Theorem 1. *Let V be a vector space and let $\mathcal{S}(V)$ be the set of all subspaces of V. If $S, T \in \mathcal{S}(V)$, then*

$$S \cap T = \text{glb}\,\{S, T\}$$

and

$$S + T = \text{lub}\,\{S, T\}.$$

Proof. It is clear that $S \cap T$ is a lower bound of $\{S, T\}$. If $U \subseteq S$ and $U \subseteq T$, then $U \subseteq S \cap T$. Similarly, $S + T$ is an upper bound of $\{S, T\}$. If $U \supseteq S$ and $U \supseteq T$, then $s + t \in U$ whenever $s \in S$ and $t \in T$. Therefore $U \supseteq S + T$. □

Theorem 2. *Let V be a vector space and let S, T be subspaces of V with bases B_1, B_2 respectively. Then the following are equivalent:*

(1) *$B_1 \cap B_2$ is empty and $B_1 \cup B_2$ is a basis for V.*
(2) *$V = S \oplus T$.*

Proof. Suppose that $B = B_1 \cup B_2$ is a basis for V and $B_1 \cap B_2$ is empty. If $v \in V$ then we may write

$$v = a_1 v_1 + \cdots + v_n b_n$$

where $v_1, \ldots, v_n \in B$. Rearranging terms shows that $v = s + t$ for some $s \in S$ and $t \in T$, and therefore $V = S + T$. Now suppose that $S \cap T \neq \{0\}$, so that there exists $v \neq 0$ such that

$$v = a_1 v_1 + \cdots + a_m v_m$$

and

$$v = b_1 w_1 + \cdots + b_n w_n$$

where $v_1, \ldots, v_m \in B_1$, $w_1, \ldots, w_n \in B_2$, not all $a_1, \ldots, a_m$ are zero, and not all $b_1, \ldots, b_n$ are zero. Since $B_1 \cap B_2$ is empty, all of $v_1, \ldots, v_m, w_1, \ldots, w_n$ are distinct, and

$$a_1 v_1 + \cdots + a_m v_m - b_1 w_1 - \cdots - b_n w_n = 0$$

contradicts the linear independence of B.

Conversely, suppose that $V = S \oplus T$. If $v \in B_1 \cap B_2$, then $v \in S \cap T$, which is a contradiction. Therefore $B_1 \cap B_2$ is empty. We know that $B_1 \cup B_2$ spans V since $V = S + T$. Suppose that

$$a_1 v_1 + \cdots + a_m v_m + b_1 w_1 + \cdots + b_n w_n = 0$$

where $v_1, \ldots, v_m \in B_1$, $w_1, \ldots, w_n \in B_2$, not all $a_1, \ldots, a_m$ are zero, and not all $b_1, \ldots, b_n$ are zero. Then

$$a_1 v_1 + \cdots + a_m v_m = -b_1 w_1 - \cdots - b_n w_n$$

is a nonzero element in $S \cap T$, which is a contradiction. This shows that $B_1 \cup B_2$ is linearly independent, and consequently $B_1 \cup B_2$ is a basis for V. □

Theorem 3. *Any subspace of a vector space has a complement.*

Proof. Let W be a subspace of V. There exists a basis B_0 of W and a basis B of V containing B_0, by Theorem 1.9. Therefore

$$\begin{aligned} V &= \operatorname{span}(B_0) \oplus \operatorname{span}(B \setminus B_0) \\ &= W \oplus \operatorname{span}(B \setminus B_0) \end{aligned}$$

by Theorem 2. □

Example 4. [Exercise 1.4] Suppose that V is a vector space with basis $B = \{b_i \mid i \in I\}$ and S is a subspace of V. Let $\{B_1, \ldots, B_k\}$ be a partition of B. Then is it true that

$$S = \bigoplus_{i=1}^{k} (S \cap \langle B_i \rangle)?$$

What if $S \cap \langle B_i \rangle \neq \{0\}$ for all i?

Take

$$S = \langle e_1, e_2, e_3 \rangle, \quad B_1 = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \right\}, \quad B_2 = \left\{ \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \right\}$$

in $\mathbb{R}^4$. We have

$$\bigoplus_{i=1}^{2} (S \cap \langle B_i \rangle) = \langle e_2 \rangle \oplus \langle e_1 \rangle \neq S$$

but $S \cap \langle B_i \rangle \neq \{0\}$ for all i.

Theorem 5. *[Exercise 1.6] Let $S, T, U \in \mathcal{S}(V)$. If $U \subseteq S$, then*

$$S \cap (T + U) = (S \cap T) + U.$$

Proof. If $v \in S \cap (T + U)$ then $v \in S$ and $v = t + u$ where $t \in T$ and $u \in U \subseteq S$. Therefore $t = v - u \in S$ and $v \in (S \cap T) + U$. Conversely, if $v \in (S \cap T) + U$ then $v = s + u$ where $s \in S \cap T$ and $u \in U \subseteq S$. Therefore $v \in S$ and $v \in S \cap (T + U)$. □

Theorem 6. *[Exercise 1.8] Let $v = (a_1, \ldots, a_n) \in \mathbb{R}^n$ be a strongly positive vector.*

1. *Let $M = \min\{a_1, \ldots, a_n\}$. Then any vector $w \in \mathbb{R}^n$ with $\|v - w\| < M$ is strongly positive.*
2. *If a subspace S of $\mathbb{R}^n$ contains v, then S has a basis of strongly positive vectors.*

Proof. Suppose that $w = (b_1, \ldots, b_n) \in \mathbb{R}^n$ has some $b_i \leq 0$ and $\|v - w\| < M$. Then $(a_i - b_i)^2 \geq M^2$, which contradicts

$$(a_1 - b_1)^2 + \cdots + (a_n - b_n)^2 < M^2.$$

Let S be a subspace of $\mathbb{R}^n$ that contains v, and let $\{v = v_1, v_2, \ldots, v_m\}$ be a basis of S containing v. Choose a positive number C such that

$$\left\| v - \left(v + \frac{v_i}{C} \right) \right\| = \left\| \frac{v_i}{C} \right\| < M$$

for every $i \geq 2$; then

$$\left\{ v, v + \frac{v_2}{C}, \ldots, v + \frac{v_m}{C} \right\}$$

is a basis of strongly positive vectors by (1). □

Theorem 7. *[Exercise 1.14] Let V be a finite-dimensional vector space over an infinite field F. If $S_1, \ldots, S_k$ are subspaces of V of equal dimension, then there is a subspace T of V for which $V = S_i \oplus T$ for all $i = 1, \ldots, k$.*

Proof. Let n be the dimension of V and let m be the common dimension of $S_1, \ldots, S_k$. We use induction on the value of $n - m$. If $m = n$, we can choose $T = \{0\}$. Otherwise, assume that the statement is true for k subspaces of dimension $m + 1$. By Theorem 1.2, we can choose a vector $v \in V \setminus (S_1 \cup \cdots \cup S_k)$. The subspaces $\{S_i \oplus \langle v \rangle\}$ have dimension $m + 1$, so applying the induction hypothesis gives a subspace T of V such that $V = (S_i \oplus \langle v \rangle) \oplus T$ for all i. Then $V = S_i \oplus (\langle v \rangle \oplus T)$, which shows that $\langle v \rangle \oplus T$ is a common complement of $S_1, \ldots, S_k$. □

Lemma 8. *If V contains an infinite set that is linearly independent, then V is infinite-dimensional.*

Proof. Let S be an infinite linearly independent set in V. Suppose that a finite set $\{v_1, \ldots, v_n\}$ spans V, and choose $n + 1$ elements from S. This contradicts Theorem 1.10. □

Theorem 9. *[Exercise 1.15] The vector space $\mathcal{C}$ of all continuous functions from $\mathbb{R}$ to $\mathbb{R}$ is infinite-dimensional.*

Proof. For any positive integer n, define $f_n : \mathbb{R} \to \mathbb{R}$ by

$$f_n(x) = \begin{cases} 1 - |x| & \text{if } -1 < x < 1, \\ 0 & \text{otherwise.} \end{cases}$$

The set $\{f_n \mid n \geq 1\}$ is linearly independent, and therefore $\mathcal{C}$ is infinite-dimensional by Lemma 8. □

Theorem 10. *[Exercise 1.19] Let V be an n-dimensional real vector space and suppose that S is a subspace of V with $\dim(S) = n-1$. Define an equivalence relation $\equiv$ on the set $V \setminus S$ by $v \equiv w$ if the "line segment"*

$$L(v,w) = \{rv + (1-r)w \mid 0 \leq r \leq 1\}$$

has the property that $L(v,w) \cap S = \emptyset$. Then $\equiv$ is an equivalence relation and it has exactly two equivalence classes.

Proof. Let $B_0 = (b_1, \ldots, b_{n-1})$ be an ordered basis for S and complete it to an ordered basis $B = (b_1, \ldots, b_n)$ of V. Let $\phi : V \to \mathbb{R}$ be the map that takes a vector to its nth coordinate with respect to B, and let $\operatorname{sgn} : \mathbb{R} \setminus \{0\} \to \{-1, 1\}$ be the sign function $\operatorname{sgn}(x) = x/|x|$. Define an equivalence relation $\sim$ on $V \setminus S$ by

$$v \sim w \Leftrightarrow \operatorname{sgn}(\phi(v)) = \operatorname{sgn}(\phi(w)).$$

We want to show that $\equiv$ is identical to $\sim$, i.e. $v \equiv w$ if and only if $v \sim w$. Let $v, w \in V \setminus S$ with $v \not\equiv w$. Then $rv + (1-r)w \in S$ for some $0 < r < 1$ (note $r \neq 0, 1$ since $v, w \notin S$), and $r\phi(v) + (1-r)\phi(w) = 0$. Rewriting this as

$$\phi(v) = \frac{r-1}{r}\phi(w)$$

shows that $\operatorname{sgn}(\phi(v)) = -\operatorname{sgn}(\phi(w))$ and $v \not\sim w$. Conversely, let $v, w \in V \setminus S$ with $v \not\sim w$. Put

$$r = -\frac{\phi(w)}{\phi(v) - \phi(w)}$$

so that $r\phi(v) + (1-r)\phi(w) = 0$ and therefore $rv + (1-r)w \in S$. This proves that $v \not\equiv w$, and the result follows from the fact that $\sim$ is an equivalence relation and $\sim$ partitions $\mathbb{R} \setminus \{0\}$ into the positive and negative numbers. □

Theorem 11. *[Exercise 1.22] Every subspace S of $\mathbb{R}^n$ is a closed set.*

Proof. Let $B_0 = (v_1, \ldots, v_m)$ be an ordered orthonormal basis for S and complete it to an ordered orthonormal basis $B = (v_1, \ldots, v_n)$ of $\mathbb{R}^n$. Let $x \in S^c$, and write

$$[x]_B = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

At least one of $x_{m+1}, \ldots, x_n$ is not equal to zero, for otherwise $x \in S$. Let

$$M = \min\{x_i \mid i \geq m+1 \text{ and } x_i \neq 0\}.$$

We want to show that the open ball $B(x, M)$ is wholly contained in S^c. Let $y \in \mathbb{R}^n$ with $\|x - y\| = \|[x]_B - [y]_B\| < M$, and write

$$[y]_B = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}.$$

If $y \in S$ then $y_{m+1} = \cdots = y_n = 0$ so that

$$(x_1 - y_1)^2 + \cdots + (x_m - y_m)^2 + x_{m+1}^2 + \cdots + x_n^2 < M^2$$

contradicts the fact that $x_i \geq M$ for at least one $i \geq m + 1$. This completes the proof. □

Theorem 12. *[Exercise 1.24] Let V be an n-dimensional vector space over an infinite field F and suppose that $S_1, \ldots, S_k$ are subspaces of V with $\dim(S_i) \leq m < n$. Then there is a subspace T of V of dimension $n - m$ for which $T \cap S_i = \{0\}$ for all i.*

Proof. Extend each subspace to dimension m, and apply Theorem 7. □

Theorem 13. *[Exercise 1.26] Let V be a real vector space with complexification $V^{\mathbb{C}}$ and let U be a subspace of $V^{\mathbb{C}}$. Then the following are equivalent:*

(1) *There is a subspace S of V for which $U = S^{\mathbb{C}} = \{s + it \mid s, t \in S\}$.*
(2) *U is closed under complex conjugation $\chi : V^{\mathbb{C}} \to V^{\mathbb{C}}$ defined by $u + iv \mapsto u - iv$.*

