Linear Algebra I

Kon Yi

November 12, 2025

Abstract

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

Contents

Chapter 1

Vector Space

Lecture 1

3 Sep. 10:20

1.1 Introduction to vector and vector space

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.

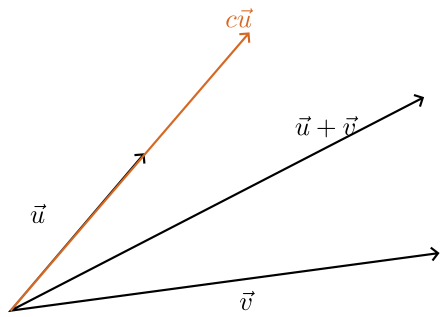

Figure 1.1: Vectors in $\mathbb{R}^2$

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

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

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

Also, if we define a space:

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

then this can also be a vector space after defining addtion 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 $\vec{a} = (a_1, a_2, \ldots, a_n) \in \mathbb{R}^n$, if we define

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

then this is also a vector space.

However, if we have

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

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

Example 1.1.5. In $V = \{(a, b) \to \mathbb{R}\}$ 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

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

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. $W_4 = \left\{\frac{\mathrm{d}^2 f}{\mathrm{d}x^2} = 0\right\}$ and $W_5 = \left\{\frac{\mathrm{d}^2 f}{\mathrm{d}x^2} = -f\right\}$ 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: $V \times V \to V$, that is, given $u, v \in V$, defining $u + v \in V$
- scalare multiplication: $\mathbb{R} \times V \to V$, that is, given $\alpha \to \mathbb{R}$ 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 filed, 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

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

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 combinaiton)**.** Given $v_1, v_2, \ldots, v_n \in V$, 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 $v_1, v_2, \ldots, v_n \in V$,

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

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

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

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

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 $v_i \in \operatorname{span}\{\hat{v_1}, \hat{v_2}, \dots, \hat{v_k}\}$ suppose the corresponding α_i of $\hat{v_1}, \dots, \hat{v_k}$ are not zeros.

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

Corollary 1.6.3. Linearly independent meeans 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 scaler of w and w is not a scalar of v.

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

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. $\operatorname{span}\{v_1, v_2, \dots, v_n\} = \langle v_1, v_2, \dots, v_n \rangle$

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

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

10 Sep 10:20

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.

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

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

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

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

Example 1.7.2. Suppose

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

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 $V = \langle v_1, v_2, \ldots, v_n \rangle$ and $\{w_1, w_2, \ldots, w_m\}$ is linearly independent, then $m \leq n$. Furthermore, one can make

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

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

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 $w_1, w_2, \ldots, w_{m+1}$ be linearly independent. In particular, $w_1, w_2, \ldots, w_m$ 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 $\langle w_1, w_2, \ldots, w_m \rangle = V$. However, by Lemma 1.7.1 and the note following it, we know

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

can not be linearly indepndent since $w_{m+1} \in V = \langle w_1, \ldots, w_m \rangle$. ⊛

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

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

Claim 1.7.2. One of $v_{m+1}, \ldots, v_n$ 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
$$\langle w_1, \ldots, w_m, w_{m+1}, v_{m+2}, \ldots, v_n \rangle = V.$$
⊛

■

Corollary 1.7.1. If $\{v_1, v_2, \ldots, v_n\}$ and $\{w_1, w_2, \ldots, w_m\}$ 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 $\langle v_1, v_2, \ldots, v_m \rangle = V$, then $m \geq n$. If $\{w_1, w_2, \ldots, w_m\}$ 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 $v_1, v_2, \ldots, v_n$ is linearly independent. If $w \notin \langle v_1, v_2, \ldots, v_n \rangle$, then
$$\{v_1, v_2, \ldots, v_n, w\}$$
is linearly independent.

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

Note 1.7.1. The reverse of Lemma 1.7.1 is still correct and is trivial. That is, if $w \notin \{v_1, \ldots, v_n\}$ and $\{v_1, v_2, \ldots, v_n, w\}$ is linearly independent, then $\{v_1, \ldots, v_n\}$ 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 $v_1, v_2, \ldots, v_m$ is linearly independent, then $\{v_1, v_2, \ldots, v_m\}$ forms a basis after adding some $v_{m+1}, \ldots, v_n$ to it.

