

1 Vector Spaces

1.1 Vector Spaces

Definition 1.1 (Vector Spaces). A vector space V over a field F consists of a set on which two operations: addition and scalar multiplication, s.t. $\forall x, y \in V, \exists! x + y \in V$ and $\forall a \in F, x \in V, \exists! ax \in V$. also,

$$(VS\ 1) \ \forall x, y \in V, x + y = y + x$$

$$(VS\ 2) \ \forall x, y, z \in V, (x + y) + z = x + (y + z)$$

$$(VS\ 3) \ \exists 0 \in V, \text{s.t.} \ \forall x \in V, x + 0 = x$$

$$(VS\ 4) \ \forall x \in V, \exists y \in V, \text{s.t.} \ x + y = 0$$

$$(VS\ 5) \ \forall x \in V, 1x = x$$

$$(VS\ 6) \ \forall a, b \in F, x \in V, (ab)x = a(bx)$$

$$(VS\ 7) \ \forall a \in F, x, y \in V, a(x + y) = ax + ay$$

$$(VS\ 8) \ \forall a, b \in F, x \in V, (a + b)x = ax + bx$$

Theorem 1.1 (Cancellation Law). Suppose $x, y, z \in V$ s.t. $x + z = y + z$. Then $x = y$.

Proof. By **Item (VS 4)**, we know $\exists w$ s.t. $z + w = 0$.

So $(x + z) + w = (y + z) + w$.

By **Item (VS 2)**, we have $x + (z + w) = y + (z + w) \implies x + 0 = y + 0$.

By **Item (VS 3)**, we have $x = y$. ▣

Corollary 1.1.1. 0 is unique.

Proof. Suppose $\exists 0'$ s.t. $\forall x \in V, 0' + x = x$.

Then $0' + 0 = 0 \implies 0 = 0'$. ▣

Corollary 1.1.2. $\forall x \in V, \exists! y$ s.t. $x + y = 0$.

Proof. Suppose for one $x \in V, \exists y, y'$ s.t. $x + y = 0, x + y' = 0$.

Then $x + y = x + y' \implies y = y'$. ▣

Theorem 1.2.

$$1. \ 0x = 0 \ \forall x \in V.$$

$$2. (-a)x = -(ax) = a(-x) \quad \forall a \in F, x \in V.$$

$$3. a0 = 0 \quad \forall a \in F.$$

Proof.

1. Consider $ax + 0x$.

By **Item (VS 8)**, $ax + 0x = (a + 0)x = ax$.

By **Corollary 1.1.1**, $0x = 0$.

2. Consider $ax + (-a)x$.

By **Item (VS 8)**, $ax + (-a)x = (a + (-a))x = 0x$.

By 1., $0x = 0$. By **Corollary 1.1.2**, $(-a)x = -(ax)$.

Consider $ax + a(-x)$.

By **Item (VS 7)**, $ax + a(-x) = a(x + (-x)) = a0$.

By 3., $a0 = 0$. By **Corollary 1.1.2**, $a(-x) = -(ax)$.

3. consider $ax + a0$.

By **Item (VS 7)**, $ax + a0 = a(x + 0) = ax$.

By **Corollary 1.1.1**, $a0 = 0$.

▣

1.2 Subspaces

Definition 1.2 (Subspace). A subset $W \subseteq V$ over F is called a subspace of V if W is also a vector space over F with the same operations as V .

Theorem 1.3. W is a subspace of $V \iff$

$$1. \quad \forall x, y \in W, x + y \in W.$$

$$2. \quad \forall x \in W, c \in F, cx \in W.$$

$$3. \quad W \text{ has } 0.$$

Proof. apparently.

▣

Theorem 1.4. If U, W are subspaces of V , then $U \cap W$ is also a subspace of V .

Proof. Since U, W 皆為 subspace, 由 Theorem 1.3 知 $0 \in U \cap W$.

$\forall x, y \in U \cap W$, consider $x + y$.

$x + y \in U$, 因為 $x \in U, y \in U$, 且 U 為 subspace.

同理 $x + y \in W$.

同理 $\forall x \in U \cap W, cx \in U \cap W$. ▣

Problem 1.3.19. Let W_1, W_2 both be subspaces of V . Show $W_1 \cup W_2$ is a subspace of $V \iff W_1 \subseteq W_2$ or $W_2 \subseteq W_1$.

Proof.

(\Leftarrow) 顯然.

(\Rightarrow) Suppose $\exists x \in W_1$ s.t. $x \notin W_2$. (i.e. $W_1 \not\subseteq W_2$, 欲證 $W_2 \subseteq W_1$)

Then 任取 $y \in W_2$. $x + y \in W_1 \cup W_2$.

if $v = x + y \in W_1, -x \in W_1, y = v - x \in W_1$. 故 $\forall y \in W_1, y \in W_2 \implies W_2 \subseteq W_1$.

if $v = x + y \in W_2, -y \in W_2, x = v - y \in W_2$. ▣

1.3 Linear Combinations and Span

Definition 1.3 (Linear Combinations). Let V be a vector space, and $S \subseteq V, S \neq \emptyset$. We say $v \in V$ is a linear combination of vectors in S if $\exists u_i \in S, i = 1, 2, \dots, n$ (with n finite) and $a_i \in F$ s.t. $v = \sum_i a_i u_i$.

Definition 1.4 (Span). Let $S \subseteq V, S \neq \emptyset$. Define $\text{span } S := \{\text{all LC of } S\}$. also $\text{span}(\emptyset) := \{0\}$.

Theorem 1.5. $\forall S \subseteq V$, $\text{span } S$ is a subspace of V . Any subspace of V that contains S must contain $\text{span } S$.

Proof.

(trivial) If $S = \emptyset, \text{span } S = \{0\}$ is a subspace.

$\forall W$ which is a subspace, we have $\emptyset \subseteq W$, so $\text{span } S \subseteq W$.

(non-trivial) if $S \neq \emptyset, 0 \in \text{span } S$, 因為 $0v = 0$.

$\forall v \in \text{span } S, v = \sum_i a_i u_i, u_i \in S$. 故 $cv = c \sum_i a_i u_i = \sum_i (ca_i) u_i \in \text{span } S$.

$\forall v_1, v_2 \in \text{span } S, v_1 = \sum_i a_i u_i, v_2 = \sum_j b_j w_j$.

故 $v_1 + v_2 = \sum_k c_k x_k \in \text{span } S$.

Suppose $\exists x \in \text{span } S$ 但 $x \notin W$, where W is a subspace of V and $S \subseteq W$. 但如此 x 必 $\in W$. ▣

✂ *Remark.* If W is a subspace of V 且 $S \subseteq W$, 則 $\text{span } S \subseteq W$. ▣

Definition 1.5. We say a subset $S \subseteq V$ generates or spans V is $\text{span } S = V$.

Problem 1.4.12. Show $W \subseteq V$ is a subspace $\iff \text{span } W = W$.

Proof.

(\Leftarrow) 由 Theorem 1.5 顯然.

(\Rightarrow) 因為 W is a subapce, 故 $\forall u_i \in W, \sum_i a_i u_i \in W \implies \text{span } W \subseteq W$.

又 $W \subseteq \text{span } W$ 顯然. 故 $\text{span } W = W$. ▣

Problem 1.4.13. Show $S_1 \subseteq S_2 \subseteq V \implies \text{span } S_1 \subseteq \text{span } S_2$. If $\text{span } S_1 = V$, then $\text{span } S_2 = V$.

Proof. $\forall v \in \text{span } S_1, v = \sum_i a_i u_i, i = 1, 2, \dots, n, u_i \in S_1$.

而 $S_1 \subseteq S_2$, 故 $u_i \in S_2$. 故 $v \in \text{span } S_2$

故 $\text{span } S_1 \subseteq \text{span } S_2$.

若 $\text{span } S_1 = V, V \subseteq \text{span } S_2$.

而 V is a vector space, 故 $\text{span } S_2 \subseteq V$, 故 $V = \text{span } S_2$. ▣

Problem 1.4.14. Show if $S_1, S_2 \subseteq V \implies \text{span}(S_1 \cup S_2) = \text{span } S_1 + \text{span } S_2$.

Proof.

(\subseteq) $\forall v \in \text{span}(S_1 \cup S_2)$,

$v = \sum_i a_i u_i + \sum_j b_j w_j + \sum_k c_k x_k, u_i \in S_1 - S_1 \cap S_2, w_j \in S_1 \cap S_2, x_k \in S_2 - S_1 \cap S_2$.

For convenience, combine u_i, w_j . Thus $v \in \text{span } S_1 + \text{span } S_2$.

(\supseteq) $\forall v \in \text{span } S_1 + \text{span } S_2, v = \sum_i a_i u_i + \sum_j b_j w_j, u_i \in S_1, w_j \in S_2$.

而 $u_i, w_j \in S_1 \cup S_2$, 故 $v \in \text{span } S_1 \cup S_2$. ▣

Problem 1.4.15. Show if $S_1, S_2 \in V \implies \text{span}(S_1 \cap S_2) \subseteq \text{span } S_1 \cap \text{span } S_2$.

Proof. $\forall v \in \text{span}(S_1 \cap S_2), v = \sum_i a_i u_i, u_i \in S_1 \cap S_2$.

So $v \in \text{span } S_1, v \in \text{span } S_2$. So $v \in \text{span } S_1 \cap \text{span } S_2$. ▣

✂ *Remark.* Equal example: $S_1 = \{(1, 0, 0)\}, S_2 = \{(0, 1, 0)\}$. $\text{span}(S_1 \cap S_2) = \text{span } \emptyset = \{0\} = \text{span } S_1 \cap \text{span } S_2$.

Unequal example: $S_1 = \{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$,
 $S_2 = \{(-1, 0, 0), (0, -1, 0), (0, 0, -1)\}$. $\text{span}(S_1 \cap S_2) = \text{span } \emptyset = \{0\}$, yet $\text{span } S_1 = \text{span } S_2 = \mathbb{R}^3$, so $\text{span } S_1 \cap \text{span } S_2 = \mathbb{R}^3$. ▣

1.4 Linear Dependence

Definition 1.6. A subset $S \subseteq V$ is linearly dependent if $\exists u_i \in S, i = 1, 2, \dots, n$ and a_i not all zero s.t. $\sum_i a_i u_i = 0$.

A subset is linearly independent if it's not linearly dependent.

Theorem 1.6. Let $S_1 \subseteq S_2 \subseteq V$. If S_1 is dependent, then S_2 is also dependent.

Proof. Given S_1 dependent $\implies \exists a_i \in F, u_i \in S_1$ s.t. $\sum_i a_i u_i = 0$, and a_i 不全為 0. 但 $u_i \in S_2$ also. 故 S_2 亦為 dependent. \square

Corollary 1.6.1. Let $S_1 \subseteq S_2 \subseteq V$. If S_2 independent, then S_1 also independent.

Proof. Suppose S_1 dependent. That means S_2 is dependent. \times \square

Theorem 1.7. Let subset $S \subseteq V$ and S independent, and let $v \in V - S$. Show $S \cup \{v\}$ is dependent $\iff v \in \text{span } S$.

Proof.

(\Leftarrow) 因為 $v \in \text{span } S$, 故 $v = \sum_i a_i u_i, u_i \in S$ 故 $\sum_i a_i u_i - v = 0, S \cup \{v\}$ dependent.

(\Rightarrow) (trivial) If $v = 0$, then apparently.

(non-trivial) If $v \neq 0$. Given $S \cup \{v\}$ dependent, that means $\exists a_i, u_i$ s.t. $a_1 u_1 + a_2 u_2 + \dots + a_n u_n + bv = 0$.

首先 a_i 不全為 0. 若 $a_i = 0 \forall i$, 則 $bv = 0$. \times

又 $b \neq 0$. 若 $b = 0$, 則 a_i 不全為 0, S dependent. \times

故 $v = b^{-1}(\sum_i a_i u_i) \in \text{span } S$. \square

Problem 1.5.9. Let $u \neq v \in V$. Show $\{u, v\}$ dependent $\iff u = cv$ or $v = cu$.

Proof.

(\Leftarrow) 顯然.

(\Rightarrow) $\exists a_1, a_2$ s.t. $a_1 u + a_2 v = 0$.

if $a_1 = 0 \implies v = 0, v = 0u$. If $a_2 = 0 \implies u = 0, u = 0v$.

if $a_1 a_2 \neq 0 \implies v = -a_1 a_2^{-1} u$. \square

Problem 1.5.13.a. Let V over $F, \text{char } F \neq 2$. Show $u, v \in V, \{u, v\}$ independent $\iff \{u + v, u - v\}$ independent.

Proof.

(\Rightarrow) Suppose $\{u + v, u - v\}$ dependent.

$\exists a_1, a_2$ s.t. $a_1(u + v) + a_2(u - v) = 0$. It's clear that $a_1 \neq 0, a_2 \neq 0$, since $u = -v$ or $u = v$, respectively.

$a_1(u + v) + a_2(u - v) = 0 \implies (a_1 + a_2)u + (a_1 - a_2)v = 0$. Similar reasoning shows $a_1 \neq a_2, a_1 \neq -a_2$, since $u = 0, v = 0$, respectively.

But if it is the case, then $u = -(a_1 - a_2)(a_1 + a_2)^{-1}v$. \times

(\Leftarrow) Suppose $\{u, v\}$ dependent. $\exists a_1, a_2$ s.t. $a_1 u + a_2 v = 0$.

則 $a'_1(u + v) + a'_2(u - v) = 0 \implies (a'_1 + a'_2)u + (a'_1 - a'_2)v = 0$.

則 $a'_1 = (a_1 + a_2)/2, a'_2 = (a_1 - a_2)/2$ 有非 0 解. \times \square

1.5 Bases and Dimension

✍ *Remark.* Show if $W = \text{span } S$ and $\nexists S' \subset S$ s.t. $W = \text{span } S' \implies S$ independent. \square

Proof. Suppose S is dependent. Then 由 Theorem 1.7, $\exists v \in S$ s.t. $v \in \text{span}(S - \{v\})$.
故 $\text{span}(S - \{v\}) = \text{span } S = W$. \square

Definition 1.7 (Basis). A basis β of V means $\beta \subseteq V$, β independent, and $\text{span } \beta = V$.

Problem 1.4.16. Let $S \subseteq V$ s.t. $v_i \in S, a_i \in F, i = 1, 2, \dots, n$ and $\sum_i a_i v_i = 0 \implies a_i = 0$. Show $\forall v \in \text{span } S, \exists! a_i$ s.t. $\sum_i a_i u_i = v$.

Proof.

(trivial) $v = 0 \implies \text{only } a_i = 0$.

(non-trivial) Suppose $\exists a'_j$ s.t. $v = \sum_i a_i u_i = \sum_j a'_j w_j, u_i, w_j \in S$.

Consider $0 = \sum_i a_i u_i - \sum_j a'_j w_j$.

Firstly, $w_j \neq \text{LC of } u_i$, since that would be linear dependent.

So only consider w_j partly equal u_i .

If $u_i \neq w_j \forall i, j$, 由題設 $a_i = a'_j = 0$ \square

If some $u_i = w_j$, then $a_i - a'_j = 0, a_i = a'_j$, others $= 0$ \square

Theorem 1.8. Let $\beta = \{u_1, u_2, \dots, u_n\} \subseteq V$. β is a basis of $V \iff \forall v \in V, \exists! a_i, i = 1, 2, \dots, n$ s.t. $v = \sum_i a_i u_i$

Proof.

(\implies) See Problem 1.4.16.

(\impliedby) Suppose β is dependent. Then $\exists b_i$ not all zero s.t. $\sum_i b_i u_i = 0$. If so, then given $v = \sum_i a_i u_i$, one can construct $v = v + 0 = \sum_i a_i u_i + \sum_i b_i u_i = \sum_i (a_i + b_i) u_i$ not unique. \square

✍ *Remark.* $v \leftrightarrow (a_1, a_2, \dots, a_n) \implies V$ is like F^n . This correspondence is an isomorphism, which we'll show in Theorem 2.14 \square

Theorem 1.9. If $V = \text{span } S$ with S finite, then some $\beta \subseteq S$ is a basis of V . So V has a finite basis.

Proof. If S is independent, then choose $\beta = S$ is a basis.

If S is dependent, then by Theorem 1.7, $\exists S' \subset S$ s.t. $V = \text{span } S'$. Repeat this process until S is independent. \square

✂ *Remark.* This proof is done by myself. I don't know if it's rigorous or not. The only place that is not rigorous is that I do not promise there is a point where this process will stop. See the proof below. \square

Proof.

(trivial) If $S = \emptyset$ or $S = \{0\} \implies V = \{0\}$ so $\emptyset \subseteq S$ and $\text{span } \emptyset = V$.

(non-trivial) If $S \neq \emptyset$, then some $\{u_1\} \subseteq S$.

choose $\beta = \{u_1, u_2, \dots, u_l\}$ s.t. $\forall v \in S - \beta, \beta \cup \{v\}$ is dependent. Then we claim that β is a basis of V .

Because β is already independent, we need only to prove $V = \text{span } \beta$.

By the interpretation of **Theorem 1.5**, we see if $S \subseteq \text{span } \beta$, then $\text{span } S = V \subseteq \text{span } \beta \implies V = \text{span } \beta$.

Let $v \in S$. If $v \in \beta$, apparently $v \in \text{span } \beta$.

if $v \notin \beta$, then by the requirement that $\beta \cup \{v\}$ is dependent, we see $v \in \text{span } \beta$. Thus $S \subseteq \text{span } \beta$. \square

Theorem 1.10 (Replacement Theorem). *Let $V = \text{span } G, |G| = n$. Let $L \subseteq V$ be an independent subset, and $|L| = m$. Then $m \leq n$, and $\exists H \subseteq G, |H| = n - m$ s.t. $L \cup H$ spans V .*

Proof. The proof is done by induction on m .

($m = 0$) If $m = 0, L = \emptyset$, then 取 $H = G$.

($m < n$) Suppose $m = k < n$ 時成立. 則 $|L_k| = k, \exists H_k \subseteq G$ s.t. $|H_k| = n - k$ and $V = \text{span}(L_k \cup H_k)$.

Consider $m = k + 1$. Let $L_{k+1} = L_k \cup \{v\}$. Since $L_k \cup H_k$ spans V , we have $v = \sum_i a_i u_i + \sum_j b_j w_j, u_i \in L_k, w_j \in H_k$.

But b_j 不全為 0. If so, $v - \sum_i a_i u_i = 0$, then L_{k+1} would be dependent. \times

Let $b_\ell \neq 0$, and let $H_{k+1} = H_k - \{w_\ell\}$. We see $w_\ell = vb_\ell^{-1} - \sum_i (a_i b_\ell^{-1}) u_i - \sum_{j \neq \ell} (b_j b_\ell^{-1}) w_j$.

So we see $w_\ell \in \text{span}(L_{k+1} \cup H_{k+1})$.

Also, $\forall u_i \in L_k, u_i \in \text{span}(L_{k+1} \cup H_{k+1})$;

$\forall w_j \neq w_\ell \in H_k, w_j \in \text{span}(L_{k+1} \cup H_{k+1})$.

So $L_k \cup H_k \subseteq \text{span}(L_{k+1} \cup H_{k+1})$, so $\text{span}(L_{k+1} \cup H_{k+1}) = V$ by **Theorem 1.5**.

($m = n$) We now show that the induction will end. Suppose $m = n, |L_n| = n, H_n = \emptyset$.

Now $V = \text{span}(L_n \cup H_n) = \text{span } L_n$, so any $v \in V$ also $\in \text{span } L_n$. By **Theorem 1.7**, L_{n+1} would be dependent. \times \square

Corollary 1.10.1. *Let V having a finite basis β . Then every basis of V has same number of vectors.*

Proof. Let β' be another basis for V .

We first show there would be no infinite basis. For $|\beta'| > n$, $\exists S \subseteq \beta'$ s.t. $|S| = n+1$, and S is independent, by **Corollary 1.6.1**. So **Theorem 1.10** says $|S| = n+1 \leq n = |\beta|$.

※

By **Theorem 1.10**, $V = \text{span } \beta \implies |\beta'| \leq |\beta|$.

But same reasoning applies to β' . So $|\beta| \leq |\beta'|$. So $|\beta| = |\beta'|$. ▣

Definition 1.8. We say V is finite-dimensional if \exists a basis containing a finite number of vectors. This unique number is called dimension, denoted by $\dim V$. V is called infinite dimensional if it is not finite dimensional.

✍ *Remark.* Dimension may depend on field. For example, $V = \mathbb{C}$ over \mathbb{C} has $\dim V = 1$; $V = \mathbb{C}$ over \mathbb{R} has $\dim V = 2$. □

Corollary 1.10.2. Let $\dim V = n$. (Which promised there would be a basis.)

1. Any finite S s.t. $V = \text{span } S$ 滿足 $|S| \geq n$. If $|S| = n$, then S is a basis.
2. Any $S \subseteq V$, S independent, $|S| = n$ is a basis.
3. Any $S \subseteq V$, S independent can be extended to form a basis.

Proof.

1. (1) If $\exists S$ s.t. $V = \text{span } S$ and $|S| < n$, then by **Theorem 1.9** some $\beta \subseteq S$ is a basis.

But $|\beta| \leq |S| < n$ contradicts $\dim V = n$. ※

- (2) If $|S| = n$, $V = \text{span } S$ but S not a basis, then S must be dependent.

By **Theorem 1.9**, $\beta \subseteq S$ is a basis. But S is dependent, so $\beta \subset S$. But $|\beta| < |S| = n$, contradicts $\dim V = n$. ※

2. By **Theorem 1.10**, if $L \subseteq V$, L independent, $|L| = n$, then $\exists H = \emptyset$ s.t. $L \cup H$ spans $V \implies V = \text{span } L \implies L$ is a basis.

3. By **Theorem 1.10**, $\forall S \subseteq V$, S independent, $\exists H \subseteq G$, $|H| = n - m$ s.t. $V = \text{span}(S \cup H)$.

By (1), $|S \cup H| \geq n$. 但由排容, $|S \cup H| \leq m + n - m = n. \implies |S \cup H| = n$. 又由 (1), $S \cup H$ is a basis.

▣

Theorem 1.11. Let W be a subspace of V , with $\dim V \neq \infty$. Then $\dim W \neq \infty$ and $\dim W \leq \dim V$. If $\dim W = \dim V$, then $W = V$.

Proof.

(trivial) If $W = \{0\}$, basis = \emptyset , $\dim W \leq \dim V$. ($W = \emptyset$ is not a subspace.)

(non-trivial) If $W \neq \{0\}$, then W contains some $\{u_1\}$, $u_1 \neq 0$.

Depend on the dimension of W , one continue expands the set independently till $\dim W = n$, when the set becomes a basis for V .

At that point, any $v \in V$ also $\in \text{span } S$, so by **Theorem 1.7**, it cannot be a basis anymore. So $\dim W \not> \dim V$, 且 $\dim W \neq \infty$.

Also at that point, $W = \text{span } S = V$. ▣

Corollary 1.11.1. *If W is a subspace of V , $\dim V \neq \infty$, then any basis W can be extended to a basis of V .*

Proof. Let S be a basis of W . If $W = V$ 則顯然. 由 **Corollary 1.10.2** 顯然. ▣

Problem 1.6.20. *Let $\dim V = n$, $S \subseteq V$, $\text{span } S = V$. ($|S|$ may not be finite.) Show*

1. $\exists S' \subseteq S$ s.t. S' is a basis of V .

2. $|S| \geq n$.

Proof.

1. Choose $u_i \in S, i = 1, 2, \dots, n$ independently.

We now show this can be done. Suppose there's a point where $i < n$ and $\forall v \in S, S' \cup \{v\}$ is dependent, then by **Theorem 1.7** $v \in \text{span } S'$.

Thus $S \subseteq \text{span } S'$, by **Theorem 1.5**, $V = \text{span } S \subseteq \text{span } S' \implies \text{span } S' = V$. *

2. trivial. ▣

Problem 1.6.21. *Show $\dim V = \infty \iff$ it contains an infinite independent set.*

Proof.

(\implies) $\forall S \subseteq V, S$ finite and independent, $\exists v \in V$ s.t. $v \notin \text{span } S$.

By **Theorem 1.7**, $v \notin \text{span } S \iff S \cup \{v\}$ independent.

The process can continue indefinitely. For if it stops, then we have a finite basis. *

(\impliedby) Suppose $\dim V = n$. Let the infinite set be S .

By **Corollary 1.6.1**, any $L \subseteq S$ is also independent.

取 $|L| = n + 1$. Then by **Theorem 1.10**, $n + 1 \leq n$. * ▣

Problem 1.6.22. Let W_1, W_2 be subspaces of V . Determine necessary and sufficient conditions on W_1, W_2 so that $\dim(W_1 \cap W_2) = \dim W_1$.

Proof.

Initially I guessed $W_1 = W_2$, which I quickly realized is not necessary. Then I guessed $W_1 \subseteq W_2$, which I'll prove below.

(\Rightarrow) Trivial.

(\Leftarrow) $\exists S \subseteq W_1 \cap W_2, \text{span } S = W_1 \cap W_2, S' \subseteq W_1, \text{span } S' = W_1$ s.t. $|S| = |S'|, S, S'$ independent.

Suppose $\exists v \in W_1$ s.t. $v \notin W_2$. (i.e. $W_1 \not\subseteq W_2$.)

Let $\dim(W_1 \cap W_2) = n$, and β a basis of $W_1 \cap W_2$.

We see $v \notin W_1 \cap W_2 = \text{span } \beta \iff \beta \cup \{v\}$ independent. By [Theorem 1.10](#), choose $L = \beta \cup \{v\} \subseteq W_1$, we have $n + 1 \leq \dim W_1$. \square

Problem 1.6.24. Let $f(x) \in P_n(\mathbb{R}), \deg f = n$. Show $\forall g(x) \in P_n(\mathbb{R}), \exists c_i \in \mathbb{R}, i = 0, 1, \dots, n$ s.t. $g(x) = \sum_{i=0}^n c_i f^{(i)}(x)$.

Proof. 已知 $\dim P_n(\mathbb{R}) = n + 1$, let $S = \{f, f', \dots, f^{(n)}\}, |S| = n + 1$. Only need to show S is independent.

Consider $0 = \sum_{i=0}^n c_i f^{(i)}(x)$.

In the term $i = 0, \deg f = n$, so x^n term $\neq 0$ yet $\deg 0 = 0 \implies c_0 = 0$.

$i = 1, \deg f' = n - 1, [x^{n-1}] \neq 0 \implies c_1 = 0$.

Same reasoning shows $c_i = 0$. \square

1.6 Cosets and Quotient Space

1.6.1 Definition

Definition 1.9 (Cosets). Let W be a subspace of V over F . $\forall v \in V$, we define coset of W containing v as

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

Definition 1.10 (Quotient Space). On the set $S := \{v + W \mid v \in V\}$, define the following operations for all $v_1, v_2, v \in V, a \in F$:

$$\text{addition: } (v_1 + W) + (v_2 + W) := (v_1 + v_2) + W,$$

$$\text{scalar multiplication: } a(v + W) := (av) + W.$$

The set S is called the quotient space of V mod W , denoted by V/W .

Problem 1.3.31.1.

1. Show $v + W$ is a subspace of $V \iff v \in W$.

2. Show $v_1 + W = v_2 + W \iff v_1 - v_2 \in W$.

Proof.

1. (\Leftarrow) 顯然.

(\Rightarrow) Suppose $v \notin W$. Then $\forall u \in W, v \neq (-u)$. 故 $v + u \neq 0, 0 \notin v + W$. *

2. (\Leftarrow) Let $v_1 - v_2 = u \in W$.

(To prove $v_1 + W = v_2 + W$, show $v_1 + W \subseteq v_2 + W$ and $v_2 + W \subseteq v_1 + W$.)

$\forall x \in v_1 + W, \exists w_1 \in W$ s.t. $x = v_1 + w_1$.

故 $x = v_1 + w_1 = v_2 + (v_1 - v_2 + w_1) = v_2 + (u + w_1) \in v_2 + W$.

同理得 $v_2 + W \subseteq v_1 + W \implies v_1 + W = v_2 + W$.

(\Rightarrow) $v_1 + W = v_2 + W$.

表示 $\forall x \in v_1 + W, \exists w_1$ s.t. $x = v_1 + w_1$,

且同時 $x \in v_2 + W, \exists w_2$ s.t. $x = v_2 + w_2$.

故 $v_1 + w_1 = v_2 + w_2 \implies v_1 - v_2 = w_2 - w_1 \in W$.

▣

Problem 1.3.31.2.

1. Show the operations of cosets are well-defined. i.e. if $v_1 + W = v'_1 + W, v_2 + W = v'_2 + W \implies (v_1 + W) + (v_2 + W) = (v'_1 + W) + (v'_2 + W)$, and $a(v_1 + W) = a(v'_1 + W)$.

2. Show S is a vector space with addition and scalar multiplication defined in *Definition 1.10*.

Proof.

1. By *Item 2*及題設, 知 $v_1 - v'_1, v_2 - v'_2 \in W$.

故加法部分欲證 $\iff (v_1 + v_2) - (v'_1 + v'_2) \in W$.

但 $(v_1 - v'_1) + (v_2 - v'_2) = (v_1 + v_2) - (v'_1 + v'_2) \in W$. 而乘法部分欲證 \iff

$av_1 - av'_1 = a(v_1 - v'_1) \in W$.