Proof. (1) $\Rightarrow$ (2) since $S^{\mathbb{C}}$ is closed under complex conjugation. Suppose that (2) holds and let $S = \{u \mid u + iv \in U\}$. We want to show that $U = S^{\mathbb{C}}$. If $u + iv \in U$, then $u \in S$ and $v \in S$ since $v - iu = -i(u + iv) \in U$. Conversely, if $s + it \in S^{\mathbb{C}}$ then $s + iv_1 \in U$ and $t + iv_2 \in U$ for some v_1, v_2. Therefore

$$s + it = \left(\frac{s + iv_1}{2} + \frac{s - iv_1}{2}\right) + i\left(\frac{t + iv_2}{2} + \frac{t - iv_2}{2}\right) \in U.$$

□

Chapter 2. Linear Transformations

Theorem 14. *[Universal property of a free object] Let V and W be vector spaces, let B be a basis for V and let $\tau_0 : B \to W$ be any function. Then τ_0 extends uniquely to a linear map $\tau : V \to W$. That is, there exists a unique linear map $\tau : V \to W$ such that $\tau i = \tau_0$, where $i : B \to V$ is the canonical injection.*

Proof. For any $v \in V$ write $v = a_1 b_1 + \cdots + a_n b_n$ where $b_1, \ldots, b_n \in B$, and define $\tau v = a_1 \tau_0 b_1 + \cdots + a_n \tau_0 b_n$. It is clear that τ is well-defined and linear. Suppose that $\tau' : V \to W$ is a linear map that extends τ_0. Let $v \in V$ and write $v = a_1 b_1 + \cdots + a_n b_n$ where $b_1, \ldots, b_n \in B$. Then

$$\tau' v = a_1 \tau_0 b_1 + \cdots + a_n \tau_0 b_n = \tau v,$$

which shows that $\tau' = \tau$. □

Theorem 15. *[Universal property of a free object, converse] Let B be a subset of V such that for every vector space W and every $\tau_0 : B \to W$ there exists a unique linear map $\tau : V \to W$ that extends τ_0. Then B is a basis for V.*

Proof. We may assume that $V \neq \{0\}$, for otherwise the result follows trivially. Suppose that B is linearly dependent, i.e.

$$a_1 b_1 + \cdots + a_n b_n = 0$$

where $b_1, \ldots, b_n$ are distinct elements from B and not all $a_1, \ldots, a_n$ are zero. We may write without loss of generality

$$b_1 = -\frac{1}{a_1}(a_2 b_2 + \cdots + a_n b_n),$$

and we may choose $\tau_0 : B \to V$ such that

$$\tau_0 b_1 \neq -\frac{1}{a_1}(a_2 \tau_0 b_2 + \cdots + a_n \tau_0 b_n).$$

This produces a contradiction when we extend τ_0 to a linear map, and therefore B is linearly independent. If $V = \{0\}$ then we are done. Otherwise if B does not span V, then we may choose a vector $v \in V \setminus \operatorname{span}(B)$ so that $B \cup \{v\}$ is linearly independent. Setting $\tau b = b$ for each $b \in B$ and either $\tau v = 0$ or $\tau v \neq 0$ shows that there is no unique linear map that extends the canonical injection $i : B \to V$. This is a contradiction, so B spans V. □

Theorem 16. *Let $\tau \in \mathcal{L}(V, W)$ be an isomorphism. Let $S \subseteq V$. Then*

1. *S spans V if and only if τS spans W.*
2. *S is linearly independent in V if and only if τS is linearly independent in W.*
3. *S is a basis for V if and only if τS is a basis for W.*

Proof. We will only prove (1) and (2) in one direction, for the same argument can be applied with the isomorphism τ^{-1}. Suppose that τS spans W, and let $v \in V$. Then $\tau v \in W$, so we may write

$$\begin{aligned} \tau v &= a_1 \tau s_1 + \cdots + a_n \tau s_n \\ &= \tau(a_1 s_1 + \cdots + a_n s_n) \end{aligned}$$

for some $s_1, \ldots, s_n \in S$. Since τ is an isomorphism, applying τ^{-1} gives

$$v = a_1 s_1 + \cdots + a_n s_n,$$

which shows that S spans V. Now suppose that τS is linearly independent in W and

$$a_1 s_1 + \cdots + a_n s_n = 0$$

where $s_1, \ldots, s_n \in S$. Then

$$\begin{aligned} 0 &= \tau(a_1 s_1 + \cdots + a_n s_n) \\ &= a_1 \tau s_1 + \cdots + a_n \tau s_n, \end{aligned}$$

and $a_1 = \cdots = a_n = 0$ by the linear independence of τS. (3) follows immediately from (1) and (2). □

Theorem 17. *Let V be a vector space and let $\rho \in \mathcal{L}(V)$.*

(1) *If $V = S \oplus T$ then*

$$\rho_{S,T} + \rho_{T,S} = \iota.$$

(2) *If $\rho = \rho_{S,T}$ then*

$$\operatorname{im}(\rho) = S \quad \text{and} \quad \ker(\rho) = T$$

and so

$$V = \operatorname{im}(\rho) \oplus \ker(\rho).$$

In other words, ρ is a projection onto its image along its kernel. Moreover,

$$v \in \operatorname{im}(\rho) \quad \Leftrightarrow \quad \rho v = v.$$

(3) *If $\sigma \in \mathcal{L}(V)$ has the property that*

$$V = \operatorname{im}(\sigma) \oplus \ker(\sigma) \quad \text{and} \quad \sigma|_{\operatorname{im}(\sigma)} = \iota$$

then σ is a projection onto $\operatorname{im}(\sigma)$ along $\ker(\sigma)$.

Proof. (1) and (2) are obvious. Let $v = v_1 + v_2 \in V$ where $v_1 \in \operatorname{im}(\sigma)$ and $v_2 \in \ker(\sigma)$. Then

$$\begin{aligned} \sigma(v) &= \sigma(v_1 + v_2) \\ &= \iota(v_1) + \sigma(v_2) \\ &= v_1, \end{aligned}$$

which proves (3). □

Theorem 18. *Let $V = S \oplus T$. Then (S, T) reduces $\tau \in \mathcal{L}(V)$ if and only if τ commutes with $\rho_{S,T}$.*

Proof. By Theorem 2.23, S and T are τ-invariant if and only if

$$(*) \qquad \rho_{S,T}\tau\rho_{S,T} = \rho_{S,T}\tau$$

and

$$(**) \qquad (\iota - \rho_{S,T})\tau(\iota - \rho_{S,T}) = (\iota - \rho_{S,T})\tau$$

since $\rho_{S,T} + \rho_{T,S} = \iota$. Given (*), (**) is equivalent to

$$\begin{aligned} \tau - \tau\rho_{S,T} - \rho_{S,T}\tau + \rho_{S,T}\tau\rho_{S,T} &= \tau - \rho_{S,T}\tau \\ \tau\rho_{S,T} &= \rho_{S,T}\tau. \end{aligned}$$

Both conditions taken together are equivalent to just $\tau\rho_{S,T} = \rho_{S,T}\tau$. □

Theorem 19. *[Exercise 2.1] Let $A \in \mathcal{M}_{m,n}$ have rank k.*

1. *There exist matrices $X \in \mathcal{M}_{m,k}$ and $Y \in \mathcal{M}_{k,n}$, both of rank k, such that $A = XY$.*
2. *A has rank 1 if and only if it has the form $A = x^T y$ where x and y are row matrices.*

Proof. If A has rank k, then

$$A = P \begin{bmatrix} I_k & 0 \\ 0 & 0 \end{bmatrix} Q$$

by performing elementary row and column operations, where P and Q are invertible. But

$$A = P'Q'$$

where P' contains the first k columns of P and Q' contains the first k rows of Q. This proves (1). Suppose now that A has rank 1; (1) shows that $A = x^T y$ for row matrices x and y. Conversely,

$$x^T y = \begin{bmatrix} x^T & 0 \\ & I_{k-1} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} & y \\ 0 & I_{k-1} \end{bmatrix},$$

which proves (2). □

Theorem 20. *[Exercise 2.2] Let $\tau \in \mathcal{L}(V, W)$, where $\dim(V) = \dim(W) < \infty$. Then τ is injective if and only if τ is surjective. This does not hold if the finiteness condition is removed.*

Proof. By Theorem 2.8, $\dim(\ker(\tau)) = 0$ if and only if $\dim(\mathrm{im}(\tau)) = \dim(W)$.

Let V be the set of all functions $f : \mathbb{R} \to \mathbb{R}$ such that $f(x) = 0$ whenever $x < -1$ or $x > 1$, as a real vector space. Define the linear map $\tau : V \to V$ by $(\tau f)(x) = f(2x)$; then τ is injective but not surjective. □

Theorem 21. *[Exercise 2.3] Let $\tau \in \mathcal{L}(V, W)$. Then τ is an isomorphism if and only if it carries a basis for V to a basis for W.*

Proof. Theorem 16 shows that if τ is an isomorphism, then it carries a basis for V to a basis for W. Conversely, suppose that τ carries a basis B for V to a basis B' for W. Clearly τ is surjective. If

$$\begin{aligned}\tau(a_1v_1 + \cdots + a_nv_n) &= a_1\tau v_1 + \cdots + a_n\tau v_n \\ &= 0\end{aligned}$$

where $v_1, \ldots, v_n \in B$, then $a_1 = \cdots = a_n = 0$ since $\tau v_1, \ldots, \tau v_n \in B'$. This shows that τ is a bijection. □

Theorem 22. *[Exercise 2.5] If $V = S \oplus T$, then $V \cong S \boxplus T$.*

Proof. Let $\varphi : S \boxplus T \to S \oplus T$ be given by $(s, t) \mapsto s + t$. This map is easily seen to be an isomorphism. □

Remark 23. [Exercise 2.6] Let $V = A + B$ and define the linear map $\tau : A \boxplus B \to V$ by $(v, w) \mapsto v + w$. This map is always surjective, but is injective if and only if $V = A \oplus B$.

Theorem 24. *[Exercise 2.7] Let $\tau \in \mathcal{L}_F(V)$ where $\dim(V) = n < \infty$. Let $A \in \mathcal{M}_n(F)$. Suppose that there is an isomorphism $\sigma : V \to F^n$ with the property that $\sigma\tau = A\sigma$. Then there exists an ordered basis B for which $A = [\tau]_B$.*

Proof. Choose $B = (\sigma^{-1}e_1, \ldots, \sigma^{-1}e_n)$, which is a basis by Theorem 16. Then $A = \sigma\tau\sigma^{-1} = [\tau]_B$. □

Definition 25. Let $\mathcal{T}$ be a subset of $\mathcal{L}(V)$. A subspace S of V is **$\mathcal{T}$-invariant** if S is τ-invariant for every $\tau \in \mathcal{T}$. Also, V is **$\mathcal{T}$-irreducible** if the only $\mathcal{T}$-invariant subspaces of V are $\{0\}$ and V.

Theorem 26. *[Exercise 2.8] Suppose that $\mathcal{T}_V \subseteq \mathcal{L}(V)$, $\mathcal{T}_W \subseteq \mathcal{L}(W)$, V is $\mathcal{T}_V$-irreducible and W is $\mathcal{T}_W$-irreducible. Let $\alpha \in \mathcal{L}(V, W)$ satisfy $\alpha\mathcal{T}_V = \mathcal{T}_W\alpha$, that is, for any $\mu \in \mathcal{T}_V$ there is a $\lambda \in \mathcal{T}_W$ such that $\alpha\mu = \lambda\alpha$ and for any $\lambda \in \mathcal{T}_W$ there is a $\mu \in \mathcal{T}_V$ such that $\alpha\mu = \lambda\alpha$. Then $\alpha = 0$ or α is an isomorphism.*

Proof. Suppose first that $\ker\alpha \neq \{0\}$, let $v \in \ker\alpha$, and let $\mu \in \mathcal{T}_V$. There exists a $\lambda \in \mathcal{T}_W$ such that $\alpha\mu = \lambda\alpha$, so $\alpha\mu v = \lambda\alpha v = 0$ and $\mu v \in \ker\alpha$. This shows that $\ker\alpha$ is a $\mathcal{T}_V$-invariant subspace and therefore $\ker\alpha = V$, i.e. $\alpha = 0$. Similarly, suppose that

$\operatorname{im}\alpha \neq V$, let $w \in \operatorname{im}\alpha$ so that $w = \alpha v$ for some $v \in V$, and let $\lambda \in \mathcal{T}_W$. There exists a $\mu \in \mathcal{T}_V$ such that $\alpha\mu = \lambda\alpha$, so $\lambda w = \lambda\alpha v = \alpha\mu v \in \operatorname{im}\alpha$. This shows that $\operatorname{im}\alpha$ is a $\mathcal{T}_W$-invariant subspace and therefore $\operatorname{im}\alpha = \{0\}$, i.e. $\alpha = 0$. We conclude that if α fails to be injective or surjective, then $\alpha = 0$. □

Theorem 27. *[Exercise 2.9] Let $\tau \in \mathcal{L}(V)$ where $\dim(V) < \infty$. If $\operatorname{rk}(\tau^2) = \operatorname{rk}(\tau)$ show that $\operatorname{im}(\tau) \cap \ker(\tau) = \{0\}$.*

Proof. If $v \in \operatorname{im}(\tau) \cap \ker(\tau)$ is nonzero, there exists some nonzero $x \in V$ such that $\tau x = v \neq 0$ and $\tau^2 x = \tau v = 0$. We know that $\ker(\tau) \subseteq \ker(\tau^2)$, and since $x \neq \ker(\tau)$ but $x \in \ker(\tau^2)$, the nullity of τ must be strictly less than the nullity of τ^2. By Theorem 2.8, $\operatorname{rk}(\tau^2) \neq \operatorname{rk}(\tau)$. □

Theorem 28. *[Exercises 2.10,2.11] Let $\tau \in \mathcal{L}(U, V)$ and $\sigma \in \mathcal{L}(V, W)$.*

(1) $\operatorname{rk}(\sigma\tau) \leq \min\{\operatorname{rk}(\tau), \operatorname{rk}(\sigma)\}$.
(2) $\operatorname{null}(\sigma\tau) \leq \operatorname{null}(\tau) + \operatorname{null}(\sigma)$.

