$  as a function which takes  $n$  row vectors as its input.

**Lemma 3.0.1.** If  $D_1, D_2$  are  $n$ -linear, then  $D_1 + \alpha D_2$  is also  $n$ -linear. If  $D$  is  $n$ -linear, then  $D$  is determined by  $D(v_1, \dots, v_n)$  with  $v_i \in \{e_i\}_{i=1}^n$ .

**Note 3.0.2.**  $D$  is a function determined by  $n^n$  values since each  $v_i$  has  $n$  choices.

**Definition 3.0.4 (Alternating).** Suppose  $D$  is  $n$ -linear, then  $D$  is alternating if

$$D(v_1, \dots, v_n) = 0$$

if  $v_i = v_j$  for some  $i \neq j$ .

**Lemma 3.0.2.** If  $D$  is alternating, then

(1)

$$D(\dots, \overbrace{v_i + \alpha v_j}^{i\text{-th position}}, \dots) = D(\dots, \overbrace{v_i}^{i\text{-th position}}, \dots).$$

(2) If  $\{v_1, v_2, \dots, v_n\}$  is linearly dependent, then  $D(v_1, v_2, \dots, v_n) = 0$ .

(3)

$$D(v_1, \dots, v_i, \dots, v_j, \dots, v_n) = -D(v_1, \dots, v_j, \dots, v_i, \dots, v_n).$$

**proof of (2).** WLOG, say  $v_i = \sum_{j \neq i} \alpha_j v_j$ , then

$$D(v_1, \dots, v_n) = D\left(v_1, \dots, \sum_{j \neq i} \alpha_j v_j, \dots, v_n\right) = \sum_{j \neq i} \alpha_j D(v_1, \dots, \overbrace{v_j}^{i\text{-th position}}, \dots, v_n) = 0$$

since  $D$  is alternating. ■

**proof of (3).** Since

$$\begin{aligned} 0 &= D(\dots, v_i + v_j, \dots, v_i + v_j, \dots) \\ &= D(\dots, v_i, \dots, v_i, \dots) + D(\dots, v_i, \dots, v_j, \dots) + D(\dots, v_j, \dots, v_i, \dots) + D(\dots, v_j, \dots, v_j, \dots) \\ &= D(\dots, v_i, \dots, v_j, \dots) + D(\dots, v_j, \dots, v_i, \dots), \end{aligned}$$

so this is true. ■

**Proposition 3.0.1.** If  $D$  is  $n$ -linear and alternating, then it is determined by

$$D(e_{\sigma(1)}, e_{\sigma(2)}, \dots, e_{\sigma(n)}),$$

where  $\sigma : \{1, 2, \dots, n\} \rightarrow \{1, 2, \dots, n\}$  is any permutation on  $[n]$ .

**Remark 3.0.2.** In this case, there is at most one  $n$ -linear alternating  $D$  satisfying  $D(e_1, \dots, e_n) = 1$ .

**Proof.** Since  $D$  is alternating, so swapping  $e_i$  and  $e_j$  just turn the original value to negative. Thus, if  $D(e_1, \dots, e_n) = 1$ , then we know

$$D(e_{\sigma(1)}, e_{\sigma(2)}, \dots, e_{\sigma(n)})$$

is uniquely defined for all permutation  $\sigma$ . Now since  $D$  is determined by  $D(e_{\sigma(1)}, e_{\sigma(2)}, \dots, e_{\sigma(n)})$ , so  $D$  is uniquely defined. ■

### Another approach/inductive construction

**Theorem 3.0.1.** There exists a function

$$\det_n : M_n(F) \rightarrow F,$$

s.t.  $\det_n$  is  $n$ -linear(on rows) and alternating(on rows) and  $\det(I_n) = 1$ .

We can just define

$$\begin{cases} \det_1(a) = a \\ \det_n(A) = \sum_{i=1}^n (-1)^{i+j} a_{ij} \det_{n-1}(A(i|j)) \end{cases},$$



where  $A(i | j)$  is  $A$  deleting  $i$ -th row and  $j$ -th column.

**Note 3.0.3.** The definition given above is the expansion along  $j$ -th column.

**Note 3.0.4.** Since we know there is at most one  $n$ -linear, alternating  $D$  satisfying  $D(e_1, e_2, \dots, e_n) = 1$ , and we have constructed such  $D$ , and thus we can define this  $D$  to be the determinant function.

## Lecture 13

Actually determinant can be defined on ring (we defined it on field before).