但已知 $v_1 - v'_1 \in W$ 故得證.

2. 顯然.

▣

1.6.2 Properties

Problem 1.6.35. Let W be a subspace of V , $\dim V \neq \infty$. Let $\beta = \{u_1, u_2, \dots, u_k\}$ be a basis of W . Let $\beta' = \{u_1, u_2, \dots, u_k, u_{k+1}, \dots, u_n\}$ be a basis of V . Show $\gamma = \{u_{k+1} + W, u_{k+2} + W, \dots, u_n + W\}$ is a basis for V/W , and $\dim(V/W) = \dim V - \dim W$.

Proof. Independence: if $\sum_{i=k+1}^n a_i(u_i + W) = 0 + W \iff \sum_{i=k+1}^n a_i u_i = w \in W = \text{span } \beta \implies \beta'$ is dependent. \ast

Generacy: Let $\alpha = \beta' - \beta = \{u_{k+1}, \dots, u_n\}$, and $W' = \text{span } \alpha$. By **Problem 1.6.33**, we see $V = W \oplus W'$.

So $\forall v \in V, \exists! w \in W, w' \in W'$ s.t. $v = w' + w \in w' + W$. $\overline{w'} \in \text{span } \alpha \implies w' + W \in \text{span } \gamma$.

After constructing a basis, it's trivial that the dimension relation is true. \blacksquare

Problem 2.1.40. Let W be a subspace of V . Define $\eta : V \rightarrow V/W$ by $\eta(v) = v + W \forall v \in V$.

1. Show η is linear and $N(\eta) = W$.

2. Suppose $\dim V \neq \infty$. Show again that $\dim(V/W) = \dim V - \dim W$.

Proof. (1) $\eta(cv + u) = (cv + u) + W = c(v + W) + (u + W) = c\eta(v) + \eta(u)$.

$\forall w \in W, \eta(w) = w + W = 0 + W \implies W \subseteq N(\eta)$.

If $\eta(v) = 0 + W = v + W \implies v - 0 \in W \implies v \in W \implies N(\eta) \subseteq W$. Thus $N(\eta) = W$.

(2) By **Theorem 2.3**, $\dim V = \dim N(\eta) + \dim R(\eta) \implies \dim R(\eta) = \dim V - \dim W$. By definition, $V/W = \{v + W \mid v \in V\}$, so it's obvious that $R(\eta) = V/W \implies \dim(V/W) = \dim V - \dim W$. \blacksquare

Problem 2.4.24. Let $T : V \rightarrow Z$ linear and onto. Define $\bar{T} : V/N(T) \rightarrow Z$ by $\bar{T}(v + N(T)) = T(v), \forall v + N(T) \in V/N(T)$.

1. Show \bar{T} is well-defined. i.e. $v + N(T) = v' + N(T) \implies T(v) = T(v')$.

2. Show \bar{T} is linear.

3. Show \bar{T} is an isomorphism.

4. Show $T = \bar{T}\eta$, where $\eta : V \rightarrow V/W$ by $\eta(v) = v + N(T) \forall v \in V$.

Proof. (1) $v + N(T) = v' + N(T) \iff v - v' \in N(T)$. Let $v - v' = u$. Then $T(v - v') = T(u) = 0 \implies T(v) = T(v')$.

(2) $\bar{T}(c(v + N(T)) + (u + N(T))) = \bar{T}((cv + u) + N(T)) = T(cv + u) = cT(v) + T(u) = c\bar{T}(v + N(T)) + \bar{T}(u + N(T))$.

(3) If $\bar{T}(v + N(T)) = T(v) = 0 \implies v \in N(T)$. But in this case $v + N(T) = 0 + N(T)$. So $N(\bar{T}) = 0$. Given T is onto, it is easy to see that $\dim Z = \dim R(T) = \dim V/N(T)$. So \bar{T} one-to-one $\iff \bar{T}$ onto, so \bar{T} is an isomorphism.

(4) Let $V = W \oplus N(T)$. Then $\forall x \in V, x = w + v, w \in W, v \in N(T)$. Then $T(x) = T(w + v) = T(w)$. But $\bar{T}\eta(x) = \bar{T}((w + v) + N(T)) = \bar{T}(w + N(T) + v + N(T))$. But $v + N(T) = 0 + N(T)$ so $\bar{T}\eta(x) = \bar{T}(w + N(T)) = T(w)$. \square

1.7 Direct Sum

1.7.1 Definition

Definition 1.11 (Sum of Sets). if $S_1, \dots, S_k \neq \emptyset, S_i \subseteq V, 1 \leq i \leq k$, where V is a vector space. Then define

$$\sum_{i=1}^k S_i := \left\{ \sum_{i=1}^k x_i \mid x_i \in S_i \right\}.$$

Definition 1.12 (Direct Sum). We say $V = \bigoplus_{i=1}^k W_i$ (V is direct sum of $W_i, 1 \leq i \leq k$), if W_i are subspaces of V s.t.

$$V = \sum_{i=1}^k W_i$$

and

$$W_j \cap \sum_{i \neq j} W_i = \{0\} \text{ for all } 1 \leq j \leq k.$$

Problem 1.3.23. Let W_1 and W_2 be subspaces of V . Show

1. $W_1 + W_2$ is a subspace of V containing both W_1 and W_2 .
2. Any subspace U of V s.t. $W_1, W_2 \subseteq U$ must have $W_1 + W_2 \subseteq U$.

Proof.

1. 顯然 $W_1, W_2 \subseteq W_1 + W_2$, 且 $0 \in W_1 + W_2$. 因此只需證明 $W_1 + W_2$ 為 subspace.
 $\forall u, v \in W_1 + W_2$, 令 $u = x + y, v = z + w$, 其中 $x, z \in W_1, y, w \in W_2$.
 則 $u + v = (x + y) + (z + w) = (x + z) + (y + w)$.
 而因為 $x + z \in W_1, y + w \in W_2$, 故 $u + v \in W_1 + W_2$.

2. 顯然.



1.7.2 Properties

Theorem 1.12. *Let $W_i, 1 \leq i \leq k$ be subspaces of V , $\dim V = n$. Then the following statements are equivalent.*

1. $V = \bigoplus_{i=1}^k W_i$.
2. $V = \sum_{i=1}^k W_i$ and for all $v_i \in W_i, 1 \leq i \leq k$, if $\sum_{i=1}^k v_i = 0$, then $v_i = 0 \forall i$.
3. $\forall v \in V, \exists! v_i \in W_i$ s.t. $v = \sum_{i=1}^k v_i$.
4. If γ_i is an ordered basis of $W_i, 1 \leq i \leq k$, then $\gamma = \bigcup_{i=1}^k \gamma_i$ is an ordered basis of V .
5. For each $1 \leq i \leq k, \exists \gamma_i$ an ordered basis of W_i s.t. $\gamma = \bigcup_{i=1}^k \gamma_i$ is an ordered basis of V .

Proof. We prove the theorem in the sequence $1 \iff 2, 2 \iff 3, 2 \iff 4, 4 \implies 5, 5 \implies 1$.

(1 \implies 2) $V = \sum_{i=1}^k W_i$ by definition. With out loss of generality rearrange v_i s.t. $v_i \neq 0, 1 \leq i \leq m, v_i = 0, m+1 \leq i \leq k$. Now this means $v_m \in \text{span}(\{v_1, \dots, v_{m-1}\})$, so $v_m \in W_m \cap \sum_{i \neq m}^k W_i, v_m \neq 0$. *

(2 \implies 1) With out loss of generality, rearrange W_i s.t. Suppose $v_1 \in W_1 \cap \sum_{i=2}^k W_i, v_1 \neq 0$, which is a subspace. This means $v_1 = \sum_{i=2}^k v_i, v_i \in W_i$. Now $(-v_1) \in W_1 \cap \sum_{i=2}^k W_i$, too, so $\sum_{i=1}^k v_i = 0$, yet $v_1 \neq 0$.

(2 \implies 3) Suppose $\exists v_i, v'_i \in W_i$ s.t. $v = \sum_{i=1}^k v_i = \sum_{i=1}^k v'_i$. So $\sum_{i=1}^k (v_i - v'_i) = 0$. By (2), we see $v_i = v'_i$, therefore the expression is unique.

(3 \implies 2) Suppose $\exists v'_i$ not all zero, $v'_i \in W_i$ s.t. $\sum_{i=1}^k v'_i = 0$. So if $v = \sum_{i=1}^k v_i, v = \sum_{i=1}^k (v_i + v'_i)$ also. Since v'_i are not all zero, the expression is not unique. *

(2 \Rightarrow 4) By similar proof of **Theorem 6.4**, we see $\gamma = \bigcup_{i=1}^k \gamma_i$ is independent. Now $V = \sum_{i=1}^k W_i$, this means $\forall v \in V, \exists w_i \in W_i$ s.t. $v = \sum_{i=1}^k w_i$. Since γ_i is a basis of W_i , $w_i \in \text{span } \gamma_i \Rightarrow v \in \text{span } \gamma \Rightarrow V = \text{span } \gamma$. So γ is a basis of V .

(4 \Rightarrow 2) Since $\gamma = \bigcup_{i=1}^k \gamma_i$ is a basis of V , $\forall v \in V, \exists v_{ij} \in \gamma_i$ s.t. $v = \sum_{i=1}^k \sum_j a_{ij} v_{ij} = \sum_{i=1}^k w_i, w_i = \sum_j a_{ij} v_{ij} \in W_i$. Thus $V = \sum_{i=1}^k W_i$. Now consider $\sum_{i=1}^k w_i = 0, w_i \in W_i, \forall w_i, 1 \leq i \leq k, \exists v'_{ij} \in \gamma_i$ s.t. $w_i = \sum_j a'_{ij} v'_{ij}$. Since γ is independent, $a'_{ij} = 0 \forall i, j$, so $w_i = 0 \forall i$.

(4 \Rightarrow 5) Trivial.

(5 \Rightarrow 1) Clearly $V = \sum_{i=1}^k W_i$.

Now suppose $\exists v_1 \neq 0$ s.t. $v_1 \in W_1 \cap \sum_{i=2}^k W_i$. So $v_1 = \sum_j a_{1j} u_{1j}, u_{1j} \in \gamma_1$. Also, $v = \sum_{i=2}^k \sum_j a_{ij} u_{ij}, u_{ij} \in \gamma_i$. Now $0 = \sum_j a_{1j} u_{1j} - \sum_{i=2}^k \sum_j a_{ij} u_{ij}$. But since γ is independent, $a_{ij} = 0, 1 \leq i \leq k$. \times

Problem 1.3.30. Given W_1, W_2 both subspaces of V , show $V = W_1 \oplus W_2 \iff \forall v \in V, \exists! x_1 \in W_1, x_2 \in W_2$ s.t. $v = x_1 + x_2$.

Proof.

(\Rightarrow) $V = W_1 \oplus W_2 \Rightarrow \forall v \in V, \exists x_1 \in W_1, x_2 \in W_2$ s.t. $v = x_1 + x_2$, and $W_1 \cap W_2 = \{0\}$.

Suppose $\exists x'_1 \in W_1, x'_2 \in W_2$ and $x'_1 \neq x_1, x'_2 \neq x_2$ s.t. $v = x'_1 + x'_2$.

則 $0 = (x_1 + x_2) - (x'_1 + x'_2) = (x_1 - x'_1) + (x_2 - x'_2)$.

因為 $x'_1 \neq x_1, x'_2 \neq x_2$, 故 $(x_1 - x'_1) = -(x_2 - x'_2) \neq 0$.

但 $x_1, x'_1 \in W_1, x_2, x'_2 \in W_2$, 又 W_1, W_2 皆為 subspaces, 故 $0 \neq (x_1 - x'_1) = -(x_2 - x'_2) \in W_1 \cap W_2 \times$

(\Leftarrow) $\forall v \in V, \exists! x_1 \in W_1, x_2 \in W_2$ s.t. $v = x_1 + x_2 \in W_1 + W_2$.

So $V \subseteq W_1 + W_2$, 但 $W_1 + W_2 \subseteq V$ 顯然, 故 $V = W_1 + W_2$.

Suppose $x \in W_1 \cap W_2$ 且 $x \neq 0$, then $\forall v \in V, v = x_1 + x_2 = (x_1 + x) + (x_2 - x)$ 不唯一. \times

Problem 1.6.29.1. Show if W_1, W_2 are subspaces of V , $\dim W_1, \dim W_2 \neq \infty$, then $\dim(W_1 + W_2) \neq \infty$, and $\dim(W_1 + W_2) = \dim W_1 + \dim W_2 - \dim(W_1 \cap W_2)$.

Proof.

(trivial) For the case $W_1 \subseteq W_2$ (or $W_1 \supseteq W_2$), in this case $\dim(W_1 + W_2) = \dim W_2 = \dim W_1 + \dim W_2 - \dim(W_1 \cap W_2)$.

(non-trivial) Let $\beta = \{u_1, u_2, \dots, u_k\}$ be a basis of $W_1 \cap W_2$.

By **Corollary 1.11.1**, β can be extended to form a basis γ for W_1 , $\gamma = \{u_1, \dots, u_k, v_1, \dots, v_p\}$ and a basis δ for W_2 , $\delta = \{u_1, \dots, u_k, w_1, \dots, w_q\}$.

At this point, we claim $\epsilon = \{u_1, \dots, u_k, v_1, \dots, v_p, w_1, \dots, w_q\} = \gamma \cup \delta$ is a basis of $W_1 + W_2$.

First of all, ϵ is independent. Given β, γ, δ all independent, we now show $\gamma \cup \{w_i\}$ and $\delta \cup \{v_i\}$ are independent.

If, say, $\gamma \cup \{w_i\}$ is dependent, that is equivalent to $w_i \in \text{span } \gamma = W_1$ yet $w_i \in W_2$.

So $w_i \in W_1 \cap W_2 = \text{span } \beta \iff \delta$ is dependent. \times

Same reasoning shows $\delta \cup \{v_i\}$ is independent.

Now $\text{span } \epsilon = \text{span}(\gamma \cup \delta) = \text{span } \gamma + \text{span } \delta$ by **Problem 1.4.14**. 故 $\text{span } \epsilon = W_1 + W_2$.
故 $\dim(W_1 + W_2) = |\epsilon| = k + p + q = \dim W_1 + \dim W_2 - \dim(W_1 \cap W_2)$. \square

Problem 1.6.29.2. Let $V = W_1 + W_2$, W_1, W_2 are subspaces. Show $V = W_1 \oplus W_2 \iff \dim V = \dim W_1 + \dim W_2$.

Proof.

$(\Rightarrow) V = W_1 \oplus W_2 \iff V = W_1 + W_2, W_1 \cap W_2 = \{0\}. \implies \dim V = \dim(W_1 + W_2) = \dim W_1 + \dim W_2 - \dim\{0\} = \dim W_1 + \dim W_2$.

(\Leftarrow) 已知 $V = W_1 + W_2$ 及 $\dim V = \dim W_1 + \dim W_2 = \dim(W_1 + W_2) \implies \dim(W_1 \cap W_2) = 0 \implies W_1 \cap W_2 = \{0\}$. \square

Problem 1.6.33. Let β_1, β_2 are bases of W_1, W_2 , respectively. If $V = W_1 \oplus W_2$, show $\beta_1 \cap \beta_2 = \emptyset$, and $\beta_1 \cup \beta_2$ is a basis of V .

Also prove the converse, i.e. Let $\beta_1 \cap \beta_2 = \emptyset$, $\beta_1 \cup \beta_2$ is a basis of V , show $V = W_1 \oplus W_2$.

Remark. I want to first say something about the requirement $\beta_1 \cap \beta_2 = \emptyset$. From the theorem, we see it promises that $W_1 \cap W_2 = \{0\}$. We now rule out the possibility that $\exists v \in \beta_1, w \in \beta_2$ s.t. $v = cw$, and $\exists u \in W_1 \cap W_2, u \neq 0$. If this is the case, then $\beta_1 \cup \beta_2$ would not be a basis. So this could not be true. \square

Proof.

(\Rightarrow) First of all, $\beta_1 \cup \beta_2$ is independent.

If it's not, then suppose $\exists v \in \beta_2 \subseteq W_2$ s.t. $v \in \text{span } \beta_1 = W_1 \implies v \in W_1 \cap W_2$.

But $v \neq 0$, for if so, β_2 would be dependent. So $\exists v \neq 0$ s.t. $v \in W_1 \cap W_2$. Similar reasoning can be done for $v \in \beta_1$. \times

Moreover, $\text{span}(\beta_1 \cup \beta_2) = \text{span } \beta_1 + \text{span } \beta_2 = W_1 + W_2 = V$, by **Problem 1.4.14**. So $\beta_1 \cup \beta_2$ is indeed a basis.

Secondly, $\beta_1 \cap \beta_2 = \emptyset$. For if $0 \in \beta_1 \cap \beta_2$, then β_1, β_2 dependent. If $v \neq 0, v \in \beta_1 \cap \beta_2 \implies v \in W_1 \cap W_2$. \times

(\Leftarrow) We would show this by ruling out the possibilities of $\exists v \in W_1 \cap W_2, v \neq 0$ and $V \neq W_1 + W_2$.

Suppose $\exists v \neq 0, v \in W_1 \cap W_2$.

Then $v = \sum_i a_i u_i, u_i \in \beta_1, v = \sum_j b_j w_j, w_j \in \beta_2$.

Since $\beta_1 \cup \beta_2$ is independent, $0 = \sum_i a_i u_i - \sum_j b_j w_j$ must have all zero solution. But $v \neq 0$, so a_i, b_j not all zero. In this way we see non-zero a_i, b_j corresponds to $u_i = w_j$, which leads to $\beta_1 \cap \beta_2 \neq \emptyset$. \times

Now $\text{span}(\beta_1 \cup \beta_2) = V = \text{span } \beta_1 + \text{span } \beta_2 = W_1 + W_2$ by **Problem 1.4.14**. \square

Problem 1.6.34. Show $\forall W_1$ is a subspace of V , $\exists W_2$ also a subspace s.t. $V = W_1 \oplus W_2$. $\dim V \neq \infty$.

Proof.

(trivial) If $W_1 = V$, then choose $W_2 = \emptyset$.

(non-trivial) Let $\dim V = n, G$ 為 V 的一個 basis. Let $\dim W_1 = n, \beta$ 為 W_1 的一個 basis.

By **Theorem 1.10**, $\exists \beta' \subseteq G, |\beta'| = n - m$ s.t. $\beta \cup \beta'$ spans V . Let $W_2 = \text{span } \beta'$.

Firstly, $\beta' \subseteq G$, which is independent. Then by **Corollary 1.6.1**, β' is independent.

Secondly, $\beta \cap \beta' = \emptyset$. If not, then $|\beta \cap \beta'| = |\beta| + |\beta'| - |\beta \cup \beta'| < m + (n - m) = n = \dim V$, which is impossible by **Corollary 1.10.2** (1).

Since $|\beta| = m, |\beta'| = n - m, \beta \cap \beta' = \emptyset \implies |\beta \cup \beta'| = n$, by **Corollary 1.10.2** (1), we have $\beta \cup \beta'$ is a basis of V . By **Problem 1.6.33**, we have $V = W_1 \oplus W_2$. \square

Problem 2.1.35. Let $T : V \rightarrow V$ linear, $\dim V \neq \infty$.

1. Suppose $V = R(T) + N(T)$. Show $V = R(T) \oplus N(T)$.

Also show this is not true if $\dim V = \infty$.

2. Suppose $R(T) \cap N(T) = \{0\}$. Show $V = R(T) \oplus N(T)$.

Also show this is not true if $\dim V = \infty$.

Proof.

1. By **Theorem 2.3**, $\dim V = \dim R(T) + \dim N(T)$. By **Problem 1.6.29.1**, $\dim V = \dim(R(T) + N(T)) = \dim R(T) + \dim N(T) - \dim(R(T) \cap N(T)) \implies R(T) \cap N(T) = \{0\}$.

2. By similar constructions in **Theorem 2.3**, we see $\beta_r \cap \beta_n = \emptyset, \beta_r \cup \beta_n$ is a basis of $V \implies V = R(T) \oplus N(T)$ by **Problem 1.6.33**.

\square

1.7.3 Examples

Definition 1.13 (Skew-symmetric). A matrix M is called skew-symmetric if $M^\top = -M$.

Problem 1.3.28. Let F be a field, $\text{char } F \neq 2$. Prove the set W_1 of all $n \times n$ skew-symmetric matrices with entries from F is a subspace of $M_{n \times n}(F)$. Let W_2 be the set of all symmetric matrices with entries from F . Prove that $M_{n \times n}(F) = W_1 \oplus W_2$.

Proof. $M_{n \times n} = \{ (A + A^\top) + (A - A^\top) \mid A \in M_{n \times n} \}$.

But $(A + A^\top)^\top = A^\top + A$ is symmetric, and

$(A - A^\top)^\top = -(A - A^\top)$ is skew-symmetric.

又若 $(A + A^\top) = (A - A^\top)$, $A = O_{n \times n}$, 故得證。 ▣

2 Linear Transformations

2.1 Linear Transformations

2.1.1 Definition

Definition 2.1 (Linear Transformation). Let V and W be vector spaces over field F . We say $T : V \rightarrow W$ is a linear transformation from V to W if $\forall x, y \in V, c \in F$,

1. $T(x + y) = T(x) + T(y)$
2. $T(cx) = cT(x)$

Definition 2.2 (Identity Transformation). $I : V \rightarrow V$ by $I(x) = x, \forall x \in V$.

Definition 2.3 (Zero Transformation). $T_0 : V \rightarrow W$ by $T_0(x) = 0, \forall x \in V$.

Problem 2.1.7. Show some immediate consequences of the definitions of linear transformations.

1. If T is linear, then $T(0) = 0$.
2. T is linear $\iff T(cx + y) = cT(x) + T(y) \forall x, y \in V, c \in F$.
3. T is linear $\implies T(x - y) = T(x) - T(y)$.
4. T is linear $\iff T(\sum_{i=1}^n a_i x_i) = \sum_{i=1}^n a_i T(x_i), \forall x_i \in V, a_i \in F$.

Proof.

1. $T(0 + x) = T(0) + T(x) = T(x)$. By **Corollary 1.1.1**, $T(0) = 0$.
2. $(\Rightarrow) T(cx + y) = T(cx) + T(y) = cT(x) + T(y)$.
 (\Leftarrow) Let $c = 1$, we have $T(x + y) = T(x) + T(y) \Rightarrow T(0) = 0$.
 So let c vary, we have $T(cx + 0) = T(cx) = cT(x) + T(0) = cT(x)$.
3. $T(x - y) = T(x + (-y)) = T(x) + T(-y) = T(x) - T(y)$.
4. (\Rightarrow) Trivial. (\Leftarrow) Let $n = 2, a_1 = a_2 = 1 \Rightarrow T(x + y) = T(x) + T(y)$. Let $n = 1, a_1 \in F \Rightarrow T(cx) = cT(x)$.

▣

2.1.2 Kernel, Image and their Properties

Definition 2.4 (Kernel and Image). Let $T : V \rightarrow W$ linear. The kernel (null space) $N(T) = \{x \in V \mid T(x) = 0\}$. The image (range) $R(T) = \{T(x) \mid x \in V\}$.

Definition 2.5. Let $T : V \rightarrow W$ linear. If $\dim N(T), \dim R(T) \neq \infty$, we call $\text{nullity}(T) = \dim N(T)$, and $\text{rank}(T) = \dim R(T)$.

Definition 2.6. We say a function $f : A \rightarrow B$ is one-to-one if $\forall x, y \in A, f(x) = f(y) \Rightarrow x = y$. A function is onto if $R(f) = B$, denoted by $f(A) = B$.

Theorem 2.1. Let $T : V \rightarrow W$ linear. Then $N(T)$ is a subspace of V , and $R(T)$ is a subspace of W .

Proof. (Kernel) By **Problem 2.1.7**, $T(0) = 0$, so $0 \in N(T)$.

$\forall x, y \in N(T), c \in F, T(x + y) = T(x) + T(y) = 0 \Rightarrow x + y \in N(T)$;

$T(cx) = cT(x) = c \cdot 0 = 0 \Rightarrow cx \in N(T)$.

(Image) $T(0) = 0$, so $0_W \in R(T)$.

$\forall T(x), T(y) \in R(T), T(x) + T(y) = T(x + y)$. Since V is a vector space, $x + y \in V$, so $T(x + y) \in R(T)$.

By similar reasoning $cT(x) = T(cx) \in R(T)$. ▣

Theorem 2.2. Let $T : V \rightarrow W$ linear. If $\beta = \{v_1, v_2, \dots, v_n\}$ is a basis of V , then $R(T) = \text{span}(T(\beta)) = \text{span}(\{T(v_1), T(v_2), \dots, T(v_n)\})$.

Proof.

$(\subseteq) \forall y \in R(T), \exists x \in V$ s.t. $y = T(x)$.

Let $x = \sum_{i=1}^n a_i v_i, v_i \in \beta$. So $y = T(\sum_{i=1}^n a_i v_i) = \sum_{i=1}^n a_i T(v_i) \in \text{span } T(\beta)$. So

$R(T) \subseteq \text{span } T(\beta)$.

(\supseteq) It is trivial that $T(\beta) \subseteq R(T)$. By **Theorem 2.1**, $R(T)$ is a subspace. So by **Theorem 1.5**, $\text{span } T(\beta) \subseteq R(T)$. Thus $R(T) = \text{span } T(\beta)$. \square

Theorem 2.3 (Dimension Theorem). *Let $T : V \rightarrow W$ linear. If $\dim V \neq \infty$, then $\dim N(T) + \dim R(T) = \dim V$.*

Proof. Let $\beta_n = \{n_1, n_2, \dots, n_p\}$ be a basis of $N(T)$. Since by **Theorem 2.1**, $N(T)$ is a subspace of V , by **Corollary 1.11.1**, β_n can be extended to be $\beta = \{n_1, \dots, n_p, r_1, \dots, r_q\}$, β a basis of V .

By **Theorem 2.2**, $R(T) = \text{span } T(\beta)$. However, $T(n_i) = 0, i = 1, 2, \dots, p$, so removing $T(n_i)$ from β doesn't change the span.

So $R(T) = \text{span } T(\beta_r), \beta_r = \{r_1, \dots, r_q\}$. Since $\beta_r \subseteq \beta$, by **Corollary 1.6.1**, β_r is independent. So β_r is a basis of $R(T)$.

Now $\dim N(T) = p, \dim R(T) = q, \dim V = p + q$. \square

Theorem 2.4. *Let $T : V \rightarrow W$ linear. Then T is one-to-one $\iff N(T) = \{0\}$.*

Proof.

(\Leftarrow) $N(T) = \{0\}$ means if $T(x) = 0 \implies x = 0$. Then $T(x) = T(y) \implies T(x - y) = 0 \implies x - y = 0 \implies x = y$.

(\Rightarrow) $T(x) = T(y) \implies x = y$. Suppose $\exists v \neq 0, v \in N(T)$. This means $T(v) = 0$. But $T(0) = 0$. So $T(v) = T(0)$ yet $v \neq 0$, so we've found a counterexample. \times \square

Theorem 2.5. *Let $T : V \rightarrow W$ linear, and $\dim V = \dim W$. Then the following statements are equivalent:*

(a) T is one-to-one \iff (b) T is onto \iff (c) $\text{rank}(T) = \dim V$.

Proof. By **Theorem 2.4**, T is one-to-one $\iff N(T) = \{0\} \iff \dim N(T) = 0 \iff \dim V = \dim N(T) + \dim R(T)$.

((b) \Rightarrow (c)) T is onto $\implies R(T) = W \implies \dim R(T) = \dim W = \dim V$.

((b) \Leftarrow (c)) $\dim R(T) = \dim V = \dim W$. By **Theorem 1.11**, $R(T) = W \implies T$ is onto. \square

\nearrow **Remark.** If $\dim V = \infty$, $T : V \rightarrow V$ linear, one to one is not equivalent to onto. \square

Theorem 2.6. *Let V, W be vector spaces over field F , and $\{v_1, v_2, \dots, v_n\}$ is a basis of V . For $w_1, w_2, \dots, w_n \in W$, $\exists! T : V \rightarrow W$ linear s.t. $T(v_i) = w_i$ for $i = 1, 2, \dots, n$.*

Proof. (existence) Given $v = \sum_{i=1}^n a_i v_i$, define $T(v) = \sum_{i=1}^n a_i w_i$.
 T is linear, since for $u, v \in V, k \in F, u = \sum_{i=1}^n b_i v_i, v = \sum_{i=1}^n c_i v_i$, $T(ku + v) = T(\sum_{i=1}^n (kb_i + c_i)v_i) = \sum_{i=1}^n (kb_i + c_i)w_i = kT(u) + T(v)$.
(uniqueness) Suppose $\exists U : V \rightarrow W$ linear s.t. $U(v_i) = w_i$.
We say $U = T$ if $\forall v \in V, U(v) = T(v)$. We now say this is the case.
 $\forall v \in V, v = \sum_{i=1}^n a_i v_i, U(v) = \sum_{i=1}^n a_i U(v_i) = \sum_{i=1}^n a_i w_i = T(v)$. \square

Corollary 2.6.1. Suppose V has finite basis $\{v_1, v_2, \dots, v_n\}$. If $U, T : V \rightarrow W$ linear, $U(v_i) = T(v_i) \forall i = 1, 2, \dots, n$, then $U = T$.