Theorem 1.7.3 (Dual version)**.** If $\langle v_1, v_2, \ldots, v_n \rangle = V$, then $\{v_1, v_2, \ldots, v_m\}$ 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 ∞-dim vector space)**.**

- $$V = \{\text{all polynomials over } F\} = F[x] = \{a_0 + a_1 x + \cdots + a_n x^n \text{ for some } n \text{ where } a_i \in F\}.$$

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

  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 $\langle\varnothing\rangle = \{0\}$, which means $\varnothing$ 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
$$\sum_{i\in S} W_i = \{w_{i_1} + w_{i_2} + \cdots + w_{i_n} \text{ for some } i_j \in S\}$$
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, $f : \mathbb{R} \to \mathbb{R}$ is called continuous if $f(\lim_{x\to a} x) = \lim_{x\to a} f(x)$. 12 Sep 10:20

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

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

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

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

then if we have $T_A : F^n \to F^m$ 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.

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

Example 1.8.3. Consider row of vector space,

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

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

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

is a linear map.

Obeserve that a linear map $T : V \to W$ 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 $T : V \to W$ 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 $T : V \to W$, and for any $v \in V$, we know

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

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 $T : V \to W$, 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

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

Remark 1.8.4. Kernel and Image are subspaces.

Lecture 5

As previously seen. Given such a linear map $T : V \to W$, we define 17 Sep. 10:20

$$\begin{aligned} \ker T &= T^{-1}(0) \quad \text{kernel/null space of } T \\ \operatorname{Im} T &= T(V) \quad \text{image/range of } T, \end{aligned}$$

and $\ker T$ is a subspaces of V, and $\operatorname{Im} T$ 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 $\operatorname{rank} T = \dim \operatorname{Im} T$.

Example 1.8.5. Suppose $T : F^n \to F^m$, 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,

$$T_A \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} c_1 & \dots & c_n \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = x_1c_1 + \cdots + x_nc_n$$

since we know

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

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

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

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

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

Hence, we have to solve

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

Given $A = (c_1, \dots, c_n)_{m\times n}$, then the column rank is $\dim\langle c_1, \dots, c_m\rangle$. 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 $S_A : F^m \to F^n$, 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 $T : V \to W$ is a linear map, then

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

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

Claim 1.8.1. $S = \{v_1, \ldots, v_\nu, u_1, \ldots, u_r\}$ 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 $T(v) \in \operatorname{Im} T$, 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 α_i's, and we're done. ⊛

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

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

As previously seen. A map $f : X \to Y$ 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 $f^{-1} : Y \to X$ with $y \mapsto x$ if $f(x) = y$.

Proposition 1.8.3. Let $T : V \to W$ 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 $T : V \to W$ 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 $v_1, v_2, \ldots, v_n$ is a basis of V and we want to show $T(v_1), \ldots, T(v_n)$ is linearly independent. Suppose $\sum \alpha_i T(v_i) = 0$, then $T\left(\sum \alpha_i v_i\right) = 0$, so $\sum \alpha_i v_i = 0$, and thus $\alpha_i = 0$ for all i.

($\Leftarrow$) T sends one particular basis $v_1, \ldots, v_n$ 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 $T : V \to W$ 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 $T : V \to W$ is linear and bijective, then there is the inverse map $T^{-1} : W \to V$, and T^{-1} is also linear. In this case, $T : V \to W$ 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 $B = \{v_1, \ldots, v_n\}$ 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 $T : V \to W$, 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

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

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

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

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

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

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 $f : U \to V$ is a linear map, then if we define a map

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

then this map is linear. Similarly,

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

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

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

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
$$\mathcal{L}(F^n, F^m) \times \mathcal{L}(F^p, F^n) \to \mathcal{L}(F^p, F^m),$$
we know this is a linear map, and by Example 1.9.1, we know
$$M_{m\times n}(F) \times M_{n\times p}(F) \to M_{m\times p}(F)$$
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

$$\begin{tikzcd}
U \arrow[d, "f"] \arrow[dr, "T\circ F"] & \\
V \arrow[r, "T"'] \arrow[ur, phantom, "\circlearrowleft"] & W
\end{tikzcd}$$

linear

There is a special case,
$$\mathcal{L}(V, V) := \mathcal{L}(V) = \{\text{linear } T : V \to V\},$$
which is the space of linear operators on V.

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

$$\begin{tikzcd}
F^p \arrow[d, "T_B"'] \arrow[dr, "T_A \circ T_B = T_{AB}"] & \\
F^n \arrow[r, "T_A"'] & F^m
\end{tikzcd}$$

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,

$$\begin{tikzcd}
F^n \arrow[r, shift left, "T_p"] & F^m \arrow[l, shift left, "T_Q"]
\end{tikzcd}$$

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

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

by Proposition 1.8.6.