17 Oct. 10:20

**Theorem 3.0.2.** There is the determinant function

$$\det : M_n(R) \rightarrow R.$$

Now we talk more about expansion. We do expansion along a column. Suppose we have

$$\delta : M_{n-1}(R) \rightarrow R,$$

which is  $(n-1)$ -linear and alternating and  $\delta(I_{n-1}) = 1$ , then if we define  $D_j = D : M_n(R) \rightarrow R$ , which is the expansion along the  $j$ -th column, and it has

$$D(A = (a_{ij})) = \sum_{i=1}^n (-1)^{i+j} a_{ij} \delta(A(i | j)).$$

**Note 3.0.5.**  $C_{ij} = (-1)^{i+j} \delta(A(i | j))$  is called the  $(i, j)$ -cofactor.

**Theorem 3.0.3.**  $D$  is  $n$ -linear and alternating, and  $D(I_n) = 1$ .

**Proof.** ■

DIY

**Note 3.0.6.** In the proof of alternating, we may need to use [Lemma 3.0.2](#).

**Note 3.0.7.** We still regard  $D$  as a function taking  $n$  row vectors as its input.

**As previously seen.** If  $D : M_n(R) \rightarrow R$  is  $n$ -linear, alternating, then

$$D((a_{ij})) = \sum_{\sigma} \prod_{i=1}^n a_{i\sigma(i)} D \begin{pmatrix} e_{\sigma(1)} \\ \vdots \\ e_{\sigma(n)} \end{pmatrix}$$

**Proof.** Suppose  $A = (a_{ij})_{n \times n}$ 's rows are  $r_1, r_2, \dots, r_n$ , then we know  $r_i = \sum_{j=1}^n a_{ij} e_{j_i}$ , so we know

$$\begin{aligned} D(A) &= \sum_{j_1=1}^n a_{1j_1} D(e_{j_1}, r_2, \dots, r_n) = \sum_{j_1=1}^n a_{1j_1} \left( \sum_{j_2=1}^n a_{2j_2} D(e_{j_1}, e_{j_2}, r_3, \dots, r_n) \right) \\ &= \sum_{j_1=1}^n \sum_{j_2=1}^n a_{1j_1} a_{2j_2} D(e_{j_1}, e_{j_2}, r_3, \dots, r_n) \\ &= \dots = \sum_{j_1=1}^n \sum_{j_2=1}^n \dots \sum_{j_n=1}^n a_{1j_1} a_{2j_2} \dots a_{nj_n} D(e_{j_1}, e_{j_2}, \dots, e_{j_n}) \\ &= \sum_{\sigma \in S_n} \left( \prod_{i=1}^n a_{i\sigma(i)} \right) D(e_{\sigma(1)}, e_{\sigma(2)}, \dots, e_{\sigma(n)}) \end{aligned}$$

since if  $j_p = j_q$  for some  $p \neq q$ , then since  $D$  is alternating, so we know that term will be 0, and thus we just need to consider the terms with  $j_p \neq j_q$  for any  $p \neq q$ . ■

Now we put things together:

#### Theorem 3.0.4.

- (i) There is a function  $\det : M_n(R) \rightarrow R$  satisfying  $n$ -linear, alternating, and  $\det(I_n) = 1$ .
- (ii) If  $D : M_n(R) \rightarrow R$  is  $n$ -linear, alternating, then  $D(A) = D(I) \cdot \det(A)$ .
- (iii) For a permutation  $\sigma$ , if  $\sigma = t_1 t_2 \dots t_n = t'_1 t'_2 \dots t'_m$ , where  $t_i, t'_i$ 's are transpositions, then  $(-1)^n = (-1)^m$ .

**Remark 3.0.3.** (ii) needs the fact that

$$D(e_{\sigma(1)}, e_{\sigma(2)}, \dots, e_{\sigma(n)}) = (-1)^m D(e_1, e_2, \dots, e_n)$$

if  $\sigma$  is the composition of  $m$  transpositions.

**Remark 3.0.4.** (i) and (ii) hold for any  $R$ .

Now we introduce two formulas:

(1)

$$\det(A) = \sum_{i=1}^n (-1)^{i+j} a_{ij} \det(A(i | j)).$$

(2)

$$\text{sgn} : \{\text{permutation}\} \rightarrow \{\pm 1\}, \quad \sigma \mapsto (-1)^m$$

if  $\sigma = t_1 t_2 \dots t_m$  if  $t_i$ 's are transpositions.

Thus, we know

$$\det(A) = \sum_{\sigma} \text{sgn}(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \dots a_{n\sigma(n)}$$

by the proof above and [Remark 3.0.3](#).

## Lecture 14

**As previously seen.** There is a unique function

$$\det : M_n(R) \rightarrow R$$

29 Oct. 10:20

satisfying  $n$ -linear in rows, alternating, and  $\det(I_n) = 1$ . Also, if  $D : M_n(R) \rightarrow R$  satisfies  $n$ -linear and alternating, then  $D(A) = D(I) \cdot \det(A)$ . Besides,  $\det$  can be constructed inductively:

$$\det(A) = \sum_{i=1}^n a_{ij} c_{ij}$$

where  $c_{ij} = (-1)^{i+j} \det(A(i | j))$  is the  $(i, j)$ -cofactor.

If  $\sigma \in S_n$ , and let  $\sigma(I) = (e_{\sigma(1)}, e_{\sigma(2)}, \dots, e_{\sigma(n)})$  (permuting the rows), then  $\det(\sigma(I)) = (-1)^m$  if  $\sigma = \tau_1 \tau_2 \dots \tau_m$  where  $\tau_i$  is a transposition since  $\det$  is alternating, so exchange two rows in the function input change the sign of the output.