Problem 2.1.14.a. Show T is one-to-one $\iff \forall S \subseteq V, S$ independent, $T(S) \subseteq W, T(S)$ independent.

Proof.

$(\Rightarrow) N(T) = \{0\}$. Suppose $\exists S$ independent s.t. $T(S)$ dependent. Then $\exists a_i$ not all zero, $\sum_i a_i T(u_i) = 0$.
Since $N(T) = \{0\} \implies \sum_i a_i u_i = 0$.
 (\Leftarrow) Suppose $\exists v \neq 0, v \in N(T)$. Then 取 $S = \{v\}, S$ independent obviously. Yet $T(S) = \{0\}$ is dependent. \times \square

Problem 2.1.14.b. Suppose T one-to-one, $S \subseteq V$. Show S independent $\iff T(S)$ independent.

Proof. S dependent $\iff \exists a_i$ not all zero, $u_i \in S$ s.t. $\sum_i a_i u_i = 0 \iff T(\sum_i a_i u_i) = 0$ given one-to-one.
 $\iff \sum_i a_i T(u_i) = 0 \iff T(S)$ dependent. \square

Problem 2.1.14.c. Let $\beta = \{v_1, v_2, \dots, v_n\}$ a basis of V , T is one-to-one and onto. Show $T(\beta) = \{T(v_1), T(v_2), \dots, T(v_n)\}$ is a basis of W .

Proof. Since T is one-to-one and β independent, by (b), $T(\beta)$ is also independent. Also, T is onto $\implies R(T) = W$. By **Theorem 2.2**, $R(T) = \text{span } T(\beta) = W$. \square

Problem 2.1.17. Let $T : V \rightarrow W$ be linear. Show

1. If $\dim V < \dim W \implies T$ cannot be onto.
2. If $\dim V > \dim W \implies T$ cannot be one-to-one.

Proof.

1. **Theorem 2.3** says $\dim N(T) + \dim R(T) = \dim V$.
If T is onto $\implies R(T) = W \implies \dim R(T) = \dim W \implies \dim N(T) = \dim V - \dim W < 0$. ✖
2. If T is one-to-one, $N(T) = \{0\}$ by **Theorem 2.4**.
By **Theorem 2.3**, $\dim V = \dim R(T) > \dim W$. But since $R(T)$ is a subspace of W by **Theorem 2.1**, and by **Theorem 1.11**, this cannot be true. ✖

▣

Problem 2.3.12. Let $T : V \rightarrow W, U : W \rightarrow Z$ linear.

1. Show if UT is one-to-one, then T is also one-to-one, and $R(T) \cap N(U) = \{0\}$.
2. Show if UT is onto, then U is onto, and $U(R(T)) = Z$.
3. Show if U, T are both one-to-one and onto, then UT is also.

Proof.

1. $N(UT) = \{0\}$. Since $U(0) = 0$, if $\exists x \neq 0, x \in N(T)$, then $UT(x) = U(0) = 0$.
✖
For U , only need $R(T) \cap N(U) = \{0\}$. If not, then we would have $x \neq 0, T(x) \neq 0$ yet $UT(x) = 0, x \neq 0$. ✖
2. $R(UT) = Z$. Suppose $R(U) \subset Z, \exists z \in Z$ s.t. $z \neq U(w) \forall w \in W$. Now $R(T) \subseteq W$, so $z \notin R(UT)$. ✖
3. If $UT(x) = UT(y) \implies U(T(x)) = U(T(y)) \implies T(x) = T(y) \implies x = y$.
 $R(UT) = U(R(T)) = U(W) = Z$.

▣

2.1.3 Space of Linear Transformations

Definition 2.7. Let $T, U : V \rightarrow W, V, W$ are vector spaces over $F, a \in F$. Define $T + U : V \rightarrow W$ by $(T + U)(x) = T(x) + U(x) \forall x \in V$, $aT : V \rightarrow W$ by $(aT)(x) = aT(x) \forall x \in V$.

Theorem 2.7. Let $T, U : V \rightarrow W$ be linear. Show $\forall a \in F, aT + U$ is linear. Furthermore, the collection of all linear transformations $V \rightarrow W$ is a vector space over F .

Proof. $(aT+U)(cx+y) = aT(cx+y) + U(cx+y) = a(cT(x) + T(y)) + cU(x) + U(y) = c(aT(x) + U(x)) + aT(y) + U(y) = c((aT+U)(x)) + (aT+U)(y).$

We note that T_0 is linear, and $aT+U$ is also linear. The rest is trivial. \square

Definition 2.8. We denote the vector space of all linear transformations $T : V \rightarrow W$ by $\mathcal{L}(V, W)$. In the case of $V = W$, we wrote $\mathcal{L}(V)$.

2.1.4 Composition and Inverse

Theorem 2.8. Let $T : V \rightarrow W, U : W \rightarrow Z$ be linear. Then $UT : V \rightarrow Z$ is also linear.

Proof. $UT(cx+y) = U(T(cx+y)) = U(cT(x) + T(y)) = cUT(x) + UT(y).$ \square

Theorem 2.9. Let $T, U_1, U_2 \in \mathcal{L}(V)$. Then

1. $T(U_1 + U_2) = TU_1 + TU_2, (U_1 + U_2)T = U_1T + U_2T.$
2. $T(U_1U_2) = (TU_1)U_2.$
3. $TI = IT = T.$
4. $a(U_1U_2) = (aU_1)U_2 = U_1(aU_2) \forall a \in F.$

Proof.

1. $(T(U_1+U_2))(x) = T((U_1+U_2)(x)) = T(U_1(x)+U_2(x)) = T(U_1(x))+T(U_2(x)) = TU_1(x) + TU_2(x) \forall x \implies T(U_1+U_2) = TU_1 + TU_2.$
 $((U_1+U_2)T)(x) = (U_1+U_2)(T(x)) = U_1(T(x)) + U_2(T(x)) = U_1T(x) + U_2T(x).$
2. Trivial.
3. $T(I(x)) = T(x) = I(T(x)).$
4. $a(U_1U_2)(x) = a(U_1(U_2(x))) = (aU_1)(U_2(x)).$
 $U_1(aU_2(x)) = aU_1(U_2(x))$ by linearity.

\square

Definition 2.9 (Inverse). Let $T : V \rightarrow W$ linear. $U : W \rightarrow V$ is an inverse of T if $TU = I_W, UT = I_V$. If T has an inverse, its inverse is unique, denoted by T^{-1} , and we call T invertible.

✂ *Remark.* $(TU)^{-1} = U^{-1}T^{-1}$; $(T^{-1})^{-1} = T$. \square

✂ *Remark.* Invertibility \iff one-to-one and onto.

If $T : V \rightarrow W$ is linear, $\dim V = \dim W$, then T invertible $\iff \dim R(T) = \dim V$. \square

Theorem 2.10. *Let $T : V \rightarrow W$ linear and invertible. Then $T^{-1} : W \rightarrow V$ is also linear.*

Proof. $T^{-1}(T(cx + y)) = cx + y = T^{-1}(cT(x) + T(y)) = cT^{-1}(T(x)) + T^{-1}(T(y))$. \square

Problem 2.3.11. *Let $T : V \rightarrow V$ linear. Show $T^2 = T_0 \iff R(T) \subseteq N(T)$.*

Proof.

(\Leftarrow) Trivial. (\Rightarrow) Suppose $\exists T(x) \in R(T)$ yet $T(x) \notin N(T)$.
Then $T^2(x) = T(T(x)) \neq 0$. \times \square

2.1.5 Examples

Problem 2.1.15. *Let $T : P(\mathbb{R}) \rightarrow P(\mathbb{R})$, $T(f(x)) = \int_0^x f(t) dt$. Prove T is linear and one-to-one, but not onto.*

Proof. $T(cf + g) = \int_0^x (cf + g) dt = c \int_0^x f(t) dt + \int_0^x g(t) dt = cT(f) + T(g)$.

One-to-one: if $T(f) = 0 \implies \int_0^x f(t) dt = 0 \implies f(x) = 0 \implies N(T) = \{0\}$.

Not onto: 取 $h(x) = 1$. $\nexists f$ s.t. $T(f) = h$. \square

Problem 2.1.16. *Let $T : P(\mathbb{R}) \rightarrow P(\mathbb{R})$, $T(f(x)) = f'(x)$. Obviously T is linear. Show T is onto, but not one-to-one.*

Proof. Onto: $\forall g(x) = \sum_{i=0}^n a_i x^i$, $\exists f(x) = \sum_{i=0}^n \frac{a_i}{i+1} x^{i+1}$ s.t. $T(f(x)) = g$.

Not one-to-one: $N(T) = \text{span}(\{1\})$. \square

2.2 Isomorphism of Linear Transformation and Matrices

2.2.1 Foreword

This subsection is rather long, for we are going to establish the isomorphism of (1) vector spaces and F^n , (2) the space of all linear transformations and matrices.

2.2.2 Definition of Isomorphism

Definition 2.10 (Isomorphism). We say two vector spaces V is isomorphic to W if $\exists T : V \rightarrow W$ s.t. T is linear, one-to-one, and onto. In this case, T is called an isomorphism from T onto W .

✍ *Remark.* A relation is bijective or invertible if it's simultaneously one-to-one and onto.

In the context of linear transformations, we add the restriction of linearity to our isomorphisms. \square

Theorem 2.11. *Isomorphism is an equivalence relation. i.e., we have*

1. $V \cong V$ (reflexivity)
2. If $V \cong W$, then $W \cong V$ (symmetry)
3. If $V \cong W$, $W \cong Z$, then $V \cong Z$ (transitivity)

Theorem 2.12. *If $V \cong W$, then $\dim V \neq \infty \iff \dim W \neq \infty$. In this case, $\dim V = \dim W$.*

Proof. Let β be a basis of V , $|\beta| = \dim V = n$, and a linear isomorphism $T : V \rightarrow W$.

Since β is independent, and T is one-to-one, by [Problem 2.1.14.b](#), $T(\beta)$ is independent. So by [Theorem 2.2](#), $T(\beta)$ is a basis of $R(T)$. And T is onto, so $W = R(T)$, so $T(\beta)$ is a basis of W .

Thus $\dim W \neq \infty, \dim V = \dim W$.

The same reasoning applies for T^{-1} , which says if $\dim W \neq \infty, \implies \dim V \neq \infty$. \square

Theorem 2.13. *Let $\dim V, \dim W \neq \infty$, then $V \cong W \iff \dim V = \dim W$.*

Proof.

(\implies) Trivial by [Theorem 2.12](#).

(\impliedby) Let $\beta = \{v_1, \dots, v_n\}, \gamma = \{w_1, \dots, w_n\}$ be the bases of V, W , respectively. We claim the linear transformation $T : V \rightarrow W$ by $T(v_i) = w_i$ is an isomorphism.

Since β is basis of V , then $R(T) = \text{span}(T(\beta))$. Yet $T(\beta) = \gamma$, so $R(T) = \text{span } \gamma = W$. Thus T is onto. By [Theorem 2.5](#), T is also one-to-one. So T is an isomorphism. \square

Problem 2.4.17. Let $\dim V, \dim W \neq \infty$, $T : V \rightarrow W$ an isomorphism, V_0 be a subspace of V .

1. Show $T(V_0)$ is a subspace of W .

2. Show $\dim V_0 = \dim T(V_0)$.

Proof. (1) $\forall x' = T(x), y' = T(y) \in T(V_0), x, y \in V_0, cx' + y' = cT(x) + T(y) = T(cx + y)$. But $cx + y \in V_0 \implies T(cx + y) \in T(V_0)$. Moreover, $T(0) = 0 \in T(V_0)$.

(2) Trivial by **Theorem 2.12**. ▣

2.2.3 Standard representation of Vectors

Definition 2.11. An ordered basis is a basis with an order. That is, changing the orders of vectors in the basis is considered different.

Definition 2.12. We define the set $\{e_1, e_2, \dots, e_n\}$ to be the standard ordered basis of F^n .

Definition 2.13. Let $\beta = \{u_1, u_2, \dots, u_n\}$ be ordered basis of V . Let $x = \sum_{i=1}^n a_i u_i$. We define coordinate of x relative to β , denoted $[x]_\beta$ by

$$[x]_\beta = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}.$$

The standard representation of V with respect to β is the function $\phi_\beta : V \rightarrow F^n$ defined by $\phi_\beta(x) = [x]_\beta, \forall x \in V$.

Theorem 2.14. For all finite dimensional vector space V , β being a basis of it, the function ϕ_β is linear, and is an isomorphism.

Proof. (linearity) $\phi_\beta(cx + y) = [cx + y]_\beta$. Let $x = \sum_{i=1}^n a_i u_i, y = \sum_{i=1}^n b_i u_i$. Then $cx + y = \sum_{i=1}^n (ca_i + b_i)u_i \implies \phi_\beta(cx + y) = c[x]_\beta + [y]_\beta = c\phi_\beta(x) + \phi_\beta(y)$.

(isomorphism) From **Theorem 1.8**, we see $\phi_\beta(x)$ is unique, so $\phi_\beta(x)$ is one-to-one. By **Theorem 2.5**, ϕ_β is also onto. So ϕ_β is an isomorphism by definition. ▣

2.2.4 Matrix Representation of Linear Transformations

Definition 2.14 (Matrix Representation). Suppose $\dim V, \dim W \neq \infty$, $\beta = \{v_1, v_2, \dots, v_n\}$, $\gamma = \{w_1, w_2, \dots, w_m\}$ be their bases, respectively. Let $T : V \rightarrow W$ linear, then for $1 \leq j \leq n$, $\exists a_{ij} \in F$ s.t.

$$T(v_j) = \sum_{i=1}^m a_{ij} w_i, \quad 1 \leq j \leq n$$

We call $A_{ij} = a_{ij}$ the matrix representation of T in the ordered bases β and γ , denoted by $A = [T]_{\beta}^{\gamma}$. If $V = W$ and $\beta = \gamma$, then we write $A = [T]_{\beta}$.

✍ *Remark.* j -th column of A is $[T(v_j)]_{\gamma}$.

If $[U]_{\beta}^{\gamma} = [T]_{\beta}^{\gamma}$, then $U = T$ by **Theorem 2.6**. □

Theorem 2.15. Let V, W be vector spaces with finite dimension, with β, γ be their ordered bases, respectively, and $T, U : V \rightarrow W$ be linear. Then

1. $[T + U]_{\beta}^{\gamma} = [T]_{\beta}^{\gamma} + [U]_{\beta}^{\gamma}$.
2. $[aT]_{\beta}^{\gamma} = a[T]_{\beta}^{\gamma} \quad \forall a \in F$.

Proof. Let $v_j \in \beta, 1 \leq j \leq n, w_i \in \gamma, 1 \leq i \leq m$.

$$\begin{aligned} 1. \quad [T + U]_{\beta}^{\gamma} &\longleftrightarrow (T + U)(v_j) = \sum_{i=1}^m a_{ij} w_i = T(v_j) + U(v_j) = \sum_{i=1}^m b_{ij} w_i + \sum_{i=1}^m c_{ij} w_i \longleftrightarrow [T]_{\beta}^{\gamma} + [U]_{\beta}^{\gamma}. \end{aligned}$$

▣

✍ *Remark.* Now we associate a particular matrix to a linear transformation, and we have shown how to add and scalar multiply. We now want to define matrix multiplication such that the "multiplication" corresponds to compositions of linear transformations.

$A = [U]_{\beta}^{\gamma}, B = [T]_{\alpha}^{\beta}, \alpha = \{v_1, \dots, v_n\}, \beta = \{w_1, \dots, w_m\}, \gamma = \{z_1, \dots, z_p\}$ be bases of V, W, Z , respectively. For $1 \leq j \leq n$,

$$\begin{aligned} (UT)(v_j) &= U(T(v_j)) = U\left(\sum_{k=1}^m B_{kj} w_k\right) = \sum_{k=1}^m B_{kj} U(w_k) \\ &= \sum_{k=1}^m B_{kj} \left(\sum_{i=1}^p A_{ik} z_i\right) = \sum_{i=1}^p \left(\sum_{k=1}^m A_{ik} B_{kj}\right) z_i = \sum_{i=1}^p C_{ij} z_i. \end{aligned}$$

□

Definition 2.15. Let A be $m \times n$, B be $n \times p$. Then AB is $m \times p$, $(AB)_{ij} = \sum_{k=1}^n A_{ik}B_{kj}$, $1 \leq i \leq m, 1 \leq j \leq p$.

Remark. $(AB)_{ij}^\top = (AB)_{ji} = \sum_{k=1}^n A_{jk}B_{ki}$; $(B^\top A^\top)_{ij} = \sum_{k=1}^n (B^\top)_{ik}(A^\top)_{kj} = \sum_{k=1}^n B_{ki}A_{jk}$. \square

Theorem 2.16. Let $\dim V, \dim W, \dim Z \neq \infty$, their bases be α, β, γ , respectively. $T : V \rightarrow W, U : W \rightarrow Z$ linear. Then $[UT]_\alpha^\gamma = [U]_\beta^\gamma [T]_\alpha^\beta$.

Corollary 2.16.1. Let $\dim V \neq \infty, \beta$ is a basis, and $T, U \in \mathcal{L}(V)$. Then $[UT]_\beta = [U]_\beta [T]_\beta$.

Theorem 2.17. Let $\dim V, \dim W \neq \infty$, β, γ are bases of V, W , respectively. Let $T : V \rightarrow W$ linear. Then $\forall u \in V, [T(u)]_\gamma = [T]_\beta^\gamma [u]_\beta$.

Proof. Suppose $\beta = \{v_1, \dots, v_n\}, \gamma = \{w_1, \dots, w_m\}$. Let $u = \sum_{i=1}^n b_i v_i$, $T(v_j) = \sum_{i=1}^m a_{ij} w_i$, then regarding u as an $n \times 1$ matrix, $[T]_\beta^\gamma [u]_\beta = \sum_{k=1}^n a_{ik} b_k$.

$$\text{And } T(u) = \sum_{j=1}^n b_j T(v_j) = \sum_{j=1}^n b_j \sum_{i=1}^m a_{ij} w_i = \sum_{i=1}^m (\sum_{j=1}^n a_{ij} b_j) w_i.$$

$$\text{Thus } ([T(u)]_\gamma)_i = \sum_{j=1}^n a_{ij} b_j = ([T]_\beta^\gamma [u]_\beta)_i. \quad \blacksquare$$

Definition 2.16. Let A be a $n \times n$ matrix. Then A is invertible if $\exists B$ $n \times n$ s.t. $AB = BA = I$.

Theorem 2.18. Let $T : V \rightarrow W$ linear, β, γ are bases of V, W , respectively. Then T is invertible $\iff [T]_\beta^\gamma$ is invertible. In particular, $[T^{-1}]_\gamma^\beta = ([T]_\beta^\gamma)^{-1}$.

Proof.

(\implies) By [Theorem 2.12](#), we know $[T]_\beta^\gamma$ is $n \times n$, $\dim V = n = \dim W$. T is invertible means $\exists T^{-1}$ s.t. $TT^{-1} = I_W, T^{-1}T = I_V$.

By [Theorem 2.16](#), we see $[TT^{-1}]_\gamma = [I_W]_\gamma = I_n = [T]_\beta^\gamma [T^{-1}]_\gamma^\beta$, and $[T^{-1}T]_\beta = [I_V]_\beta = I_n = [T^{-1}]_\gamma^\beta [T]_\beta^\gamma$. So we see $[T^{-1}]_\gamma^\beta [T]_\beta^\gamma = [T]_\beta^\gamma [T^{-1}]_\gamma^\beta = I_n$, so $[T^{-1}]_\gamma^\beta = ([T]_\beta^\gamma)^{-1}$ by uniqueness.

(\impliedby) $[T]_\beta^\gamma$ invertible $\implies [T]_\beta^\gamma$ $n \times n, \exists A$ $n \times n$ s.t. $A[T]_\beta^\gamma = [T]_\beta^\gamma A = I_n$. Let $U = \sum_{i=1}^n a_{ij} v_j, v_j \in \gamma$. Then $A = [U]_\gamma^\beta$. So we see $[U]_\gamma^\beta [T]_\beta^\gamma = [T]_\beta^\gamma [U]_\gamma^\beta = I_n \implies [UT]_\beta = [TU]_\gamma = I_n$. So $UT = I_V, TU = I_W$. So $U = T^{-1}$, and T is invertible. \blacksquare

2.2.5 Properties of Matrix multiplication

Theorem 2.19. Let A, B, C, D, E be matrices such that their multiplication is defined. Then

1. $A(B + C) = AB + AC, (D + E)A = DA + EA.$
2. $a(AB) = (aA)B = A(aB).$
3. $I_m A = A = A I_n.$
4. Let $\dim V = n$, basis β , then $[I_V]_\beta = I_n.$
5. $(AB)C = A(BC).$

Proof. (1)

$$(A(B + C))_{ij} = \sum_{k=1}^n A_{ik}(B + C)_{kj} = \sum_{k=1}^n A_{ik}B_{kj} + \sum_{k=1}^n A_{ik}C_{kj} = (AB)_{ij} + (AC)_{ij}.$$

The other follows similarly.

(2)

$$(a(AB))_{ij} = a \sum_{k=1}^n A_{ik}B_{kj} = \sum_{k=1}^n (aA)_{ik}B_{kj} = \sum_{k=1}^n A_{ik}(aB)_{kj}.$$

(3)

$$(I_m A)_{ij} = \sum_{k=1}^m \delta_{ik}A_{kj} = A_{ij}.$$

(4)

$$I_V(v_j) = v_j = \sum_{k=1}^n \delta_{kj}v_j. \implies (I_n)_{kj} = \delta_{kj}.$$

(5)

$$((AB)C)_{ij} = \sum_{k=1}^p \left(\sum_{\ell=1}^n a_{i\ell}b_{\ell k} \right) c_{kj} = \sum_{\ell=1}^n a_{i\ell} \left(\sum_{k=1}^p b_{\ell k}c_{kj} \right) = (A(BC))_{ij}$$

▣

Theorem 2.20. Let A $m \times n$, B $n \times p$, $1 \leq j \leq p$, u_j, v_j, w_j are the j -th column of AB, B, A , respectively, and $z = \sum_{j=1}^p a_j e_j$. Then,

1. $u_j = Av_j.$
2. $v_j = Be_j.$
3. $Bz = \sum_{j=1}^p a_j v_j.$

$$4. u_j = \sum_{i=1}^n v_{ij} w_i.$$

Proof. (1) $(AB)_{ij} = \sum_{k=1}^n A_{ik} B_{kj}$. Let $B_{kj} = v_{kj}$ then we are done.

$$(2) B'_{ij} = \sum_{k=1}^n B_{ik} e_{kj} = B_{ij}.$$

$$(3) Bz = B(\sum_{j=1}^p a_j e_j) = \sum_{j=1}^p a_j B(e_j). \text{ By (2) we have } = \sum_{j=1}^p a_j v_j.$$

$$(4) \text{ by (1) we have } u_j = Av_j = \sum_{i=1}^n v_{ij} w_i. \quad \blacksquare$$

✍ *Remark.* When A left multiplies x , then the j -th column of Ax is a linear combination of the columns of A , with coefficients from x .

Column j of AB is linear combination of columns of A with coefficients from column j of B .

Row i of AB is linear combination of rows of B with coefficients from row i of A .
 \square

Problem 2.3.13. Given A, B $n \times n$, show $\text{tr}(AB) = \text{tr}(BA)$, $\text{tr}(A) = \text{tr}(A^\top)$.

Proof. $\text{tr}(A) = \text{tr}(A^\top)$ is trivial.

$$\text{tr}(BA) = \sum_{k=1}^n (\sum_{i=1}^n B_{ki} A_{ik}); \text{tr}(AB) = \sum_{k=1}^n (\sum_{i=1}^n A_{ki} B_{ik}).$$

Switching the order of addition, we see $\text{tr}(AB) = \sum_{i=1}^n (\sum_{k=1}^n A_{ki} B_{ik}) = \text{tr}(BA)$ by renaming the variables. \square

Problem 2.4.16. Let B $n \times n$, invertible. Let $\Phi : M_{n \times n}(F) \rightarrow M_{n \times n}(F)$ by $\Phi(A) = B^{-1}AB$. Show Φ is an isomorphism.

Proof. (linearity) $\Phi(cA + D) = B^{-1}(cA + D)B = B^{-1}(cAB + DB) = cB^{-1}AB + B^{-1}DB = c\Phi(A) + \Phi(D)$.

(one-to-one) If $\Phi(A) = O \implies B^{-1}AB = O \implies BB^{-1}ABB^{-1} = BOB^{-1} \implies A = O$, thus $N(\Phi) = \{0\}$. Thus Φ is one-to-one.

Since the dimensions of domain and codomain are equal, Φ is onto, and therefore an isomorphism. \square

Problem 4.3.21. Prove block multiplication. i.e.,

$$M_1 = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, M_2 = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix},$$

where M_1, M_2 are $(m+n) \times (m+n)$, A_{11}, B_{11} are $n \times n$, A_{12}, B_{12} are $n \times m$, A_{21}, B_{21} are $m \times n$, A_{22}, B_{22} are $m \times m$. Then

$$M_1 M_2 = \begin{pmatrix} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \end{pmatrix}$$

Proof.

$$\begin{aligned}(M_1 M_2)_{ij} &= \sum_{k=1}^{m+n} (M_1)_{ik} (M_2)_{kj} \\ &= \sum_{k=1}^n (M_1)_{ik} (M_2)_{kj} + \sum_{k=n+1}^{m+n} (M_1)_{ik} (M_2)_{kj}\end{aligned}$$

Consider the case where $1 \leq i, j \leq n$.

$$\begin{aligned}(M_1 M_2)_{ij} &= \sum_{k=1}^n (A_{11})_{ik} (B_{11})_{kj} + \sum_{k=n+1}^{m+n} (A_{12})_{ik} (B_{21})_{kj} \\ &= (A_{11} B_{11} + A_{12} B_{21})_{ij}\end{aligned}$$

The other cases follows similarly. ▣

2.2.6 Left-multiplication Transformation

Definition 2.17 (Left multiplication). Given matrix A $m \times n$, define $L_A : F^n \rightarrow F^m$ by $L_A(x) = Ax \ \forall x \in F^n$. We say L_A is a left-multiplication transformation.

Theorem 2.21. Let A $m \times n$. $L_A : F^n \rightarrow F^m$ linear. B $m \times n$, β, γ are standard bases of F^n, F^m , respectively. Then

1. $[L_A]_{\beta}^{\gamma} = A$.
2. $L_A = L_B \iff A = B$.
3. $L_{A+B} = L_A + L_B$, $L_{aA} = aL_A \ \forall a \in F$.
4. If $T : F^n \rightarrow F^m$ linear, $\exists!$ C $m \times n$ s.t. $T = L_C$, $C = [T]_{\beta}^{\gamma}$.
5. $L_{AE} = L_A L_E$ if defined.
6. $m = n \implies L_{I_n} = I_{F^n}$.

Proof. The linearity of L_A is guaranteed by [Theorem 2.19](#).

(1) $L_A(e_j) = Ae_j = v_j = \sum_{i=1}^m a'_{ij} e'_i$. So A and A' have all columns equal, so $A = A'$.

(2) (\Leftarrow) Trivial. (\Rightarrow) $L_A = L_B \implies Ae_j = Be_j \implies u_j = v_j \implies A$ and B all columns equal. So $A = B$.

(3) By [Theorem 2.19](#) trivial.

(4) Existence is given by $C = [T]_\beta^\gamma$. So $[T(x)]_\gamma = [T]_\beta^\gamma[x]_\beta \implies T(x) = Cx = L_C(x)$. Uniqueness is given by (2). ▮

(5) Trivial. ▮

✍ *Remark.* Left multiplication transformations can also be used to prove that matrix multiplications are associative. But I think that's quite cheating. ▮

Theorem 2.22. *Given matrix A $n \times n$. Show A is invertible $\iff L_A$ is invertible. In particular, $(L_A)^{-1} = L_{A^{-1}}$.*

Proof.

(\implies) A invertible $\implies A^{-1}A = AA^{-1} = I_n$. Then $(A^{-1}A)(x) = I_n x$, $(AA^{-1})(x) = I_n x$. Then $L_{A^{-1}}(L_A(x)) = I(x)$; $L_A(L_{A^{-1}}(x)) = I(x) \implies L_{A^{-1}} = (L_A)^{-1}$.

(\impliedby) Let $L_B = (L_A)^{-1}$, $L_B = Bx$. We see $L_B L_A = I$, $L_A L_B = I \implies BA = I$, $AB = I$. ▮

✍ *Remark.* Left-multiplication transformations can be used to tackle matrix properties when you don't want to resort to brute force. See the following problem. ▮

Problem 2.4.7. *Let A $n \times n$.*

1. *Suppose $A^2 = O$, show A is not invertible.*
2. *Suppose $AB = O$, B $n \times n$, $B \neq O$, could A be invertible?*

Proof. (1) Let $L_A = Ax$, then $L_A^2 = T_0 \iff R(L_A) \subseteq N(L_A)$ by [Problem 2.3.11](#).