1.10 Map/matrix correspondence

$$\begin{tikzcd}
V \arrow[r, "T"] \arrow[d, "{[\cdot]_B}"'] \arrow[dr, phantom, "\circlearrowleft"] & W \arrow[d, "{[\cdot]_{B'}}"] \\
F^n \arrow[r, "\text{What is this?}"'] & F^m
\end{tikzcd}$$

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

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

and then we called A the martix 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 $[\cdot]_{B'}^{B} : \mathcal{L}(V, W) \to M_{m \times n}(F)$, ans 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.

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

$$\begin{tikzcd}
v_j \arrow[dd, leftrightarrow] & V \arrow[r] \arrow[d] & W \arrow[d] & \sum_{i=1}^{n} \alpha_{ij} w_i \arrow[dd, leftrightarrow] \\
& F^n \arrow[r] & F^m & \\
e_j \arrow[rrr, mapsto] & & & c_j = (\alpha_{1j}, \cdots, \alpha_{mj})^t
\end{tikzcd}$$

Special case:

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

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

Corollary 1.10.2. Given $T : V \to W$. There are $B = \{v_1, \ldots, v_n\}$ and $B' = \{w_1, \ldots, w_m\}$ 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 $p = \operatorname{rank}(T)$.

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

Example 1.10.1. Suppose $V = \{\text{polynomials with degree} \le k\}$ and W is the space of polynomials with degree $\le k+1$, then if $T : V \to W$ and $p(x) \mapsto \int_0^x p(t)\,dt$, then we know an ordered basis $B = \{1, x, x^2, \ldots, x^k\}$ and $B' = \{1, x, x^2, \ldots, x^{k+1}\}$, 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 $\le k$, and $B = \{1, x, x^j, \ldots, x^k\}$, and $B' = \{1, y, y^2, \ldots, y^k\}$ with $y = x - 1$. Then, if T is the identity transformation, note that

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

Hence, we have

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

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

26 Sep. 10:20

Lemma 1.10.1. Consider

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

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

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 $T : \mathbb{F}^n \to \mathbb{F}^m$, 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 $\operatorname{rank}(A) = \operatorname{rank}(B)$.

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

Now if $\operatorname{rank} A = \operatorname{rank} B$, 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 $T : V \to W$ 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

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

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

The matrix A is given by

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

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

$$v = c_1v_1 + c_2v_2 + \cdots + c_nv_n$$

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

$$T(v) = c_1T(v_1) + c_2T(v_2) + \cdots + c_nT(v_n).$$

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

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

and hence

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

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

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

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

The coordinate mapping $[\cdot]_C : W \to \mathbb{F}^m$ 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 $\operatorname{Im}(T)$ and $\operatorname{Col}(A)$:

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

Step 3: Compare dimensions.

Since isomorphic vector spaces have the same dimension,

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

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

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

Combining these equalities, we obtain

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

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 $\dim(\operatorname{Im}(T))$ 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 $P(A \mid Y) = (PA \mid PY)$. Hence, we can do row operations on $(X \mid Y)$ 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 $x_1 v_1 + \cdots + x_n v_n = 0$, where

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

then we have

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

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 $x_1 = x_2 = \cdots = x_n = 0$.

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 $f : V \to F$.
- $V^* = \mathcal{L}(V, F)$ 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 $(a_1, a_2, \ldots, a_n)$, then $f(e_i) = a_i$.

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

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

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

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

then we can define ev_s, the evaluation at s, by

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

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

$$I(f) = \int_p^q f(x)\,\mathrm{d}x,$$

then 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 $B = \{v_1, v_2, \ldots, v_n\}$ is a basis of V, and define $B^* = \{v_1^*, v_2^*, \ldots, v_n^*\}$ 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

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

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

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

Now if we know $T : V \to W$ is a linear map, then we can define $T^* : W^* \to V^*$ 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 $1 \le i \le n$ and $1 \le j \le m$.