**Corollary 3.0.1.** For  $\sigma \in S_n$ , if  $\sigma = \tau_1 \tau_2 \dots \tau_p = \tau'_1 \tau'_2 \dots \tau'_q$ , then  $p$  and  $q$  are both even or both odds.

**Definition 3.0.5.**  $\sigma \in S_n$  is called an even (resp. odd) permutation if  $\sigma = \tau_1 \tau_2 \dots \tau_m$  for  $m$  even (resp. odd). Thus, we can define

$$\text{sgn} : S_n \rightarrow \{\pm 1\}, \quad \sigma \mapsto \det(\sigma(I)).$$

Hence, we can give a second method to construct  $\det$ :

$$\det((a_{ij})_{n \times n}) = \sum_{\sigma \in S_n} \text{sgn}(\sigma) \prod_{i=1}^n a_{i, \sigma(i)}.$$

**Example 3.0.2.** If we want to calculate

$$\det \begin{pmatrix} 0 & 0 & & a_n \\ a_1 & 0 & & 0 \\ & & \ddots & \\ 0 & \cdots & a_{n-1} & 0 \end{pmatrix},$$

then we have two ways:

- (1) expand along the last column.
- (2) Suppose  $A = (a_{ij})_{n \times n}$ , where  $a_{ii} = a_i$  for all  $i$  and  $a_{ij} = 0$  for all  $i \neq j$ , then  $\det A = a_1 a_2 \dots a_n$ , and the matrix given in the problem is from exchanging first row and second row of  $A$ , then exchange second row and third row, and keep going until exchanging the  $n-1$ -th row and  $n$ -th row, so the answer is  $(-1)^{n-1} a_1 a_2 \dots a_n$  since it takes  $n-1$  times exchange. (exchange rows in the input of an alternating function will change the sign of output.)

**Example 3.0.3.** Companion form of  $f(x) = x^n + a_1 x^{n-1} + \dots + a_n$ :

$$A_f = \begin{pmatrix} 0 & 0 & \cdots & -a_n \\ 1 & 0 & \cdots & -a_{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 1 & -a_1 \end{pmatrix}.$$

We can calculate  $\det(xI - A_f) = f(x)$ .

**Theorem 3.0.5.** Suppose  $A, B \in M_n(R)$ , where  $R$  is a ring with identity, then

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

Thus, we have  $\det(P^{-1}) = \det(P)^{-1}$ .

**Proof.** Let  $D(A) = \det(AB)$ , then we can check that  $D$  satisfies  $n$ -linear and alternating. If this were true, then  $D(A) = D(I) \det(A)$ , and  $D(I) = \det(IB) = \det(B)$ , so  $D(A) = \det(A) \det(B)$  and thus we have

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

**Note 3.0.8.** Note that

$$D \begin{pmatrix} u_1 \\ \vdots \\ v + \alpha w \\ \vdots \\ u_n \end{pmatrix} = \det \left( \begin{pmatrix} u_1 \\ \vdots \\ v + \alpha w \\ \vdots \\ u_n \end{pmatrix} B \right) = \det \left( \begin{pmatrix} u_1 B \\ \vdots \\ vB + \alpha wB \\ \vdots \\ u_n B \end{pmatrix} \right) = D \begin{pmatrix} u_1 \\ \vdots \\ v \\ \vdots \\ u_n \end{pmatrix} + \alpha D \begin{pmatrix} u_1 \\ \vdots \\ w \\ \vdots \\ u_n \end{pmatrix},$$

and alternating can be proved similarly. ■

**Theorem 3.0.6.** If  $A \sim B$ , then  $\det A = \det B$ .

**Theorem 3.0.7.**  $\det A^t = \det A$ .

**Proof.** Note that

$$a_{1\sigma(1)} \cdots a_{n\sigma(n)} = a_{\sigma^{-1}(1),1} \cdots a_{\sigma^{-1}(n),n},$$

and  $\operatorname{sgn}(\sigma) = \operatorname{sgn}(\sigma^{-1})$ . Hence, if we suppose  $B = A^t$ , then

$$\begin{aligned} \det(B) &= \sum_{\sigma} \operatorname{sgn}(\sigma) \prod b_{i,\sigma(i)} \\ &= \sum_{\sigma} \operatorname{sgn}(\sigma) \prod a_{\sigma(i),i} \\ &= \sum_{\tau: \tau = \sigma^{-1}} \operatorname{sgn}(\tau) \prod a_{i,\tau(i)} = \det(A). \end{aligned}$$
■

**Exercise 3.0.1.** Show that

$$\det \begin{pmatrix} A & B \\ 0 & D \end{pmatrix} = \det(A) \det(D).$$

**Theorem 3.0.8.** Let  $A \in M_n(R)$ , then we can define the (classical) adjoint

$$\operatorname{adj}(A) = \tilde{A} = (\widetilde{a_{ij}}),$$

where

$$\widetilde{a_{ij}} = (j, i)\text{-cofactor } c_{j,i} = (-1)^{i+j} \det(A(j \mid i)),$$

then  $A\tilde{A} = \tilde{A}A = \det(A)I$ . This means if  $A$  is invertible, then  $A^{-1} = \frac{1}{\det(A)}\tilde{A}$ .