If $N(L_A) = \{0\}$, then $R(L_A) \subseteq N(L_A) \implies R(L_A) = \emptyset$ or $\{0\}$. *

Thus L_A could not be one-to-one thus not invertible, so A is not invertible.

(2) $L_A L_B = T_0 \implies R(L_A L_B) = \{0\} \implies R(L_B) \subseteq N(L_A)$. If A is invertible, $N(L_A) = \{0\} \implies R(L_B) = 0 \implies B = O$. * ▮

Problem 2.4.9. *Let A, B $n \times n$ s.t. AB is invertible. Show A, B is also invertible. Show by example that if A, B come in arbitrary sizes, then they may not be invertible.*

Proof. Let $L_A, L_B : F^n \rightarrow F^n$ by $L_A = Ax$, $L_B = Bx$. By [Problem 2.3.12](#), we see if $L_A L_B$ is one-to-one, then L_B is one-to-one, therefore invertible since the dimensions of domain and codomain are equal.

Moreover, $L_A L_B$ is onto, this means that L_A is onto, and therefore invertible. ▮

Problem 2.4.10. Let A, B $n \times n$, $AB = I_n$. By the previous problem, we see A, B are invertible. Prove $A = B^{-1}$, $B = A^{-1}$, i.e. for square matrices, one-sided inverse is also two-sided inverse.

Proof. Let $B' = A^{-1}$, $A' = B^{-1}$. Then $AB = I_n = AB' = B'A = A'B = BA'$.

Thus $B'AB = B'AB' \implies B = B'$. The other side goes similarly. \square

2.2.7 Isomorphism of Linear Transformation and Matrices

Theorem 2.23. Let $\dim V = n$, $\dim W = m$, β, γ are bases of V, W , respectively. Then $\Phi : \mathcal{L}(V, W) \rightarrow M_{m \times n}(F)$ defined by $\Phi(T) = [T]_{\beta}^{\gamma}$ for $T \in \mathcal{L}(V, W)$ is an isomorphism.

Proof. By **Theorem 2.15**, we see Φ is linear. We still need to show it is one-to-one and onto.

$\forall A \in M_{m \times n}(F)$, define $T(v_j) = \sum_{i=1}^m a_{ij}w_i$, $1 \leq j \leq n$, $v_j \in \beta$, $w_i \in \gamma$. By **Theorem 2.6**, we see such a T is unique, so Φ is one-to-one. For all $A \in M_{m \times n}(F)$, we can find such a correspondence, so Φ is onto. So Φ is an isomorphism. \square

Remark. Notice we cannot use linearity to say one-to-one is equivalent to onto, since we were not yet sure if their dimensions are equal. \square

Corollary 2.23.1. Let $\dim V = n$, $\dim W = m$, then $\dim \mathcal{L}(V, W) = nm$.

Theorem 2.24. Let $T : V \rightarrow W$, $\dim V = n$, $\dim W = m$, β, γ be their bases, respectively. Also let $L_A : F^n \rightarrow F^m$, $A = [T]_{\beta}^{\gamma}$. Then,

$$L_A \phi_{\beta} = \phi_{\gamma} T, \quad \forall x \in V.$$

Proof. By **Theorem 2.17**, $[T(x)]_{\gamma} = [T]_{\beta}^{\gamma}[x]_{\beta}$. \square

Remark. The above isomorphism do this: We have two abstract vector spaces, $\dim V = n$, $\dim W = m$, and a linear relationship $T : V \rightarrow W$.

By the isomorphisms, we can isomorph V into F^n , W into F^m , and associate T with L_A , $A = [T]_{\beta}^{\gamma}$, then work only with concrete tuples. \square

Problem 2.4.20. Let $\dim V = n$, $\dim W = m$, $T : V \rightarrow W$ linear. Let β, γ be ordered bases for V, W , respectively. Show $\dim R(T) = \dim R(L_A)$, $\dim N(T) = \dim N(L_A)$, where $A = [T]_{\beta}^{\gamma}$.

Proof. Let $W = \phi_\gamma R(T)$. Then $\dim W = \dim R(T)$. We claim $W = \phi_\gamma R(T) = R(L_A)$.

$$\forall T(v) \in R(T), v \in V, y = \phi_\gamma T(v) = L_A \phi_\beta(v) \in R(L_A) \implies W \subseteq R(L_A).$$

$$\forall L_A(x) \in R(L_A), x \in F^n, w = L_A(x) = \phi_\gamma T \phi_\beta^{-1}(x) \in \phi_\gamma R(T) \implies R(L_A) \subseteq W.$$

Since $\dim R(T) + \dim N(T) = \dim V = n = \dim F^n = \dim R(L_A) + \dim N(L_A)$, we have $\dim N(T) = \dim N(L_A)$. \square

2.3 Projections

2.3.1 Definition

Definition 2.18. Let $V = W_1 \oplus W_2$. $T : V \rightarrow V$ is called projection on W_1 along W_2 if $\forall x = x_1 + x_2, x_1 \in W_1, x_2 \in W_2$, we have $T(x) = x_1$.

Problem 2.1.26.a. Suppose $T : V \rightarrow V$ is projection on W_1 along W_2 . Show T is linear and $W_1 = \{x \in V \mid T(x) = x\}$.

Proof. $T(cx + x') = T(c(x_1 + x_2) + (x'_1 + x'_2)) = T((cx_1 + x'_1) + (cx_2 + x'_2)) = cx_1 + x'_1 = cT(x) + T(x')$.

(\subseteq) Trivial. (\supseteq) If $x \notin W_1, x = x_1 + x_2, x_2 \neq 0$. Then $T(x) = x_1 \neq x$. Thus $W_1 = S$. \square

2.3.2 Properties

Problem 2.1.26.b. Show $W_1 = R(T), W_2 = N(T)$.

Proof.

$$(1) (\supseteq) \forall x \in V, x = x_1 + x_2. T(x) = x_1 \in W_1 \implies R(T) \subseteq W_1.$$

$$(\subseteq) \forall x_1 \in W_1, x_1 \in V \text{ and } T(x_1) = x_1 \implies W_1 \subseteq R(T). \text{ Thus } W_1 = R(T).$$

$$(2) (\subseteq) \forall x \in W_2, x = 0 + x_2. T(x) = 0. W_2 \subseteq N(T).$$

$$(\supseteq) \text{ Suppose } \exists x \in N(T) \text{ s.t. } x \notin W_2. \implies x = x_1 + x_2, x_1 \neq 0. \quad \square$$

\nearrow *Remark.* If $W_1 = V, T = I$; if $W_1 = \{0\}, T = T_0$. \square

Problem 2.1.27.a. Suppose W is a subspace of $V, \dim V \neq \infty$. Show $\exists W', T : V \rightarrow V$ s.t. T is a projection on W along W' .

Proof. (trivial) If $W = V$, 取 $W' = \{0\}, T = I$.

(non-trivial) If $W \neq V$, by **Problem 1.6.34**, $\exists W'$ s.t. $V = W \oplus W'$. So $\forall x \in V, x = x_1 + x_2$, define $T(x) = x_1$. ▣

Problem 2.1.27.b. Give example of W , a subspace of V , s.t. \exists 2 projections on W along 2 distinct subspaces.

Proof. In \mathbb{R}^2 , chose y axis along x axis / $y=x$. ▣

✍ *Remark.* The two subspaces in direct sum need not be orthogonal. ▣

Problem 2.1.30. Suppose T is projection on W along W' . Prove W is T -invariant and $T_W = I_W$.

Proof. Trivial. ▣

Problem 2.3.17. Determine all $T : V \rightarrow V$ s.t. $T = T^2$.

Proof. Let $S = \{ y \mid T(y) = y \}$. Clearly S is a subspace. Since $T(0) = 0$, $S \cap N(T) = \{0\}$.

If $T = T^2$, then $T(x) = T(T(x)) \implies T(x) \in S$.

And $T(x - T(x)) = T(x) - T^2(x) = 0 \implies x - T(x) \in N(T)$.

And $\forall x \in V, x = T(x) + (x - T(x))$, so $V = S \oplus N(T)$. So this corresponds to all projection transformations. ▣

2.4 Invariants and Restrictions

2.4.1 Definition

Definition 2.19. Let $T : V \rightarrow V$ linear. A subspace W of V is called T -invariant if $T(W) \subseteq W$ (i.e. $\forall x \in W, T(x) \in W$). If W is T -invariant, define restriction of T on W to be $T_W : W \rightarrow W, T_W(x) = T(x) \forall x \in W$.

✍ *Remark.* 一個線性算子若是可以 restrict 的話，那就可以把整個空間切成 Direct sum 之後再用 restrict 處理，會更簡單。之後在研究對角化的時候也有用處。 ▣

Problem 2.1.29. If W is T -invariant, prove T_W is linear.

Proof. Trivial. ▣

2.4.2 Properties

Problem 2.1.29. Let $T : V \rightarrow V$ linear. Show $\{0\}, V, R(T), N(T)$ are T -invariant.

Proof. (1) $T(0) = 0 \in \{0\}$.

(2) V is a vector space, so $T(V) \subseteq V$.

(3) $R(T) = \{T(x) \mid x \in V\}$. $\forall x \in V, T(x) \in R(T), R(T) \subseteq V \implies T(R(T)) \subseteq R(T)$. (4) $T(N(T)) = \{0\} \subseteq N(T)$ ▣

Problem 2.1.31. Sps $V = R(T) \oplus W$ and W is T -invariant. $T : V \rightarrow V$ linear. Show

1. $W \subseteq N(T)$
2. If $\dim V \neq \infty, W = N(T)$.
3. $W = N(T)$ may not be true if $\dim V = \infty$.

Proof.

1. Suppose $\exists w \in W$ s.t. $T(w) \neq 0$ (i.e. $w \notin N(T)$). Since W is T -invariant, then $T(w) \in W$, and $T(w) \in R(T)$ by definition. Then $T(w) \in W \cap R(T)$ yet $T(w) \neq 0$. *
2. By **Theorem 2.3**, $\dim V = \dim R(T) + \dim N(T) = \dim R(T) + \dim W \implies \dim N(T) = \dim W$.
By (1) we know W is a subspace of $N(T)$, but their dimensions are equal, so by **Theorem 1.11**, $W = N(T)$.
3. Left shift is an example. $W = \{0\} \neq N(T) = \{c(1, 0, \dots)\}$.

▣

Problem 2.1.32. Suppose W is T -invariant. Show $N(T_W) = N(T) \cap W$ and $R(T_W) = T(W)$.

Proof. $R(T_W) = \{T(x) \mid \forall x \in W\} = T(W)$. $N(T_W) = N(T) \cap W$ trivial. ▣

2.4.3 Examples

Problem 2.1.30. Suppose T is projection on W along W' . Prove W is T -invariant and $T_W = I_W$.

Proof. Trivial. ▣

2.5 Change of Coordinate Matrices

Theorem 2.25. Let β, β' be two ordered bases for V , $\dim V \neq \infty$. Let $Q = [I_V]_{\beta'}^{\beta}$. Then Q is invertible, and $\forall v \in V$, $[v]_{\beta} = Q[v]_{\beta'}$.

Proof. By [Theorem 2.18](#), I_V invertible $\iff Q$ invertible.

By [Theorem 2.17](#), $[v]_{\beta} = [I_V(v)]_{\beta} = [I_V]_{\beta'}^{\beta}[v]_{\beta'} = Q[v]_{\beta'}$. ▣

Remark. If $\beta = \{x_1, \dots, x_n\}$, $\beta' = \{x'_1, \dots, x'_n\}$, then $x'_j = \sum_{i=1}^n Q_{ij}x_i$, column j of Q is $[x'_j]_{\beta}$. In other words, the columns of Q is the old bases in terms of the new coordinates. ▣

Definition 2.20. A linear operator of a vector space V is some $T \in \mathcal{L}(V)$.

Theorem 2.26. If T is a linear operator over V , $[T]_{\beta'} = Q^{-1}[T]_{\beta}Q$, where Q is the change of coordinate matrix from $\beta' \rightarrow \beta$.

Proof. $Q^{-1}[T]_{\beta}Q = [I_V]_{\beta'}^{\beta'}[T]_{\beta}^{\beta}[I_V]_{\beta}^{\beta'} = [I_V]_{\beta}^{\beta'}[T]_{\beta}^{\beta} = [T]_{\beta'}$. ▣

Corollary 2.26.1. Let $A \in M_{n \times n}(F)$, γ is a basis of F^n , then $[L_A]_{\gamma} = Q^{-1}AQ$, Q $n \times n$, column j of Q is the j -th vector of γ .

Proof. Let β be the standard basis of F^n . Column j of Q is $[\gamma_j]_{\beta} \implies Q = [I_{F^n}]_{\gamma}^{\beta}$. So $[L_A]_{\gamma} = [I_{F^n}]_{\beta}^{\gamma}[L_A]_{\beta}^{\beta}[I_{F^n}]_{\gamma}^{\beta}$. ▣

2.6 Dual Spaces

Definition 2.21 (Dual Spaces). The dual space of V is the space of functionals $V^* := \mathcal{L}(V, F)$ which takes a vector in V to the field of scalars.

Remark. $\dim(V^*) = \dim(\mathcal{L}(V, F)) = \dim V \cdot \dim F = \dim V$. ▣

Theorem 2.27. Suppose $\dim V = n$, $\beta = \{x_1, \dots, x_n\}$ be a basis of V . $f_i(x) = ([x]_{\beta})_i$. Let $\beta^* = \{f_1, \dots, f_n\}$. Then β^* is a basis of V^* , and $\forall f \in V^*$, $f = \sum_{i=1}^n f(x_i)f_i$.

Proof. Let $\sum_{i=1}^n a_i f_i = g = 0$. Then $g(x_j) = a_j = 0, 1 \leq j \leq n$. This implies $a_j = 0$. So β^* is independent. And $\dim V^* = n$. So β^* is a basis.

Let $x = \sum_{i=1}^n a_i x_i, f(x) = \sum_{i=1}^n a_i f(x_i)$. And $a_i = f_i(x)$ by definition, so $f(x) = \sum_{i=1}^n f(x_i) f_i(x) \forall x$, so $f = \sum_{i=1}^n f(x_i) f_i$. \square

Definition 2.22. The basis $\beta^* = \{f_1, \dots, f_n\}$ of V^* which $f_i(x_j) = \delta_{ij}$ is called the dual basis of β .

Theorem 2.28. Let $\dim V, \dim W \neq \infty, \beta, \gamma$ are their bases, respectively. $\forall T : V \rightarrow W$ linear, $T^\top : W^* \rightarrow V^*$ by $T^\top(g) = gT \forall g \in W^*$ is a linear transformation with the property that $[T^\top]_{\gamma^*}^{\beta^*} = ([T]_\beta^\gamma)^\top$.

Proof. (linearity) $\forall x \in V$, consider $T^\top(cf + g)(x) = (cf + g)(T(x)) = c(fT)(x) + (gT)(x) = cT^\top(f)(x) + T^\top(g)(x) \implies T^\top(cf + g) = cT^\top(f) + T^\top(g)$.

(transposition) Let $\beta = \{x_1, \dots, x_n\}, \gamma = \{y_1, \dots, y_m\}, \beta^* = \{f_1, \dots, f_n\}, \gamma^* = \{g_1, \dots, g_m\}$, and $T(x_j) = \sum_{k=1}^m a_{kj} y_k, T^\top(g_j) = \sum_{k=1}^n b_{kj} f_k$. Thus

$$T^\top(g_j)(x_i) = (g_j T)(x_i) = g_j\left(\sum_{k=1}^m a_{ki} y_k\right) = a_{ji}$$

and

$$T^\top(g_j)(x_i) = \sum_{k=1}^n b_{kj} f_k(x_i) = b_{ij}$$

so we have $a_{ji} = b_{ij}$, thus $B = A^\top$. \square

\blacktriangleright *Remark.* Now we want to establish an isomorphism between V and V^{**} .

$\forall x \in V$, define $\hat{x} : V^* \rightarrow F$ by $\hat{x}(f) = f(x) \forall f \in V^*$. Clearly $\hat{x} \in V^{**}$. \blacksquare

Theorem 2.29. Let $\dim V \neq \infty, x \in V$. If $\hat{x}(f) = 0 \forall f \in V^*$, then $x = 0$.

Proof. Suppose $x \neq 0$, choose $\beta = \{x_1, \dots, x_n\}$ basis for V s.t. $x_1 = x$. Let β^* be the dual basis. Then $f_1(x_1) = 1 \neq 0$, letting $f = f_1$ shows $\forall x \neq 0, \exists f$ s.t. $f(x) \neq 0$. \square

\blacktriangleright *Remark.* The above proof shows that the correspondence between V and V^{**} is one-to-one. \blacksquare

Theorem 2.30. Let $\dim V \neq \infty$ define $\psi : V \rightarrow V^{**}$ by $\psi(x) = \hat{x}$. Show ψ is an isomorphism.

Proof. (linearity) $\psi(cx + y)(f) = \widehat{(cx + y)}(f) = f(cx + y)$. Yet $f \in V^* = \mathcal{L}(V, F)$, so $f(cx + y) = cf(x) + f(y) = c\hat{x}(f) + \hat{y}(f) = c\psi(x)(f) + \psi(y)(f) \forall f$. This says $\psi(cx + y) = c\psi(x) + \psi(y)$.

(isomorphism) Consider $\psi(x) = 0$. Then $\hat{x}(f) = 0 \forall x \in V^*$. By **Theorem 2.29**, $x = 0$, so $N(\psi) = \{0\}$. Now $\dim V = \dim V^{**}$, so ψ is onto, and therefore isomorphic. \square

Corollary 2.30.1. *Let $\dim V \neq \infty$, and its dual V^* . Then every ordered basis of V^* is the dual basis for some basis in V .*

Proof. Let $\beta^* = \{f_1, \dots, f_n\}$ be a basis of V^* . Let $\beta^{**} = \{\hat{x}_1, \dots, \hat{x}_n\}$ be the dual basis of β^* . Then $\hat{x}_i(f_j) = \delta_{ij}$.

Define $x_i = \psi^{-1}(\hat{x}_i)$, $\beta = \{x_1, \dots, x_n\}$. By **Problem 2.1.14.c**, we see β is a basis of V .

Now since ψ is an isomorphism, there is one and only one x_i for \hat{x}_i to behave like it. So the \tilde{x}_i that makes $\hat{x}_i(f_j) = \delta_{ij} = f_j(\tilde{x}_i)$ is the same x that makes $x_i = \psi^{-1}(\hat{x}_i)$. \square

3 Linear Equations

3.1 Elementary Matrix Operations

Definition 3.1 (Elementary Matrix Operations). Let $A \in M_{n \times n}(F)$. Elementary row operations are:

1. Interchanging two rows of A .
2. Scale any row of A by $c \neq 0$.
3. Add any scalar multiple of one row to another row of A .

Definition 3.2 (Elementary Matrix). Elementary matrix is obtained by elementary row operating on I_n .

Theorem 3.1. $A \in M_{m \times n}(F)$, and B comes from doing elementary row operations of A . Then $\exists E \in M_{m \times m}(F)$ s.t. $B = EA$, where E is obtained by doing the same elementary row operations on I_m .

Theorem 3.2. Elementary matrices are invertible, and their inverses are also the same type.


3.2 Rank of a Matrix

3.2.1 Definition and Properties

Definition 3.3. Let $A \in M_{m \times n}(F)$, the rank of A , $\text{rank}(A) = \dim R(L_A)$, $L_A : F^n \rightarrow F^m$.

✂ *Remark.* If $A \in M_{n \times n}(F)$, then A is invertible $\iff \text{rank}(A) = n$, by [Theorem 2.5](#).


Theorem 3.3. Let $T : V \rightarrow W$ linear, $\dim V = \dim W$, β, γ are bases of V, W , respectively. Then $\text{rank}(T) = \text{rank}([T]_\beta^\gamma)$.

Proof. Trivial by [Problem 2.4.20](#), since $\text{rank}([T]_\beta^\gamma) = \text{rank } A = \text{rank}(L_A) = \text{rank } T$.



Theorem 3.4. Let $A \in M_{m \times n}(F)$, $P \in M_{m \times m}(F)$, $Q \in M_{n \times n}(F)$, P, Q are invertible. Then

1. $\text{rank}(AQ) = \text{rank } A$.
2. $\text{rank}(PA) = \text{rank } A$.
3. $\text{rank}(PAQ) = \text{rank } A$.

Proof. Let $L_A : F^n \rightarrow F^m$, $L_P : F^m \rightarrow F^m$, $L_Q : F^n \rightarrow F^n$.


(1) $\text{rank}(AQ) = \dim R(L_A L_Q)$. But L_Q is invertible, so $R(L_Q) = F^n \implies R(L_A L_Q) = L_A(R(L_Q)) = R(L_A)$.

(2) $\text{rank}(PA) = \dim R(L_P L_A) = \dim(L_P(R(L_A)))$. Now $R(L_A)$ is a subspace, so by [Problem 2.4.17](#), $\dim(L_P(R(L_A))) = \dim R(L_A) = \text{rank } A$.

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


Theorem 3.5. The rank of a matrix equals the maximum number of linearly independent columns. Let S be the set of columns of A , then $\text{rank } A = \dim(\text{span}(S))$.

Proof. Let $A \in M_{m \times n}(F)$, $L_A : F^n \rightarrow F^m$, β, γ be standard bases of F^n, F^m , respectively. Then $A = [L_A]_\beta^\gamma$, $R(L_A) = \text{span}(L_A(\beta))$ by [Theorem 2.2](#).

By [Theorem 2.20](#), $L_A e_i = A e_i = v_i$, $1 \leq i \leq n$, v_i is the i -th column of A . So $\dim R(L_A) = \dim \text{span}(\{v_1, \dots, v_n\})$.


3.2.2 Elimination and Important Consequences

Theorem 3.6. *Let $A \in M_{m \times n}(F)$, $\text{rank } A = r$. Then $r \leq m, r \leq n$, and by a finite number of elementary row and column operations, A can be transformed into*

$$D = \begin{pmatrix} I_r & O_1 \\ O_2 & O_3 \end{pmatrix}, O_1, O_2, O_3 \text{ are zero matrices.}$$

In other words, $D_{ii} = 1$ for $1 \leq i \leq r$, $D_{ij} = 0$ otherwise.

Proof. (trivial) If $A = O_{m \times n}$, then $r = 0$, let $D = A$.

(dimension) Since $R(L_A)$ is a subspace of F^m , $\dim R(L_A) = r \leq m$. Since $\dim F^n = \dim N(L_A) + \dim R(L_A) = n \implies r \leq n$.

(form) Suppose $A \neq O, r = \text{rank } A, r > 0$, induction on m .

If $m = 1$,

$$A = (a_{11} \ a_{12} \ \dots \ a_{1n})$$

If $a_{11} \neq 0$, by column operations we can get

$$\begin{aligned} A &= (1 \ a_{11}^{-1}a_{12} \ a_{11}^{-1}a_{13} \ \dots \ a_{11}^{-1}a_{1n}) \\ &\implies (1 \ 0 \ 0 \ \dots \ 0) \end{aligned}$$

Suppose $m = k$ is true, consider $m = k + 1$. Via the same operations on A_k , we can arrive at

$$D'_{k+1} = \begin{pmatrix} I_{r_k} & O_{n-r_k} \\ O_{k-r_k} & O_3 \\ a_{k+1,1} & \dots & a_{k+1,n} \end{pmatrix}$$

If $a_{k+1,r_k+1}, \dots, a_{k+1,n} = 0$, then by row operations, we can cancel $a_{k+1,1}, \dots, a_{k+1,r_k}$ to zero, so the last row is zero.

If not, then we can do column operations just like the $m = 1$ case, and get a 1 standing in $a_{k+1,r_k+1}, \dots, a_{k+1,n} = 0$. Cancelling out $a_{k+1,1}, \dots, a_{k+1,r_k} = 0$ similarly, we get

$$D_{k+1} = \begin{pmatrix} I_{r_k+1} & O_{n-r_k-1} \\ O_{k-r_k} & O_3 \end{pmatrix}$$

▣

✍ *Remark.* This proof is somewhat different from the one in the book, but I think it's still valid. \square

Corollary 3.6.1. *Let $A \in M_{m \times n}(F)$, $\text{rank } A = r$. Then $\exists B \in M_{m \times m}(F), C \in M_{n \times n}(F)$, B, C invertible s.t. $D = BAC$, and*

$$D = \begin{pmatrix} I_r & O_1 \\ O_2 & O_3 \end{pmatrix}$$

Proof. Let B be the product of all elementary row operation matrices, and C be the product of all elementary column operation matrices in the process of **Theorem 3.6**. Then B, C are invertible, and $D = BAC$. \square

Corollary 3.6.2. *Let $A \in M_{m \times n}(F)$.*

1. $\text{rank}(A^\top) = \text{rank } A$.
2. $\text{rank } A = \text{number of independent rows of } A$.
3. *Rows and columns of any matrix span subspaces of the same dimension, numerically equal to the rank of the matrix.*

Proof. (1) By **Corollary 3.6.1**, $D = BAC, D^\top = C^\top A^\top B^\top$. It's easy to see that $\text{rank}(D^\top) = r$, so $\text{rank}(D^\top) = r = \text{rank}(A^\top) = \text{rank } A$.

For (2)(3), apply **Theorem 3.5** to A^\top . \square

Corollary 3.6.3. *Every invertible matrix is a product of elementary matrices.*

Proof. If $A \in M_{n \times n}(F)$ invertible, then $\text{rank } A = n, D = I, I = BAC \implies A = B^{-1}C^{-1}$. \square

✂ *Remark.* **Corollary 3.6.3** is a very important theorem, and it is used in many unexpected places. \square

3.2.3 Rank of Composition

Theorem 3.7. *Let $T : V \rightarrow W, U : W \rightarrow Z$ be linear, $\dim V, \dim W, \dim Z \neq \infty, AB$ is defined. Then*

1. $\text{rank}(UT) \leq \text{rank } U$.
2. $\text{rank}(UT) \leq \text{rank } T$.
3. $\text{rank}(AB) \leq \text{rank } A$.
4. $\text{rank}(AB) \leq \text{rank } B$.

Proof. (1) Since $R(T) \subseteq W$, $U(R(T)) \subseteq U(W) \implies \text{rank}(UT) \leq \text{rank } U$.

$$(3) \text{rank}(AB) = \text{rank}(L_{AB}) = \text{rank}(L_A L_B) \leq \text{rank}(L_A) = \text{rank } A.$$

$$(4) \text{rank}(AB) = \text{rank}((AB)^\top) = \text{rank}(B^\top A^\top) \leq \text{rank}(B^\top) = \text{rank } B.$$

(2) Let α, β, γ be bases of V, W, Z , respectively. $A' = [U]_\beta^\gamma, B' = [T]_\alpha^\beta, A'B' = [UT]_\alpha^\gamma$. Then $\text{rank}(UT) = \text{rank}(A'B') \leq \text{rank}(B') = \text{rank } T$. \square

✍ *Remark.* I didn't think of this proof. I guess a Goddess must have revealed this proof to the author. \square

3.3 Solving Linear Equations

3.3.1 The Form of Solutions

Theorem 3.8. *Let $Ax = 0$ be a homogeneous system of m linear equations in n unknowns over F . $K = \{x \mid Ax = 0, x \in F^n\}$. Then $K = N(L_A)$, K is a subspace of F^n , $\dim K = n - \text{rank } A$.*

Corollary 3.8.1. *If $m < n$, then $Ax = 0$ has a non-zero solution.*

Proof. Let $L_A : F^n \rightarrow F^m, \dim R(L_A) + \dim N(L_A) = n$. But $\dim R(L_A) \leq m < n \implies \dim N(L_A) \geq n - m > 0$. \square

Theorem 3.9. *Let $K = \{x \mid Ax = b\}, K_H = \{x \mid Ax = 0\}$. Then for any $As = b$, $K = \{s\} + K_H = \{s + k \mid k \in K_H\}$.*

Proof.

(\supseteq) $\forall x \in \{s\} + K_H, x = s + k, k \in K_H$. $Ax = A(s + k) = As + Ak = b$. So $x \in K, \{s\} + K_H \subseteq K$.

(\subseteq) $\forall x \in K$, consider $x - s$. $A(x - s) = 0 \implies x - s \in K_H$. \square

Theorem 3.10. *Let $A \in M_{n \times n}(F), Ax = b$. If A is invertible, then $\exists! x = A^{-1}b$ s.t. $Ax = b$. If $\exists! x$ s.t. $Ax = b \implies A$ is invertible.*

Proof.

(\implies) Trivial.

(\impliedby) $K = \{s\} + K_H$. If $|K| = 1$, this means $K_H = \{0\}$, $\text{rank } A = n$, which means A is invertible. \square

3.3.2 Procedure of Solving Linear Equations

Theorem 3.11. *Let $Ax = b$. The system has a solution $\iff \text{rank } A = \text{rank}(A|b)$.*

Proof. The system has a solution $\iff b \in \text{span}(\{v_i\})$, where v_i are the columns of A . This is also equivalent to $\dim \text{span } S = \dim(\text{span}(S \cup \{b\}))$ ▣

Definition 3.4. Two sets of linear equations are equivalent if they have the same solution set.