Proof. Obvious. □

Theorem 29. *[Exercise 2.12] Let $\tau, \sigma \in \mathcal{L}(V)$ where τ is invertible. Then $\operatorname{rk}(\tau\sigma) = \operatorname{rk}(\sigma\tau) = \operatorname{rk}(\sigma)$.*

Proof. Obvious. □

Theorem 30. *[Exercise 2.13] Let $\tau, \sigma \in \mathcal{L}(V, W)$. Then $\operatorname{rk}(\tau + \sigma) \leq \operatorname{rk}(\tau) + \operatorname{rk}(\sigma)$.*

Proof. $\operatorname{im}(\tau + \sigma) \subseteq \operatorname{im}(\tau) + \operatorname{im}(\sigma)$, and $\dim(A + B) \leq \dim(A \oplus B)$. □

Theorem 31. *[Exercise 2.14] Let S be a subspace of V.*

(1) *There exists a $\tau \in \mathcal{L}(V)$ such that $\ker(\tau) = S$.*
(2) *There exists a $\sigma \in \mathcal{L}(V)$ such that $\operatorname{im}(\sigma) = S$.*

Proof. Choose a complement T so that $V = S \oplus T$. Define τ to be the projection onto T along S, and define σ to be the projection onto S along T. By Theorem 2.21, τ and σ satisfy (1) and (2). Note that τ and σ as defined here are orthogonal. □

Theorem 32. *[Exercise 2.15] Let $\tau, \sigma \in \mathcal{L}(V)$.*

(1) *$\sigma = \tau\mu$ for some $\mu \in \mathcal{L}(V)$ if and only if $\operatorname{im}(\sigma) \subseteq \operatorname{im}(\tau)$.*
(2) *$\sigma = \mu\tau$ for some $\mu \in \mathcal{L}(V)$ if and only if $\ker(\tau) \subseteq \ker(\sigma)$.*

Proof. If $\sigma = \tau\mu$ and $v = \sigma x$ for some $x \in V$, then $v = \tau\mu x \in \operatorname{im}(\tau)$. Conversely, suppose that $\operatorname{im}(\sigma) \subseteq \operatorname{im}(\tau)$. Let $P = \tau^{-1}(\operatorname{im}(\tau))$ and $Q = \tau^{-1}(\operatorname{im}(\sigma))$ be subspaces of V so that $Q \subseteq P$. Define $\mu \in \mathcal{L}(V)$ as a projection onto Q along some complement of Q. Then for any $v \in V$, write $v = q + r$ where $q \in Q$, $r \notin Q \setminus \{0\}$, and

$$\begin{aligned}\tau\mu v &= \tau\mu q + \tau\mu r \\ &= \tau q \\ &= \tau(v - r) \\ &= \tau v - \tau r \\ &= \tau v\end{aligned}$$

since $r \notin Q \subseteq P$ or $r = 0$. The proof for (2) is similar. □

Theorem 33. *[Exercise 2.16] Let* $\dim(V) < \infty$ *and suppose that* $\tau \in \mathcal{L}(V)$ *satisfies* $\tau^2 = 0$. *Then* $2\operatorname{rk}(\tau) \leq \dim(V)$.

Proof. Since $\tau^2 = 0$, we have $\ker(\tau) \subseteq \operatorname{im}(\tau)$ and $\operatorname{null}(\tau) \leq \operatorname{rk}(\tau)$. By Theorem 2.8,

$$2\operatorname{rk}(\tau) \leq \operatorname{null}(\tau) + \operatorname{rk}(\tau) = \dim(V).$$

□

Theorem 34. *[Exercise 2.18] Let* V *have basis* $B = \{v_1, \ldots, v_n\}$ *and assume that the base field* F *for* V *has characteristic 0. Suppose that for each* $1 \leq i, j \leq n$ *we define* $\tau_{i,j} \in \mathcal{L}(V)$ *by*

$$\tau_{i,j}(v_k) = \begin{cases} v_k & \text{if } k \neq i, \\ v_i + v_j & \text{if } k = i. \end{cases}$$

Then the $\tau_{i,j}$ *are invertible and form a basis for* $\mathcal{L}(V)$.

Proof. Define $\epsilon_{i,j} \in \mathcal{L}(V)$ by $\epsilon_{i,j}(v_i) = v_j$ and $\epsilon_{i,j}(v_k) = 0$ for $k \neq i$; then $\tau_{i,j} = \iota + \epsilon_{i,j}$. We can compute inverses

$$\tau_{i,j}^{-1} = \begin{cases} 2^{-1}\epsilon_{i,i} + \sum_{k \neq i} \epsilon_{k,k} & \text{if } i = j, \\ \iota - \epsilon_{i,j} & \text{if } i \neq j, \end{cases}$$

since $[\tau_{i,i}]_B$ is diagonal and

$$(\iota + \epsilon_{i,j})(\iota - \varepsilon_{i,j}) = \iota - \varepsilon_{i,j}^2 = \iota.$$

To see that the $\tau_{i,j}$ form a basis for $\mathcal{L}(V)$, suppose that

$$a_{1,1}\tau_{1,1} + \cdots + a_{n,n}\tau_{n,n} = \sum_{i,j} a_{i,j}\iota + \sum_{i,j} a_{i,j}\epsilon_{i,j} = 0.$$

For each k, we have

$$(*) \qquad \sum_{i,j} a_{i,j}\iota v_k + \sum_{i,j} a_{i,j}\epsilon_{i,j}v_k = \sum_{i,j} a_{i,j}v_k + \sum_j a_{k,j}v_j = 0$$

so that $a_{k,j} = 0$ for all (k, j) with $j \neq k$. Equation (*) then shows that

$$\sum_i a_{i,i} + a_{k,k} = 0$$

for each k; viewing this as n equations in n unknowns, $a_{k,k} = 0$ for each k since $U + I$ is invertible where U is a matrix with all entries set to 1. □

Remark 35. [Exercise 2.19] Let $\tau \in \mathcal{L}(V)$. If S is a τ-invariant subspace of V must there be a subspace T of V for which (S, T) reduces τ? No, unless τ is an isomorphism. In that case, let T be a complement of S in V. Since $\tau|_S : S \to S$ is an isomorphism, $\mathrm{im}(\tau|_T)$ must be a subspace of T.

Example 36. [Exercise 2.20] Find an example of a vector space V and a proper subspace S of V for which $V \cong S$. Let F_a denote the set of all functions $f : \mathbb{R} \to \mathbb{R}$ such that $f(x) = 0$ whenever $x < -a$ or $x > a$. Fix $0 < \alpha < \beta$; then F_α is a proper subspace of F_β, but $\varphi : F_\alpha \to F_\beta$ defined by $(\varphi f)(x) = f(\beta x/\alpha)$ is an isomorphism.

Theorem 37. *[Exercise 2.21] Let* $\dim(V) < \infty$. *If* $\tau, \sigma \in \mathcal{L}(V)$ *then* $\sigma\tau = \iota$ *implies that* τ *and* σ *are invertible and that* $\sigma = p(\tau)$ *for some polynomial* $p(x) \in F[x]$.

Proof. The first assertion is clear since $\ker(\sigma) \neq \{0\}$ or $\ker(\tau) \neq \{0\}$ implies that $\ker(\sigma\tau) \neq \{0\}$. Since $\mathcal{L}(V)$ is finite-dimensional, there exists a nonzero polynomial $p(x) \in F[x]$ such that $p(\tau) = 0$. Such a polynomial can be taken to have nonzero constant term, for otherwise we may apply τ^{-1} to both sides. Then

$$\begin{aligned} -a_0\iota &= \sum_{k\geq 1} a_k\tau^k \\ -a_0\sigma &= \sum_{k\geq 0} a_{k+1}\tau^k \\ \sigma &= -\sum_{k\geq 0} \frac{a_{k+1}}{a_0}\tau^k. \end{aligned}$$

□

Theorem 38. *[Exercise 2.22] Let* $\tau \in \mathcal{L}(V)$. *The following are equivalent:*

(1) $\tau\sigma = \sigma\tau$ *for all* $\sigma \in \mathcal{L}(V)$.
(2) $\tau = a\iota$ *for some* $a \in F$.

Proof. (2) $\Rightarrow$ (1) is obvious. Suppose that $\tau \neq a\iota$ for all $a \in F$. Then there exists some $v \in V$ such that $\{v, \tau v\}$ is linearly independent. Complete this to a (possibly infinite) basis B of V, and choose $\sigma \in \mathcal{L}(V)$ such that $\sigma v = v$, $\sigma\tau v = v + \tau v$, and $\sigma b = b$ for all $b \in B \setminus \{v, \tau v\}$. Then $\tau\sigma v = \tau v \neq \sigma\tau v$ which shows that τ and σ do not commute. This proves (1) $\Rightarrow$ (2). $\square$

Theorem 39. *[Exercise 2.23] Let V be a vector space over a field F of characteristic other than 2, and let ρ, σ be projections.*

(1) *The difference $\rho - \sigma$ is a projection if and only if*

$$\rho\sigma = \sigma\rho = \sigma$$

in which case

$$\operatorname{im}(\rho - \sigma) = \operatorname{im}(\rho) \cap \ker(\sigma) \quad \text{and} \quad \ker(\rho - \sigma) = \ker(\rho) \oplus \operatorname{im}(\sigma).$$

(2) *If ρ and σ commute, then $\rho\sigma$ is a projection, in which case*

$$\operatorname{im}(\rho\sigma) = \operatorname{im}(\rho) \cap \operatorname{im}(\sigma) \quad \text{and} \quad \ker(\rho\sigma) = \ker(\rho) + \ker(\sigma).$$

Proof. The difference $\rho - \sigma$ is a projection if and only if $\iota - (\rho - \sigma) = (\iota - \rho) + \sigma$ is a projection; this sum is a projection if and only if $\iota - \rho \perp \sigma$, i.e. $\sigma = \rho\sigma = \sigma\rho$. In this case, $\operatorname{im}(\rho) \cap \ker(\sigma) \subseteq \operatorname{im}(\rho - \sigma)$ since $v = \rho x$ and $v \in \ker(\sigma)$ implies that $(\rho - \sigma)v = \rho^2 x = v$. If $v = (\rho - \sigma)x$ then $\sigma v = \sigma\rho x - \sigma^2 x = 0$ and $v = (\rho - \rho\sigma)x = \rho(\iota - \sigma)x$, which shows the reverse inclusion. If $v \in \ker(\rho)$ then $(\rho - \sigma)v = \rho v - \sigma\rho v = 0$, and if $v \in \operatorname{im}(\sigma)$ then $(\rho - \sigma)v = \rho(v - \sigma v) = 0$. Therefore $\ker(\rho) + \operatorname{im}(\sigma) \subseteq \ker(\rho - \sigma)$. Conversely, if $v \in \ker(\rho - \sigma)$ then $v = (v - \rho v) + \sigma v \in \ker(\rho) + \operatorname{im}(\sigma)$. Finally, $v \in \ker(\rho) \cap \operatorname{im}(\sigma)$ implies that $v = \sigma x = \sigma\rho x = 0$, which shows that the sum is direct.