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

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

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

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

and now we claim $\left\{v_{m+1}^*, \ldots, v_n^*\right\}$ 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 $f \in \langle v_{m+1}^*, \ldots, v_n^* \rangle$. Thus, $W^0 \subseteq \langle v_{m+1}^*, \ldots, v_n^* \rangle$ Also, $v_i^*(w) = 0$ for all $w \in W$, so we know $\langle v_{m+1}^*, \ldots, v_n^* \rangle \subseteq W^0$, and we're done. ■

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

Proof. First we show that $\ker T^* = (\operatorname{Im} T)^0$. 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 $w \in \operatorname{Im} T$, so $f \in (\operatorname{Im} T)^0$. Conversely, we can similarly show that $(\operatorname{Im} T)^0 \subseteq \ker T^*$, and we're done. Note that

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

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

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 $\operatorname{rank} T = \operatorname{rank} T^*$, 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

$$\mathrm{ev} : V \to V^{**}, \quad x \mapsto \widetilde{x},$$

where $\widetilde{x}$ is the functional

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

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

Proof. We can check $\widetilde{x}, \mathrm{ev}$ 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 $B = \{v_1 = v, v_2, \ldots, v_n\}$ as a basis of V and take dual B^*, then $v_1^*(v) = 1$. ■

Claim 2.1.1. $\mathrm{ev} : V \to V^{**}$ is injective.

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

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

Corollary 2.1.1. If $T : V \to W$ is a linear map with inverse $S : W \to V$, then $T^* : W^* \to V^*$'s inverse is $S^* : V^* \to W^*$, 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

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

Chapter 3

Eigenvalue and Eigenvector

Lecture 12

15 Oct. 10:20

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

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

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 λ, 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 λ, 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 λ 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 $\operatorname{adj}(A) \in M_n(F)$ where

$$(\operatorname{adj}(A))_{ij} = (-1)^{i+j} \det \left(A\left(j \mid i\right)\right),$$

where $A\left(j \mid i\right)$ 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

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

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 $D(v_1, v_2, \ldots, v_n) \in F$, then D is called multilinear or n-linear if

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

Example 3.0.1. If we suppose $A \in M_n(F)$, and r_i is the i-th row of A, where $r_i = (a_{i1}, a_{i2}, \ldots, a_{in})$, then If we define $D(A) = a_{ak_1} a_{2k_2} \ldots a_{nk_n}$, 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 $D(v_1, \ldots, v_n)$ 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

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

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

Lemma 3.0.2. If D is alternating, then

(1)
$$D(\ldots, \overbrace{v_i + \alpha v_j}^{i\text{-th position}}, \ldots) = D(\ldots, \overbrace{v_i}^{i\text{-th position}}, \ldots).$$

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

(3)
$$D(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_n) = -D(v_1, \ldots, v_j, \ldots, v_i, \ldots, v_n).$$

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

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

since D is alternating. ■

proof of (3). Since

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

so this is true. ■

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

$$D\left(e_{\sigma(1)}, e_{\sigma(2)}, \ldots, e_{\sigma(n)}\right),$$

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

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

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

$$D\left(e_{\sigma(1)}, e_{\sigma(2)}, \ldots, e_{\sigma(n)}\right)$$

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

Another approach/inductive construction

Theorem 3.0.1. There exists a function

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

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

We can just define

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

where $A(i \mid j)$ 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 $D(e_1, e_2, \ldots, e_n) = 1$, 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

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

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

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

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

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

Note 3.0.5. $C_{ij} = (-1)^{i+j} \delta\left(A(i \mid j)\right)$ 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 $D : M_n(R) \to R$ is n-linear, alternating, then

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

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

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

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 $\det : M_n(R) \to R$ satisfying n-linear, alternating, and $\det(I_n) = 1$.

(ii) If $D : M_n(R) \to R$ is n-linear, alternating, then $D(A) = D(I) \cdot \det(A)$.

(iii) For a permutation σ, if $\sigma = t_1 t_2 \ldots t_n = t'_1 t'_2 \ldots t'_m$, where t_i, t'_i's are transpositions, then $(-1)^n = (-1)^m$.

Remark 3.0.3. (ii) needs the fact that

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

if σ is the composition of m traspositions.

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

Now we introduce two formulas:

(1)

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

(2)

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

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

Thus, we know

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

by the proof above and Remark 3.0.3.

Lecture 14

29 Oct. 10:20

As previously seen. There is a unique function

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

satisfying n-linear in rows, alternating, and $\det(I_n) = 1$. Also, if $D : M_n(R) \to R$ 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 $c_{ij} = (-1)^{i+j} \det(A(i \mid j))$ is the (i, j)-cofactor.

If $\sigma \in S_n$, and let $\sigma(I) = (e_{\sigma(1)}, e_{\sigma(2)}, \ldots, e_{\sigma(n)})$ (permuting the rows), then $\det(\sigma(I)) = (-1)^m$ if $\sigma = \tau_1\tau_2 \ldots \tau_m$ where τ_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 $\sigma = \tau_1\tau_2 \ldots \tau_p = \tau_1'\tau_2' \ldots \tau_q'$, 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 $\sigma = \tau_1\tau_2 \ldots \tau_m$ for m even(resp. odd). Thus, we can define

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

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

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

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 $\det A = a_1a_2 \ldots a_n$, 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 $(-1)^{n-1}a_1a_2 \ldots a_n$ since it takes $n-1$ times exchangement. (exchange rows in the inoput of an alternating function will change the sign of output.)