**Proof.** Note that the  $(i, i)$ -entry of  $A\tilde{A}$  is

$$\sum_{k=1}^n a_{ik} \widetilde{a_{ki}} = \sum_{k=1}^n (-1)^{i+k} a_{ik} \det(A(i \mid k)) = \det(A),$$

while the  $(i, j)$ -entry for  $i \neq j$  is

$$\begin{aligned} \sum_{k=1}^n a_{ik} \widetilde{a}_{kj} &= \sum_{k=1}^n (-1)^{j+k} a_{ik} \det(A(j \mid k)) \\ &= \det \begin{pmatrix} a_{i1} & a_{i2} & \dots & a_{in} \end{pmatrix} (j\text{-th row}) = 0 \end{aligned}$$

since  $\det$  is alternating. Thus,  $A\tilde{A} = \det(A)I$ . Similarly, we can show  $\tilde{A}A = \det(A)I$ . ■

**Theorem 3.0.9.** Suppose  $A \in M_n(F)$  is invertible, then consider the system

$$A \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix},$$

then  $x_i = \frac{1}{\det(A)} \det(C_i)$ , where  $C_i$  is the matrix  $A$  but replace the  $i$ -th column with  $(c_1, c_2, \dots, c_n)^t$ .

**Proof.** In fact,

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \frac{1}{\det(A)} \tilde{A} \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix},$$

and by comparing the entries, we know

$$\det(A)x_i = \sum_{j=1}^n (-1)^{i+j} c_j \det(A(j \mid i)) = \det(C_i).$$

■

**Exercise 3.0.2.** If  $v_1, v_2, \dots, v_n \in \mathbb{R}^n$ , then

$$\det(v_1, v_2, \dots, v_n) = \pm \text{volumn}.$$

**Definition 3.0.6.** For finite dimensional vector space  $V$ , suppose  $T \in \mathcal{L}(V)$ , then one can define  $\det(T)$  by choosing an ordered basis  $B$  of  $V$ , and define

$$\det(T) := \det([T]_B).$$

**Remark 3.0.5.** This  $\det(T)$  does not depend on the choice of  $B$  since

$$[T]_B \sim [T]_{B'}$$

for any two basis  $B, B'$  of  $V$ . This is because

$$[T]_{B'} = [id]_{B'}^B [T]_B [id]_B^{B'}.$$

## Lecture 15

**Definition 3.0.7.** Let  $T \in \mathcal{L}(V)$  (or a matrix  $A \in M_n(F)$ ). A scalar  $\lambda \in F$  is called an eigenvalue of  $T$  if  $\exists v \neq 0$  s.t.  $Tv = \lambda v$ . Equivalently,  $T - \lambda I$  is singular, or  $\det(T - \lambda I) = 0$  or  $\nu(T - \lambda I) > 0$ . In

31 Oct. 10:20

this case,  $E(\lambda) = \ker(T - \lambda I)$  is called the eigenspace and any vector in  $E(\lambda)$  is called an eigenvector (for  $\lambda$ ).

**Remark 3.0.6.** If  $A$  is not invertible, then  $\det(A) = 0$  since there is a row of  $A$  is the linear combination of other rows, and  $\det$  is  $n$ -linear and alternating.

**Remark 3.0.7.** Eigenvalues are also called characteristic values, proper value, spectral value.

If  $A \in M_n(F)$  is the matrix representation of  $T$ , then

$$\det(T - \lambda I) = \det(A - \lambda I) = (-1)^n \det(\lambda I - A).$$

**Definition 3.0.8.** The polynomial  $f(x) = \det(xI - A)$  is called the characteristic polynomial of  $T$ .

**Remark 3.0.8.**  $f(x)$  does not depend on the choice of matrix representation since if we choose another  $B = P^{-1}AP$ , then

$$\begin{aligned} \det(xI - B) &= \det(xI - P^{-1}AP) = \det(P^{-1}(xI)P - P^{-1}AP) \\ &= \det(P^{-1}(xI - A)P) = \det(P^{-1}) \det(xI - A) = \det(P) = \det(xI - A). \end{aligned}$$

**Remark 3.0.9.** One can verify that for two similar matrices  $A, B$ , we have  $\text{Tr}(A) = \text{Tr}(B)$ .

**Remark 3.0.10.** Note that

$$f(x) = x^n - \text{Tr}(T)x^{n-1} + \cdots + (-1)^n \det(T).$$

This is because  $x^n$  and  $x^{n-1}$  terms come from  $(x - a_{11})(x - a_{22}) \cdots (x - a_{nn})$ , and by Vieta's theorem, we know the coefficient of  $x^{n-1}$  is  $\text{Tr}(T)$ . Also,  $f(0) = \det(-A) = (-1)^n \det(A)$  is trivial.

**Remark 3.0.11.** For the coefficient of  $x^{n-1}$ , suppose  $B = xI - A$ , then we know

$$\det B = \sum_{\sigma \in S_n} \text{sgn}(\sigma) b_{1\sigma(1)} b_{2\sigma(2)} \cdots b_{n\sigma(n)},$$

so if some term contributes  $x^{n-1}$ , then at least  $n-1$  of  $\sigma(i)$  is equal to  $i$ , which means all  $n$  of  $\sigma(i)$ 's are  $i$ , and thus the only term contributes  $x^{n-1}$  is  $(x - a_{11})(x - a_{22}) \cdots (x - a_{nn})$ .