If ρ and σ commute then $\rho\sigma\rho\sigma = \rho^2\sigma^2 = \rho\sigma$. Clearly $\operatorname{im}(\rho\sigma) \subseteq \operatorname{im}(\rho)$ and $\operatorname{im}(\rho\sigma) = \operatorname{im}(\sigma\rho) \subseteq \operatorname{im}(\sigma)$. If $v \in \operatorname{im}(\rho) \cap \operatorname{im}(\sigma)$ then $v = \rho x = \sigma x'$ so that

$$\rho\sigma v = \sigma\rho v = \sigma\rho x = \sigma v = \sigma x' = v,$$

which shows that $v \in \operatorname{im}(\rho\sigma)$. Clearly $\ker(\rho) + \ker(\sigma) \subseteq \ker(\rho\sigma)$. If $v \in \ker(\rho\sigma)$ then $v = (v - \rho v) + \rho v \in \ker(\rho) + \ker(\sigma)$, which proves that $\ker(\rho\sigma) = \ker(\rho) + \ker(\sigma)$. $\square$

Theorem 40. *[Exercise 2.24] Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuous function with the property that $f(x + y) = f(x) + f(y)$. Then f is a linear functional on $\mathbb{R}^n$.*

Proof. By induction, $f(nx) = nf(x)$ for all integers n, and consequently $f(rx) = rf(x)$ for all $r \in \mathbb{Q}$. Let $a \in \mathbb{R}$, and let $\{a_n\}$ be a sequence in $\mathbb{Q}$ such that $a_n \to a$. Then

$$\begin{aligned} f(ax) &= f\left(\lim_{n\to\infty} a_n x\right) \\ &= \lim_{n\to\infty} f(a_n x) \end{aligned}$$

$$\begin{aligned} &= \lim_{n\to\infty} a_n f(x) \\ &= af(x) \end{aligned}$$

since f is continuous, which completes the proof. □

Theorem 41. *[Exercise 2.25] Any linear functional $f : \mathbb{R}^n \to \mathbb{R}$ is a continuous map.*

Proof. Let $M = \max\{|f(e_1)|, \ldots, |f(e_n)|\}$. Let $x \in \mathbb{R}$ and $\varepsilon > 0$. For all $y \in \mathbb{R}$ such that

$$\|x - y\| < \frac{\varepsilon}{Mn},$$

we have

$$\begin{aligned} |f(x) - f(y)| &= |f(x - y)| \\ &= \|x - y\| \left| f\left(\frac{x-y}{\|x-y\|}\right)\right|. \end{aligned}$$

Write

$$\frac{x-y}{\|x-y\|} = a_1 e_1 + \cdots + a_n e_n$$

with $a_1^2 + \cdots + a_n^2 = 1$. Then

$$\begin{aligned} |f(x) - f(y)| &= \|x - y\| \, |a_1 f(e_1) + \cdots + a_n f(e_n)| \\ &= M \, \|x - y\| \, (|a_1| + \cdots + |a_n|) \\ &= Mn \, \|x - y\| \max\{|a_1|, \ldots, |a_n|\} \\ &< \varepsilon \end{aligned}$$

since $|a_i| \leq 1$ for all i. □

Theorem 42. *[Exercise 2.32] Let V be a real vector space with complexification $V^{\mathbb{C}}$ and let $\sigma \in \mathcal{L}(V^{\mathbb{C}})$. Then the following are equivalent:*

(1) *σ is a complexification, i.e. σ has the form $\tau^{\mathbb{C}}$ for some $\tau \in \mathcal{L}(V)$.*
(2) *σ commutes with the conjugate map $\chi : V^{\mathbb{C}} \to V^{\mathbb{C}}$ defined by $\chi(u+iv) = u - iv$.*

Proof. Suppose that $\sigma = \tau^{\mathbb{C}}$ for some $\tau \in \mathcal{L}(V)$. Then

$$\begin{aligned} \sigma\chi(u + iv) &= \sigma(u - iv) \\ &= \tau u - i\tau v \\ &= \chi(\tau u + i\tau v) \\ &= \chi\sigma(u + iv). \end{aligned}$$

Conversely, suppose that σ commutes with χ. Let $\tau : V \to V$ be given by $v \mapsto \mathrm{Re}(\sigma(v + 0i))$; then $\sigma(u + iv) = x + iy$ where $x, y \in V$. We have

$$\begin{aligned} x &= \frac{x+iy}{2} + \frac{x-iy}{2} \\ &= \frac{\sigma(u+iv)}{2} + \frac{\sigma(u-iv)}{2} \\ &= \sigma u \\ &= \tau u \end{aligned}$$

and similarly $y = \tau v$. This shows that $\sigma = \tau^{\mathbb{C}}$. □

Chapter 3. The Isomorphism Theorems

Remark 43. [Exercise 3.1] If V is infinite-dimensional and S is an infinite-dimensional subspace of V, must the dimension of V/S be finite? No, take $V = \mathbb{R}^2$ and $S = \mathbb{R} \times \{0\}$. Then $V/S \cong \mathbb{R}$, which is not finite-dimensional.

Theorem 44. *[Exercise 3.2] Let S be a subspace of V. Then the function that assigns to each intermediate subspace $S \subseteq T \subseteq V$ the subspace T/S of V/S is an order-preserving (with respect to set inclusion) bijection between the set of all subspaces of V containing S and the set of all subspaces of V/S.*

Proof. Let ψ denote the correspondence $T \mapsto T/S$, which is clearly order-preserving. Suppose that $T_1/S = T_2/S$; we want to show that $T_1 = T_2$. If $v \in T_1$, then $v + S \in T_1/S = T_2/S$ and we may choose an element $v' \in v + S \subseteq T_2$. We have $v' - v = s$ for some $s \in S$, so $v = v' - s \in T_2$. Therefore $T_1 \subseteq T_2$, and similarly $T_2 \subseteq T_1$. This shows that ψ is injective. Now let

$$W = \{u + S \mid u \in U\}$$

be a subspace of V/S, where $U \subseteq V$. Let T be the union of all cosets in W:

$$T = \bigcup_{u \in U} (u + S).$$

If $x, y \in T$ then $x + S, y + S \in W$; since W is a subspace of V/S,

$$rx + S, (x + y) + S \in W,$$

which implies that $rx, x + y \in T$. Hence, T is a subspace of V containing S. It is clear that $T/S = W$; this proves that ψ is surjective. □

Theorem 45. *[Exercise 3.3] Let $\tau : V \to W$ be a linear map. Then* $\mathrm{im}(\tau) \cong V/\ker(\tau)$.

Proof. By Theorem 3.4, there exists a map $\tau' : V/\ker(\tau) \to W$ such that $\ker(\tau') = \ker(\tau)/\ker(\tau) \cong \{0\}$. Then τ' is injective and

$$\operatorname{im}(\tau) = \operatorname{im}(\tau') \cong V/\ker(\tau).$$

□

Remark 46. [Exercise 3.5] Let S be a subspace of V. Starting with a basis $\{s_1, \ldots, s_k\}$ for S, how would you find a basis for V/S?

Let F be the base field of V. Complete the given basis of S to a basis $\{s_1, \ldots, s_k, v_{k+1}, \ldots, v_n\}$ of V. We have

$$V = \langle s_1 \rangle \oplus \cdots \oplus \langle s_k \rangle \oplus \langle v_{k+1} \rangle \oplus \cdots \oplus \langle v_n \rangle$$

and

$$S = \langle s_1 \rangle \oplus \cdots \oplus \langle s_k \rangle$$

so that $\varphi : \langle v_{k+1} \rangle \oplus \cdots \oplus \langle v_n \rangle \to V/S$ defined by $v \mapsto v + S$ is an isomorphism. Then φ carries the basis $\{v_{k+1}, \ldots, v_n\}$ to a basis $\{v_{k+1} + S, \ldots, v_n + S\}$ of V/S.

Remark 47. [Exercise 3.6] The rank-nullity theorem follows from the first isomorphism theorem.

Let $\tau \in \mathcal{L}(V, W)$ with $\dim(V) < \infty$; we have $\operatorname{im}(\tau) \cong V/\ker(\tau)$, so $\dim(\operatorname{im}(\tau)) = \dim(V) - \dim(\ker(\tau))$. However, the main part of this proof applies the dimension formula for quotient spaces (Corollary 3.7). The formula follows from taking a complement of $\ker(\tau)$, which is the same technique used in the original proof of the rank-nullity theorem.

Remark 48. [Exercise 3.7] Let $\tau \in \mathcal{L}(V)$ and let S be a subspace of V. Define a map $\tau' : V/S \to V/S$ by $v + S \mapsto \tau v + S$. When is τ' well-defined? If τ' is well-defined, is it a linear transformation? What are $\operatorname{im}(\tau')$ and $\ker(\tau')$?

τ' is well-defined if and only $\tau v - \tau v' = \tau(v - v') \in S$ whenever $v, v' \in V$ and $v - v' \in S$. That is, τ' is well-defined if and only if S is a τ-invariant subspace. In that case τ' is a linear map, and we may compute

$$\operatorname{im}(\tau') = \{\tau v + S \mid v \in V\} = (\operatorname{im}(\tau) + S)/S \cong \operatorname{im}(\tau)/(\operatorname{im}(\tau) \cap S)$$

and

$$\ker(\tau') = \{v \in V \mid \tau v \in S\} = \tau^{-1}(S).$$

Theorem 49. *[Exercise 3.8] For any nonzero vector $v \in V$, there exists a linear functional $f \in V^*$ for which $f(v) \neq 0$.*

Proof. Let F be the base field of V. Choose a basis B of V that contains v, and define $f : V \to F$ by $f(v) = 1$ and $f(b) = 0$ for every $b \in B \setminus \{v\}$. □

Corollary 50. *[Exercise 3.9] A vector $v \in V$ is zero if and only if $f(v) = 0$ for all $f \in V^*$. Two vectors $v, w \in V$ are equal if and only if $f(v) = f(w)$ for all $f \in V^*$.*

Theorem 51. *[Exercise 3.10] Let S be a proper subspace of a finite-dimensional vector space V and let $v \in V \setminus S$. Then there exists a linear functional $f \in V^*$ for which $f(v) = 1$ and $f(s) = 0$ for all $s \in S$.*

Proof. Choose a basis B_0 of S and choose a basis B of V that contains $B_0 \cup \{v\}$. Define $f \in V^*$ by $f(v) = 1$ and $f(x) = 0$ for all $x \in B \setminus \{v\}$. This functional has the desired properties. □

Example 52. [Exercise 3.11] Find a vector space V and decompositions

$$V = A \oplus B = C \oplus D$$

with $A \cong C$ but $B \ncong D$. Hence $A \cong C$ does not imply that $A^c \cong C^c$.

Consider the vector space V over $\mathbb{R}$ of all sequences $f : \mathbb{N} \to \mathbb{R}$ (where $0 \in \mathbb{N}$). We say that f is an even sequence if $f(2k+1) = 0$ for all $k \in \mathbb{N}$, and that f is an odd sequence if $f(2k) = 0$ for all $k \in \mathbb{N}$. Choose A to be the space of all even sequences, choose B to be the space of all odd sequences, choose $C = V$, and choose $D = \{0\}$. Then $A \cong C$ since $\varphi : A \to C$ given by $(\varphi f)(k) = f(2k)$ is an isomorphism. But clearly $B \ncong \{0\}$.

Example 53. [Exercise 3.12] Find isomorphic vector spaces V and W with

$$V = S \oplus B \quad \text{and} \quad W = S \oplus D$$

but $B \ncong D$. Hence $V \ncong W$ does not imply that $V/S \cong W/S$.

Using the terminology of Example 52, choose S to be the space of all even sequences, choose B to be the space of all odd sequences, and choose $D = \{0\}$. Then V is the space of all sequences, which is isomorphic to $W = S \oplus D = S$, the space of all even sequences.