Theorem 3.12. *Let $A \in M_{m \times n}(F)$, $Ax = b$. Let $C \in M_{m \times m}(F)$ invertible. Then $CAx = Cb$ is equivalent to $Ax = b$.*

Proof. Let the solution set of $(CA)x = Cb$ be K' , $Ax = b$ be K .

Since $N(C) = \{0\}$, all the v' that makes $(CA)v' = 0$ must have $Av' = 0$. So $K'_H \subseteq K_H$.

And obviously all the v that makes $Av = 0$ also makes $(CA)v = 0$. So $K_H \subseteq K'_H$. So $K'_H = K_H$.

Moreover, let $K = \{s\} + K_H$, $As = b$, $Av = 0 \forall v \in K_H$.

Now $s \in K'$, since $(CA)s = C(As) = Cb$;

By **Theorem 3.9**, any s s.t. $As = b$ satisfies $K' = \{s\} + K'_H$. Since $K_H = K'_H$, $K = K'$. ▣

Corollary 3.12.1. *Let $A \in M_{m \times n}(F)$, $Ax = b$. If $(A'|b')$ is obtained from $(A|b)$ by a finite number of elementary row operations, then $A'x = b' \iff Ax = b$.*

Proof. Trivial since elementary row operation matrices are invertible. ▣

4 Determinants

4.1 Foundations

4.1.1 The Algebraic Object of Study

Definition 4.1 (Ring). A ring is a set K together with two operations: addition $(x, y) \rightarrow x + y$ and multiplication $(x, y) \rightarrow xy$ satisfying

1. K is a commutative group under addition.
2. $\forall x, y, z \in K, (xy)z = x(yz)$.
3. $\forall x, y, z \in K, x(y + z) = xy + xz$, and $(x + y)z = xz + yz$.

✂ *Remark.* A ring, unlike a field, does not promise the existence of multiplicative inverse. So all the properties of vector spaces with multiplicative inverses need to be excluded when concerned with a ring. We loosen the requirements mainly because we want to include polynomials into our discussion. \square

4.1.2 Axioms of Determinant

Definition 4.2 (N-linearity). Let K be a commutative ring with identity, $n \in \mathbb{Z}^+$, $D : M_{n \times n}(K) \rightarrow K$. Then D is n -linear if D is a linear function of row i when the other rows are fixed, for all $1 \leq i \leq n$.

Theorem 4.1. Any linear combination of n -linear functions is n -linear.

Definition 4.3 (Alternating). Let D be an n -linear function. D is alternating if

1. $D(A) = 0$ whenever two rows of A are equal.
2. If B is A with two rows exchanged, then $D(B) = -D(A)$.

✂ *Remark.* $1 \Rightarrow 2$ in any field, as we will show below. However, $2 \not\Rightarrow 1$ in general. Let A have two equal rows, $a_r = a_s$. Exchange these rows to obtain B . Then $\det A = \det B = -\det A$. Then $\det A + \det A = 0$. In a field of characteristic 2, for example, the identity is also a possible solution. \square

Definition 4.4. A determinant function is $D : M_{n \times n}(K) \rightarrow K$ which is n -linear, alternating, and $D(I) = 1$.

Theorem 4.2. Let $D : M_{n \times n}(K) \rightarrow K$, and D is n -linear. Suppose $D(A) = 0$ whenever A has two equal adjacent rows. Then D is alternating.

Proof. Let

$$A = \begin{pmatrix} a_1 \\ \vdots \\ a_r \\ a_{r+1} \\ \vdots \\ a_n \end{pmatrix}, A' = \begin{pmatrix} a_1 \\ \vdots \\ a_{r+1} \\ a_r \\ \vdots \\ a_n \end{pmatrix},$$

where a_i are row vectors. Then

$$\begin{aligned}
D \begin{pmatrix} a_1 \\ \vdots \\ a_r \\ a_r \\ \vdots \\ a_n \end{pmatrix} + D \begin{pmatrix} a_1 \\ \vdots \\ a_r \\ a_{r+1} \\ \vdots \\ a_n \end{pmatrix} + D \begin{pmatrix} a_1 \\ \vdots \\ a_{r+1} \\ a_r \\ \vdots \\ a_n \end{pmatrix} + D \begin{pmatrix} a_1 \\ \vdots \\ a_{r+1} \\ a_{r+1} \\ \vdots \\ a_n \end{pmatrix} = D \begin{pmatrix} a_1 \\ \vdots \\ a_r + a_{r+1} \\ a_r + a_{r+1} \\ \vdots \\ a_n \end{pmatrix} \\
= 0 = 0 + D(A) + D(A') + 0.
\end{aligned}$$

Then consider the two exchanging rows be a_r, a_s , where $s > r$. If $s = r + 1$, then we are done. If not, then consider

$$A = \begin{pmatrix} a_1 \\ \vdots \\ a_r \\ \vdots \\ a_s \\ \vdots \\ a_n \end{pmatrix}, B = \begin{pmatrix} a_1 \\ \vdots \\ a_s \\ \vdots \\ a_r \\ \vdots \\ a_n \end{pmatrix}.$$

Exchange a_s in A and a_r in B simultaneously, until in a position similar to the $r + 1$ case. We have done the same number of exchanges on each side, so it cancels out eventually. At this point, by similar reasoning to the $r + 1$ case, we see they have opposite sign. So $D(A) = -D(B)$.

For arbitrary identical rows a_r, a_s , exchange the adjacent rows repeatedly until a_s takes the place of a_{r+1} . Then $D(A) = (-1)^m D(A') = 0$. ▣

Theorem 4.3. Let $n > 1$, $D : M_{(n-1) \times (n-1)}(K) \rightarrow K$, show

$$E_j(A) = \sum_{i=1}^n (-1)^{i+j} A_{ij} D(\tilde{A}_{ij})$$

is an alternating n -linear function on $M_{n \times n}(K)$, where \tilde{A} is the submatrix obtained from A by deleting row i and col j . If D is a determinant function, then so does E_j .

Proof. (n-linearity) Define row i of A be a_i , B be A with row r replaced by b_r , and A' be A with row r replaced by $a_r + kb_r$. Consider $E_j(a_1, \dots, a_r + kb_r, \dots, a_n)$

$$\begin{aligned}
&= (-1)^{r+j} A'_{rj} D(\tilde{A}'_{rj}) + \sum_{i \neq r} (-1)^{i+j} A'_{ij} D(\tilde{A}'_{ij}) \\
&= (-1)^{r+j} A_{rj} D(\tilde{A}_{rj}) + k(-1)^{r+j} B_{rj} D(\tilde{B}_{rj}) \sum_{i \neq r} (-1)^{i+j} A'_{ij} D(\tilde{A}'_{ij}).
\end{aligned}$$

Now $\tilde{A}_{rj} = \tilde{B}_{rj} = \tilde{A}'_{rj}$ because row r is deleted; $A_{ij} = B_{ij} = A'_{ij}, i \neq r$ are unaffected; $D(\tilde{A}'_{ij}) = D(\tilde{A}_{ij}) + kD(\tilde{B}_{ij}), i \neq r$ by n-linearity. So $E_j(A') = E_j(A) + kE_j(B)$.

(alternativity) Let $a_r = a_{r+1}$, consider $E_j(A)$

$$= \sum_{i \neq r, r+1} (-1)^{i+j} A_{ij} D(\tilde{A}_{ij}) + (-1)^{r+j} A_{rj} D(\tilde{A}_{rj}) + (-1)^{r+1+j} A_{r+1,j} D(\tilde{A}_{r+1,j}).$$

Now $D(\tilde{A}_{ij}) = 0$ by alternativity, and we also have $\tilde{A}_{rj} = \tilde{A}_{r+1,j}$ and $A_{rj} = A_{r+1,j}$ because $a_{r+1} = a_r$. Thus, the last two terms cancel.

(determinant) $E_j(I) = D(I) = 1$. ▣

4.1.3 The Permutation Formulation of Determinant

Definition 4.5 (Parity). Given permutation $\sigma \in S_n$, where S_n is the symmetry group of order n , we define the signature (parity) of a permutation $\text{sgn} : S_n \rightarrow \{1, -1\}$ by $\text{sgn } \sigma = (-1)^m$, where m is the number of transpositions in the decomposition of σ into transpositions. In a moment we will show it is well-defined.

Theorem 4.4. *Prove that the parity of permutations is well-defined. That is, given permutation $\sigma \in S_n$, where S_n is the symmetry group of order n , and $\sigma = \tau_1 \cdots \tau_k = \rho_1 \cdots \rho_\ell$, where τ_i, ρ_i are all transpositions, show k and ℓ are both even or both odd.*

Proof. Consider x_1, \dots, x_n and the quantity

$$\Delta = \prod_{1 \leq i < j \leq n} (x_j - x_i).$$

Define $f_\sigma : \{\Delta, -\Delta\} \rightarrow \{\Delta, -\Delta\}$ by

$$f_\sigma(\Delta) = \prod_{1 \leq i < j \leq n} (x_{\sigma(j)} - x_{\sigma(i)}).$$

Consider τ a transposition, $\tau = (a \ b), 1 \leq a < b \leq n$. Observe that

$$\Delta = (x_b - x_a) \left(\prod_{i \neq a, b} (x_i - x_a)(x_i - x_b) \right) P,$$

where P is a polynomial independent of x_a, x_b . Then we see

$$f_\tau(\Delta) = (x_a - x_b) \left(\prod_{i \neq a, b} (x_i - x_b)(x_i - x_a) \right) P = -\Delta.$$

Moreover, we see

$$f_{\sigma_2 \sigma_1}(\Delta) = \prod_{1 \leq i < j \leq n} (x_{\sigma_2(\sigma_1(j))} - x_{\sigma_2(\sigma_1(i))}) = (f_{\sigma_2} \circ f_{\sigma_1})(\Delta)$$

and $f_\epsilon(\Delta) = \Delta$, where ϵ is the identity permutation.

So if $\epsilon = \alpha_1 \cdots \alpha_k$, $f_\epsilon(\Delta) = \Delta = f_{\alpha_1} \circ \cdots \circ f_{\alpha_k}(\Delta) = (-1)^k \Delta$, which means k is even.

Then $\forall \sigma = \tau_1 \cdots \tau_k = \rho_1 \cdots \rho_\ell$, $\tau_1 \cdots \tau_k \rho_1^{-1} \cdots \rho_\ell^{-1} = \epsilon$, which says $k + \ell$ is even. \square

Theorem 4.5. *If $D : M_{n \times n}(K) \rightarrow K$, D n -linear and alternating, then*

$$D(A) = D(I) \sum_{\sigma} (\text{sgn } \sigma) A_{1,\sigma(1)} \cdots A_{n,\sigma(n)}.$$

Therefore, there is only one determinant function, which is given by

$$\det(A) = \sum_{\sigma} (\text{sgn } \sigma) A_{1,\sigma(1)} \cdots A_{n,\sigma(n)}.$$

And they satisfy

$$D(A) = (\det A) D(I).$$

Proof. Let $A \in M_{n \times n}(K)$ with rows $a_i, 1 \leq i \leq n$. Then $a_i = \sum_{j=1}^n A_{ij} e_j, 1 \leq i \leq n$, e_j is row j of the identity matrix. Then by n -linearity,

$$\begin{aligned} D(A) &= D\left(\sum_{j=1}^n A_{1j} e_j, a_2, \dots, a_n\right) \\ &= \sum_{j=1}^n A_{1j} D(e_j, a_2, \dots, a_n) \\ &= \sum_{j,k} A_{1j} A_{2k} D(e_j, e_k, \dots, a_n) \\ &= \sum_{k_1, k_2, \dots, k_n} A_{1,k_1} A_{2,k_2} \cdots A_{n,k_n} D(e_{k_1}, e_{k_2}, \dots, e_{k_n}). \end{aligned}$$

If $k_i = k_j, 1 \leq i \neq j \leq n$, then $D(e_{k_1}, e_{k_2}, \dots, e_{k_n}) = 0$ by alternativity. So we quickly see that we don't need to add all the k_i , only the permutations of $1, \dots, n$. Thus,

$$D(A) = \sum_{\sigma} A_{1,\sigma(1)} \cdots A_{n,\sigma(n)} D(e_{\sigma(1)}, e_{\sigma(2)}, \dots, e_{\sigma(n)}).$$

Since $D(A) = -D(B)$ under row-exchange, we see transpositions make the sign of D reverse. Thus, $D(e_{\sigma(1)}, e_{\sigma(2)}, \dots, e_{\sigma(n)}) = (\text{sgn } \sigma) D(I)$. Thus

$$D(A) = D(I) \sum_{\sigma} (\text{sgn } \sigma) A_{1,\sigma(1)} \cdots A_{n,\sigma(n)}.$$

Since determinant functions require $D(I) = 1$, we see there is only one such function. \square

4.2 Properties

Theorem 4.6. *Let $A, B \in M_{n \times n}(K)$, then $\det(AB) = (\det A)(\det B)$.*

Proof. Let $D(A) = \det(AB)$. When viewed in rows, $D(A) = \det(a_1 B, \dots, a_n B)$ is obviously an n -linear alternating function. So $D(A) = (\det A)D(I) = (\det A)(\det B)$. \square

Theorem 4.7. *Let $A, B \in M_{n \times n}(K)$, and B is obtained from A by adding a scalar multiple of a row to another row. Show $\det A = \det B$.*

Proof. Follows immediately from n -linearity and alternativity. \square

Theorem 4.8. *Show $\det(A^\top) = \det A$.*

Proof.

$$\det(A^\top) = \sum_{\sigma} (\text{sgn } \sigma) A_{1, \sigma(1)}^\top A_{2, \sigma(2)}^\top \cdots A_{n, \sigma(n)}^\top.$$

Add it another way. If originally we add it in the sequence of $\sigma_1, \sigma_2, \dots, \sigma_{n!}$, we now add it in $\sigma_1^{-1}, \sigma_2^{-1}, \dots, \sigma_{n!}^{-1}$. Then

$$\begin{aligned} \det(A^\top) &= \sum_{\sigma^{-1}} A_{\sigma^{-1}(1), 1}^\top A_{\sigma^{-1}(2), 2}^\top \cdots A_{\sigma^{-1}(n), n}^\top \\ &= \sum_{\sigma^{-1}} A_{1, \sigma^{-1}(1)} A_{2, \sigma^{-1}(2)} \cdots A_{n, \sigma^{-1}(n)} \\ &= \det A. \end{aligned}$$

\square

Problem 5.1.7. *Let $T \in \mathcal{L}(V)$. Define $\det T = \det([T]_\beta)$. Show $\det([T]_\beta) = \det([T]_\gamma)$.*

Proof. $\det([T]_\beta) = \det([I_V]_\gamma^\beta [T]_\gamma [I_V]_\beta^\gamma) = \det Q^{-1} \det [T]_\gamma \det Q = \det [T]_\gamma$. \square

4.3 Determinants of Special Matrices

Problem 4.3.0. *The determinant of an upper triangular matrix is the product of its diagonal entries. i.e., if A $n \times n$ satisfies $a_{ij} = 0 \forall i > j$, then $\det A = \prod_{i=1}^n a_{ii}$.*

Proof. Induction on n .

$n = 2$ is trivial.

Suppose it holds for $n = k$, consider $n = k + 1$. The result follows immediately if we expand the determinant along the first col. \square

Problem 4.3.9. *An upper triangular $n \times n$ matrix is invertible \iff all its diagonal entries are non-zero.*

Definition 4.6. A matrix $M \in M_{n \times n}(\mathbb{C})$ is nilpotent if for some positive integer k , $M^k = O_{n \times n}$. It's obvious from the multiplicative property of determinants (**Theorem 4.6**) that $\det M = 0$.

Problem 4.3.11. *Prove if M is skew-symmetric and n is odd, then $\det M = 0$.*

Proof. $M^\top = -M \implies \det(M^\top) = (-1)^n \det M$. If $n \in \text{odd} \implies \det M = -\det M \implies \det M = 0$. \square

Problem 4.3.13. *Let $M \in M_{n \times n}(\mathbb{C})$, and define $(\overline{M})_{ij} = \overline{(M_{ij})}$.*

1. *Show $\det \overline{M} = \overline{\det M}$.*
2. *$Q \in M_{n \times n}(\mathbb{C})$ is unitary if $QQ^* = I$, $Q^* = \overline{Q}^\top$. Show if Q is unitary, then $|\det Q| = 1$.*

Proof. (1) Induction on n . $n = 1$ is trivial.

Suppose it holds for $n = k$, consider $n = k + 1$.

From the definition of determinant, we have $\det \overline{M} = \sum_{j=1}^n (-1)^{i+j} \overline{A_{ij}} \det(\overline{\tilde{A}_{ij}})$.

But by induction hypothesis, since \tilde{A}_{ij} is $(k-1) \times (k-1)$, $\det(\overline{\tilde{A}_{ij}}) = \overline{\det(\tilde{A}_{ij})}$.

So $\det \overline{M} = \sum_{j=1}^n \overline{(-1)^{i+j} A_{ij} \det(\tilde{A}_{ij})} = \overline{\det M}$.

(2) $QQ^* = I \implies \det Q \det \overline{Q}^\top = 1 \implies \det Q \det \overline{Q} = 1 \implies |\det Q|^2 = 1 \implies |\det Q| = 1$. \square

Problem 4.3.20. *Suppose $M \in M_{n \times n}(F)$ can be written in the form*

$$M = \begin{pmatrix} A & B \\ O & I \end{pmatrix}$$

where A is square. Show $\det M = \det A$.

Proof. Doing row operations using the rows of I , we can eliminate B ultimately to get

$$\det M = \det \begin{pmatrix} A & O_1 \\ O_2 & I \end{pmatrix} = \det A$$

by cofactor expansion along any column of I . ▣

Problem 4.3.21. Suppose $M \in M_{n \times n}(F)$ can be written in the form

$$M = \begin{pmatrix} A & B \\ O & C \end{pmatrix}$$

where A, C is square. Show $\det M = \det A \cdot \det C$.

Proof. Using block multiplication (**Problem 4.3.21**), we have

$$\begin{pmatrix} A & B \\ O & C \end{pmatrix} = \begin{pmatrix} I & O \\ O & C \end{pmatrix} \cdot \begin{pmatrix} A & B \\ O & I \end{pmatrix}$$

Apply multiplicity of determinants (**Theorem 4.6**), we get the desired result. ▣

Proof. Let C be $s \times s$, and

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

We see D is s -linear in the rows of C , therefore $D(A, B, C) = (\det C)D(A, B, I)$. By the previous problem we quickly see that $D(A, B, C) = (\det C)(\det A)$. ▣

Problem 4.3.24. Let

$$A = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & a_0 \\ -1 & 0 & 0 & \cdots & 0 & a_1 \\ 0 & -1 & 0 & \cdots & 0 & a_2 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & a_{n-1} \end{pmatrix}$$

Compute $\det(A + tI_n)$.

Proof. ($n = 3$) Consider the end case.

$$A + tI_3 = \begin{pmatrix} t & 0 & a_0 \\ -1 & t & a_1 \\ 0 & -1 & a_2 + t \end{pmatrix}$$

So $\det(A + tI) = a_0 + t(a_1 + t(a_2 + t))$.

($n = n$) Consider the general case. Expanding along the first row, we have

$$\det(A + tI_n) = t \det \begin{pmatrix} t & \cdots & a_1 \\ -1 & t & \cdots & a_2 \\ & -1 & t & \cdots & a_3 \\ & & \ddots & \ddots & \vdots \\ & & & a_{n-1} + t \end{pmatrix} \\ + (-1)^{1+n} a_0 \det \begin{pmatrix} -1 & t & & & \\ & -1 & t & & \\ & & \ddots & \ddots & \\ & & & -1 & t \\ & & & & -1 \end{pmatrix}$$

Where the determinant in the second term can be worked out using row operation to be $\det(-I_{n-1})$, so the second term is $a_0(-1)^{n+1}(-1)^{n-1} = a_0$.

So we see the whole thing is

$$\det(A + tI_n) = a_0 + t(a_1 + t(a_2 + t(\cdots + t(a_{n-1} + t)))) \\ = t^n + a_{n-1}t^{n-1} + a_{n-2}t^{n-2} + \cdots + a_1t + a_0$$

▣

Problem Hoffman 5.4.12. Let $V = M_{n \times n}(F)$, $B \in V$. Define $L_B, R_B \in \mathcal{L}(V)$ by $L_B(A) = BA$, $R_B(A) = AB$. Show $\det L_B = (\det B)^n = \det R_B$.

Proof. Let $\beta = e_{ij}$, $1 \leq i, j \leq n$, where e_{ij} is the matrix s.t. it has a 1 in row i , col j but 0 otherwise.

We see $L_B(e_{ij}) = Be_{ij} = \sum_{k=1}^n B_{ki}e_{kj}$. The j doesn't change, by virtue of **Theorem 2.20**. This motivates us to define $\beta_i = \{e_{1i}, e_{2i}, \dots, e_{ni}\}$ and $W_i = \text{span } \beta_i$. We see W_i is L_B -invariant, and therefore it's obvious that $V = \bigoplus_{i=1}^n W_i$ and that $[L_B]_{\beta_i} = B$.

Thus $[L_B]_{\beta} = \bigoplus_{i=1}^n B$, $\det[L_B]_{\beta} = (\det B)^n$. ▣

4.3.1 Wronskian

Definition 4.7. Let y_1, y_2, \dots, y_n be linearly independent functions in C^∞ . $\forall y \in C^\infty$, define the Wronskian $T(y) \in C^\infty$ by

$$(T(y))(t) = \det \begin{pmatrix} y(t) & y_1(t) & y_2(t) & \cdots & y_n(t) \\ y'(t) & y_1'(t) & y_2'(t) & \cdots & y_n'(t) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y^{(n)}(t) & y_1^{(n)}(t) & y_2^{(n)}(t) & \cdots & y_n^{(n)}(t) \end{pmatrix}$$

Problem 4.3.28.1. Prove $T : C^\infty \rightarrow C^\infty$ is linear.

Proof. Trivial by n-linearity. ▣

Problem 4.3.28.2. Prove $N(T) = \text{span}(\{y_1, y_2, \dots, y_n\})$.

Proof.

(\supseteq) Noting that $S = \{y_1, \dots, y_n\} \subseteq N(T)$ by alternativity. By [Theorem 1.5](#), $\text{span } S \subseteq N(T)$.

(\subseteq) Noting that $\det W = 0 \iff$ cols dependent. Since $S = \{y_1, y_2, \dots, y_n\}$ is independent, and $S \cup \{y\}$ dependent, by [Theorem 1.7](#), $y \in \text{span } S$. So $N(T) \subseteq \text{span } S$. ▣

4.4 Classical Adjoint

4.4.1 Definition

Definition 4.8 (Classical Adjoint). The classical adjoint $\text{adj } A$ of a matrix A is defined as the transpose of the cofactor matrix C , where $C_{ij} = c_{ij}$ is the cofactor of A at position (i, j) .

✂ *Remark.* The name of classical adjoint is used to distinguish it from the adjoint operator, which for a matrix means the conjugate transpose.

The classical adjoint is also known as the adjugate matrix. ▣

4.4.2 Classical Adjoint and Inverse

Theorem 4.9. Show

$$\sum_{i=1}^n A_{ik} C_{ij} = \delta_{jk} \det A,$$

where C_{ij} is the cofactor of A at position i, j .

Proof. Define a_i to be col i of A , and let B be A with col j replaced by col k . That is, $A = (a_1, \dots, a_j, \dots, a_k, \dots, a_n)$, $B = (a_1, \dots, a_k, \dots, a_k, \dots, a_n)$.

By cofactor expansion, we see $\det A = \sum_{i=1}^n A_{ij} C_{ij}$.

By cofactor expansion along col j , we see $\det B = 0 = \sum_{i=1}^n B_{ij} C'_{ij} = \sum_{i=1}^n A_{ik} C_{ij}$, where $C'_{ij} = C_{ij}$ because col j is deleted. Combine the two cases, we have the desired result. ▣

Theorem 4.10. Let $A \in M_{n \times n}(K)$. Then A is invertible over $K \iff \det A$ is invertible in K . When A is invertible, $A^{-1} = (\det A)^{-1} \operatorname{adj} A$.

Proof.

$$\begin{aligned} ((\operatorname{adj} A)A)_{ij} &= \sum_{k=1}^n (\operatorname{adj} A)_{ik} A_{kj} \\ &= \sum_{k=1}^n A_{kj} C_{ki} \\ &= \delta_{ij} \det A. \end{aligned}$$

To prove the other side, take the transpose,

$$((\operatorname{adj} A)^\top)_{ij} = (\operatorname{adj} A)_{ji} = C_{ij} = C'_{ji} = (\operatorname{adj}(A^\top))_{ij}.$$

Thus $A^\top \operatorname{adj}(A^\top) = (\det A^\top)I$, let $M = A^\top$ completes the proof. \square

Theorem 4.11 (Cramer's Rule). Let $Ax = b$, where $A \in M_{n \times n}(K)$ and is invertible over K . Then the k -th component of x is given by,

$$x_k = (\det A)^{-1} (\det M_k),$$

where M_k is A with col k replaced by b .

Proof. $Ax = b$, thus $((\operatorname{adj} A)A)x = (\operatorname{adj} A)b$. Then $(\det A)x = (\operatorname{adj} A)b$. So the k -th component is given by

$$\begin{aligned} (\det A)x_k &= ((\operatorname{adj} A)b)_k = \sum_{i=1}^n (\operatorname{adj} A)_{ki} b_i \\ &= \sum_{i=1}^n C_{ik} b_i = \det M_k. \end{aligned}$$

where the last equality comes from cofactor expanding M_k along the k -th column. \square

Problem 4.3.27.1. Let $C = \operatorname{adj} A$, $A \in M_{n \times n}(F)$. Then $\det C = (\det A)^{n-1}$.

Proof. Since $A \cdot (\frac{1}{\det A} C) = I \implies \det(A) \cdot \det(\frac{1}{\det A} C) = 1 \implies \det(A) \cdot \frac{1}{(\det A)^n} \det C = 1 \implies \det C = (\det A)^{n-1}$. \square

Problem 4.3.27.2. Let $C = \operatorname{adj} A$, $A \in M_{n \times n}(F)$. Then $C^\top = \operatorname{adj} A^\top$.

Proof. Given $A \cdot (\frac{1}{\det A} C) = I \implies (\frac{1}{\det A^\top} C^\top) A^\top = I$. By [Problem 2.4.10](#), we see $A^\top (\frac{1}{\det A^\top} C^\top) = I$. \square

Problem 4.3.27.3. If A is invertible, upper triangular, then $C = \text{adj } A$ and A^{-1} are also upper triangular.

Proof. If A is upper triangular, then \tilde{A}_{ji} is also upper triangular. When $i > j$, \tilde{A}_{ji} has at least one zero on the diagonal, so $\det \tilde{A}_{ji} = 0 \implies C_{ij}$ is upper triangular, so does A^{-1} . \blacksquare

5 Polynomials

5.1 Foreword

In this chapter we will define polynomial over a field as formal power series, and establish an isomorphism between polynomials and polynomial functions. Most importantly, we are concerned with the polynomial ideal, which facilitates the development of factorization and common divisors.

5.2 The Algebraic Object of Study

Definition 5.1 (Linear Algebra). Let F be a field. A linear algebra over the field F is a vector space V over F with an additional operation called multiplication $(\alpha, \beta) \rightarrow \alpha\beta \in V$ called the product s.t.

1. $\alpha(\beta\gamma) = (\alpha\beta)\gamma$.
2. $\alpha(\beta + \gamma) = \alpha\beta + \alpha\gamma$, $(\alpha + \beta)\gamma = \alpha\gamma + \beta\gamma$.
3. $\forall c \in F, c(\alpha\beta) = (c\alpha)\beta = \alpha(c\beta)$.

Definition 5.2. The algebra of formal power series F^∞ is the set of all functions $f: \mathbb{Z}^+ \rightarrow F$. We represent a vector in F^∞ by $f = (f_0, f_1, f_2, \dots)$, implying that $f(i) = f_i, i \in \mathbb{Z}^+$. We define linear combination to be component-wise, and multiplication by convolution. I.e.,

$$af + bg = (af_0 + bg_0, af_1 + bg_1, \dots)$$

$$(fg)_n = \sum_{i=0}^n f_i g_{n-i}, n = 0, 1, 2, \dots$$

In such a way, the algebra of formal power series is commutative. We also define $1 = (1, 0, \dots), x = (0, 1, 0, \dots)$.

Definition 5.3. Let $F[x]$ be the subspace of F^∞ spanned by the vectors $1, x, x^2, \dots$. An element of $F[x]$ is called a polynomial over F .

Definition 5.4. The degree of a polynomial is the unique integer $n \geq 0$ s.t. $f_n \neq 0$ and $f_k = 0 \forall k > n$. If $f = cx^0$, then f is called a scalar polynomial, and we do not assign a degree to the zero polynomial. If $\deg f = n$ and $f_n = 1$, then f is a monic polynomial.

Theorem 5.1. Let $f, g \in F[x]$, $f, g \neq 0$.

1. $fg \neq 0$.
2. $\deg(fg) = \deg f + \deg g$.
3. fg is monic if both f and g are monic.
4. fg is scalar \iff both f and g are scalar.
5. if $f + g \neq 0$, $\deg(f + g) \leq \max(\deg f, \deg g)$.