Example 3.0.3. Companion form of $f(x) = x^n + a_1x^{n-1} + \cdots + a_n$:

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

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 $\operatorname{sgn}(\sigma) = \operatorname{sgn}\left(\sigma^{-1}\right)$.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

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

where

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

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

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

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

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

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 $(c_1, c_2, \ldots, c_n)^t$.

Proof. In fact,

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

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 $v_1, v_2, \ldots, v_n \in \mathbb{R}^n$, then

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

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

$$\det(T) \coloneqq \det\left([T]_B\right).$$

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 λ).

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

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

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

Remark 3.0.10. Note that

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

This is because x^n and x^{n-1} terms come from $(x - a_{11})(x - a_{22})\ldots(x - a_{nn})$, and by Vieta's theorem, we know the coefficient of x^{n-1} is $\mathrm{Tr}(T)$. 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

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

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 $(x - a_{11})(x - a_{22})\ldots(x - a_{nn})$.

Theorem 3.0.10. λ is an eigenvalue of T iff λ 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

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

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 been seen by observing the rank of matrix $[T]_B - \lambda_i I$. Also, we can observe that $V = E(\lambda_1) + E(\lambda_2) + \cdots + E(\lambda_r)$, 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 λ is an eigenvalue of T and characteristic polynomial $f(x) = (x-\lambda)^m g(x)$ with $g(\lambda) \neq 0$. The algebraic multiplicity of λ, $\text{a-mult}(\lambda) = m$, 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 λ as a root, so the algebraic multiplicity of $\lambda \geq$ the geometric multiplicity of λ.

Note 3.1.1. If A is not diagonalizable, then we know $\det(xI-D)$ may have λ 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 $w_1 + w_2 + \cdots + w_r = 0$ for $w_i \in W_i$, then $w_i = 0$ for all i.

Proposition 3.1.2. Let $W = W_1 + W_2 + \cdots + W_r$, 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)

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

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

(i) to (ii),(iii),(iv). ■

DIY

(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

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

for all i, so $w_i = 0$ for all i. ■

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

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

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

Hence, by this thought, suppose $v_1 + \cdots + v_m = 0$ 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

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

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 $F[x], \mathbb{Z}, \ldots$, the principal ideal domains(PID). Note that $F[x], \mathbb{Z}$ are Euclidean domain, which means that there is the degree map

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

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

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

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

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

Note 3.2.1. There exists a non-trivial relation among $T^0, T^1, \ldots, T^{n_2}$ 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 $\mathrm{Ann}_T(V)$ with minimal degree. Then,

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

Proof. For any $g(x) \in \mathrm{Ann}_T(V)$, 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 beacause $Ce_k = \lambda e_k$ 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 $\mathrm{ch}_T(x) \in \mathrm{Ann}_T(V)$(Cayley-Hamilton Theorem).

Proposition 3.2.2. If λ is some element of F, then

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

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 $(x-\lambda) \mid \mathrm{ch}_T(x)$.

■

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 $\mathrm{Im}(S)$ 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 $\mathrm{ch}_S(x) \mid \mathrm{ch}_T(x)$ and $m_S(x) \mid m_T(x)$.

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

$$\widetilde{B} = \{w_1, \ldots, w_m, w_{m+1}, \ldots, w_n\},$$

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

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

so we know $\det\left(xI - \widetilde{A}\right) = \det(xI - A)\det(D - xI)$, which gives $\mathrm{ch}_{\widetilde{A}}(x) = \mathrm{ch}_A(x)\mathrm{ch}_D(x)$, 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

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

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

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

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

Proof. Take $g \in \mathrm{Ann}_T(V/W)$, 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

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

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

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

and WLOG we can suppose $p(x) = (x - \lambda_1)q(x)$ since $m_T(x) \in \mathrm{Ann}_T(V/W)$, 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

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

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.

■

Lecture 17

7 Nov. 10:20

As previously seen. T is diagonalizable if and only if

$$\begin{cases} \mathrm{ch}_T(x) = \prod_{i=1}^{r}(x-\lambda_i)^{m_i}, \\ \dim E(\lambda_i) = m_i \quad \forall i \end{cases},$$

and we've learned that $m_T(x) = \prod_{i=1}^{r}(x-\lambda_i)$ for $\lambda_i \neq \lambda_j$.