Lemma 54. *Let S, T be subspaces of V. Then there exist subspaces S', T', U such that*

$$\begin{aligned} S &= S' \oplus (S \cap T), \\ T &= T' \oplus (S \cap T), \\ V &= S' \oplus (S \cap T) \oplus T' \oplus U. \end{aligned}$$

Proof. Since $S \cap T$ is a subspace of both S and T, we may choose complements S' and T' of $S \cap T$ in S and T respectively. Choose a complement U of $S + T$ so that $V = (S + T) \oplus U$. It is clear that $S + T = S' + (S \cap T) + T'$; it remains to show that the sum is direct. Suppose that $v \in S' \cap [T' + (S \cap T)] = S' \cap T$. Then $v \in S'$ and $v \in S \cap T$, so $v = 0$. Similarly, $v \in T' \cap [S' + (S \cap T)]$ implies that $v = 0$. Finally, if $v \in (S \cap T) \cap (S' + T')$ then $v = s + t$ where $s \in S'$ and $t \in T'$. Since $t = v - s \in S$ we have $t = 0$ and since $s = v - t \in T$ we have $s = 0$. This completes the proof. □

Theorem 55. *[Exercise 3.13] Let V be a vector space with*

$$V = S_1 \oplus T_1 = S_2 \oplus T_2.$$

If S_1 and S_2 have finite codimension in V, then so does $S_1 \cap S_2$ and

$$\operatorname{codim}(S_1 \cap S_2) \leq \dim(T_1) + \dim(T_2).$$

Proof. Applying Lemma 54 gives a decomposition

$$\begin{aligned} S_1 &= S_1' \oplus (S_1 \cap S_2), \\ S_2 &= S_2' \oplus (S_1 \cap S_2), \\ V &= S_1' \oplus (S_1 \cap S_2) \oplus S_2' \oplus U. \end{aligned}$$

We can compute

$$V/(S_1 \cap S_2) \cong S_1' \oplus S_2' \oplus U = (S_1' \oplus U) + (S_2' \oplus U).$$

Then $V/(S_1 \cap S_2)$ must be finite-dimensional, being the sum of the finite-dimensional vector spaces $S_2' \oplus U \cong V/S_1$ and $S_1' \oplus U \cong V/S_2$. Furthermore,

$$\begin{aligned} \dim(V/(S_1 \cap S_2)) &= \dim((S_1' \oplus U) + (S_2' \oplus U)) \\ &\leq \dim(S_1' \oplus U) + \dim(S_2' \oplus U) \\ &= \dim(V/S_1) + \dim(V/S_2) \\ &= \dim(T_1) + \dim(T_2) \end{aligned}$$

since $T_1 \cong V/S_1$ and $T_2 \cong V/S_2$. □

Remark 56. [Exercise 3.15] Let B be a basis for an infinite-dimensional vector space V and define, for all $b \in B$, the map $b' \in V^*$ by $b'(c) = 1$ if $c = b$ and 0 otherwise, for all $c \in B$. Does $B' = \{b' \mid b \in B\}$ form a basis for V^*? What do you conclude about the concept of a dual basis?

No, B' is not a basis for V^*. For example, the functional $f \in V^*$ that sends all elements of B to 1 is not a linear combination of elements from B', since such combinations must be finite. Therefore the concept of a dual basis only applies to finite-dimensional vector spaces.

Theorem 57. *[Exercise 3.16] If S and T are subspaces of V, then $(S \oplus T)^* \cong S^* \boxplus T^*$.*

Proof. Define $\varphi : S^* \boxplus T^* \to (S \oplus T)^*$ by

$$(\varphi(f, g))(s + t) = f(s) + g(t).$$

This map is easily seen to be linear. Given a functional $f \in (S \oplus T)^*$,

$$\varphi(f|_S, f|_T)(s + t) = f(s) + f(t) = f(s + t),$$

which shows that φ is surjective. If $\varphi(f, g) = 0$, then $f(s) + g(t) = 0$ for every $s \in S$ and $t \in T$. Therefore f and g are both zero, which shows that φ is injective. □

Theorem 58. *[Exercise 3.17]* $0^\times = 0$ *and* $\iota^\times = \iota$*, where* 0 *is the zero operator and* $\iota : V \to V$ *is the identity.*

Proof. For all $f \in W^*$,

$$(0^\times f)(v) = (f0)(v) = f(0) = 0,$$

and for all $f \in V^*$,

$$(\iota^\times f)(v) = (f\iota)(v) = f(v).$$

□

Theorem 59. *[Exercise 3.18] Let* S *be a subspace of* V*. Then* $(V/S)^* \cong S^0$*.*

Proof. Define $\varphi : (V/S)^* \to S^0$ by $(\varphi f)(v) = f(v + S)$, which is clearly linear. Given any $g \in S^0$ define $f \in (V/S)^*$ by $f(v+S) = g(v)$; this is well-defined since $v+S = v'+S$ implies

$$\begin{aligned} f(v + S) - f(v' + S) &= g(v - v') \\ &= 0 \end{aligned}$$

since $v - v' \in S$. Then we have $\varphi f = g$, which shows that φ is surjective. If $\varphi f = 0$ then $f(v + S) = 0$ for all $v \in V$, i.e. $f = 0$. □

Theorem 60. *[Exercise 3.19] Let* V *and* W *be vector spaces.*

(1) $(\tau + \sigma)^\times = \tau^\times + \sigma^\times$ *for* $\tau, \sigma \in \mathcal{L}(V, W)$*.*
(2) $(r\tau)^\times = r\tau^\times$ *for any* $r \in F$ *and* $\tau \in \mathcal{L}(V, W)$*.*

Proof. For all $f \in W^*$,

$$[(\tau + \sigma)^\times f](v) = f(\tau + \sigma)v = f\tau v + f\sigma v = (\tau^\times f)(v) + (\sigma^\times f)(v),$$

and for all $r \in F$, $f \in W^*$,

$$[(r\tau)^\times f](v) = f(r\tau)v = rf\tau v = r(\tau^\times f)(v).$$

□

Theorem 61. *[Exercise 3.20] Let* $\tau \in \mathcal{L}(V, W)$*, where* V *and* W *are finite-dimensional. Then* $\mathrm{rk}(\tau) = \mathrm{rk}(\tau^\times)$*.*

Proof. By Theorem 3.19 we have $\mathrm{im}(\tau^\times) = \ker(\tau)^0 \cong S^*$ where S is any complement of $\ker(\tau)$ in V. Then

$$\mathrm{rk}(\tau^\times) = \dim(S^*) = \dim(S) = \mathrm{rk}(\tau).$$

□

Theorem 62. *Let* V *be a vector space over a field* F *and let* $X, Y \subseteq V$*.*

(1) X^0 *is a subspace of* V*.*

(2) *If* $X \subseteq Y$ *then* $Y^0 \subseteq X^0$.
(3) $X \subseteq \text{span}(X) \subseteq X^{00}$.

Proof. Let $f, g \in X^0$, $a \in F$ and $x \in X$. Clearly $0 \in X^0$; we also have $f(x) = g(x) = 0$ so $(f+g)(x) = 0$ and $(af)(x) = 0$. Therefore $f + g, af \in X^0$, which proves (1). For (2), let $f \in Y^0$ so that $f(y)$ for every $y \in Y$. In particular, $f(x)$ for every $x \in X$, so $f \in X^0$.

If $v \in \text{span}(X)$ then $v = a_1 v_1 + \cdots + a_n v_n$ where $v_1, \ldots, v_n \in X$. If $f \in X^0$ then

$$f(v) = a_1 f(v_1) + \cdots + a_n f(v_n) = 0,$$

so $v \in X^{00}$. This proves (3). □

Chapter 4. Modules I: Basic Properties

Theorem 63. *[Universal property of a free object] Let* F *and* M *be* R*-modules, let* B *be a basis for* F *and let* $\tau_0 : B \to M$ *be any function. Then* τ_0 *extends uniquely to an* R*-map* $\tau : F \to M$*. That is, there exists a unique* R*-map* $\tau : F \to M$ *such that* $\tau i = \tau_0$*, where* $i : B \to V$ *is the canonical injection.*

Theorem 64. *[Universal property of a free object, converse] Let* F *be an* R*-module. Let* B *be a subset of* F *such that for every* R*-module* M *and every* $\tau_0 : B \to M$ *there exists a unique* R*-map* $\tau : F \to M$ *that extends* τ_0*. Then* B *is a basis for* F*.*

Proof. Let $F(B)$ be the free module on B. There exists a unique R-map $f : F \to F(B)$ that extends the inclusion map $i : B \to F(B)$. There also exists a unique R-map $g : F(B) \to F$ that extends the inclusion map $i_1 : B \to F$. Since $fg : F(B) \to F(B)$ agrees with $\text{id}_{F(B)}$ on B and B is a basis for $F(B)$, we must have $fg = \text{id}_{F(B)}$. On the other hand, $gf : F \to F$ agrees with id_F on B, and by the given conditions on B we have $gf = \text{id}_F$. Therefore g is an isomorphism, and g carries the basis B of $F(B)$ to an identical basis of F. □

Theorem 65. *[Exercise 4.2] Let* $S = \{v_1, \ldots, v_n\}$ *be a subset of a module* M*. Then* $N = \langle S \rangle$ *is the smallest submodule of* M *containing* S*. That is,* N *is a submodule of every module that contains* S*.*

Proof. Any module that contains S must also contain all linear combinations of elements from S, i.e. $\langle S \rangle$. □

Theorem 66. *[Exercise 4.3] Let* M *be an* R*-module and let* I *be an ideal in* R*. Let* IM *be the set of all finite sums of the form*

$$r_1 v_1 + \cdots + r_n v_n$$

here $r_i \in I$ *and* $v_i \in M$*. Then* IM *is a submodule of* M*.*

Proof. It is clear that $IM \subseteq M$ and that IM is closed under addition. If $r \in R$, $r_i \in I$ and $v_i \in M$, then

$$r(r_1v_1 + \cdots + r_nv_n) = (rr_1)v_1 + \cdots + (rr_n)v_n \in IM$$

since I is an ideal in R. □

Theorem 67. *[Exercise 4.4] Let M be an R-module and let $\mathcal{S}(M)$ be the set of all submodules of M. If $S, T \in \mathcal{S}(M)$, then*

$$S \cap T = \operatorname{glb}\{S, T\}$$

and

$$S + T = \operatorname{lub}\{S, T\}.$$

Proof. Same as Theorem 1. □

Theorem 68. *[Exercise 4.5] Let $S_1 \subseteq S_2 \subseteq \cdots$ be an ascending sequence of submodules of an R-module M. Then the union $\bigcup S_i$ is a submodule of M.*

Proof. If $r \in R$ and $v \in \bigcup S_i$, then $v \in S_k$ for some k. Therefore $rv \in S_k \subseteq \bigcup S_i$. If $v, v' \in \bigcup S_i$, then $v \in S_j$ and $v' \in S_k$ for some j, k. Since $v, v' \in S_{\max(j,k)}$, we have $v + v' \in S_{\max(j,k)} \subseteq \bigcup S_i$. □

Example 69. [*Exercise 4.6] Give an example of a module M that has a finite basis but with the property that not every spanning set in M contains a basis and not every linearly independent set in M is contained in a basis.*

Consider $\mathbb{Z}^2$ as a $\mathbb{Z}$-module, which has a basis $\{(1,0),(0,1)\}$. The spanning set

$$\{(2,0),(3,0),(0,2),(0,3)\}$$

does not contain a basis, and the linearly independent set $\{(2,0)\}$ is not contained in a basis.

Theorem 70. *[Exercise 4.8] Let $\tau \in \operatorname{Hom}_R(M, N)$ be an R-isomorphism. If B is a basis for M, then τB is a basis for N.*

Proof. Same as Theorem 16. □

Theorem 71. *[Exercise 4.9] Let M be an R-module and let $\tau \in \operatorname{Hom}_R(M, M)$ be an R-endomorphism. Then τ is idempotent, that is, $\tau^2 = \tau$, if and only if*

$$M = \ker(\tau) \oplus \operatorname{im}(\tau) \quad \text{and} \quad \tau|_{\operatorname{im}(\tau)} = \iota.$$

Proof. Suppose that $\tau^2 = \tau$ and let $v \in M$. Since $\tau(v - \tau v) = \tau v - \tau^2 v = 0$, we have $v = (v - \tau v) + \tau v \in \ker(\tau) + \mathrm{im}(\tau)$. If $v \in \ker(\tau)$ and $v \in \mathrm{im}(\tau)$ then $v = \tau x$ for some $x \in M$, and $v = \tau^2 x = \tau v = 0$. This shows that $M = \ker(\tau) \oplus \mathrm{im}(\tau)$, and clearly $\tau|_{\mathrm{im}(\tau)} = \iota$. Conversely, suppose that $M = \ker(\tau) \oplus \mathrm{im}(\tau)$ and $\tau|_{\mathrm{im}(\tau)} = \iota$. Let $v \in M$ so that $v = x + y$ where $x \in \ker(\tau)$ and $y \in \mathrm{im}(\tau)$. Then

$$\tau^2 v = \tau^2 y = \tau y = \tau(x + y) = \tau v.$$

□

Example 72. [Exercise 4.10] Consider the ring $R = F[x, y]$ of polynomials in two variables. Show that the set M consisting of all polynomials in R that have zero constant term is an R-module. Show that M is not a free R-module.

The first claim is obvious. For the second claim, suppose that B is a basis for M. Suppose that B contains at least two elements $f(x, y)$ and $g(x, y)$. Then

$$g(x, y)f(x, y) - f(x, y)g(x, y) = 0$$

which contradicts the linear independence of $f(x, y)$ and $g(x, y)$. Therefore B contains exactly one element $f(x, y)$, for an empty set cannot span M. Write

$$f(x, y) = \sum_i a_i x^{m_i} y^{n_i},$$

and choose i such that $m_i + n_i \neq 0$ is minimized. We must have $m_i \leq 1$ for otherwise $x \notin \langle B \rangle$, and similarly $n_i \leq 1$ for otherwise $y \notin \langle B \rangle$. If $m_i = n_i = 1$ then $x \notin \langle B \rangle$, so $m_i = 1$ or $n_i = 1$, but not both. If $m_i = 1$ then $y \notin \langle B \rangle$; if $n_i = 1$ then $x \notin \langle B \rangle$. This shows that B cannot span M. Therefore M is not free.