Corollary 5.1.1. Suppose $f, g, h \in F[x]$ s.t. $f \neq 0$, $fg = fh$. Then $g = h$.

Definition 5.5. Let V be a linear algebra with identity over F . Denote the identity of V by 1 and order that $\alpha^0 = 1 \forall \alpha \in V$. Then $\forall f = \sum_{i=0}^n f_i x^i \in F[x]$, $\alpha \in V$, we define

$$f(\alpha) = \sum_{i=0}^n f_i \alpha^i.$$

Theorem 5.2. Suppose $f, g \in F[x]$, $\alpha \in V$, $c \in F$. Then

1. $(cf + g)(\alpha) = cf(\alpha) + g(\alpha)$.
2. $(fg)(\alpha) = f(\alpha)g(\alpha)$.

Proof. (2) Let $f = \sum_{i=0}^m f_i x^i$, $g = \sum_{j=0}^n g_j x^j$.

Thus $fg(\alpha) = \sum_{i,j} f_i g_j \alpha^{i+j} = (\sum_{i=0}^m f_i \alpha^i) (\sum_{j=0}^n g_j \alpha^j)$. ▣

Theorem 5.3. Let $T \in \mathcal{L}(V)$, \mathcal{A} be the linear algebra of powers of T over F , $f, g \in F[x]$, and $\alpha \in V$. Then

$$((fg)(T))(\alpha) = f(T)((g(T))(\alpha)) = f(T) \circ g(T)(\alpha)$$

Proof. Let $f = \sum_{i=0}^m f_i x^i$, $g = \sum_{i=0}^n g_i x^i$. By definition, $((fg)(T))(\alpha) = (f(T) \cdot$

$g(T))(\alpha)$. Now $\forall k \in \mathbb{Z}^+ \cup \{0\}$, $(T^k \cdot g(T))(\alpha) = T^k((g(T))(\alpha))$. So

$$\begin{aligned}
(f(T) \cdot g(T))(\alpha) &= \left(\left(\sum_{i=1}^m f_i T^i \right) \cdot g(T) \right) (\alpha) \\
&= \left(\sum_{i=1}^m f_i T^i g(T) \right) (\alpha) \\
&= \sum_{i=1}^m f_i (T^i g(T)) (\alpha) \\
&= \sum_{i=1}^m f_i T^i ((g(T))(\alpha)) \\
&= f(T) ((g(T))(\alpha)).
\end{aligned}$$

▣

5.3 Ideals

Definition 5.6 (Ideals). An ideal in $F[x]$ is a subspace M of $F[x]$ s.t. $fg \in M \forall f \in F[x], g \in M$.

✂ *Remark.* Should we not in commutative algebra, we will distinguish between a left ideal (given above), a right ideal $gf \in M$, and a two-sided ideal. \mathfrak{M}

Definition 5.7. The principle ideal generated by d is the set $M = dF[x]$. The ideal generated by d_1, d_2, \dots, d_n is the sum of subspaces $M_i = d_i F[x]$.

Theorem 5.4. Suppose $f, d \in F[x], f, d \neq 0, \deg d \leq \deg f$. Then $\exists g \in F[x]$ s.t. either $f - dg = 0$ or $\deg(f - dg) < \deg f$.

Proof. Let $f = \sum_{i=0}^n f_i x_i, d = \sum_{i=0}^m d_i x_i$. Let $g = f_n (d_m)^{-1} x^{n-m}$ completes the proof. \square

Theorem 5.5. Let $f, d \in F[x], d \neq 0$, then $\exists q, r \in F[x]$ s.t. $f = dq + r$ and either $r = 0$ or $\deg r < \deg d$. If $r = 0$, then we say d divides f .

Proof. (1) If $\deg f < \deg d$, then let $q = 0, r = f$.

(2) If $\deg f \geq \deg d$, then by **Theorem 5.4**, either $f - dq = 0 = r$ or $\deg(f - dq) := \deg f_2 < \deg f$. Repeatedly apply this until $\deg f_n < \deg d$. Let $r = f_n$ completes the proof.

(uniqueness) Let $f = dq + r = dq' + r'$. Then $d(q - q') + (r - r') = 0$. But

$$\begin{aligned}\deg(r - r') &\leq \max(\deg r, \deg r') \\ &< \deg d \\ &< \deg d + \deg(q - q').\end{aligned}$$

So $(r - r')$ cannot cancel the high order terms of $d(q - q')$, but the sum is zero, so $d(q - q') = (r - r') = 0$. \square

Theorem 5.6. *Let M be any non-zero ideal in $F[x]$. Then $\exists! d \in F[x], d$ monic s.t. $M = dF[x]$.*

Proof. Let $d \in M$ s.t. $\deg d \leq \deg g \forall g \in M$. We claim $M = dF[x]$.

Obviously $dF[x] \subseteq M$. $\forall f \in M, \exists q, r \in F[x]$ s.t. $f = dq + r, \deg r < \deg d$. Since $d \in M, dq \in M$ also. Then $f - dq = r \in M$ also, since M is a subspace. But this contradicts d is the polynomial of minimal degree. We conclude that $r = 0, f = dq$.

(uniqueness) If $M = dF[x] = gF[x]$, then $\exists p, q \neq 0$ s.t. $d = gp, g = dp$. Thus $d = dpq \implies \deg d = \deg d + \deg p + \deg q \implies \deg p = \deg q = 0$. Since d, g are monic, $p = q = 1 \implies d = g$. \square

Corollary 5.6.1. *Let $p_1, \dots, p_n \in F[x]$ not all zero, then $\exists! d \in F[x]$ s.t.*

1. $d \in p_1F[x] + \dots + p_nF[x]$.
2. d divides $p_i, 1 \leq i \leq n$.
3. (1) and (2) implies d is divisible by any polynomial that divides each of the polynomial p_1, \dots, p_n .

Proof. (3) If $p_i = fq'_i, f \in F[x]$, then $\exists s_i$ s.t.

$$d = \sum_i p_i s_i = \sum_i fq'_i s_i = f \sum_i q'_i s_i.$$

\square

Problem Hoffman 4.3.3. *Let $A \in M_{n \times n}(F)$. Show $M = \{ f \in F[x] \mid f(A) = 0 \}$ is an ideal.*

Proof. (1) $\forall f, g \in M, f(A) = g(A) = 0$. Then $(cf + g)(A) = cf(A) + g(A) = 0$.

(2) $\forall f \in F[x], g \in M, (fg)(A) = f(A)g(A) = 0$. \square

6 Diagonalization

6.1 Foreword

This section concerns primarily about what is diagonalization and when can a linear operator be diagonalized. The diagonal form of linear operators is often called the elementary canonical form.

6.2 Definition of Diagonalization

Definition 6.1 (Diagonalizability). A linear operator T on finite-dimensional vector space V is diagonalizable if $\exists \beta$ a basis of V s.t. $[T]_\beta$ is a diagonal matrix.

A matrix is diagonalizable if L_A is diagonalizable.

Definition 6.2 (Eigenvalue and eigenvector). Let $T \in \mathcal{L}(V)$, $v \neq 0$ is called an eigenvector if $T(v) = \lambda v$. λ is called the corresponding eigenvalue.

✂ *Remark.* Do not confuse eigenvalue and diagonalizable. An operator can exist eigenvalues and eigenvectors but may not be diagonalizable.

When doing problems, pay special attention to whether the operator is diagonalizable, or it just have eigenvalues.

The zero vector is never an eigenvector, but zero can be an eigenvalue. \mathfrak{M}

Theorem 6.1. Let $A \in M_{n \times n}(F)$. Then a scalar λ is an eigenvalue of $A \iff \det(A - \lambda I_n) = 0$.

Proof. $\exists v$ s.t. $Av = \lambda v \iff Av - \lambda Iv = 0 \iff (A - \lambda I)v = 0 \iff$ cols of $A - \lambda I$ dependent $\iff \det(A - \lambda I) = 0$. \blacksquare

Theorem 6.2. Let $T \in \mathcal{L}(V)$, λ be an eigenvalue of T . $v \in V$ is an eigenvector of T corresponding to $\lambda \iff v \in N(T - \lambda I_V)$, $v \neq 0$.

Proof. $T(v) = \lambda v \iff (T - \lambda I_V)(v) = 0 \iff v \in N(T - \lambda I_V)$ \blacksquare

Problem 5.1.15. Let $V \in \mathcal{L}(V)$, x be a eigenvector of T corresponding to the eigenvalue λ . Show $\forall m \in \mathbb{Z}^+$, x is an eigenvector of T^m corresponding to the eigenvalue λ^m .

6.3 Characteristic Polynomial

6.3.1 Definition and Invariance

Definition 6.3 (Characteristic Polynomial). Let $T \in \mathcal{L}(V)$, β be a basis of V . $f(t) = \det(tI_n - [T]_\beta)$ is called the characteristic polynomial of operator T . It is independent of the choice of basis, as we will show in a moment.

Let $A \in M_{n \times n}(F)$. $f(t) = \det(tI_n - A)$ is called the characteristic polynomial of the matrix A .

Problem 5.1.12. *Show that the characteristic polynomial of an operator $T \in \mathcal{L}(V)$ is independent of the choice of basis.*

Proof. Let $T \in \mathcal{L}(V)$, β, γ be two bases of V , $A = [T]_\beta$, $Q = [I_V]_\gamma^\beta$. Then by **Theorem 2.26**, we see $[T]_\gamma = Q^{-1}AQ$.

So the characteristic polynomial in the basis β is $f_\beta(t) = \det(tI_n - A)$, while in the basis γ is $f_\gamma(t) = \det(tI_n - Q^{-1}AQ) = \det(Q^{-1}(tI_n)Q - Q^{-1}AQ) = \det(Q^{-1}(tI_n - A)Q) = \det(Q^{-1}) \det(tI_n - A) \det(Q) = f_\beta(t)$. \square

Problem 5.1.14. *Show $\forall A \in M_{n \times n}(F)$, the characteristic polynomial of A and A^\top are the same.*

Proof. $f_A(t) = \det(tI - A) = \det((tI - A)^\top) = \det(tI - A^\top) = f_{A^\top}(t)$. \square

6.3.2 Coefficients

Theorem 6.3. *Let $A \in M_{n \times n}(F)$. The characteristic polynomial of A is a monic polynomial of degree n , so A has at most n distinct eigenvalues.*

Proof. Induction on n . $n = 1$ is trivial.

Suppose $n = k$ is true, consider $n = k + 1$.

$$\det(tI_{k+1} - A_{k+1}) = (t - a_{11}) \det(tI_k - \tilde{A}_{k+1}) + a_{21}(\dots) + \dots$$

When cofactoring on an off-diagonal element, we at least deleted two of the diagonal elements, so besides the first term, all the remaining terms are of degree $k - 1$ or lower.

But by the induction hypothesis, the first term is

$$\det(tI_{k+1} - A_{k+1}) = (t - a_{11})(t^k + \dots) + \dots = t^{k+1} + \dots$$

and therefore completed the induction. ▣

Problem 5.1.21.1. Show if $A \in M_{n \times n}(F)$, the characteristic polynomial of A satisfies

$$f(t) = \prod_{i=1}^n (t - A_{ii}) + q(t)$$

where $\deg q \leq n - 2$.

Proof. Induction on n . $n = 2$ is trivial. Suppose $n = k$ is true, consider $n = k + 1$.

$$\det(tI_{k+1} - A_{k+1}) = (t - A_{11}) \det(tI_k - \tilde{A}_1) + A_{21}q_{21}(t) + \cdots + A_{n1}q_{n1}(t).$$

When cofactoring on an off-diagonal element, we delete at least two t 's from the diagonal. So $\deg q_{n1}(t) \leq k + 1 - 2 = k - 1$. So by the induction hypothesis,

$$\begin{aligned} \det(A_{k+1} - tI_{k+1}) &= (t - A_{11})((t - A_{22}) \cdots (t - A_{nn}) + q(t)) + \sum_{j=2}^n A_{j1}q_{j1}(t) \\ &= \prod_{j=1}^n (t - A_{jj}) - A_{11}q(t) + tq(t) + \sum_{j=2}^n A_{j1}q_{j1}(t) \end{aligned}$$

▣

Problem 5.1.20. Let $A \in M_{n \times n}(F)$, its characteristic polynomial be $f(t) = t^n + a_{n-1}t^{n-1} + \cdots + a_1t + a_0$. Show $\det A = (-1)^n a_0$, and $\text{tr}(A) = -a_{n-1}$.

Proof. $a_0 = f(0) = \det A$ trivial.

By **Problem 5.1.21.1**, we see $f(t) = \prod_{i=1}^n (A_{ii} - t) + q(t)$, $\deg q \leq n - 2$. So the coefficient a_{n-1} comes entirely from the product term. By Vieta, we see $\text{tr}(A) = -\sum_{i=1}^n A_{ii} = -a_{n-1}$. ▣

Problem 5.1.16. Show similar matrices have the same trace. Therefore on a finite-dimensional vector space V , it is possible to define the trace of a linear operator by $\text{tr}(T) = \text{tr}([T]_\beta)$, where β is a basis of V .

Proof. $\text{tr}(B^{-1}AB) = \text{tr}((B^{-1})(AB)) = \text{tr}(ABB^{-1}) = \text{tr}(A)$. ▣

6.4 Diagonalizability

Problem 5.1.22. Let $T \in \mathcal{L}(V)$ over F , and $g(x) \in F[x]$. Show if v is an eigenvector of T with corresponding eigenvalue λ , then $g(T)(v) = g(\lambda)v$, which means v is an eigenvector of $g(T)$ with eigenvalue $g(\lambda)$.

Proof. Let $g(x) = \sum_{j=0}^n g_j x^j$.

Then $(g(T))(v) = \left(\sum_{j=0}^n g_j T^j\right)(v) = \sum_{j=0}^n g_j (T^j(v)) = \sum_{j=0}^n g_j \lambda^j v = g(\lambda)v$ by **Problem 5.1.15**. ▮

Theorem 6.4. Let $T \in \mathcal{L}(V)$, c_1, \dots, c_k be distinct eigenvalues of T , and W_i be the space of eigenvectors corresponding to c_i . If $W = \sum_{i=1}^k W_i$, show

$$\dim W = \sum_{i=1}^k \dim W_i.$$

So that if β_i is a basis of W_i , then $\beta = \bigcup_{i=1}^k \beta_i$ is a basis of W .

Proof. Yes I know direct sums, but I want to show the elegant proof of Hoffman.

Suppose $\sum_{i=1}^k \beta_i = 0$, we will show (elegantly) that $\beta_i = 0, 1 \leq i \leq k$. Let $f \in F[x]$,

$$\begin{aligned} 0 &= f(T)(0) = f(T)(\beta_1) + \dots + f(T)(\beta_k) \\ &= f(c_1)(\beta_1) + \dots + f(c_k)(\beta_k). \end{aligned}$$

Since we don't restrict the degree of the polynomial, we can easily construct $f_i(c_j) = \delta_{ij}$. Applying $f_i, 1 \leq i \leq k$ on both sides, we get the desired result.

So the eigenspaces are all independent, and since W is a sum of all spaces, we arrive easily at $\beta = \bigcup_{i=1}^k \beta_i$ is a basis of W , and the dimension condition. ▮

Theorem 6.5. Let $T \in \mathcal{L}(V)$, c_1, \dots, c_k be distinct eigenvalues of T , and W_i be $N(T - c_i I)$. Then the following are equivalent.

1. T is diagonalizable.
2. $f(x) = \prod_{i=1}^k (x - c_i)^{d_i}$, $\dim W_i = d_i, 1 \leq i \leq k$.
3. $\dim V = \sum_{i=1}^k \dim W_i$.

Proof. (1 \Rightarrow 2) If T is diagonalizable, choose β s.t. $[T]_\beta$ is diagonal. Since c_i are distinct, clearly $\dim N(T - c_i I) = d_i$.

(2 \Rightarrow 3) By [Theorem 6.4](#), we see the subspace $W = \sum_{i=1}^k W_i$ has dimension $\dim W = \sum_{i=1}^k \dim W_i = \sum_{i=1}^k d_i = n = \dim V$ by the degree of characteristic polynomial, [Theorem 6.3](#). Therefore (3) is true.

(3 \Rightarrow 1) Since by [Theorem 6.4](#), eigenspaces are independent, and by (2) their dimensions sum up to n , we have $V = \sum_{i=1}^k W_i$. \square

✍ *Remark.* For another criteria in terms of minimal polynomials, see [Theorem 7.7](#). \square

6.5 Upper Triangular and Scalar Matrices

Definition 6.4 (Scalar Matrix). Matrix $A \in M_{n \times n}(F)$ is called scalar matrix if $A = \lambda I$ for some λ .

Problem 5.1.9. Show the eigenvalues of an upper triangular matrix M are the diagonal entries of M .

Problem 5.1.10. Show that for any basis β , $[\lambda I_V]_\beta = \lambda I$, and show that λI_V is diagonalizable and has only one eigenvalue.

Proof. $f(t) = \det([\lambda I_V]_\beta - tI) = \det(\lambda I - tI) = (-1)^n(t - \lambda)^n$. \square

Problem 5.1.11.1. Show if some square matrix A is similar to a scalar matrix λI , then $A = \lambda I$.

Proof. $B^{-1}AB = \lambda I \Rightarrow AB = B\lambda I = \lambda B \Rightarrow A = \lambda I$. \square

Problem 5.1.11.2. Show a diagonalizable matrix having only one eigenvalue is a scalar matrix.

Proof. Suppose A is diagonalizable, that means $\exists \gamma$ s.t. $[L_A]_\gamma$ is diagonal matrix. Let β be the standard basis of F^n . $A = [L_A]_\beta$. Thus $[L_A]_\gamma = [I_V]_\beta^\gamma [L_A]_\beta [I_V]_\gamma^\beta := Q^{-1}AQ$.

Now $f(t) = \det([L_A]_\gamma - tI)$ has only one solution. Then $f(t) = (a_{11} - t)(a_{22} - t) \cdots (a_{nn} - t) \Rightarrow a_{11} = a_{22} = \cdots = a_{nn} = \lambda$. In this way, we see $[L_A]_\gamma = \lambda I$, so $A = [L_A]_\beta$ is similar to a scalar matrix, so $A = \lambda I$. \square

Problem 5.1.11.3. Show $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ is not diagonalizable.

Proof. Using [Problem 5.1.11.2](#), since all the eigenvalues of $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ is 1 yet A is not a scalar matrix, we see it must not be diagonalizable. \square

6.6 Examples

Problem 5.1.17. Let T be the linear operator on $M_{n \times n}(\mathbb{R})$ by $T(A) = A^\top$. Show ± 1 are the only eigenvalues of T . Describe the eigenvectors, and diagonalize T .

Proof. If $A^\top = kA$, then consider $(A^\top)^\top = k^2A = A$, so $k^2 = 1$.

$\lambda = 1$ corresponds to symmetric matrices; $\lambda = -1$ corresponds to skew-symmetric matrices. Together they form a direct sum decomposition of V , by [Problem 1.3.28](#). Therefore, any union of their bases form a basis of $M_{n \times n}(\mathbb{R})$ by [Problem 1.6.33](#).

For 2×2 case, let

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

therefore the coordinate representation of any matrix is

$$[A]_\beta = (a, b, c, d) \implies A = \begin{pmatrix} a & c+d \\ c-d & b \end{pmatrix}$$

therefore

$$T(A) = A^\top = \begin{pmatrix} a & c-d \\ c+d & b \end{pmatrix} \implies [T(A)]_\beta = (a, b, c, -d)$$

thus the matrix representation of T is

$$[T]_\beta = \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ & & & -1 \end{pmatrix}.$$

▣

7 Annihilating Polynomials

7.1 Minimal Polynomial

Definition 7.1 (Annihilators). The annihilators of $T \in \mathcal{L}(V)$ is the set $M = \{f \in F[x] \mid f(T) = 0\}$. By [Problem Hoffman 4.3.3](#), we see M is an ideal.

Definition 7.2 (Minimal Polynomial). The minimal polynomial of $T \in \mathcal{L}(V)$ is the unique monic generator of the ideal of annihilators of T . Its existence is guaranteed by [Theorem 5.6](#).

✎ *Remark.* Since $f(P^{-1}AP) = P^{-1}f(A)P$, the minimal polynomial is independent of the choice of basis. \mathfrak{M}

7.2 Cayley-Hamilton

Theorem 7.1. *Let $T \in \mathcal{L}(V)$, $\dim V = n$. Then the characteristic polynomial f and minimal polynomial p of T have the same roots, except for multiplicities.*

Proof. ($f(c) = 0 \implies p(c) = 0$) Since $f(c) = 0 \iff T(v) = cv$, then $p(T)(v) = p(c)v = 0$ by [Problem 5.1.22](#). But $v \neq 0$, so $p(c) = 0$.

($p(c) = 0 \implies f(c) = 0$) $p(c) = 0 \iff p = (x - c)f$ and $f(T) \neq 0$. Let $\alpha = f(T)\beta \neq 0$. Then

$$\begin{aligned} (p(T))(\beta) &= 0 = ((x - c)(T) \cdot f(T))(\beta) \\ &= (T - cI)((f(T))(\beta)) \\ &= T(\alpha) - c\alpha, \end{aligned}$$

by [Theorem 5.3](#). So α is an eigenvector, and $f(c) = 0$. ▣

Theorem 7.2 (Cayley-Hamilton). *Let $T \in \mathcal{L}(V)$, f be the characteristic polynomial of T , p be the minimal polynomial of T . Then $f(T) = 0$, i.e., p divides f .*

Proof. Let $K = F[T]$. That is, we consider the commutative ring consisting of polynomials, but this time over the linear algebra of powers of T . We sort of directly "plug T into" $f(x) = \det(xI - T)$ the right way.

We consider $T\alpha_j = \sum_{i=1}^n A_{ij}\alpha_i$, $1 \leq j \leq n$. In this regard, $A = [T]_\alpha$. We see

$$\sum_{i=1}^n (\delta_{ji}T - A_{ij})\alpha_i = 0.$$

We define the matrix $B \in M_{n \times n}(K)$ (over the ring of polynomials of T) by $B_{ij} = \delta_{ji}T - A_{ij}I$. We also define the multiplication of B with vector α to be $(B\alpha)_i = \sum_{j=1}^n B_{ij}\alpha_j$. By the equation above, we see $\sum_{i=1}^n B_{ij}\alpha_i = 0 \implies B^\top \alpha = 0$.

Since $(xI - A)_{ij} = \delta_{ij}x - A_{ij} = \delta_{ji}x - A_{ij}$, we see $\det B = f(T)$.

Now by $B^\top \alpha = 0 \implies (\text{adj } B^\top)B^\top \alpha = (\text{adj } B^\top) \cdot 0 \implies (\det B)\alpha = 0 \implies (f(T))(\alpha) = 0$. ▣

7.3 Criteria of Triangulability and Diagonalizability

7.3.1 Restriction Operators

Theorem 7.3. *Let W be T -invariant. Then the characteristic polynomial of T_W divides the characteristic polynomial of T ; the minimal polynomial of T_W divides the minimal polynomial of T .*

Proof. Since W is T -invariant, we have

$$[T]_\beta = \begin{pmatrix} [T_W]_{\beta'} & C \\ O & D \end{pmatrix},$$

so $\det(xI - [T]_\beta) = \det(xI - [T_W]_{\beta'}) \det(xI - D)$, by [Problem 4.3.21](#). By [Problem 4.3.21](#), we have

$$[T]_\beta^k = \begin{pmatrix} [T_W]_{\beta'}^k & C_k \\ O & D^k \end{pmatrix}.$$

So if $[T]_\beta^k = 0$, $[T_W]_{\beta'}^k = 0$. So any polynomial that annihilates T also annihilates T_W . ▣

7.3.2 Conductors

Definition 7.3 (Conductors). Let W be T -invariant, and $\alpha \in V$. The T -conductor of α into W may refer to the set $S_T(\alpha; W) = \{g \in F[x] \mid g(T)(\alpha) \in W\}$ or its monic generator, depending on the context. If $W = 0$, then the conductor reduces to the annihilator.

Theorem 7.4. *Let W be T -invariant. $\forall \alpha \in V$, $S(\alpha; W)$ is an ideal in the polynomial algebra $F[x]$.*

Proof. If W is T -invariant, then W is also T^k -invariant, $\forall k \in \mathbb{Z}^+$, and also $f(T)$ -invariant. $S(\alpha; W) = \{g \in F[x] \mid g(T)(\alpha) \in W\}$. Let $f, g \in S(\alpha; W)$, $c \in F$. Then $((cf + g)(T))(\alpha) = (cf(T) + g(T))(\alpha) = cf(T)(\alpha) + g(T)(\alpha) \in W$, since W is a subspace.

Let $g \in S(\alpha; W)$, $f \in F[x]$. Then $((fg)(T))(\alpha) = (f(T)g(T))(\alpha) = f(T) \circ g(T)(\alpha)$. Since $g(T)(\alpha) \in W$, then $f(T)(g(T)(\alpha)) \in W$ also. ▣

Remark. The conductors $S(\alpha; W)$ always contain the minimal polynomial of T , since $\{0\} \subseteq W$ always. Thus, every T -conductor divides the minimal polynomial of T . ▣

Theorem 7.5. *Let $T \in \mathcal{L}(V)$, p be the minimal polynomial of T , and $p = \prod_{i=1}^k (x - c_i)^{r_i}$, $c_i \in F$. Let $W \subset V$, and W T -invariant, show $\exists \alpha \in V$ s.t. (1) $\alpha \notin W$, (2) $(T - c_i I)\alpha \in W$ for some i .*

Proof. Let $\beta \in V - W$. Consider the monic generator g of the ideal $S(\beta; W)$. Since g divides the minimal polynomial, g is of the form $g = \prod_{i=1}^k (x - c_i)^{e_i}$, $0 \leq e_i \leq r_i$.

Since β is not in W , $g \neq 1$, so $\exists e_j > 0$. So $g = (x - c_j)q$. By [Theorem 5.3](#), $g(T)(\beta) = (T - c_j I) \circ q(T)(\beta)$. Also $q(T)(\beta) \notin W$ since that would contradict g is generator. Thus $(T - c_j I)(\beta) \in W$. ▣

7.3.3 Criteria of Triangulability and Diagonalizability

Theorem 7.6. *Let $T \in \mathcal{L}(V)$. T is triangulable \iff the minimal polynomial p is a product of linear polynomials over F .*

Proof.

(\Rightarrow) If T is triangulable, then its characteristic polynomial is itself a product of linear factors. By [Theorem 7.2](#), its minimal polynomial divides characteristic polynomial, so p is also a product of linear factors.

(\Leftarrow) If the minimal polynomial $g = \prod_{i=1}^k (x - c_i)^{e_i}$, repeatedly apply [Theorem 7.5](#).

Start from $W_1 = \{0\}$. $\exists \alpha_1 \in V - W_1$ s.t. $(T - c_1 I)\alpha_1 \in W_1 \implies T(\alpha_1) = c_1 \alpha_1$.

Suppose we've got to a point where $W_k = \text{span}(\{\alpha_1, \dots, \alpha_{k-1}\})$, $\alpha_1, \dots, \alpha_{k-1}$ are independent. Then $\exists \alpha_k \in V - W_k$ s.t. $(T - c_k I)\alpha_k \in W_k \implies T(\alpha_k) = c_k \alpha_k + \sum_{i=1}^{k-1} a_i \alpha_i$.

Continue until $\mathfrak{B} = \{\alpha_1, \dots, \alpha_n\}$ forms a basis of V . By the equation $T(\alpha_k) = c_k \alpha_k + \sum_{i=1}^{k-1} a_i \alpha_i$, we immediately see that $[T]_{\mathfrak{B}}$ is a triangular matrix. \blacksquare

Corollary 7.6.1. *Let F be an algebraically closed field (i.e., every polynomial can be factored into linear polynomials on that field, for example the complex field), then every $A \in M_{n \times n}(F)$ is similar to a triangular matrix.*

Theorem 7.7. *Let $T \in \mathcal{L}(V)$. T is diagonalizable \iff the minimal polynomial p has the form*

$$p = \prod_{i=1}^k (x - c_i)$$

, where c_1, \dots, c_k are distinct.

Proof.

(\Rightarrow) If T is diagonalizable, let $[T]_{\mathfrak{B}}$ be diagonal matrix. Then the characteristic polynomial $f(x) = \det(xI - [T]_{\mathfrak{B}}) = \prod_{i=1}^k (x - c_i)^{d_i}$. Let $\mathfrak{B} = \{\beta_1, \dots, \beta_n\}$, where β_i corresponding to c_i , and c_i may repeat.

By the diagonal form, we see $(T - c_i I)\beta_i = 0$ implies

$$p(T)(\beta_i) = \left(\prod_{i=1}^k (T - c_i I) \right) (\beta_i) = 0, 1 \leq i \leq n,$$

so $p(T) = 0$. Now since minimal polynomial and characteristic polynomial have the same roots, and the expression given above is the minimal possible degree with the same roots, we see $p(x)$ is the minimal polynomial.