3.4 Triangulanization and Cayley-Hamilton theorem

Definition 3.4.1. We call $T \in \mathcal{L}(V)$ triangulanizable if $\exists B = \{v_1, \dots\}$ s.t.

$$[T]_B = \begin{pmatrix} a_1 & \cdots & * \\ \vdots & \ddots & \vdots \\ 0 & \cdots & a_n \end{pmatrix},$$

i.e. $[T]_B$ is upper triangular. In particular, $T(v_k) \in \langle v_1, \dots, v_k \rangle$.

Corollary 3.4.1. If T is triangulanizable, then there exists a chain of T-invariant subspace $0 = W_0 \subseteq W_1 \subseteq W_2 \subseteq \dots \subseteq W_n = V$ where $\dim W_k = k$ for all k.

Corollary 3.4.2. If T is triangulanizable, and

$$[T]_B = \begin{pmatrix} a_1 & \cdots & * \\ \vdots & \ddots & \vdots \\ 0 & \cdots & a_n \end{pmatrix},$$

then $\mathrm{ch}_T(x) = \prod_{i=1}^{n}(x-a_i)$ splits (e.g. This always holds for $F = \mathbb{C}$). (And by Cayley-Hamilton we know $m_T(x)$ splits completely).

Lemma 3.4.1. Suppose $m_T(x)$ splits. If W is a T-invariant proper subspace of V, then $\exists u \notin W$ (i.e. u and W are linearly independent), and $\lambda \in F$ s.t. $(T-\lambda)(u) \in W$.

Proof. Since we have

$$\mathrm{Ann}_T(V/W) = \{g(x) \in F[x] \mid g(T)(V) \subseteq W\} = F[x] \cdot p(x),$$

and we know $m_T(x) \in \mathrm{Ann}_T(V/W)$, so $p(x) = (x-\lambda)q(x)$ for some $\lambda \in F$, where $x - \lambda \mid m_T(x)$

since $W \neq V$ and $p(x) \mid m_T(x)$. Hence, there exists $v \notin W$ s.t. $u = q(T)(v) \notin W$. Thus, we know

$$(T-\lambda)(u) = (T-\lambda)q(T)(v) = p(T)(v) \in W.$$

■

Theorem 3.4.1. Suppose $m_T(x)$ splits, then T is triangulanizable.

Proof. Use induction (for finding a T-invariant chain). Suppose we have

$$0 = W_0 \subseteq W_1 \subseteq W_2 \subseteq \cdots \subseteq W_k \neq V,$$

where $\dim W_i = i$ for all i. Then, $\exists v_{k+1} \notin W_k$ and $\lambda \in F$ s.t.

$$(T-\lambda)(v_{k+1}) = \sum_{i=1}^{k} a_{i,k+1} v_i$$

by Lemma 3.4.1. Hence, $T(v_{k+1}) \in \langle v_1, v_2, \ldots, v_{k+1}\rangle$ and thus $\langle v_1, \ldots, v_{k+1}\rangle$ is T-invariant, so we can let $W_{k+1} = \langle v_1, \ldots, v_{k+1}\rangle$. ■

Theorem 3.4.2 (Cayley-Hamilton theorem)**.** Let $f(x) = \mathrm{ch}_T(x)$ be the characteristic polynomial of T, then $f(T) = 0$.

Proof. We consider a matrix $A = (a_{ij})$, which is a matrix representation of T. We work over the commutative ring $F[A] = \left\{\sum_{i=0}^{m} a_i A^i\right\}$. Since $Ae_k = \sum_{i=1}^{n} a_{ik} e_k$, so if we let

$$B = (B_{ij}) = \begin{pmatrix} A - a_{11} & -a_{21} & & \\ \vdots & & & \\ -a_{1k} & \ldots & A - a_{kk} & \cdots \end{pmatrix},$$

we have $B_{k1}e_1 + \cdots + B_{kn}e_n = 0$. If we let $\mathrm{adj}(B) = (C_{ij})$, then

$$\begin{aligned} \det B &= C_{11}B_{11} + C_{12}B_{21} + \cdots + C_{1n}B_{n1} \\ 0 &= C_{11}B_{12} + C_{12}B_{22} + \cdots + C_{1n}B_{n2} \\ &\vdots \end{aligned}$$

We can check that $\det(e_k) = 0$ for all k, and $\det(B) = f(A)$, so we're done. ■