Theorem 73. *[Exercise 4.11] Let R be an integral domain and let M be an R-module. If $v_1, \ldots, v_n$ is linearly independent over R, then so is $rv_1, \ldots, rv_n$ for any nonzero $r \in R$.*

Proof. If

$$a_1 r v_1 + \cdots + a_n r v_n = 0,$$

then $a_1 r = \cdots = a_n r = 0$. Since $r \neq 0$ and R is an integral domain, we must have $a_1 = \cdots = a_n = 0$. □

Theorem 74. *[Exercise 4.12] If a nonzero commutative ring R with identity has the property that every finitely generated R-module is free, then R is a field.*

Proof. Let I be an ideal of R. Then $R/I = \langle 1 + I \rangle$, so R/I is free with some basis B. Suppose that B is not empty and $I \neq \{0\}$, i.e. $\{0\} \subset I \subset R$, choose $b + I \in B$, and choose some nonzero $i \in I$. Then $i(b + I) = I$, which contradicts the linear independence of B. Therefore $I = \{0\}$ or $I = R$, which shows that R is simple and therefore a field. □

Theorem 75. *[Exercise 4.13] Let M and N be R-modules. If S is a submodule of M and T is a submodule of N then*

$$(M \oplus N)/(S \oplus T) \cong (M/S) \boxplus (N/T).$$

Proof. The R-map $\varphi : M \oplus N \to (M/S) \oplus (N/T)$ given by $m + n \mapsto (m + S, n + T)$ is surjective and $\ker(\varphi) = S \oplus T$. □

Remark 76. [Exercise 4.14] If R is a commutative ring with identity and I is an ideal of R, then I is an R-module. What is the maximum size of a linearly independent set in I? Under what conditions is I free?

Any linearly independent set in I contains at most 1 element, for if $a, b \in I$ are nonzero elements then $ba - ab = 0$, which is a contradiction. Therefore if I is free then I is principal. If R is an integral domain, then I is free if and only if I is principal.

Theorem 77. *[Exercise 4.15] Let R be an integral domain and let M be an R-module.*

(1) *M_{tor} is a submodule of M.*
(2) *M/M_{tor} is torsion-free.*
(3) *If R is not an integral domain, then M_{tor} may not be a submodule of M.*

Proof. If $x \in M_{\text{tor}}$ then $ax = 0$ for some nonzero $a \in R$. Therefore $a(rx) = 0$ and $rx \in M_{\text{tor}}$. If $x, y \in M_{\text{tor}}$ then $ax = by = 0$ for some nonzero $a, b \in R$. Therefore $ab(x + y) = 0$ and $x + y \in M_{\text{tor}}$, noting that $ab \neq 0$ since R is an integral domain. This shows that M_{tor} is a submodule of M. For (2), suppose that $r(x + M_{\text{tor}}) = M_{\text{tor}}$ for some $x + M_{\text{tor}} \in M/M_{\text{tor}}$ and some nonzero $r \in R$. Then $rx \in M_{\text{tor}}$, and $srx = 0$ for some nonzero $s \in R$. Since R is an integral domain and $sr \neq 0$, we must have $x = 0 \in M_{\text{tor}}$. This shows that M/M_{tor} is torsion-free. For (3), consider $\mathbb{Z}_6$ as a $\mathbb{Z}$-module; 2 and 3 are torsion elements but $2 + 3 = 5$ is not. □

Example 78. [Exercise 4.16]

(1) Find a module M that is finitely generated by torsion elements but for which $\text{ann}(M) = \{0\}$.
(2) Find a torsion module M for which $\text{ann}(M) = \{0\}$.

Proof. For (1), choose $M = \langle 2, 3 \rangle$ as a submodule of the $\mathbb{Z}$-module $\mathbb{Z}_6$. Then $\text{ann}(1) = \{0\}$, so $\text{ann}(M) = \{0\}$. For (2), let $\{p_i\}_{i \geq 1}$ be the sequence of all prime numbers and let M be the $\mathbb{Z}$-module defined by the infinite external direct sum

$$\mathbb{Z}_2 \boxplus \mathbb{Z}_3 \boxplus \cdots \boxplus \mathbb{Z}_{p_i} \boxplus \cdots .$$

If $x = (k_1, \ldots, k_n, 0, 0, \ldots) \in M$, then $\prod_{k_i \neq 0} k_i \in \text{ann}(x)$ and x is a torsion element. However, $\text{ann}(M) = 0$, for if $r \in \text{ann}(M)$ is nonzero and p_i is the largest prime number dividing r, then $r \notin \text{ann}(x)$ where x has a 1 in the $(i + 1)$th position. □

Theorem 79. *[Exercise 4.19] Any R-module M is isomorphic to the R-module* $\mathrm{Hom}_R(R, M)$.

Proof. Define $\varphi : \mathrm{Hom}_R(R, M) \to M$ by $f \mapsto f(1)$. If $f, g \in \mathrm{Hom}_R(R, M)$ and $r \in R$, we have

$$\begin{aligned} \varphi(f+g) &= (f+g)(1) = f(1) + g(1) = \varphi(f) + \varphi(g), \\ \varphi(rf) &= rf(1) = r\varphi(f), \end{aligned}$$

which shows that φ is an R-map. If $\varphi(f) = f(1) = 0$ then for any $r \in R$,

$$\varphi(r) = \varphi(1)\varphi(r) = 0.$$

This shows that φ is injective. If $v \in M$ then we may define an R-map $f : R \to M$ by $f(r) = rv$. Since $f(1) = v$, we have $\varphi(f) = v$, which shows that φ is surjective. Therefore $\mathrm{Hom}_R(R, M) \cong M$. □

Theorem 80. *[Exercise 4.20] Let R and S be commutative rings with identity and let $f : R \to S$ be a ring homomorphism. Then any S-module is also an R-module under the scalar multiplication $rv = f(r)v$.*

Proof. Obvious. □

Lemma 81. *Let $f : \mathbb{Z}_n \to \mathbb{Z}_m$ be a map with $f(j) \equiv kj \pmod{m}$ for some k. Then f is a $\mathbb{Z}$-map if and only if $m/\gcd(m, n)$ divides $f(1) = k$.*

Theorem 82. *[Exercise 4.21]* $\mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}_n, \mathbb{Z}_m) \cong \mathbb{Z}_d$ *where* $d = \gcd(n, m)$.

Proof. Define the $\mathbb{Z}$-map $\varphi : \mathbb{Z}_m \to \mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}_n, \mathbb{Z}_m)$ by $k \mapsto \psi_k$ where $\psi_k : \mathbb{Z}_n \to \mathbb{Z}_m$ is given by $j \mapsto k(m/d)j$. This is well-defined by Lemma 81. Given $f \in \mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}_n, \mathbb{Z}_m)$, we know by Lemma 81 that m/d divides $f(1)$, and $\psi_{f(1)/(m/d)} = f$ since

$$\psi_{f(1)/(m/d)}(j) = f(1)j = f(j).$$

Therefore $\varphi(f(1)) = f$, which shows that φ is surjective. Suppose that $\varphi(k) = \psi_k = 0$. This is true if and only if

$$\psi_k(1) = \frac{km}{d} \equiv 0 \pmod{m} \Leftrightarrow d \mid k.$$

Then $\ker(\varphi) = d\mathbb{Z}_m$ and $\mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}_n, \mathbb{Z}_m) \cong \mathbb{Z}_m/d\mathbb{Z}_m \cong \mathbb{Z}_d$. □

Theorem 83. *[Exercise 4.22] Let R be a commutative ring with identity. If I, J are ideals of R for which $R/I \cong R/J$ as R-modules, then $I = J$. If $R/I \cong R/J$ only as rings, then $I = J$ may be false.*

Proof. Let $\varphi : R/I \to R/J$ be an R-isomorphism. Let $i \in I$. Then

$$\begin{aligned} i + J &= i(1 + J) \\ &= i\varphi(\varphi^{-1}(1 + J)) \\ &= \varphi(i\varphi^{-1}(1 + J)) \\ &= \varphi(I) \\ &= J \end{aligned}$$

since $\varphi^{-1}(1 + J) = a + I$ for some $a \in R$ and $i(a + I) = I$. Therefore $i \in J$ and $I \subseteq J$. Similarly, $J \subseteq I$.

Alternatively, $\text{ann}(R/I) = \text{ann}(R/J)$ since $R/I \cong R/J$, but $\text{ann}(R/I) = I$ and $\text{ann}(R/J) = J$.

If $R/I \cong R/J$ only as rings, the conclusion may not hold. For example, take $R = \mathbb{Z}_3[x]$, $I = \langle x^3 - x + 1\rangle$ and $J = \langle x^3 - x^2 + 1\rangle$. We have $R/I \cong R/J$ as finite fields, but $I \neq J$. □

Chapter 5. Modules II: Free and Noetherian Modules

Remark 84. [Exercise 5.1] If M is a free R-module and $\tau : M \to N$ is an epimorphism, then must N also be free? No. Take $M = \mathbb{Z}^2$ as a $\mathbb{Z}$-module with basis $\{(1, 0), (0, 1)\}$, and $N = \mathbb{Z}_2^2$ with the epimorphism $(m, n) \mapsto ([m], [n])$.

Theorem 85. *[Exercise 5.2] Let I be an ideal of R. If R/I is a free R-module, then $I = \{0\}$ or $I = R$.*

Proof. If $R/I = \{0\}$, then $I = R$ and we are done. Otherwise, let B be a (nonempty) basis for R/I. Let $i \in I$ and choose any $r + I \in B$. Then $i(r + I) = I$, and since B is linearly independent, $i = 0$. □

Theorem 86. *[Exercise 5.4] Let S be a submodule of an R-module M. If M is finitely generated, so is the quotient module M/S.*

Proof. Let $P = \{p_1, \ldots, p_n\}$ be a finite subset of M such that $M = \langle P\rangle$. Then $\{p_1 + S, \ldots, p_n + S\}$ spans M/S, for if $m + S \in M/S$ then

$$m = a_1p_1 + \cdots + a_np_n$$

and

$$m + S = a_1(p_1 + S) + \cdots + a_n(p_n + S).$$

□

Theorem 87. *[Exercise 5.5] Let S be a submodule of an R-module M. If both S and M/S are finitely generated, then so is M.*

Proof. Let $P = \{p_1 + S, \ldots, p_n + S\}$ be a finite subset of M/S such that $M/S = \langle P \rangle$, and let $Q = \{q_1, \ldots, q_n\}$ be a finite subset of S such that $S = \langle Q \rangle$. Let $v \in M$ so that $v + S \in M/S$ and

$$v + S = a_1 p_1 + \cdots + a_n p_n + S.$$

Then $v - a_1 p_1 + \cdots + a_n p_n \in S$, so

$$v - a_1 p_1 + \cdots + a_n p_n = b_1 q_1 + \cdots + b_n q_n.$$

This shows that $M = \langle P \cup Q \rangle$. □

Theorem 88. *[Exercise 5.7] Let $\tau : M \to N$ be an R-homomorphism.*

(1) *If M is finitely generated, then so is* $\operatorname{im}(\tau)$.
(2) *If* $\ker(\tau)$ *and* $\operatorname{im}(\tau)$ *are finitely generated, then $M = \ker(\tau) + S$ where S is a finitely generated submodule of M. Hence, M is finitely generated.*

Proof. For (1), $\operatorname{im}(\tau) \cong M/\ker(\tau)$ which is finitely generated by Theorem 86. (2) follows from Theorem 87. □

Theorem 89. *[Exercise 5.8] If R is Noetherian and I is an ideal of R, then R/I is also Noetherian.*

Proof. Let

$$J_1 \subseteq J_2 \subseteq J_3 \subseteq \cdots$$

be an ascending sequence of ideals of R/I. By the correspondence theorem, for each i we can write $J_i = S_i/I$ where S_i is an ideal of R, and

$$S_1 \subseteq S_2 \subseteq S_3 \subseteq \cdots$$

is an ascending sequence of ideals of R. Since R is Noetherian, there exists a k such that

$$S_k = S_{k+1} = S_{k+2} = \cdots,$$

and therefore

$$J_k = J_{k+1} = J_{k+2} = \cdots.$$

□

Theorem 90. *[Exercise 5.9] If R is Noetherian, then so is $R[x_1, \ldots, x_n]$.*

Proof. Follows from induction on n and the Hilbert basis theorem. □

Example 91. [Exercise 5.10] Find an example of a commutative ring with identity that does not satisfy the ascending chain condition.

The ring of polynomials in infinitely many variables is not Noetherian. Specifically,

$$\mathbb{Z}[x_1] \subseteq \mathbb{Z}[x_1, x_2] \subseteq \mathbb{Z}[x_1, x_2, x_3] \subseteq \cdots$$

is an ascending chain that does not become constant.