(\Leftarrow) Let W be the span of all eigenvectors, and suppose $W \neq V$. By [Theorem 7.5](#), $\exists \alpha \in V - W$ s.t. $\beta = (T - c_i I)\alpha \in W$. Since $\beta \in W$, $\beta = \sum_{i=1}^k \beta_i$, $\beta_i \in W_i$, where W_i is the eigenspace corresponding to c_i . Thus

$$h(T)(\beta) = \sum_{i=1}^k h(c_i)\beta_i \in W, \quad \forall h \in F[x].$$

Now $p = (x - c_j)q$ for some $q \in F[x]$. Also $q - q(c_j) = (x - c_j)h$, so

$$q(T)(\alpha) - q(c_j)\alpha = h(T)(T - c_j I)(\alpha) = h(T)(\beta) \in W.$$

Also since

$$0 = p(T)(\alpha) = (T - c_j I)q(T)(\alpha)$$

, we see $q(T)(\alpha)$ is an eigenvector therefore lies in W , so $q(c_j)\alpha \in W$ also. Since α is not in W , $q(c_j) = 0$, a contradiction since p has no repeated roots. \blacksquare

\nearrow *Remark.* For another more elegant proof, see the following section about projections. \mathfrak{M}

7.4 The Primary Decomposition Theorem

7.4.1 Direct Sum and Projections

\nearrow *Remark.* By [Problem 2.3.17](#), we see the set of all projections is equal to the set $S = \{E \in \mathcal{L}(V) \mid E^2 = E\}$. We might as well take that as the definition. \mathfrak{M}

Definition 7.4 (Projections). A projection of V is $E \in \mathcal{L}(V)$ s.t. $E^2 = E$.

Theorem 7.8. Suppose E is a projection of V . Then

1. $\beta \in R(E) \iff E(\beta) = \beta$.
2. $\forall \alpha \in V, \alpha = E(\alpha) + (\alpha - E(\alpha))$.
3. $V = R(E) \oplus N(E)$.
4. Given $V = W_1 \oplus W_2$, E exists and is the only projection that have $R(E) = W_1, N(E) = W_2$. Such a projection is given by: $\alpha = \alpha_1 + \alpha_2, \alpha_1 \in W_1, \alpha_2 \in W_2, E(\alpha) = \alpha_1$.

5. If $\mathfrak{A} = \{\alpha_1, \dots, \alpha_r\}$ is a basis of $R(E)$, $\mathfrak{B} = \{\alpha_{r+1}, \dots, \alpha_n\}$ is a basis of $N(E)$, then $\mathfrak{C} = \mathfrak{A} \cup \mathfrak{B}$ diagonalizes E , with $[E]_{\mathfrak{C}} = \begin{pmatrix} I & O \\ O & O \end{pmatrix}$.

Proof. (1) If $\beta \in R(E)$, $\beta = E\alpha$, then $E\beta = E^2\alpha = E\alpha = \beta$. If $E\beta = \beta$, then $\beta \in R(E)$ by definition.

(2) is obviously true.

(3) By (2), we see $E(E(\alpha)) = E(\alpha)$, so by (1) $E(\alpha) \in R(E)$. By similar reasoning we see $\alpha - E(\alpha) \in N(E)$. So $V = R(E) + N(E)$. Now suppose $\beta \in R(E) \cap N(E)$, then $\beta = E(\alpha)$ for some $\alpha \in V$. But

$$E(\beta) = 0 = E^2(\alpha) = E(\alpha) = \beta,$$

so $\beta = 0$, $R(E) \cap N(E) = \{0\}$.

(4) It is easy to see that such a construction is a projection. By (1) we quickly see that it is the only projection. ▣

Theorem 7.9. If $V = \bigoplus_{i=1}^k W_i$, then $\exists E_1, \dots, E_k \in \mathcal{L}(V)$ s.t.

1. $E_i^2 = E_i$.
2. $E_i E_j = 0$ if $i \neq j$.
3. $I = \sum_{i=1}^k E_i$.
4. $W_i = R(E_i)$.

The converse is true also. That is, given E_1, \dots, E_k which satisfy (1), (2) and (3), and let $W_i = R(E_i)$, then $V = \bigoplus_{i=1}^k W_i$.

Proof.

(\Rightarrow) By (4) of **Theorem 7.8**, we construct E_j by $\alpha = \sum_{i=1}^k \alpha_i$, $\alpha_i \in W_i$, $E_j \alpha = \alpha_j$. It is easy to verify (1), (2), and (4). Now

$$\alpha = \sum_{i=1}^k E_i(\alpha) \Rightarrow I = \sum_{i=1}^k E_i.$$

(\Leftarrow) From (3) we know

$$I(V) = \sum_{i=1}^k E_i(V) \Rightarrow V = \sum_{i=1}^k W_i.$$

Consider $\sum_{i=1}^k \beta_i = 0, \beta_i \in W_i$. Then by (3), it is equivalent to $\sum_{i=1}^k E_i(\beta_i) = 0$. Applying E_1 to both sides,

$$E_1^2(\beta_1) + E_1 E_2(\beta_2) + \cdots + E_1 E_k(\beta_k) = E_1^2(\beta_1) = E_1(\beta_1) = 0 \iff \beta_1 = 0.$$

Same for all the other β_i . So all the W_i are independent, therefore they form a direct sum. ▣

7.4.2 Direct Sum of Restricted Operators

Definition 7.5. Given $V = \bigoplus_{i=1}^k W_i$, each W_i being T -invariant, we restrict T on W_i , producing T_i . Then

$$\alpha = \sum_{i=1}^k \alpha_i, \alpha_i \in W_i \implies T(\alpha) = \sum_{i=1}^k T_i(\alpha_i).$$

We define $T = \bigoplus_{i=1}^k T_i$.

Definition 7.6. For the corresponding matrices $[T]_{\mathfrak{B}} = A, [T_i]_{\mathfrak{B}_i} = A_i$, we say $A = \bigoplus_{i=1}^k A_i$, and

$$A = \begin{pmatrix} A_1 & O & \cdots & O \\ O & A_2 & \cdots & O \\ \vdots & \vdots & \ddots & \vdots \\ O & O & \cdots & A_k \end{pmatrix}.$$

Theorem 7.10. Let $T \in \mathcal{L}(V)$, $V = \bigoplus_{i=1}^k W_i$, E_i is the projection on W_i . Then every W_i is T -invariant $\iff T$ commute with each E_i ,

$$TE_i = E_i T, \quad 1 \leq i \leq k.$$

Proof.

(\Rightarrow) Suppose W_i is T -invariant, for all $1 \leq i \leq k$. $\forall w_i \in W_i, T(w_i) \in W_i$.

$\forall x \in V, x = \sum_{i=1}^k w_i$. $T(w_i) = w'_i \in W_i$. Then $E_i(T(x)) = w'_i = T(E_i(x))$. (\Leftarrow) Let $TE_i = E_i T, 1 \leq i \leq k, \forall \alpha \in V$. Let $\alpha \in W_j$. Then $E_j(\alpha) = \alpha$, and

$$T(\alpha) = T \circ E_j(\alpha) = E_j \circ T(\alpha).$$

Thus, by (1) of [Theorem 7.8](#), $T(\alpha) \in R(E_j) = W_j$. ▣

7.4.3 Projection Decomposition of Diagonalizable Operators

Theorem 7.11. *Let $T \in \mathcal{L}(V)$. If T is diagonalizable, c_1, \dots, c_k are distinct eigenvalues, then $\exists E_1, \dots, E_k$ projection on V s.t.*

1. $T = \sum_{i=1}^k c_i E_i$.
2. $I = \sum_{i=1}^k E_i$.
3. $E_i E_j = 0$ if $i \neq j$.
4. $E_i^2 = E_i$.
5. $R(E_i)$ is the eigenspace of T corresponding to c_i .

Conversely, if $\exists c_1, \dots, c_k$ distinct and E_1, \dots, E_k which satisfy (1), (2), and (3), then T is diagonalizable, c_i are eigenvalues of T , and (4), (5) are also satisfied.

Proof.

(\Rightarrow) If T is diagonalizable, then $V = \bigoplus_{i=1}^k W_i$. By [Theorem 7.9](#), $\exists E_i, 1 \leq i \leq k$ s.t. $E_i^2 = E_i, E_i E_j = 0, I = E_1 + \dots + E_k$, and $R(E_i) = W_i$. Let $\alpha = \sum_{i=1}^k w_i, w_i \in W_i$. So

$$T(\alpha) = T\left(\sum_{i=1}^k w_i\right) = \sum_{i=1}^k c_i w_i = \sum_{i=1}^k c_i E_i(\alpha).$$

(\Leftarrow) We have

$$\begin{aligned} T &= c_1 E_1 + \dots + c_k E_k \\ I &= E_1 + \dots + E_k \\ E_i E_j &= 0 \text{ if } i \neq j. \end{aligned}$$

And we want to prove

1. $E_i^2 = E_i$.
2. $R(E_i) = N(T - c_i I)$.
3. T is diagonalizable.

(1) Since $I = E_1 + \dots + E_k$, apply E_i on both sides to obtain $E_i I = E_i^2 = E_i$.

(2)(\subseteq) We see

$$T - c_i I = \sum_{j \neq i} (c_j - c_i) E_j.$$

$\forall w_i \in R(E_i), w_i = E_i w_i$. Consider

$$(T - c_i I)(w_i) = \sum_{j \neq i} (c_j - c_i) E_j E_i w_i = 0.$$

Thus $R(E_i) \subseteq N(T - c_i I)$. Since $E_i \neq 0$, $R(E_i) \subseteq N(T - c_i I) \neq 0$. So the value c_i has at least an eigenvector.

(2)(\supseteq) If $\alpha \in N(T - c_i I)$, then $(T - c_i I)(\alpha) = 0$,

$$\sum_{j \neq i} (c_j - c_i) E_j(\alpha) = 0.$$

We already know $(c_j - c_i) E_j(\alpha) \in R(E_j) \subseteq N(T - c_j I)$. By [Theorem 6.4](#),

$$\sum_{j \neq i} (c_j - c_i) E_j(\alpha) = 0 \implies (c_j - c_i) E_j(\alpha) = 0, \forall j \neq i.$$

Since $c_j \neq c_i$, $E_j(\alpha) = 0 \forall j \neq i$. This could only be that $v \in R(E_i)$. So $N(T - c_i I) \subseteq R(E_i)$.

(4) Since $I = E_1 + \dots + E_k$ and $R(E_i) = N(T - c_i I) = W_i$, clearly $V = \bigoplus_{i=1}^k W_i$. \blacksquare

Theorem 7.12. *Let $g \in F[x]$, T diagonalizable and has distinct eigenvalues c_1, \dots, c_k . Then*

$$g(T) = g(c_1)E_1 + \dots + g(c_k)E_k.$$

Proof. Consider T^r . We show

$$T^r = \sum_{i=1}^k c_i^r E_i$$

by induction on r . $r = 2$ is true, since

$$\begin{aligned} T^2 &= \left(\sum_{i=1}^k c_i E_i \right) \left(\sum_{j=1}^k c_j E_j \right) \\ &= \sum_{i=1}^k \sum_{j=1}^k c_i c_j E_i E_j \\ &= \sum_{i=1}^k c_i^2 E_i^2 \\ &= \sum_{i=1}^k c_i^2 E_i, \end{aligned}$$

by $E_i E_j = 0$.

Now suppose $r - 1$ is true, $T^{r-1} = \sum_{i=1}^k c_i^{r-1} E_i$. Consider

$$\begin{aligned} T^r &= \left(\sum_{i=1}^k c_i E_i \right) \left(\sum_{j=1}^k c_j^{r-1} E_j \right) \\ &= \sum_{i,j} c_i c_j^{r-1} E_i E_j \\ &= \sum_{i=1}^k c_i^r E_i. \end{aligned}$$

Thus,

$$\begin{aligned} f(T) &= \sum_{i=1}^m f_i T^i \\ &= \sum_{i=1}^m \left(f_i \sum_{j=1}^k c_j^i E_j \right) \\ &= \sum_{j=1}^k \left(\sum_{i=1}^m (f_i c_j^i) E_j \right) \\ &= \sum_{j=1}^k f(c_j) E_j. \end{aligned}$$

▣

✍ *Remark.* Now we give another proof of [Theorem 7.7](#) using the projection decomposition of diagonalizable operator given above. This proof is intimately related to the proof of the Primary Decomposition Theorem. \square

Theorem 7.13. *Let $T \in \mathcal{L}(V)$. T is diagonalizable \iff the minimal polynomial p has the form*

$$p = \prod_{i=1}^k (x - c_i),$$

c_1, \dots, c_k are distinct.

Proof.

(\Rightarrow) Let T be diagonalizable. By [Theorem 7.11](#), we see $T = \sum_{i=1}^k c_i E_i$. Let $p' = (x - c_1) \cdots (x - c_k)$. By [Theorem 7.12](#),

$$p'(T) = p'(c_1) E_1 + \cdots + p'(c_k) E_k = 0.$$

Now by [Theorem 7.1](#), p and f have the same roots, and p' is the polynomial with lowest possible degree that shares the same roots with f , thus $p = p'$.

(\Leftarrow) Here comes the elegant part. Let $p = (x - c_1) \cdots (x - c_k)$ be the minimal polynomial of T . Since p and f have the same roots, we see c_1, \dots, c_k are the only eigenvalues of T . Define

$$q_j = \prod_{i \neq j} \frac{(x - c_i)}{(c_j - c_i)}, E_j = q_j(T).$$

We see $p = d_j(x - c_j)q_j$.

We would like to show E_j has the properties of projection. Firstly, by [Theorem 5.3¹](#),

$$E_i \circ E_j = (q_i q_j)(T) = \frac{p^2}{(x - c_i)(x - c_j)d_i d_j} = 0, \quad i \neq j.$$

Since the expression is divisible by p . If $i = j$, then the factor $(x - c_i)$ would be cancelled out, so $q_i^2(T) \neq 0$.

At this point, we want to use [Theorem 7.11](#) to conclude that T is diagonalizable, so we want to prove the following properties, which have very elegant polynomial analogues,

$$\begin{aligned} T &= \sum_{i=1}^k c_i E_i \iff x = \sum_{i=1}^k c_i q_i(x). \\ I &= \sum_{i=1}^k E_i \iff 1 = \sum_{i=1}^k q_i(x). \end{aligned}$$

Consider the sum $f = \sum_{i=1}^k q_i(x)$. It has degree $k - 1$, which means it lives in a vector space of dimension k . Moreover, by the definition of q_j , we have $q_j(c_i) = \delta_{ij}$. Thus we have found k values c_1, \dots, c_k s.t. $f(c_i) = 1$. Therefore, $f(x) = 1, f(T) = I$.

The proof for $g = \sum_{i=1}^k c_i q_i(x) = x$ follows similarly. We've found k values c_1, \dots, c_k s.t. $g(c_i) = c_i$, thus $g(x) = x, g(T) = T$. \blacksquare

7.4.4 The Primary Decomposition Theorem

Theorem 7.14 (Primary Decomposition Theorem). *Let $T \in \mathcal{L}(V)$, p be the minimal polynomial.*

$$p = p_1^{r_1} \cdots p_k^{r_k}$$

where p_i are distinct, irreducible. Let $W_i = N(p_i^{r_i}(T))$, then

¹Please remember there's no division in a ring. I wrote fraction for clarity; the correct expression need to use product notation.

1. $V = \bigoplus_{i=1}^k W_i$.
2. W_i are T -invariant, $1 \leq i \leq k$.
3. T_i is the restricted operator of T on W_i , then the minimal polynomial of T_i is $p_i^{r_i}$.

Proof. Define

$$q_i = d_i \frac{p}{p_i^{r_i}}.$$

We see q_i, q_j are coprime. By the polynomial version of Bezout's identity, $\exists f_i$ s.t. $\sum_{i=1}^k f_i q_i = 1$. Let

$$E_i = (f_i q_i)(T) = f_i(T) \circ q_i(T) \implies I = \sum_{i=1}^k E_i.$$

We want to show that E_i has the properties of projection, and use [Theorem 7.9](#) to say they form a direct sum. So we need to show

1. $I = \sum_{i=1}^k E_i$. (done)
2. $E_i E_j = 0$.
3. $E_i^2 = E_i$. (follows from 1 and 2)
4. $W_i = R(E_i)$.

So firstly,

$$q_i(T) \circ q_j(T) = (q_i q_j)(T) = \left(d_i d_j \frac{p^2}{p_i^{r_i} p_j^{r_j}} \right) (T) = 0.$$

Since it is divisible by p . So $E_i E_j = 0, i \neq j$.

If $w_i \in W_i = N(p_i^{r_i}(T))$, then

$$q_j(w_i) = d_i \prod_{k \neq j} p_k^{r_k}(T)(w_i) = \delta_{ij},$$

so

$$I(w_i) = \sum_{j=1}^k E_j(w_i) = E_i(w_i) = w_i.$$

Thus, $W_i = N(p_i^{r_i}(T)) \subseteq R(E_i)$.

If $w_i \in R(E_i)$, then $w_i = E_i(w_i)$. Thus

$$p_i^{r_i}(w_i) = p_i^{r_i}(E_i(w_i)) = ((p_i^{r_i} f_i q_i)(T))(w_i) = 0.$$

So $R(E_i) \subseteq N(p_i^{r_i}(T))$. Combine the result above, we have $R(E_i) = W_i$. By **Theorem 7.9**, we have $V = \bigoplus_{i=1}^k W_i$.

Now we need to show they are invariant and give the minimal polynomial. $\forall w_i \in W_i$, $p_i^{r_i}(T)(w_i) = 0$. Thus

$$p_i^{r_i}(T)(T(w_i)) = p_i^{r_i}(T) \circ T(w_i) = ((p_i^{r_i} x)(T))(w_i) = T(p_i^{r_i}(w_i)) = 0.$$

So $T(w_i) \in W_i$, which means W_i are T -invariant.

$\forall w_i \in W_i = N(p_i^{r_i}(T))$, $(p_i^{r_i}(T))(w_i) = 0$. Let T_i be the restricted operator of T on W_i , that means $p_i^{r_i}(T_i)(w_i) = 0$, $\forall w_i \in W_i$. Thus $p_i^{r_i}(T_i) = 0$. If g is the minimal polynomial of T_i , then g divides $p_i^{r_i}$. Since p_i is irreducible, $g = p_i^{e_i}$. Consider $g(T)q_i(T)$,

$$\begin{cases} g(T)(\alpha) = 0, & \text{if } \alpha \in W_i \\ q_i(T)(\alpha) = 0, & \text{if } \alpha \notin W_i. \end{cases}$$

Thus, $g(T)q_i(T) = 0$. But now p divides gq_i , so $p_i^{r_i}$ divides $p_i^{e_i}$. Thus $e_i = r_i$. \square

7.4.5 Diag-Nilpotent Decomposition

Definition 7.7 (Nilpotent). Let $N \in \mathcal{L}(V)$. N is nilpotent if $\exists r > 0$ s.t. $N^r = 0$.

Theorem 7.15. Let $T \in \mathcal{L}(V)$. Suppose $p = \prod_{i=1}^k (x - c_i)^{r_i}$, then $\exists!$ D, N s.t. D diagonalizable, N nilpotent and

1. $T = D + N$.
2. $DN = ND$.
3. N, D are polynomials in T .

Proof. (Existence) Let $R(E_i) = W_i = N((T - c_i I)^{r_i})$ be projection, then construct D as below, and consider $N = T - D$,

$$\begin{aligned} T &= TE_1 + \cdots + TE_k \\ D &= c_1 E_1 + \cdots + c_k E_k \\ N &= T - D = (T - c_1 I)E_1 + \cdots + (T - c_k I)E_k. \end{aligned}$$

By similar proof of [Theorem 7.12](#), we see

$$N^r = (T - c_1 I)^r E_1 + \cdots + (T - c_k I)^r E_k.$$

If $r \geq \max_i r_i$, then $N^r = 0$.

By the construction of projections in [Theorem 7.14](#), we see both D, N are polynomials of T , therefore they commute.

(Uniqueness) Suppose $T = D + N = D' + N'$, $DN = ND$, $D'N' = N'D'$, and $N^r = 0, N'^k = 0$. Then we have $(D - D') + (N - N') = 0$. So

$$\begin{aligned} (D - D')^{r+k} &= (-1)^{r+k} (N - N')^{r+k} \\ &= (-1)^{r+k} \sum_{i=0}^{r+k} (-1)^i N^i (N')^{r+k-i} \\ &= \begin{cases} 0 & \text{if } 0 \leq i < r, \text{ since } (N')^{r+k-i} = 0. \\ 0 & \text{if } r \leq i \leq r+k, \text{ since } N^i = 0. \end{cases} = 0. \end{aligned}$$

Since $D - D'$ is also diagonalizable, we have $([D]_{\mathfrak{B}} - [D']_{\mathfrak{B}})_{ii}^{r+k} = 0 \implies D - D' = 0, N - N' = 0$. ▣

8 Rational and Jordan Form

8.1 Cyclic Subspaces

8.1.1 Definition

Definition 8.1. Let $\alpha \in V$. The T -cyclic subspace generated by α is $Z(\alpha; T) = g(T)\alpha$, $g \in F[x]$. If $Z(\alpha; T) = V$, then α is a cyclic vector of T .

Definition 8.2. Let $\alpha \in V$. The T -annihilator of α can refer to the ideal $M(\alpha; T) = \{g \in F[x] \mid g(T)(\alpha) = 0\}$ or its monic generator, depending on the context. The T -annihilator of α p_α divides the minimal polynomial.

Theorem 8.1. Let $\alpha \in V$ and p_α be the T -annihilator of α . Then

1. $\deg p_\alpha = \dim Z(\alpha; T)$.
2. Let $\deg p_\alpha = k$, then $\mathfrak{B} = \{\alpha, T(\alpha), \dots, T^{k-1}(\alpha)\}$ forms a basis of $Z(\alpha; T)$.
3. Let U be the restricted operator of T on $Z(\alpha; T)$, then p_α is the minimal polynomial of U .

Proof. My proof was to show that $T^n(\alpha) \in \text{span}(\{\alpha, T(\alpha), \dots, T^{k-1}(\alpha)\})$, and the set is independent. But I want to show Hoffman's elegant proof.

Let $g \in F[x]$, and $g = p_\alpha q + r$. So $g(T)(\alpha) = r(T)(\alpha)$, and most importantly $\deg r < \deg p_\alpha = k$, thus $g(T)(\alpha)$ can be written as linear combinations of $\{\alpha, T(\alpha), \dots, T^{k-1}(\alpha)\}$. Thus $\text{span } \mathfrak{B} = Z(\alpha; T)$. Now if \mathfrak{B} is dependent, then there exist $g'(T)(\alpha) = 0$ and $\deg g' < \deg p_\alpha$.

Consider

$$\begin{aligned} p_\alpha(U)g(T)(\alpha) &= p_\alpha(T)g(T)(\alpha) \\ &= g(T)p_\alpha(T)(\alpha) \\ &= 0. \end{aligned}$$

So $p_\alpha(U) = 0$. There can be no $h(T) = 0$ s.t. $\deg h < \deg p_\alpha$, for if so, $h(T)(\alpha) = 0$, contradicting the definition of annihilator. \blacksquare

Corollary 8.1.1. *If T has a cyclic vector, then characteristic polynomial and minimal polynomial of T are equal.*

✍ *Remark.* Actually, T has a cyclic vector \iff characteristic polynomial and minimal polynomial of T are equal. We'll prove that later. \blacksquare

8.1.2 Companion Matrix

Theorem 8.2. $U \in \mathcal{L}(W)$ has a cyclic vector \iff there is some basis \mathfrak{B} for W s.t.

$$[U]_{\mathfrak{B}} = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & -c_0 \\ 1 & 0 & 0 & \cdots & 0 & -c_1 \\ 0 & 1 & 0 & \cdots & 0 & -c_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -c_{k-1} \end{pmatrix},$$

where $p(x) = c_0 + c_1x + \cdots + c_{k-1}x^{k-1} + x^k$.

Proof.

Let $\alpha_i = U^{i-1}(\alpha)$, $1 \leq i \leq k$. (\Rightarrow) If U has a cyclic vector, then $\mathfrak{B} = \{\alpha_1, \dots, \alpha_k\}$ is a basis of W . Then clearly

$$U(\alpha_i) = \alpha_{i+1}, \quad 1 \leq i \leq k-1,$$

and this accounts for the off-diagonal 1's.

Now since $p_\alpha(U) = 0$, thus

$$\begin{aligned} U^k(\alpha) + c_{k-1}U^{k-1}(\alpha) + \cdots + c_1U(\alpha) + c_0\alpha &= 0 \\ \implies U(\alpha_k) &= -c_0\alpha_1 - c_1\alpha_2 - \cdots - c_{k-1}\alpha_k. \end{aligned}$$

And this completes the proof.

(\Leftarrow) Given the matrix form, it is easy to see that α_1 is a cyclic vector. \square

Corollary 8.2.1. *If A is the companion matrix of monic p , then p is both the minimal and characteristic polynomial of A .*

Proof. By [Problem 4.3.24](#), we see the characteristic polynomial is just p . By [Corollary 8.1.1](#), we see the characteristic and minimal polynomial are equal. \square

8.2 Cyclic Decomposition Theorem

8.2.1 Motivation

Definition 8.3 (Admissible). Let $T \in \mathcal{L}(V)$, W a subspace of V . If W is T -admissible, then

1. W is T -invariant.
2. If $f(T)(\beta) \in W$, $\exists \gamma \in W$ s.t. $f(T)(\beta) = f(T)(\gamma)$.

\nearrow *Remark.* If W is invariant and $V = W \oplus W'$, then W is T -admissible. For, $\forall \alpha \in V = W \oplus W'$, $\alpha = w + w'$, $w \in W$, $w' \in W'$. $f(T)(\alpha) = f(T)(w) + f(T)(w')$. Since $f(T)(w) \in W$, if given $f(T)(\alpha) \in W$, then $f(T)(w') = 0$. Thus, $f(T)(\alpha) = f(T)(w)$. \square

Theorem 8.3.

1. $f(T)(Z(\alpha; T)) = Z(f(T)(\alpha); T)$.
2. If $V = \bigoplus_{i=1}^k V_i$, where V_i are T -invariant, then $f(T)(V) = \bigoplus_{i=1}^k f(T)(V_i)$.
3. If α and γ have the same T -annihilator, then $f(T)(\alpha)$ and $f(T)(\gamma)$ have the same T -annihilator, so $\dim Z(f(T)(\alpha); T) = \dim Z(f(T)(\gamma); T)$.

Proof. (1) $f(T)(Z(\alpha; T)) = f(T)(g(T)(\alpha))$, $\forall g \in F[x]$. So by [Theorem 5.3](#) it equals $(fg)(T)(\alpha) = g(f(T)(\alpha)) = Z(f(T)(\alpha); T)$.

(2) $\forall \alpha \in V, \alpha = \alpha_1 + \dots + \alpha_k$. Then $f(T)(\alpha) = f(T)(\alpha_1) + \dots + f(T)(\alpha_k)$. So $f(T)(V) = f(T)(V_1) + \dots + f(T)(V_k)$. If $\exists \beta \neq 0, \beta \in f(T)(V_1) \cap \sum_{i=2}^k f(T)(V_i)$, then $f(T)(\beta_1) - \sum_{i=2}^k f(T)(\beta_i) = 0$. Since all the V_i are T -invariant, this violates the fact that $V = \bigoplus_{i=1}^k V_i$. \times

(3) Let $p(T)(\alpha) = p(T)(\gamma) = 0$. Then

$$p(T)(f(T)(\alpha)) = (pf)(T)(\alpha) = f(T)(p(T)(\alpha)) = 0.$$

Same for $p(T)(f(T)(\gamma))$. ▣

8.2.2 Cyclic Decomposition Theorem

Theorem 8.4 (Cyclic Decomposition Theorem). *Let $T \in \mathcal{L}(V)$, $\dim V \neq \infty$, and $W_0 \subset V$ is a T -admissible subspace, e.g. $\{0\}$. Then $\exists \alpha_1, \dots, \alpha_r$ all non-zero with corresponding T -annihilators p_1, \dots, p_r s.t.*

1. $V = W_0 \oplus Z(\alpha_1; T) \oplus \dots \oplus Z(\alpha_r; T)$.
2. p_k divides p_{k-1} , $2 \leq k \leq r$.
3. r and p_1, \dots, p_r satisfying the above conditions are unique.

The polynomials p_1, \dots, p_r are called the invariant factors.

✂ *Remark.* Just believe it. Its proof is way too sophisticated for a physicist. ▣

Corollary 8.4.1. *Let $T \in \mathcal{L}(V)$. Then every T -admissible subspace has a complementary subspace ($V = W \oplus W'$) which is also T -invariant.*

Corollary 8.4.2. *Let $T \in \mathcal{L}(V)$. Then*

1. $\exists \alpha \in V$ s.t. the T -annihilator of α is the minimal polynomial of T . In fact, if $V = Z(\alpha_1; T) \oplus \dots \oplus Z(\alpha_r; T)$, then $p = p_1$.
2. T has a cyclic vector \iff its characteristic and minimal polynomials are equal.