**Theorem 3.0.10.**  $\lambda$  is an eigenvalue of  $T$  iff  $\lambda$  is a root of  $f(x)$ .

## 3.1 Diagonalization

**Definition 3.1.1.**  $T \in \mathcal{L}(V)$  is called diagonalizable if  $\exists$  matrix representation of  $T$ , which is a diagonal matrix. A matrix  $A$  is called diagonalizable if  $A$  is similar to a diagonal matrix.

If

$$[T]_B = \begin{pmatrix} \lambda_1 I_1 & & \\ & \ddots & \\ & & \lambda_r I_{m_r} \end{pmatrix}$$

and  $\lambda_i \neq \lambda_j$  for any  $i \neq j$  with

$$B = \bigcup_{i=1}^r \{v_{i1}, v_{i2}, \dots, v_{im_i}\},$$

then  $f(x) = (x - \lambda_1)^{m_1}(x - \lambda_2)^{m_2} \dots (x - \lambda_r)^{m_r}$  splits (by plugging  $[T]_B$  into  $\det(xI - A)$ ), and we have  $\dim(E(\lambda_i)) = \dim \ker(T - \lambda_i I) = m_i$ , which can be seen by observing the rank of matrix  $[T]_B - \lambda_i I$ . Also, we can observe that  $V = E(\lambda_1) + E(\lambda_2) + \dots + E(\lambda_r)$ , so  $\dim V = \sum_{i=1}^r \dim E(\lambda_i)$  since

$$E(\lambda_i) \cap E(\lambda_j) = \{0\}$$

for any  $i \neq j$ .

**Definition 3.1.2.** Suppose  $\lambda$  is an eigenvalue of  $T$  and characteristic polynomial  $f(x) = (x - \lambda)^m g(x)$  with  $g(\lambda) \neq 0$ . The algebraic multiplicity of  $\lambda$ ,  $\text{a-mult}(\lambda) = m$ , and the geometric multiplicity  $\text{g-mult}(\lambda) = \dim(E_\lambda) = \nu(T - \lambda I) \geq 1$ .

**Proposition 3.1.1.**  $\text{a-mult}(\lambda) \geq \text{g-mult}(\lambda)$ .

**Proof.** Let  $\{v_1, \dots, v_e\}$  be a basis of  $E(\lambda)$ , and extend it to a basis of  $V$ , say  $B = \{v_1, \dots, v_e, \dots, v_n\}$ . Hence,

$$A = [T]_B = \begin{pmatrix} \lambda I_e & B \\ 0 & D \end{pmatrix},$$

which gives

$$f(x) = \det(xI - A) = (x - \lambda)^e \det(xI - D),$$

note that  $\det(xI - D)$  may have  $\lambda$  as a root, so the algebraic multiplicity of  $\lambda \geq$  the geometric multiplicity of  $\lambda$ .

**Note 3.1.1.** If  $A$  is not diagonalizable, then we know  $\det(xI - D)$  may have  $\lambda$  as its root. ■

**Definition 3.1.3.** Let  $W_1, W_2, \dots, W_r$  be subspaces of  $V$ . We say  $W_i$ 's are linearly independent if  $w_1 + w_2 + \dots + w_r = 0$  for  $w_i \in W_i$ , then  $w_i = 0$  for all  $i$ .

**Proposition 3.1.2.** Let  $W = W_1 + W_2 + \dots + W_r$ , then TFAE:

- (i)  $W_i$  are linearly independent.
- (ii) Any  $w \in W$  has a unique expression

$$w = \sum_{i=1}^r w_i, \quad \forall w_i \in W_i.$$

(iii)

$$W_i \cap (W_1 + W_2 + \dots + W_{i-1} + W_{i+1} + \dots + W_r) = \{0\}.$$

(iv)  $\dim W = \sum_{i=1}^r \dim W_i$ .

**(i) to (ii),(iii),(iv).** ■

**(ii) to (i).** If  $\sum w_i = 0$ , then since  $\sum 0 = 0$  and  $0 \in W_i$  for all  $i$ , and  $0$  has unique expression, so  $w_i = 0$  for all  $i$ . ■

**(iii) to (i).** If  $\sum w_i = 0$  for  $w_i \in W_i$ , then

$$-w_i = w_1 + w_2 + \dots + w_{i-1} + w_{i+1} + \dots + w_r \in W_i \cap (W_1 + W_2 + \dots + W_{i-1} + W_{i+1} + \dots + W_r) = \{0\}$$

for all  $i$ , so  $w_i = 0$  for all  $i$ . ■

DIY

**(iv) to (i).** If  $\{v_{ij}\}_{j=1}^{m_i}$  is a basis of  $W_i$ , then  $\{v_{ij}\}_{i,j}$  generates  $W$ . Also, we know  $\dim W = \sum_{i=1}^r \dim W_i$ , so  $\{v_{ij}\}_{i,j}$  is a basis of  $W$ . Now if  $\sum_{i=1}^r w_i = 0$ , so we have  $\sum_{i,j} \alpha_{ij} v_{ij} = 0$ , and thus  $\alpha_{ij} = 0$  for all  $i, j$ . Hence,  $w_i = 0$  for all  $i$ . ■

**Proposition 3.1.3.** If  $\lambda_1, \lambda_2, \dots, \lambda_r$  are distinct eigenvalues of  $T$ , then  $\{E(\lambda_i)\}_{i=1}^r$  are linearly independent.