Theorem 92. *[Exercise 5.11]*

(1) *An R-module M is cyclic if and only if it is isomorphic to R/I where I is an ideal of R.*
(2) *An R-module M is simple if and only if it is isomorphic to R/I where I is a maximal ideal of R.*
(3) *For any nonzero commutative ring R with identity, a simple R-module exists.*

Proof. If M is cyclic, then $M = \langle a \rangle$ for some $a \in M$, and $\varphi : R \to M$ given by $r \mapsto ra$ is a surjection. Therefore $M \cong R/\ker(\varphi)$. Conversely, let $\varphi : R/I \to M$ be an isomorphism, and let $a = \varphi(1 + I)$. Then for any $v \in M$, there exists some $r \in R$ such that

$$\begin{aligned} v &= \varphi(\varphi^{-1}(v)) \\ &= \varphi(r + I) \\ &= ra, \end{aligned}$$

which shows that $M = \langle a \rangle$. This proves (1).

For (2), suppose that M is simple and choose any nonzero $v \in M$. Since $\langle v \rangle$ is a nonzero submodule of M, we must have $M = \langle v \rangle$. By (1), M is isomorphic to R/I where I is an ideal of R. Furthermore, I is a maximal ideal since $R/I \cong M$ is simple. Conversely, suppose that $M \cong R/I$ where I is a maximal ideal. Then M is simple since R/I is simple.

For (3), choose any maximal ideal I of R (which always exists). Then R/I is simple by (2). □

Example 93. [Exercise 5.12] If R is not a principal ideal domain, then there may exist an n-generated R-module that contains a submodule that is not n-generated.

Take $R = \mathbb{Z}[x]$. Then R is a 1-generated R-module since $R = \langle 1 \rangle$, but $\langle 2, x \rangle = 2\mathbb{Z} + x\mathbb{Z}[x]$ is not 1-generated.

Theorem 94. *[Exercise 5.13] Let R be a commutative ring with identity.*

(1) *R is Noetherian if and only if every finitely generated R-module is Noetherian.*
(2) *Suppose that R is a principal ideal domain. If an R-module M is n-generated, then any submodule of M is also n-generated.*

Proof. One direction for (1) is obvious. For the other direction, we use induction on the number n of generating elements of an R-module M. If $n = 0$ then $M = \{0\}$, and trivially all submodules of M are finitely generated. Now suppose that every n-generated R-module is Noetherian, and let M be an $(n + 1)$-generated R-module with $M = \langle v_1, \ldots, v_{n+1} \rangle$. Let $M' = \langle v_1, \ldots, v_n \rangle$, let S be a submodule of M, and

let $T = S \cap M'$. By the induction hypothesis, T is finitely generated since M' is n-generated. Now consider

$$S/T = S/(S \cap M') \cong (S + M')/M'.$$

If S/T is finitely generated then S is finitely generated, by Theorem 87. Therefore it remains to show that $(S + M')/M'$ is finitely generated. Let $\varphi : R^{n+1} \to S + M'$ be the epimorphism defined by $(r_1, \ldots, r_{n+1}) \mapsto r_1 v_1 + \cdots + r_{n+1} v_{n+1}$, and let $\psi : R^{n+1} \to R$ be the R-map that sends each tuple to its last coordinate. Then $U = \psi(\varphi^{-1}(S + M'))$ is a submodule of R that is generated by some finite set $G = \{g_1, \ldots, g_k\} \subseteq S$, since R is Noetherian. Let $G^* = \{g_1 v_{n+1} + M', \ldots, g_k v_{n+1} + M'\}$. For any $v + M' \in (S + M')/M'$, we can choose some $(r_1, \ldots, r_{n+1}) \in \varphi^{-1}(\{v\})$, and

$$\begin{aligned} v + M' &= r_1 v_1 + \cdots + r_{n+1} v_{n+1} + M' \\ &= r_{n+1} v_{n+1} + M' \\ &\in G^* \end{aligned}$$

since $r_{n+1} \in U$. This shows that $(S + M')/M' = \langle G^* \rangle$.

For (2), we can modify the proof by choosing some $a \in R$ such that $U = \langle a \rangle$. Then $S/T \cong (S + M')/M'$ is cyclic, and S is generated by $n + 1$ elements (see Theorem 87). □

Theorem 95. *[Exercise 5.14] Any R-module M is isomorphic to the quotient of a free module F. If M is finitely generated, then F can also be taken to be finitely generated.*

Proof. Let $F(M)$ be the free module on M and let $\varphi : F(M) \to M$ be the (unique) epimorphism that extends id_M. Then $M \cong F(M)/\ker(\varphi)$. Now suppose that $M = \langle v_1, \ldots, v_n \rangle$, and let $\varphi : R^n \to M$ be the epimorphism defined by

$$(r_1, \ldots, r_n) \mapsto r_1 v_1 + \cdots + r_n v_n.$$

Then $M \cong R^n/\ker(\varphi)$. □

Theorem 96. *[Exercise 5.15] Let S, T, T_1, T_2 be submodules of a module M, where all modules are free and have finite rank.*

(1) *If $S \cong T$, then $M/S \cong M/T$.*
(2) *If $S \oplus T_1 \cong S \oplus T_2$, then $T_1 \cong T_2$.*
(3) *The above statements do not necessarily hold if the modules are not free or do not have finite rank.*

Proof. If $S \cong T$ then $\mathrm{rk}(S) = \mathrm{rk}(T)$ and

$$\mathrm{rk}(M/S) = \mathrm{rk}(M) - \mathrm{rk}(S) = \mathrm{rk}(M) - \mathrm{rk}(T) = \mathrm{rk}(M/T).$$

Similarly, if $S \oplus T_1 \cong S \oplus T_2$ then

$$\mathrm{rk}(T_1) = \mathrm{rk}(S \oplus T_1) - \mathrm{rk}(S) = \mathrm{rk}(S \oplus T_2) - \mathrm{rk}(S) = \mathrm{rk}(T_2),$$

so $T_1 \cong T_2$.

For (3), see Example 53. □

Chapter 6. Modules over a Principal Ideal Domain

Lemma 97. *Let $\alpha_1, \ldots, \alpha_n$ be relatively prime, let $\mu = \alpha_1 \cdots \alpha_n$, let $\beta_i = \mu/\alpha_i$ for each i, and choose $a_i \in R$ such that*

$$a_1\beta_1 + \cdots + a_n\beta_n = 1.$$

Then a_i and α_i are relatively prime for each i.

Proof. Let d be a GCD of a_k and α_k. Then d clearly divides $a_k\beta_k$, and d divides $a_i\beta_i$ when $i \neq k$ since α_k divides β_i. Therefore d divides

$$a_1\beta_1 + \cdots + a_n\beta_n = 1$$

and d must be a unit. □

Theorem 98. *Let M be an R-module and suppose that*

$$M = A_1 + \cdots + A_n$$

where the submodules A_i have relatively prime orders. Then the sum is direct.

Proof. Denote the order of each A_i by α_i, and let $\mu = \alpha_1 \cdots \alpha_n$. Suppose that

$$v_1 + \cdots + v_n = 0$$

where $v_i \in A_i$. Then for each i,

$$\begin{aligned} 0 &= \frac{\mu}{\alpha_i}(v_1 + \cdots + v_n) \\ &= \frac{\mu}{\alpha_i} v_i, \end{aligned}$$

and

$$\begin{aligned} 1 &= o\left(\frac{\mu}{\alpha_i} v_i\right) \\ &= \frac{o(v_i)}{\gcd(\mu/\alpha_i, o(v_i))} \\ &= o(v_i) \end{aligned}$$

since $o(v_i)$ divides α_i and $\gcd(\mu/\alpha_i, \alpha_i) = 1$. Therefore $v_i = 0$. □

Theorem 99. *[Exercise 6.1] Any free module over an integral domain is torsion-free.*

Proof. Let R be an integral domain and let M be a free R-module. Let B be a basis for M. If $v \in M$ is nonzero and $rv = 0$ for some r, then we can write

$$r(a_1v_1 + \cdots + a_nv_n) = 0$$

where $v_1, \ldots, v_n \in B$ and at least one a_i is nonzero, say a_k. Since B is a basis, $ra_k = 0$, and since R is an integral domain, $r = 0$. □

Theorem 100. *[Exercise 6.2] Let M be a finitely generated torsion module over a principal ideal domain. The following are equivalent:*

(1) *M is indecomposable.*
(2) *M has only one elementary divisor (including multiplicity).*
(3) *M is cyclic of prime power order.*

Proof. If M has two or more elementary divisors, it must be decomposable by definition, so (1) $\Rightarrow$ (2). If M has only one elementary divisor p^e, then $M = \langle v \rangle$ where $\text{ann}(\langle v \rangle) = \langle p^e \rangle$, which shows (2) $\Rightarrow$ (3). Suppose that $M = A \oplus B$ where A and B are proper submodules of M. If $o(A)$ and $o(B)$ are relatively prime, then M cannot have prime power order. If $o(A)$ and $o(B)$ are not relatively prime, then M cannot be cyclic by Theorem 6.17. This proves (3) $\Rightarrow$ (1). □

Theorem 101. *[Exercise 6.3] Let R be a principal ideal domain and R^+ the field of quotients. Then R^+ is an R-module, and any nonzero finitely generated submodule of R^+ is a free module of rank* 1.

Proof. If $a/b \in R^+$, defining $r\,(a/b) = (ra)/b$ gives R^+ the structure of an R-module. Furthermore, R^+ is free with basis $\{1\}$, and the result follows from Theorem 6.5. □

Theorem 102. *[Exercise 6.4] Let R be a principal ideal domain. Let M be a finitely generated torsion-free R-module. Suppose that N is a submodule of M for which N is a free R-module of rank* 1 *and M/N is a torsion module. Then M is a free R-module of rank* 1.

Proof. By Theorem 6.8, we know that M is free (with a finite basis). Let $B = \{v_1, \ldots, v_n\}$ be a basis for M. Since M/N is a torsion module, for each i there exists a nonzero $r_i \in R$ such that $r_i(v_i + N) = N$ and $r_iv_i \in N$. Suppose that $n \geq 2$. Since N has rank 1, $\{r_1v_1, \ldots, r_nv_n\}$ must be linearly dependent by Theorem 5.5. Each r_i is nonzero, so this contradicts the linear independence of B. Therefore $n \leq 1$, and since $N \subseteq M$ is nonempty, we must have $n = 1$. □

Example 103. [Exercise 6.5] Show that the primary cyclic decomposition of a torsion module over a principal ideal domain is not unique (even though the elementary divisors are).

Consider $\mathbb{Z}_2 \times \mathbb{Z}_2$ as a $\mathbb{Z}$-module. We have the two decompositions

$$\begin{aligned}\mathbb{Z}_2 \times \mathbb{Z}_2 &= \langle (1,0) \rangle \oplus \langle (0,1) \rangle \\ &= \langle (1,1) \rangle \oplus \langle (0,1) \rangle .\end{aligned}$$

Example 104. [Exercise 6.6] Show that if M is a finitely generated R-module where R is a principal ideal domain, then the free summand in the decomposition $M = F \oplus M_{\text{tor}}$ need not be unique.

Consider $\mathbb{Z} \times \mathbb{Z}_2$ as $\mathbb{Z}$-module. We have the two decompositions

$$\begin{aligned}\mathbb{Z} \times \mathbb{Z}_2 &= \langle (1,0) \rangle \oplus \langle (0,1) \rangle \\ &= \langle (1,1) \rangle \oplus \langle (0,1) \rangle .\end{aligned}$$

Theorem 105. *[Exercise 6.7] If $\langle v \rangle$ is a cyclic R-module of order a then the map $\tau : R \to \langle v \rangle$ defined by $\tau r = rv$ is a surjective R-homomorphism with kernel $\langle a \rangle$ and so $\langle v \rangle \cong R/\langle a \rangle$.*

Proof. Obvious. □

Theorem 106. *[Exercise 6.8] If R is an integral domain with the property that all submodules of cyclic R-modules are cyclic, then R is a principal ideal domain.*

Proof. Observe that R is a cyclic R-module with $R = \langle 1 \rangle$. If I is an ideal of R, then I is a submodule of R and $I = \langle a \rangle$ for some $a \in I$. This shows that I is a principal ideal domain. □

Theorem 107. *[Exercise 6.9] Let F be a finite field and let F^* be the set of all nonzero elements of F.*

1. *If $p(x) \in F[x]$ is a nonconstant polynomial over F and if $r \in F$ is a root of $p(x)$, then $x - r$ is a factor of $p(x)$.*
2. *Every nonconstant polynomial $p(x) \in F[x]$ of degree n has at most n distinct roots in F.*
3. *F^* is a cyclic group under multiplication.*

Proof. Since $F[x]$ is a Euclidean domain, we can write

$$p(x) = (x - r)q(x) + c.$$

where $c \in F$. But $0 = p(r) = c$, so $x - r$ is a factor of $p(x)$ and this proves (1). For (2), if $p(x)$ has distinct roots $\{r_1, \ldots, r_m\}$ where $m > n$ then

$$\prod_{i=1}^{m} (x - r_i)$$

is a factor of $p(x)$ with degree m, which contradicts $p(x)$ having degree n.