Alternative. First, recall that for any matrix $B \in M_n(\mathbb{C})$, one has

$$B\ \mathrm{adj}(B) = \det(B) I_n.$$

Take $B = A - xI$, we get

$$(A - xI)\,\mathrm{adj}(A - xI) = \det(A - xI)I_n = p_A(x)I_n.$$

Observation

Let

$$p_A(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0.$$

Note that

$$A - xI = \begin{pmatrix} a_{11} - x & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} - x & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} - x \end{pmatrix}.$$

Any minor of $A - xI$ is a polynomial of degree $\leq n-1$. Then we can write

$$\operatorname{adj}(A - xI) = B_0 + B_1 x + B_2 x^2 + \cdots + B_{n-1}x^{n-1}.$$

For example,

$$\operatorname{adj}\begin{pmatrix} x^2-3x & 2+2x & x \\ 3+x^2 & x & 2x \\ 4x & 3x^2 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 2 & 0 \\ 3 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} + \begin{pmatrix} -3 & 2 & 1 \\ 0 & 1 & 2 \\ 4 & 0 & 0 \end{pmatrix} x + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 3 & 0 \end{pmatrix} x^2.$$

Hence,

$$(A - xI)(B_0 + B_1 x + \cdots + B_{n-1}x^{n-1}) = a_n I x^n + a_{n-1} I x^{n-1} + \cdots + a_0 I.$$

By comparing coefficients, we get:

$$\begin{cases} a_n I = -B_{n-1}, \\ a_{n-1} I = AB_{n-1} - B_{n-2}, \\ a_{n-2} I = AB_{n-2} - B_{n-3}, \\ \quad \vdots \\ a_0 I = AB_0. \end{cases}$$

Multiplying each equation successively by appropriate powers of A (First equation multiplies A^n, the second one multiplies A^{n-1}, and so on), we obtain

$$a_n A^n + a_{n-1} A^{n-1} + \cdots + a_0 I = 0.$$

Thus,

$$p_A(A) = 0.$$

■

Chapter 4

Decompositions of spaces

Lecture 18

4.1 Direct Sums

12 Nov. 10:20

As previously seen. Let $W_1, \dots, W_r$ be subspaces of V. They are called linearly independent if $\sum w_i = 0$ with $w_i \in W_i$ for all i iff $w_i = 0$ for all i.

Let $W = W_1 + \cdots + W_r$, then TFAE:

(i) W_i's are linearly independent.

(ii) Any $w \in W$ has a unique expression $w = \sum_{i=1}^r w_i$ where $w_i \in W_i$.

(iii)
$$W_i \cap [W_1 + W_2 + \cdots + W_{i-1} + W_{i+1} + \cdots + W_r] = \{0\}.$$

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

(v) If $\{v_{ij}\}_{j=1}^{m_i}$ is a basis of W_i, then $\{v_{ij}\}_{i,j}$ is a basis of W.

In this case, we write
$$W = W_1 \oplus W_2 \oplus \cdots \oplus W_r,$$
and call it the direct sum.

Example 4.1.1. Let $T \in \mathcal{L}(V)$ with eigenvalues $\lambda_1, \dots, \lambda_r$ with $\lambda_i \neq \lambda_j$. Then,
$$W = E(\lambda_1) \oplus E(\lambda_2) \oplus \cdots \oplus E(\lambda_r).$$

4.2 Projections and idempotent decompositions

Definition 4.2.1. An operator $P \in \mathcal{L}(V)$ is called a projection if $P^2 = P$.

Remark 4.2.1. Note that if P is a projection, then suppose $W_1 = \operatorname{Im} P$ and $W_2 = \ker P$, then $V = W_1 \oplus W_2$. Suppose $v = v_1 + v_2$ with $v_i \in W_i$, then $Pv = Pv_1 + Pv_2 = v_1$. (Since $v_1 \in \operatorname{Im}(P)$, we have $v_1 = Pu$ for some u, so $Pv_1 = P^2u = Pu = v_1$.) Moreover, W_i's are P-invariant with $P|_{W_1} = \mathrm{id}$, $P|_{W_2} = 0$. $1 - P$ is a projection since $(1-P)^2 = 1 - 2P + P^2 = 1 - 2P + P = 1 - P$. In this case, we say P is a projection/idempotent onto W_1 and along W_2.

Theorem 4.2.1 (Idempotent decomposition)**.** Suppose $P_i \in \mathcal{L}(V)$ satisfying $1 = \sum_{i=1}^r P_i$ and $P_iP_j = 0$ for all $i \neq j$. Let $V_i = \operatorname{Im}(P_i)$, then $V = \bigoplus_{i=1}^r V_i$, and P_1 is the projection onto V_1 along $V_2 \oplus \cdots \oplus V_r$.