**Proof.** Suppose  $v_1 + v_2 + \dots + v_r = 0$  for  $v_i \in E(\lambda_i)$ , then by applying  $T$ , we know  $\lambda_1 v_1 + \dots + \lambda_r v_r = 0$ , so we have

$$(\lambda_2 - \lambda_1)v_2 + \dots + (\lambda_r - \lambda_1)v_r = 0.$$

Hence, by this thought, suppose  $v_1 + \dots + v_m = 0$  for  $v_i \in E(\lambda_i)$  and it is a shortest equality of a non-trivial relation. Then, we can always obtain a shorter non-trivial relation by above method, so it is a contradiction. ■

**Corollary 3.1.1.** If  $\{v_{ij}\}_{j=1}^{m_i}$  is a basis of  $E(\lambda_i)$ , then  $B = \bigcup_{i=1}^r \{v_{ij}\}_{j=1}^{m_i}$  is linearly independent.

**Proof.** Suppose  $\sum_{i=1}^r \sum_{j=1}^{m_i} \alpha_{ij} v_{ij} = 0$ , then since  $\sum_{j=1}^{m_i} \alpha_{ij} v_{ij} \in W_i$ , so since  $\{E(\lambda_i)\}_{i=1}^r$  are linearly independent, so we know  $\sum_{j=1}^{m_i} \alpha_{ij} v_{ij} \in W_i = 0$  for all  $i$ , and since  $\{v_{ij}\}_{j=1}^{m_i}$  is a basis of  $E(\lambda_i)$  for all  $i$ , so they are linearly independent, and thus we know  $\alpha_{ij} = 0$  for all  $i, j$ , which shows  $B$  is linearly independent. ■

**Corollary 3.1.2.** Suppose  $T \in \mathcal{L}(V)$  and has a characteristic polynomial

$$f(x) = \prod_{i=1}^r (x - \lambda_i)^{m_i}$$

with  $\lambda_i \neq \lambda_j$  for any  $i \neq j$ , then TFAE:

- (i)  $T$  is diagonalizable.
- (ii)  $\dim E(\lambda_i) = m_i$  for all  $i$ .
- (iii)  $V = \sum_{i=1}^r E(\lambda_i)$  (or any  $v \in V$  is a linear combination of eigenvectors.)
- (iv)  $\dim V = \sum_{i=1}^r \dim E(\lambda_i)$ .

**Corollary 3.1.3.** If the characteristic polynomial of a linear operator has degree  $n$  and has  $n$  distinct roots, then  $T$  is diagonalizable.

**Proof.** By (ii) of [Corollary 3.1.2](#). ■

**Corollary 3.1.4.** If  $T^2 = T$ , then  $T$  is diagonalizable.

## Lecture 16

Suppose  $V$  is a finite dimensional vector space, then fix  $T \in \mathcal{L}(V)$ , we have

5 Nov. 10:20

$$a_0 + a_1 T + a_2 T^2 + \dots + a_n T^n \in \mathcal{L}(V),$$

which means  $f(T) \in \mathcal{L}(V)$  where  $f(x) = \sum_{k=0}^n a_k x^k \in F[x]$ . We call  $V$  is an  $F[x]$ -module. (= "vector space over a ring") What makes the classification (structure theorem) simple. The answer is something like  $F[x], \mathbb{Z}, \dots$ , the principal ideal domains (PID). Note that  $F[x], \mathbb{Z}$  are Euclidean domain, which means that there is the degree map

$$\deg : F[x] \rightarrow \mathbb{Z}_{\geq 0} \text{ or } \deg : \mathbb{Z} \rightarrow \mathbb{Z}_{\geq 0}$$

s.t. for any  $a, b \in F[x]$  and  $b \neq 0$ , there exists unique  $a = qb + r$  where  $\deg r < \deg b$ .



### 3.2 Minimal polynomial

Fix  $T \in \mathcal{L}(V)$ . For  $g(x) = b_n x^n + \cdots + b_0 \in F[x]$ , let  $g(T) = b_n T^n + \cdots + b_0 \in \mathcal{L}(V)$ . Note that

$$\begin{aligned} g(x) &= g_1(x) \cdot g_2(x) \Rightarrow g(T) = g_1(T) \cdot g_2(T). \\ g(x) &= g_1(x) + g_2(x) \Rightarrow g(T) = g_1(T) + g_2(T). \\ \text{If } T(v) &= \lambda v \Rightarrow g(T)(v) = g(\lambda)(v). \end{aligned}$$

**Definition 3.2.1.** Suppose  $T : V \rightarrow V$  is a linear operator, then we define

$$\begin{aligned} \text{Ann}_T(V) &= \{\text{annihilator of } T\} \\ &= \{g(x) \in F[x] \mid g(T) = 0\} \\ &= \{\text{linear relations of } T^0, T^1, T^2, \dots \in \mathcal{L}(V)\}. \end{aligned}$$

**Note 3.2.1.** There exists a non-trivial relation among  $T^0, T^1, \dots, T^{n^2}$  since  $\dim \mathcal{L}(V) = n^2$ .