Proof. (1) By **Theorem 8.4**, let $V = Z(\alpha_1; T) \oplus \dots \oplus Z(\alpha_r; T)$. Obviously $p(T)(\alpha_1) = 0$, so p_1 divides p . Now

$$\begin{aligned} p_1(T)(V) &= p_1(T)(Z(\alpha_1; T) \oplus \dots \oplus Z(\alpha_r; T)) \\ &= p_1(T)(Z(\alpha_1; T)) \oplus \dots \oplus p_1(T)(Z(\alpha_r; T)) \\ &= Z(p_1(T)(\alpha_1); T) \oplus \dots \oplus Z(p_1(T)(\alpha_r; T)) \\ &= \{0\}. \end{aligned}$$

Since p_k divides p_1 , $p_1(\alpha_k) = 0$. Then, $p_1(T) = 0$, so p divides p_1 . Thus $p = p_1$.

(2) (\Rightarrow) Proved in [Corollary 8.1.1](#).

(\Leftarrow) By (1), we see $\exists \alpha$ s.t. p is its T -annihilator. So the set $\mathfrak{B} = \{\alpha, T(\alpha), \dots, T^{n-1}(\alpha)\}$ is independent. If not, then it violates that p has the minimal degree. Since $|\mathfrak{B}| = n$, we see it's a basis, and therefore α is the cyclic vector. \square

8.2.3 Generalized Cayley-Hamilton Theorem

Theorem 8.5 (Generalized Cayley-Hamilton Theorem). *Let $T \in \mathcal{L}(V)$, $\dim V \neq \infty$, p minimal, f characteristic. Then*

1. p divides f .
2. p and f has the same prime factors, except for multiplicities.
3. If $p = f_1^{r_1} \dots f_k^{r_k}$ is its prime factorization, then $f = f_1^{d_1} \dots f_k^{d_k}$, where $d_i = (\deg f_i)^{-1} N(f_i^{r_i}(T))$.

Proof. By [Theorem 8.4](#), let $V = Z(\alpha_1; T) \oplus \dots \oplus Z(\alpha_r; T)$.

(1) Let U_i be the restriction of T to $Z(\alpha_i; T)$. It is obvious that α_i is the cyclic vector of U_i . Thus, p_i is both the characteristic and minimal of U_i . By [Theorem 7.3](#), we see p_i divides p , and p_i divides f . In particular, $p_1 = p$ divides f .

(2) Since $T = \bigoplus_{i=1}^r U_i$, it is easy to see $f_T = \prod_{i=1}^r f_{U_i}$ by the matrix form. Thus, $f = p_1 \dots p_r$. Since each p_i divides p_1 , decompose p easily produces the desired result.

(3) Suppose $p = f_1^{r_1} \dots f_k^{r_k}$. By [Theorem 7.14](#), we see

$$V = \bigoplus_{i=1}^k N(f_i^{r_i}(T)) := \bigoplus_{i=1}^k V_i,$$

where $V_i = N(f_i^{r_i}(T))$.

Let T_i be the restriction of T to V_i . Then the minimal polynomial of T_i is $q_i = f_i^{r_i}$, characteristic polynomial is $g_i = f_i^{d_i}$. Then

$$\deg g_i = \dim V_i = (\deg f_i) \cdot d_i.$$

\square

Corollary 8.5.1. *If $T \in \mathcal{L}(V)$, $\dim V = n$ is nilpotent, then $f = x^n$.*

Proof. We prove it explicitly as an easy application to [Theorem 8.5](#). If T is nilpotent, then $p = x^r$ for some $r \in \mathbb{Z}^+$. By (2) of [Theorem 8.5](#), we see $f = x^k$ for some k . Moreover, (3) says that $k = N(T^r) = n$, which completes the proof. \square

8.3 Rational Canonical Form

8.3.1 Definition, Existence and Uniqueness

Definition 8.4 (Rational Form). Let $T \in \mathcal{L}(V)$. By [Theorem 8.4](#), we see

$$V = \bigoplus_{i=1}^r Z(\alpha_i; T).$$

Let

$$\mathfrak{B}_i = \{\alpha_i, T(\alpha_i), \dots, T^{k_i-1}(\alpha_i)\},$$

where $k_i = \deg p_i$, p_i being the T -annihilator of α_i . Then let

$$\mathfrak{B} = \bigcup_{i=1}^r \mathfrak{B}_i,$$

we see

$$A = [T]_{\mathfrak{B}} = \begin{pmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_r \end{pmatrix},$$

where A_i is the companion matrix of p_i . This is the rational form of T .

Theorem 8.6. *Let $B \in M_{n \times n}(F)$. Then B is similar to one and only one matrix which is in rational canonical form.*

Proof. By [Definition 8.4](#), we see such a basis exists. If it is not unique, then we would have violated [Theorem 8.4](#). \square

8.3.2 Connection to Diagonalized Form

Theorem 8.7. *Let $T \in \mathcal{L}(V)$ be diagonalizable, $c_i, 1 \leq i \leq k$ be its distinct eigenvalues, $V_i, 1 \leq i \leq k$ be the eigenspaces, $\dim V_i = d_i$, and $\mathfrak{B}_i = \{\beta_1^i, \dots, \beta_{d_i}^i\}$ be the basis*

of V_i . Then the invariant factors and corresponding cyclic vectors are given by

$$\alpha_j = \sum_{d_i \geq j} \beta_j^i,$$

$$p_j = \prod_{d_i \geq j} (x - c_i), \quad 1 \leq j \leq \max_i d_i,$$

which can be understood symbolically as

$$\begin{aligned} \mathfrak{B}_1 &= \{ \boxed{\beta_1^1}, \boxed{\beta_2^1}, \boxed{\beta_3^1}, \} \\ \mathfrak{B}_2 &= \{ \boxed{\beta_1^2}, \} \\ \mathfrak{B}_3 &= \{ \boxed{\beta_1^3}, \boxed{\beta_2^3}, \} \end{aligned}$$

$\alpha_1 \quad \alpha_2 \quad \alpha_3$

Proof. Since T is diagonalizable, then $V = \bigoplus_{i=1}^k V_i$. Then, $\forall \alpha \in V$,

$$\alpha = \beta_1 + \cdots + \beta_k,$$

$$f(T)(\alpha) = f(c_1)\beta_1 + \cdots + f(c_k)\beta_k, \beta_i \in V_i.$$

So if $f(T)(\alpha) = 0, f(c_i)\beta_i = 0, 1 \leq i \leq k$. Thus, given α , its annihilator is of the form

$$\prod_{\beta_i \neq 0} (x - c_i).$$

Thus, take the definition $\alpha_j = \sum_{d_i \geq j} \beta_j^i$, it is easy to see that its annihilator is $p_j = \prod_{d_i \geq j} (x - c_i)$, and that $p_{j+1} \mid p_j$.

Now, since each β_j^i belongs to one and only one of the subspaces $Z(\alpha_1; T), \dots, Z(\alpha_r; T)$,

$$V = Z(\alpha_1; T) \oplus \cdots \oplus Z(\alpha_r; T).$$

▣

8.4 Jordan Form

8.4.1 Cyclic Decomposition of Nilpotent Operators

Theorem 8.8. *Let $N \in \mathcal{L}(V)$ which is nilpotent, $p = x^k$ be its minimal polynomial. Let its cyclic decomposition contain r non-zero vectors $\alpha_1, \dots, \alpha_r$, with corresponding*

invariant factors x^{k_1}, \dots, x^{k_r} , $k_1 \geq k_2 \geq \dots \geq k_r$. Then $r = \text{nullity}(N)$, $k = k_1$. Moreover,

$$N = \bigoplus_{i=1}^r A_i,$$

where

$$A_i = J_{0, k_i} := \begin{pmatrix} 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix} \in M_{k_i \times k_i}(F).$$

Proof. Since $p_{j+1} \mid p_j$ and [Corollary 8.4.2](#), it's easy to see that $p_1 = x^{k_1} = x^k$, that every $p_j = x^{k_j}$, and that $k_{j+1} \leq k_j$. Thus, their companion matrices are given by A_i .

We now prove that $r = \text{nullity}(N)$. It is obvious from the matrix direct sum, since every A_i has a zero column. If we combine $A = \bigoplus_{i=1}^r A_i$, there will be exactly r zero columns, and therefore equal to the nullity of N .

We give another proof by Hoffman, which gives the full basis of the nullspace. Let $\alpha \in N(N)$. Since $N(N) \subseteq V = \bigoplus_{i=1}^r Z(\alpha_r; N)$, we see

$$\alpha = f_1(N)(\alpha_1) + \dots + f_r(N)(\alpha_r), \quad \deg f_i < k_i,$$

since $N^{k_i}(\alpha_i) = 0$, a polynomial with k_i degree or more is meaningless. Since $N(\alpha) = 0$,

$$0 = N(f_i(N)(\alpha_i)) = (xf_i)(N)(\alpha_i).$$

Thus, $xf_i \in M(\alpha_i; N)$, and thus is divisible by $p_i = x^{k_i}$. Since $\deg f_i < k_i$, $f_i = c_i x^{k_i-1}$. Thus

$$\alpha = c_1(N^{k_1-1}(\alpha_1)) + \dots + c_r(N^{k_r-1}(\alpha_r)).$$

There are r such vectors, independent and spans $N(N)$, so the nullity is just r . ▣

8.4.2 The Jordan Form

Definition 8.5 (Jordan Form). Let $T \in \mathcal{L}(V)$, with its minimal and characteristic polynomial completely factored.

$$p = \prod_{i=1}^k (x - c_i)^{r_i}, \quad f = \prod_{i=1}^k (x - c_i)^{d_i}.$$

By [Theorem 7.14](#) and [Theorem 8.5](#), the following is guaranteed,

$$V = \bigoplus_{i=1}^k W_i, \quad W_i = N((T - c_i I)^{r_i}), \quad d_i = \dim W_i.$$

Let T_i be the restriction of T on W_i , and define

$$N_i = T_i - c_i I_{W_i}, \quad T = \bigoplus_{i=1}^k (N_i + c_i I_{W_i}).$$

We see N_i are nilpotent operators, and have minimal polynomial $p'_i = x^{r_i}$. Motivated by the simple rational form of nilpotent operators ([Theorem 8.8](#)), we further cyclic decompose W_i ,

$$\begin{aligned} W_i &= \bigoplus_{j=1}^{s_i} Z(\alpha_j^i; T), \quad p_j = x^{k_j^i}, \quad k_j^i \leq r_i, \\ \mathfrak{B}_j^i &= \{\alpha_j^i, T(\alpha_j^i), \dots, T^{k_j^i-1}(\alpha_j^i)\}, \\ \mathfrak{B}_i &= \bigcup_{j=1}^{s_i} \mathfrak{B}_j^i, \\ \mathfrak{B} &= \bigcup_{i=1}^k \mathfrak{B}_i, \\ [N_i]_{\mathfrak{B}_i} &= \bigoplus_{j=1}^{s_i} J_{0, k_j^i}, \end{aligned}$$

where \mathfrak{B}_j^i is the basis of $Z(\alpha_j^i; T)$, \mathfrak{B}_i is the basis of W_i , and \mathfrak{B} is the basis of V .

By the equation $T = \bigoplus_{i=1}^k (N_i + c_i I_{W_i})$, we see the full matrix is of the form

$$\begin{aligned} [T]_{\mathfrak{B}} &= \bigoplus_{i=1}^k (N_i + c_i I_{W_i})_{\mathfrak{B}_i}, \quad (N_i + c_i I_{W_i})_{\mathfrak{B}_i} \in M_{d_i \times d_i}(F), \\ (N_i + c_i I_{W_i})_{\mathfrak{B}_i} &= \bigoplus_{j=1}^{s_i} (J_{0, k_j^i} + c_i I_{k_j^i}) := \bigoplus_{j=1}^{s_i} J_{c_i, k_j^i}, \quad J_{c_i, k_j^i} \in M_{k_j^i \times k_j^i}(F), \end{aligned}$$

where J_{c_i, k_j^i} is called the elementary Jordan matrix with characteristic value c_i ,

$$J_{c_i, k_j^i} := \begin{pmatrix} c_i & 0 & \cdots & 0 & 0 \\ 1 & c_i & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ & & & c_i & 0 \\ 0 & 0 & \cdots & 1 & c_i \end{pmatrix} \in M_{k_j^i \times k_j^i}(F).$$

This decomposition is called the Jordan form, and it is unique by virtue of [Theorem 7.14](#) and [Theorem 8.4](#).

✂ *Remark.* That's a lot of information. A quick wrap-up. Given that a linear operator T whose characteristic (hence minimal) polynomial splits into k distinct linear factors, Jordan form allows us to uniquely decompose T into k sub-operators, each living in dimension d_i . Each sub-operators can be decomposed again into s_i subsub-operators, with each subsub-operators living in dimension k_j^i , $k_{j+1}^i \leq k_j^i$, and the numbers s_i, k_j^i are uniquely determined by [Theorem 8.4](#). When using the cyclic basis, the subsub-operators have a special matrix form, namely the elementary Jordan matrices with characteristic value c_i . \square

✂ *Remark.* We observe that $x^{k_i^1} = p_i = x^{r_i}$. Therefore, each sub-operator must begin with a subsub-operator of dimension r_i . However, at this point, we know nothing about the other k_j^i . The Smith normal form accounts for this problem, but when the sub-operator lives in a space with dimension $\leq 6 \times 6$, we only need to know $\dim N(N_i) = \dim N(T - c_i I)$. See the following problem. \square

Problem Hoffman 7.3.11. *Let $N_1, N_2 \in M_{6 \times 6}(F)$ be nilpotent. Suppose they have the same minimal polynomial and nullity. Prove that N_1, N_2 are similar (and hence have the same Jordan form). Show this is not true for 7×7 matrices.*

Proof. We decompose $6 = d_i$ into the sum of decreasing integers, with the number of integers determined by the nullity s_i , and the first number determined by r_i .

$$\begin{aligned}
 6 &= 1 + 1 + 1 + 1 + 1 + 1 \\
 &= 2 + 2 + 2 \\
 &= 2 + 2 + 1 + 1 \\
 &= 2 + 1 + 1 + 1 + 1 \\
 &= 3 + 3 \\
 &= 3 + 2 + 1 \\
 &= 3 + 1 + 1 + 1 \\
 &= 4 + 2 \\
 &= 4 + 1 + 1 \\
 &= 5 + 1 \\
 &= 6.
 \end{aligned}$$

$$\begin{aligned}
 7 &= 1 + 1 + 1 + 1 + 1 + 1 + 1 \\
 &= 2 + 2 + 2 + 1 \\
 &= 2 + 2 + 1 + 1 + 1 \\
 &= 2 + 1 + 1 + 1 + 1 + 1 \\
 &= 3 + 3 + 1 \\
 &= 3 + 2 + 2. \times
 \end{aligned}$$

We see that for the case of 6, there is only one possible decomposition given r_i and s_i ; however, for the case of 7, there exists two possible decompositions with the same $r_i = 3 = s_i$. \square

8.5 The Smith Normal Form

8.5.1 Definition

✍ *Remark.* In this section I want only to state the general procedure of reaching the Smith normal form without giving a rigorous proof, for the proof is quite troublesome and is not so meaningful to physicists. \square

Definition 8.6. A matrix N is equivalent to M if we can go from N to M by a finite number of row and column operations.

Definition 8.7 (Smith Normal Form). Let $N \in M_{m \times n}(F[x])$. N is in Smith normal form if

1. $M_{ij} = 0$ when $i \neq j$.
2. Let $M_{ii} = f_i$, $1 \leq i \leq \min(m, n)$, then $f_i \mid f_{i+1}$.

Theorem 8.9. Any matrix $M \in M_{m \times n}(F[x])$ is equivalent to one and only one matrix N which is in Smith normal form.

Theorem 8.10. Let $A \in M_{n \times n}(F)$, whose invariant factors are p_1, \dots, p_r . Then the matrix $xI - A$ is equivalent to the $n \times n$ diagonal matrix whose diagonal entries are $p_1, \dots, p_r, 1, \dots, 1$.

✍ *Remark.* We now state the concrete steps to get to Smith normal form, given a polynomial matrix.

First, choose the polynomial with the minimal degree in col 1, and exchange that row with the first row. Do polynomial division to every element in col 1. Then we have that the degree of the remainders are less than the original, or that they are all zero. Repeat the steps until col 1 are zero except the first element, which would be the gcd.

$$\begin{pmatrix} f_{11} & f_{12} & \cdots & f_{1n} \\ f_{21} & f_{22} & \cdots & f_{2n} \\ \vdots & \vdots & & \vdots \\ f_{n1} & f_{n2} & \cdots & f_{nn} \end{pmatrix} \Rightarrow \begin{pmatrix} p'_1 & f'_{12} & \cdots & f'_{1n} \\ 0 & f'_{22} & \cdots & f'_{2n} \\ \vdots & \vdots & & \vdots \\ 0 & f'_{n2} & \cdots & f'_{nn} \end{pmatrix}.$$

Do similar thing to row 1. The zeroes in col 1 may be disturbed, but by doing row operations again, it is guaranteed that we will eventually reach

$$\begin{pmatrix} q & 0 & \cdots & 0 \\ 0 & f''_{22} & \cdots & f''_{2n} \\ \vdots & \vdots & & \vdots \\ 0 & f''_{n2} & \cdots & f''_{nn} \end{pmatrix}.$$

Continue with the second column, until the matrix is in Smith normal form. \square

9 Inner Product Spaces

9.1 Inner Product

9.1.1 Definition and Immediate Consequences

Definition 9.1 (Inner Product). Let V be a vector space. An inner product on V is a function $\langle \cdot, \cdot \rangle : V \times V \rightarrow F$ s.t.

1. $\langle \alpha + \gamma, \beta \rangle = \langle \alpha, \beta \rangle + \langle \gamma, \beta \rangle$,
2. $\langle c\alpha, \beta \rangle = c \langle \alpha, \beta \rangle$,
3. $\overline{\langle \alpha, \beta \rangle} = \langle \beta, \alpha \rangle$,
4. $\langle \alpha, \alpha \rangle > 0$ if $\alpha \neq 0$,

which implies

5. $\langle \alpha, \beta + \gamma \rangle = \langle \alpha, \beta \rangle + \langle \alpha, \gamma \rangle$,
6. $\langle \alpha, c\beta \rangle = \bar{c} \langle \alpha, \beta \rangle$,
7. $\langle \alpha, 0 \rangle = \langle 0, \alpha \rangle = 0$,
8. $\langle \alpha, \alpha \rangle = 0 \iff \alpha = 0$,
9. If $\langle \alpha, \beta \rangle = \langle \alpha, \gamma \rangle \ \forall \alpha \in V$, then $\beta = \gamma$.

Inner product is linear in the first component, and conjugate linear in the second component. A vector space equipped with an inner product is called an inner product space. Two inner product spaces with the same underlying vector space may be different if their associated inner products are different.

9.1.2 Norms

Definition 9.2 (Norm). Let V be an inner product space. $\forall \alpha \in V$, we define the norm or length of α by

$$\|\alpha\| = (\langle \alpha, \alpha \rangle)^{1/2}.$$

Theorem 9.1. *Let V be an inner product space. Then*

1. $\|c\alpha\| = |c| \cdot \|\alpha\|$.

2. $\|\alpha\| = 0 \iff \alpha = 0$. $\|\alpha\| \geq 0$, $\forall \alpha \in V$.

Proof. (1)

$$\|c\alpha\| = (\langle c\alpha, c\alpha \rangle)^{1/2} = (c\bar{c} \langle \alpha, \alpha \rangle) = |c| \cdot \|\alpha\|.$$

(2) Trivial. ▣

Theorem 9.2 (Pythagorean). *Let V be an inner product space, and $\langle \alpha, \beta \rangle = 0$. Show*

$$\|\alpha + \beta\|^2 = \|\alpha\|^2 + \|\beta\|^2.$$

Proof.

$$\begin{aligned} \|\alpha + \beta\|^2 &= \langle \alpha + \beta, \alpha + \beta \rangle = \langle \alpha + \beta, \alpha \rangle + \langle \alpha + \beta, \beta \rangle \\ &= \langle \alpha, \alpha \rangle + \langle \beta, \alpha \rangle + \langle \alpha, \beta \rangle + \langle \beta, \beta \rangle \\ &= \|\alpha\|^2 + \|\beta\|^2. \end{aligned}$$
▣

Theorem 9.3 (Cauchy-Schwarz Inequality). *Let V be an inner product space. Then*

$$|\langle \alpha, \beta \rangle| \leq \|\alpha\| \cdot \|\beta\|.$$

Moreover, if $|\langle \alpha, \beta \rangle| = \|\alpha\| \cdot \|\beta\|$, then $\beta = c\alpha$.

Proof. (trivial) If $\beta = 0$, then the result is immediate.

(non-trivial) If $\beta \neq 0$, then $\forall c \in F$,

$$\begin{aligned} 0 &\leq \|\alpha - c\beta\|^2 = \langle \alpha - c\beta, \alpha - c\beta \rangle \\ &= \langle \alpha, \alpha - c\beta \rangle - c \langle \beta, \alpha - c\beta \rangle \\ &= \langle \alpha, \alpha \rangle - \bar{c} \langle \alpha, \beta \rangle - c \langle \beta, \alpha \rangle + c\bar{c} \langle \beta, \beta \rangle. \end{aligned}$$

Choose $c = \frac{\langle \alpha, \beta \rangle}{\langle \beta, \beta \rangle}$, then

$$0 \leq \langle \alpha, \alpha \rangle - \frac{|\langle \alpha, \beta \rangle|^2}{\langle \beta, \beta \rangle} - \frac{|\langle \alpha, \beta \rangle|^2}{\langle \beta, \beta \rangle} + \frac{|\langle \alpha, \beta \rangle|^2}{\langle \beta, \beta \rangle}$$

so

$$0 \leq \|\alpha\|^2 - \frac{|\langle \alpha, \beta \rangle|^2}{\|\beta\|^2}.$$

(Equality) If $\beta = 0$, the result is trivial. If $\beta \neq 0$ and $|\langle \alpha, \beta \rangle| = \|\alpha\| \cdot \|\beta\|$, let

$$a = \frac{\langle \alpha, \beta \rangle}{\langle \beta, \beta \rangle}, \quad \gamma = \alpha - a\beta.$$

Then

$$\langle \beta, \gamma \rangle = \langle \beta, \alpha - a\beta \rangle = \langle \beta, \alpha \rangle - \bar{a} \langle \beta, \beta \rangle = \langle \beta, \alpha \rangle - \langle \beta, \alpha \rangle = 0.$$

Moreover,

$$|a|^2 = a\bar{a} = \frac{\langle \alpha, \beta \rangle}{\langle \beta, \beta \rangle} \cdot \overline{\frac{\langle \alpha, \beta \rangle}{\langle \beta, \beta \rangle}} = \frac{|\langle \alpha, \beta \rangle|^2}{\langle \beta, \beta \rangle^2} = \frac{\|\alpha\|^2 \cdot \|\beta\|^2}{\|\beta\|^4}.$$

So $|a| = \frac{\|\alpha\|}{\|\beta\|}$. Since β and γ are orthogonal, by [Theorem 9.2](#),

$$\begin{aligned} \|\alpha\|^2 &= \|a\beta + \gamma\|^2 \\ &= \|a\beta\|^2 + \|\gamma\|^2 \\ &= \frac{\|\alpha\|^2}{\|\beta\|^2} \|\beta\|^2 + \|\gamma\|^2. \end{aligned}$$

So $\|\gamma\|^2 = 0 \implies \gamma = 0$. ▣

Theorem 9.4 (Triangle). *Let V be an inner product space. Then*

$$\|\alpha + \beta\| \leq \|\alpha\| + \|\beta\|.$$

Moreover, if $\|\alpha + \beta\| = \|\alpha\| + \|\beta\|$, then either $\alpha = 0$, $\beta = 0$, or $\beta = c\alpha$, $c > 0$.

Proof. (Inequality)

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

(Equality) If $\beta = 0$, the result is obvious. If not, by the proof of the inequality, we see the equation is true only when

$$\Re(\langle \alpha, \beta \rangle) = \langle \alpha, \beta \rangle = \|\alpha\| \cdot \|\beta\|.$$

By the equality condition of [Theorem 9.3](#), we see $\beta = c\alpha$. Now since $\langle \alpha, \beta \rangle = \bar{c} \langle \alpha, \alpha \rangle = \Re(\langle \alpha, \beta \rangle)$, we see $c \in \mathbb{R}$. Moreover, since $\langle \alpha, \beta \rangle = \bar{c} \langle \alpha, \alpha \rangle = \|\alpha\| \cdot \|\beta\| \geq 0$ and $\langle \alpha, \alpha \rangle \geq 0$, this means that $c \geq 0$. But $\beta \neq 0$, so $c > 0$. ▣

9.1.3 Standard Inner Products

The following are the default inner products when not specified.

Definition 9.3. For $\alpha = (a_1, a_2, \dots, a_n)$ and $\beta = (b_1, b_2, \dots, b_n)$ in F^n , define

$$\langle \alpha, \beta \rangle = \sum_{i=1}^n a_i \overline{b_i}.$$

Definition 9.4. For $A, B \in M_{n \times n}(F)$, define the Frobenius inner product to be

$$\langle A, B \rangle = \text{tr}(B^* A).$$

Then

$$\langle A, A \rangle = \sum_{i=1}^n \sum_{k=1}^n |A_{ki}|^2.$$

Definition 9.5. We denote the inner product space H by the vector space of continuous functions $f : [0, 2\pi] \rightarrow \mathbb{C}$ equipped with the inner product

$$\langle f, g \rangle = \frac{1}{2\pi} \int_0^{2\pi} f(t) \overline{g(t)} dt.$$

9.2 Adjoint

9.2.1 Natural Pairing

Definition 9.6 (Standard Dual Basis). For a vector space V equipped with an inner product $\langle \cdot, \cdot \rangle$, denote $|\alpha\rangle \in V$, and define $\langle \alpha| \in V^*$ by

$$\langle \alpha| = \langle \cdot, \alpha \rangle.$$

Let $\mathfrak{B} = \{e_1, \dots, e_n\}$ be an orthonormal basis of V , then

$$\alpha = \sum_{i=1}^n \langle \alpha, e_i \rangle e_i.$$

Thus,

$$e^i := \langle \cdot, e_i \rangle = f_i, \quad e^i e_j = \delta_j^i,$$

where f_i is the i -th coordinate function. Thus, by [Theorem 2.27](#), $\mathfrak{B}^* = \{e^1, \dots, e^n\}$ forms a basis of V^* . The basis \mathfrak{B}^* is called the standard dual basis of V^* , and therefore, $\forall g \in V^*, \exists a_i$ s.t. $g = \sum_{i=1}^n a_i e^i$.

Theorem 9.5. Let $|\alpha\rangle = \sum_{i=1}^n a^i e_i$, then

$$\langle\alpha| = \sum_{i=1}^n \overline{a_i} e^i.$$

We represent $\langle\alpha|$ as a row vector, and $|\alpha\rangle$ as a column vector from now on.

Proof. Let $|\alpha\rangle = \sum_{i=1}^n a^i e_i$. Then

$$\langle\alpha|(|e_i\rangle) := \langle e_i, \alpha \rangle = \sum_{j=1}^n \overline{a_j} \langle e_i, e_j \rangle = \overline{a_i}.$$

Consider $\langle\alpha| = \sum_{i=1}^n a'_i e^i$, then

$$\langle\alpha|(|e_i\rangle) = \sum_{j=1}^n a'_j e^j(|e_i\rangle) = a'_i.$$

Therefore,

$$a'_i = \overline{a_i}.$$

▣

10 Linear Differential Equations

Theorem 10.1. Let $f : \mathbb{R} \rightarrow \mathbb{R}$ be n -times differentiable function, and it satisfy

$$\frac{d^n f}{dt^n} + a_{n-1} \frac{d^{n-1} f}{dt^{n-1}} + \cdots + a_1 \frac{df}{dt} + a_0 f = 0.$$

Let

$$p = x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0 = \prod_{i=1}^k (x - c_i)^{r_i}.$$

Then the solution space V is n -dimensional. Let $\mathfrak{B}_i = \{e^{c_i t}, t e^{c_i t}, \dots, t^{r_i-1} e^{c_i t}\}$, then $\mathfrak{B} = \bigcup_{i=1}^k \mathfrak{B}_i$ forms a basis of V .

Proof. Let D denote the differentiation operator, then $N(p(D)) = V$. If we restrict D on V , then $p(D) = 0$. Let $W_j = N(D - c_j I)^{r_j}$, then by [Theorem 7.14](#), $V = \bigoplus_{i=1}^k W_i$. So any $f \in V$, which is a solution of the differential equation, can be decomposed as

$$f = f_1 + \cdots + f_k, \quad (D - c_j I)^{r_j} f_j = 0.$$

We now show that $\mathfrak{B}_i = \{e^{c_i t}, te^{c_i t}, \dots, t^{r_i-1}e^{c_i t}\}$ forms a basis of W_i . By induction,

$$(D - c_i I)^{r_i} f_i = e^{c_i t} D^{r_i} (e^{-c_i t} f_i),$$

so $(D - c_i I)^{r_i} f_i = 0 \iff D^{r_i} (e^{-c_i t} f_i) = 0$. But if $D^{r_i} (e^{-c_i t} f_i) = 0$, that means it is a polynomial with degree less than or equal to $(r_i - 1)$, then

$$f_i = e^{c_i t} (b_0 + b_1 t + \dots + b_{r_i-1} t^{r_i-1}) = \sum_{i=0}^{r_i-1} b_i t^i e^{c_i t}.$$

Since $t^i e^{c_i t}$ are linearly independent, we see they form a basis of W_i . ▣