Proof. We first show that P_i is a projection for all i. WLOG, suppose $i = 1$, then

$$P_1^2 = P_1 (1 - P_2 - P_3 - \cdots - P_r) = P_1.$$

Given any $v \in V$, then

$$v = \sum_{i=1}^{r} v_i$$

where we suppose $v_i \in V_i$. Note that if $u \in V_j$, then $u \in \ker P_i$ for $i \neq j$. Thus,

$$P_i v = P_i \sum_{j=1}^{r} v_j = P_i v_i + \sum_{j \neq i} P_i v_j = P_i v_i = v_i.$$

Thus, this prove the uniqueness. ∎

Theorem 4.2.2. Suppose $V = \bigoplus_{i=1}^{r} V_i$. Let P_i be the projection onto V_i along $V_1 + \cdots + V_{i-1} + V_{i+1} + \cdots + V_r$, then $\sum_i P_i = 1$ and $P_i P_j = 0$ for all $i \neq j$.

Proof. Explicitly, for any $v \in V$, write its unique expression

$$v = \sum_{i=1}^{r} v_i, \quad v_i \in V_i.$$

Then, $P_i v = v_i$. Hence, we have $v = \sum v_i = \sum P_i v$ and

$$P_i P_j (v_1 + \cdots + v_r) = P_i \left(\sum_{l=1}^{r} P_j v_l \right) = P_i P_j (v_j) = P_i(v_j) = 0.$$

∎

4.3 T-invariant decomposition

Proposition 4.3.1. Suppose $V = \bigoplus V_i$ and $T_i \in \mathcal{L}(V_i)$. Define a map

$$T : V \to V, \quad \sum v_i \mapsto \sum T_i(v_i),$$

then

(i) $T \in \mathcal{L}(V)$

(ii) V_i is T-invariant

(iii) Suppose $1 = \sum P_i$ is the corresponding idempotenet decomposition. Then $TP_i = P_i T = T_i$.

Proof. Check

$$T(v + \alpha w) = Tv + \alpha T(w).$$

Now if $v \in V_i$, then the unique expression of v in V is $v = \sum_{j=1}^{r} v_j$ with $v_i = v$ and $v_j = 0$ for all $j \neq i$. So $T(v) = T_i(v_i) \in V_i$. Hence, V_i is T-invariant.

∎

Proposition 4.3.2. Let $V = \bigoplus_{i=1}^{r} V_i$, corresponding to $1 = \sum_{i=1}^{r} P_i$. Let $T \in \mathcal{L}(V)$. Suppose $TP_i = P_i T$, then

(i) V_i is T-invariant.

(ii) Let $T_i = T|_{V_i}$, then $T = \bigoplus T_i$.

Proof. For $u \in V_i$, we have $u = P_i u$ and $Tu = TP_i u = P_i(Tu)$, so $Tu \in V_i = \operatorname{Im} P_i$. For any $v \in V$, we know

$$Tv = \sum T_i v_i$$

if $v = \sum v_i$ where $v_i \in V_i$ since

$$Tv = T\left(\sum P_i v\right) = \sum T(P_i v) = \sum T v_i = \sum T_i v_i.$$

In this case, we write $T = \bigoplus T_i$ and if $B_i = \{v_{ij}\}_{j=1}^{m_i}$ is an ordered basis of V_i, and $B = \{v_{ij}\}_{i,j}$ is an ordered basis of V. ■

Example 4.3.1. Let $T \in \mathcal{L}(V)$, and let $f(x) = \mathrm{ch}_T(x)$ be its characteristic polynomial. Suppose $f(x) = g(x) \cdot h(x)$ with $g(x)$ and $h(x)$ coprime, then

$$1 = p(x)g(x) + q(x)h(x)$$

for some p, q. Thus,

$$1 = p(T)g(T) + g(T)h(T),$$

and let $P = p(T)g(T)$ and $Q = g(T)h(T)$, then $PT = TP$ and $QT = TQ$. Also, $PQ = 0$. Note that $PQ = 0$ since

$$PQ = p(T)q(T)f(T) = 0$$

by Cayley-Hamilton theorem. Thus, we know this gives an idempotenet decomposition.

Remark 4.3.1. $\operatorname{Im} P = \ker Q$, and $\operatorname{Im} Q = \ker P$. If we let $W_1 = \operatorname{Im} P$ and $W_2 = \operatorname{Im} Q$, we will see the characteristic polynomial of $T|_{W_1} = h(x)$ and $T|_{W_2} = g(x)$.

Appendix