**Proposition 3.2.1.** Let  $m(x) = m_T(x)$  be a monic polynomial (leading coefficient is 1) in  $\text{Ann}_T(V)$  with minimal degree. Then,

$$\text{Ann}_T(V) = F[x] \cdot m(x).$$

**Proof.** For any  $g(x) \in \text{Ann}_T(V)$ , we have

$$g(x) = q(x) \cdot m(x) + r(x)$$

with  $\deg r < \deg m$ . Then,

$$0 = g(T) = q(T) \cdot m(T) + r(T) = r(T).$$

Since  $m$  is the "minimal degree" monic polynomial, so  $r(x) = 0$ . ■

**Definition 3.2.2.** This  $m_T(x)$  is called the minimal polynomial of  $T$ .

**Example 3.2.1.** Suppose

$$A = \begin{pmatrix} 0 & & & \\ 1 & 0 & & \\ & \ddots & \ddots & \\ & & 1 & 0 \end{pmatrix},$$

then  $m_A(x) = x^n$  since we can found that  $A^n = 0$ , so  $m_A(x) \mid x^n$ , so  $m_A(x) = x^p$  for some  $p \leq n$ , and we can find that  $n$  is the minimal  $p$  s.t.  $A^p = 0$ .

**Example 3.2.2.** Suppose

$$B = \begin{pmatrix} 0 & & & -a_0 \\ 1 & 0 & & -a_1 \\ & 1 & \ddots & -a_2 \\ & & \ddots & 0 & \vdots \\ & & & 1 & -a_{n-1} \end{pmatrix},$$

then we know  $m_B(x) = f(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_0$ .

**Remark 3.2.1.** Check that  $B(e_i) = e_{i+1}$  for  $1 \leq i \leq n-1$  and  $B(e_n) = \sum_{i=0}^{n-1} -a_i e_{i+1}$ , and thus  $f(B) = 0$  since it sends the standard basis to 0. Then, we can check that  $\deg m_B(x) \geq n$ , and we're done.

**Remark 3.2.2.**  $f(B)(e_i) = f(B)B^{i-1}(e_1) = B^{i-1}f(B)(e_1) = 0$ .

**Example 3.2.3.** Suppose

$$C = \begin{pmatrix} \lambda_1 I_{m_1} & & \\ & \ddots & \\ & & \lambda_r I_{m_r} \end{pmatrix},$$

then  $m(x) = m_C(x) = (x - \lambda_1)(x - \lambda_2) \dots (x - \lambda_r)$ . This is because  $Ce_k = \lambda e_k$  for some  $\lambda = \lambda_1, \dots, \lambda_r$ , and thus

$$(C - \lambda_1)(C - \lambda_2) \dots (C - \lambda_r)(e_i) = 0$$

for all  $i$ , and thus we know

$$m_C(x) \mid (x - \lambda_1) \dots (x - \lambda_r).$$

Also, we can check that if  $q(C) = 0$ , then  $(x - \lambda_i) \mid q(x)$  for all  $i$  by observing the matrix of  $q(C)$ .

Observe that if  $T$  is diagonalizable with distinct eigenvalues  $\lambda_1, \lambda_2, \dots, \lambda_r$ , then

$$m_T(x) = \prod_{i=1}^r (x - \lambda_i),$$

and  $\text{ch}_T(x) \in \text{Ann}_T(V)$  (Cayley-Hamilton Theorem).

**Proposition 3.2.2.** If  $\lambda$  is some element of  $F$ , then

$$\text{ch}_T(\lambda) = 0 \Leftrightarrow m_T(\lambda) = 0.$$

**Proof.**

( $\Rightarrow$ ) Since there exists  $v \neq 0$  s.t.  $T(v) = \lambda v$ , then

$$0 = m_T(T)(v) = m_T(\lambda)(v),$$

and since  $v \neq 0$ , so  $m_T(\lambda) = 0$ .

( $\Leftarrow$ ) Write  $m_T(x) = (x - \lambda)p(x)$ , then  $\exists v$  s.t.  $p(T)(v) \neq 0$ , so

$$0 = m_T(T)(v) = (T - \lambda)p(T)(v) = (T - \lambda)w,$$

and since  $w \neq 0$ , so  $E(\lambda) \neq \{0\}$ , so  $(x - \lambda) \mid \text{ch}_T(x)$ . ■

### 3.3 Invariant subspaces

**Definition 3.3.1.** Suppose  $T \in \mathcal{L}(V)$ , then a subspace  $W$  is called  $T$ -invariant if  $T(W) \subseteq W$ . In this case,  $W$  is also  $g(T)$ -invariant for  $g(x) \in F[x]$ . Besides, we know  $T$  induces an operator  $T_W = T|_W \in \mathcal{L}(W)$ .

**Example 3.3.1.** If  $ST = TS$ , then  $\ker(S)$  and  $\text{Im}(S)$  are  $T$ -invariant. In particular,  $E(\lambda) = \ker(T - \lambda)$  is  $T$ -invariant.

**Example 3.3.2.** If  $W_1, W_2$  are  $T$ -invariant, then  $W_1 + W_2$  and  $W_1 \cap W_2$  are  $T$ -invariant.

**Proposition 3.3.1.** Let  $W$  be  $T$ -invariant and  $S = T_W \in \mathcal{L}(W)$ , then we have  $\text{ch}_S(x) \mid \text{ch}_T(x)$  and  $m_S(x) \mid m_T(x)$ .

**Proof.** Let  $B = \{w_1, \dots, w_m\}$  be a basis of  $W$ , and extend it to a basis of  $V$ , say

$$\tilde{B} = \{w_1, \dots, w_m, w_{m+1}, \dots, w_n\},$$

and suppose  $A = [S]_B$  and  $\tilde{A} = [T]_{\tilde{B}}$ , then we know

$$\tilde{A} = \begin{pmatrix} A & C \\ 0 & D \end{pmatrix} \Rightarrow xI - \tilde{A} = \begin{pmatrix} xI - A & C \\ 0 & xI - D \end{pmatrix},$$

so we know  $\det(xI - \tilde{A}) = \det(xI - A) \det(D - xI)$ , which gives  $\text{ch}_{\tilde{A}}(x) = \text{ch}_A(x) \text{ch}_D(x)$ , so we proved the first part.

Now since  $m_T(S) = m_T(T_W)$ , and  $m_T(T_W)(w) = m_T(T)(w) = 0$  for all  $w \in W$ , so  $m_T(T_W) = 0$  and thus  $m_T(S) = 0$ , so  $m_S(x) \mid m_T(x)$ . ■

**Definition 3.3.2.** Let  $W$  be  $T$ -invariant, then

$$\begin{aligned} \text{Ann}_T(V/W) &= \{f(x) \in F[x] \mid f(T)(v) \in W \quad \forall v\} \\ &= \{f(x) \in F[x] \mid f(T)(V) \subseteq W\}. \end{aligned}$$

In particular, we know  $m_T(x) \in \text{Ann}_T(V/W)$ .

**Lemma 3.3.1.** Let  $p(x) \in \text{Ann}_T(V/W)$  be the monic polynomial of smallest degree, then

$$\text{Ann}_T(V/W) = F[x] \cdot p(x).$$

**Proof.** Take  $g \in \text{Ann}_T(V/W)$ , then  $g = qp + r$ , and

$$g(T)(v) = q(T)p(T)(v) + r(T)(v) \in W \quad \forall v \in V.$$

since  $p(T)(v) \in W$  and  $W$  is  $q(T)$ -invariant, then  $r(T)(v) \in W$ , so  $r(x) = 0$ . ■

**Theorem 3.3.1.**  $T$  is diagonalizable if and only if  $m_T(x) = \prod_{i=1}^r (x - \lambda_i)$  with  $\lambda_i \neq \lambda_j$  for all  $i \neq j$ .

**Proof.**

( $\Rightarrow$ ) We have shown in previous example.

( $\Leftarrow$ ) Suppose  $m_T(x) = \prod (x - \lambda_i)$  for  $\lambda_i \neq \lambda_j$  for all  $i \neq j$ , and suppose

$$W = E(\lambda_1) + E(\lambda_2) + \dots + E(\lambda_r),$$

then we know  $W \subseteq V$  and  $W$  is  $T$ -invariant. Now if  $W \neq V$ , then let

$$\text{Ann}_T(V/W) = F[x] \cdot p(x),$$

and WLOG we can suppose  $p(x) = (x - \lambda_1)q(x)$  since  $m_T(x) \in \text{Ann}_T(V/W)$ , and we can check that  $p(x)$  cannot be a constant polynomial, otherwise  $V = W$ , which is a contradiction. Thus, there exists  $v \in V$  s.t.  $q(T)(v) \notin W$ . Set

$$g(x) = \frac{m_T(x)}{x - \lambda_1} = (x - \lambda_2) \dots (x - \lambda_r),$$

then  $g(x) = (x - \lambda_1)h(x) + g(\lambda_1)$ . Note that  $g(\lambda_1) \neq 0$ , so if we pick  $u = q(T)(v) \notin W$ , then

$$g(T)(u) = h(T)(T - \lambda_1)(u) + g(\lambda_1)(u)$$

and  $h(T)(T - \lambda_1)(u) = h(T)p(v) \in W$ , and  $g(\lambda_1)(u) \notin W$ , and  $g(T)(u) \in E(\lambda_1) \subseteq W$  since

$$(T - \lambda_1)g(T)(u) = (T - \lambda_1)(T - \lambda_2) \dots (T - \lambda_r)(u) = 0,$$

so we know this is a contradiction. Hence,  $W = V$ , so  $T$  is diagonalizable. ■

# Appendix