F^* is a group under multiplication, and we can define F^* as a $\mathbb{Z}$-module with scalar multiplication $kf = f^k$. Since F^* is finite, it must be finitely generated. Then we have a primary decomposition

$$(*) \qquad F^* = M_{p_1} \oplus \cdots \oplus M_{p_n}$$

where $p_1, \ldots, p_n$ are distinct prime numbers and $o(M_{p_i}) = p_i^{e_i}$ for each i. Fix some i and let $k = p_i^{e_i}$. Choose an element $f_i \in M_{p_i}$ with $o(f_i) = k$ so that the order of the subgroup $\langle f_i \rangle$ is k. Every element of M_{p_i} is a root of the polynomial $p(x) = x^k - 1$, so $|M_{p_i}| \leq k$ by (2). This shows that $M_{p_i} = \langle f_i \rangle$ is cyclic. Applying Theorem 6.4 to (*), we have

$$\begin{aligned} F^* &= \langle f_1 \rangle \oplus \cdots \oplus \langle f_n \rangle \\ &= \langle f_1 + \cdots + f_n \rangle , \end{aligned}$$

which proves (3). □

Lemma 108. *Let R be a principal ideal domain and let M be an R-module. If $v, w \in M$ where $\langle w \rangle$ is a submodule of $\langle v \rangle$ and $o(w) = o(v)$, then $\langle w \rangle = \langle v \rangle$.*

Proof. We have $w = rv$ for some $r \in R$, and we are given that $o(w) = o(rv) = o(v)$. By Theorem 6.3, $\langle w \rangle = \langle rv \rangle = \langle v \rangle$. □

Theorem 109. *[Exercise 6.10] Let R be a principal ideal domain and let $M = \langle v \rangle$ be a cyclic R-module with order α. Then for every $\beta \in R$ such that β divides α there is a unique submodule of M of order β.*

Proof. Let $\alpha = p_1^{e_1} \cdots p_n^{e_n}$ and $\beta = p_1^{f_1} \cdots p_n^{f_n}$ be factorizations of α and β where $f_i \leq e_i$ for each i. By Theorem 6.17, we can write

$$M = \langle v_1 \rangle \oplus \cdots \oplus \langle v_n \rangle$$

where $\langle v_i \rangle$ is a primary submodule of M and $o(\langle v_i \rangle) = p_i^{e_i}$. Then $o(p_i^{e_i - f_i} \langle v_i \rangle) = p_i^{f_i}$ and

$$N = p_1^{e_1 - f_1} \langle v_1 \rangle \oplus \cdots \oplus p_n^{e_n - f_n} \langle v_n \rangle$$

is a submodule of M with order β. To prove uniqueness, suppose that N' is a submodule M of order β. Since N' is cyclic, applying Theorem 6.17 gives a decomposition

$$N' = \langle w_1 \rangle \oplus \cdots \oplus \langle w_n \rangle$$

where $o(\langle w_i \rangle) = p_i^{f_i}$. Fix some i; we want to show that $\langle w_i \rangle = p_i^{e_i - f_i} \langle v_i \rangle$. By definition,

$$\langle v_i \rangle = \{ v \in M \mid p_i^{e_i} v = 0 \}$$

is a primary submodule of M and therefore $\langle w_i \rangle$ must be a submodule of $\langle v_i \rangle$. Let $x \in \langle w_i \rangle$ so that $x = kv_i$ for some $k \in R$. We have

$$p_i^{f_i} x = p_i^{f_i} k v_i = 0,$$

so $o(v_i) = p_i^{e_i}$ divides $p_i^{f_i}k$ and $p_i^{e_i-f_i}$ must divide k. This shows that $x \in p_i^{e_i-f_i}\langle v_i\rangle$ and therefore $\langle w_i\rangle$ is a submodule of $p_i^{e_i-f_i}\langle v_i\rangle$. But $\langle w_i\rangle = p_i^{e_i-f_i}\langle v_i\rangle$ by Lemma 108, which proves the uniqueness of N. □

Theorem 110. *[Exercise 6.11] Let M be a free module of finite rank over a principal ideal domain R. Let N be a submodule of M. If M/N is torsion, then* $\mathrm{rk}(N) = \mathrm{rk}(M)$.

Proof. Let B_0 be a basis for N and let $B = \{v_1, \ldots, v_n\}$ be a basis for M. Since B_0 is a linearly independent set in M, we have

$$\mathrm{rk}(N) = |B_0| \leq \mathrm{rk}(M)$$

by Theorem 5.5. Suppose that $n > \mathrm{rk}(N)$. For each i, the element $v_i + N \in M/N$ is torsion, and there exists some nonzero $r_i \in R$ such that $r_iv_i + N = N$, i.e. $r_iv_i \in N$. By Theorem 5.5, $\{r_1v_1, \ldots, r_nv_n\}$ is a linearly dependent set, and

$$a_1r_1v_1 + \cdots + a_nr_nv_n = 0$$

for some $a_1, \ldots, a_n \in R$ where not all a_i are zero. But every r_i is nonzero, so $\{v_1, \ldots, v_n\}$ is linearly dependent, which is a contradiction. Therefore

$$\mathrm{rk}(N) \leq \mathrm{rk}(M) = n \leq \mathrm{rk}(N)$$

and $\mathrm{rk}(N) = \mathrm{rk}(M)$. □

Example 111. [Exercise 6.13] Show that the rational numbers $\mathbb{Q}$ form a torsion-free $\mathbb{Z}$-module that is not free.

It is clear that $\mathbb{Q}$ is torsion-free. Suppose that B is a basis for $\mathbb{Q}$. If r_1, r_2 are distinct elements of B, then we may write $r_1 = a_1/b_1$ and $r_2 = a_2/b_2$ so that

$$(a_2b_1)r_1 + (-a_1b_2)r_2 = a_1a_2 - a_1a_2 = 0.$$

This contradicts the linear independence of B, so B must contain exactly one element r. But $r/2$ cannot be written as an integer multiple of r, so $\mathbb{Q}$ cannot be free.

More on Complemented Submodules.

Theorem 112. *[Exercise 6.14] Let R be a principal ideal domain and let M be a free R-module.*

(1) *A submodule N of M is complemented if and only if M/N is free.*
(2) *If M is also finitely generated, then N is complemented if and only if M/N is torsion-free.*

Proof. If M/N is free, then applying Theorem 5.6 to the quotient map $\tau : M \to M/N$ shows that N is complemented. Conversely, if N is complemented then $M = N \oplus S$ for some submodule S of M, and $M/N \cong S$ by Theorem 4.16. But S is free, so M/N is free. If M is finitely generated then we know that M/N is finitely generated, by

Theorem 86. Theorem 6.8 shows that M/N is free if and only if M/N is torsion-free, which proves (2). □

Theorem 113. *[Exercise 6.15] Let M be a free module of finite rank over a principal ideal domain R.*

(1) *If N is a complemented submodule of M, then* $\mathrm{rk}(N) = \mathrm{rk}(M)$ *if and only if $N = M$.*
(2) (1) *need not hold if N is not complemented.*
(3) *N is complemented if and only if any basis for N can be extended to a basis for M.*

Proof. One direction for (1) is evident. Write $M = N \oplus S$. If $\mathrm{rk}(N) = \mathrm{rk}(M)$ then $\mathrm{rk}(S) = 0$ and $S = \{0\}$, which shows that $N = M$. To show (2), let $M = \mathbb{Z}$ and $N = 2\mathbb{Z}$ as $\mathbb{Z}$-modules. Then M is free with basis $\{1\}$ and N is free with basis $\{2\}$, but $N \neq M$.

Suppose that $M = N \oplus S$ where S is a submodule of M, and let B_1 be a basis for S. If B_0 is a basis for N, then $B_0 \cup B_1$ is a basis for M. Conversely, suppose that any basis for N can be extended to a basis for M. Let B_0 be a basis for N and extend this to a basis B of M. Then

$$M = N \oplus \langle B \setminus B_0 \rangle ,$$

which shows that N is complemented. This proves (3). □

Theorem 114. *[Exercise 6.16] Let M and N be free modules of finite rank over a principal ideal domain R. Let $\tau : M \to N$ be an R-map.*

(1) $\ker(\tau)$ *is complemented.*
(2) $\mathrm{im}(\tau)$ *need not be complemented.*
(3) $\mathrm{rk}(M) = \mathrm{rk}(\ker(\tau)) + \mathrm{rk}(\mathrm{im}(\tau)) = \mathrm{rk}(\ker(\tau)) + \mathrm{rk}(M/\ker(\tau))$.
(4) *If τ is surjective, then τ is an isomorphism if and only if* $\mathrm{rk}(M) = \mathrm{rk}(N)$.
(5) *If L is a submodule of M and if M/L is free, then*

$$\mathrm{rk}(M/L) = \mathrm{rk}(M) - \mathrm{rk}(L).$$

Proof. By considering τ as a map onto the free module $\mathrm{im}(\tau)$, Theorem 5.6 shows that $\ker(\tau)$ is complemented and that

$$M = \ker(\tau) \oplus N \tag{*}$$

where $N \cong \mathrm{im}(\tau)$. However, $\mathrm{im}(\tau)$ need not be complemented; we can take $\tau : \mathbb{Z} \to \mathbb{Z}$ defined by $k \mapsto 2k$. This proves (1) and (2). We have $\mathrm{im}(\tau) \cong M/\ker(\tau)$ by the first isomorphism theorem, so from (*) we see that

$$\begin{aligned} \mathrm{rk}(M) &= \mathrm{rk}(\ker(\tau)) + \mathrm{rk}(N) \\ &= \mathrm{rk}(\ker(\tau)) + \mathrm{rk}(\mathrm{im}(\tau)) \end{aligned}$$

$$= \mathrm{rk}(\ker(\tau)) + \mathrm{rk}(M/\ker(\tau)),$$

which proves (3). Suppose that τ is surjective. If τ is an isomorphism then $\ker(\tau) = \{0\}$ and $\mathrm{rk}(M) = \mathrm{rk}(N)$ by (3). Conversely, if $\mathrm{rk}(M) = \mathrm{rk}(N)$ then $\mathrm{rk}(\ker(\tau)) = 0$ and $\ker(\tau) = \{0\}$. This proves (4). Finally, (5) follows from taking the quotient map $\pi : M \to M/L$ in (3). □

Theorem 115. *[Exercise 6.17] Let R be a commutative ring with identity. A submodule N of an R-module M is said to be **pure in** M if whenever $v \in M \setminus N$, then $rv \notin N$ for all nonzero $r \in R$.*

(1) *N is pure if and only if $v \in N$ and $v = rw$ for nonzero $r \in R$ implies that $w \in N$.*
(2) *N is pure if and only if M/N is torsion-free.*
(3) *If R is a principal ideal domain and M is finitely generated, then N is pure if and only if M/N is free.*
(4) *If L and N are pure submodules of M, then so is $L \cap N$.*
(5) *If N is pure in M, then $L \cap N$ is pure in L for any submodule L of M.*

Proof. (1) is the contrapositive of the definition of pure. Suppose that N is pure and suppose that $r(v + N) = N$ for some $r \neq 0$. Then $rv \in N$ which implies that $v \in N$ and $v + N = N$. Conversely, suppose that M/N is torsion-free. If $v \in N$ and $v = rw$, then $r(w + N) = N$ and $w \in N$. This proves (2), and (3) follows from Theorem 6.8. Suppose that L and N are pure submodules of M. If $v \in L \cap N$ and $v = rw$ for some nonzero $r \in R$, then $w \in L$ and $w \in N$, which shows that $L \cap N$ is pure. This proves (4). Now suppose that N is pure in M and L is any submodule of M. If $v \in L \cap N$ and $v = rw$ for some nonzero $r \in R$ and $w \in L$, then $w \in N$. This proves (5). □

Theorem 116. *[Exercise 6.18] Let M be a free module of finite rank over a principal ideal domain R. Let L and N be submodules of M with L complemented in M. Then*

$$\mathrm{rk}(L + N) + \mathrm{rk}(L \cap N) = \mathrm{rk}(L) + \mathrm{rk}(N).$$

Proof. By the second isomorphism theorem,

$$(N + L)/L \cong N/(N \cap L).$$

Applying Theorem 114 gives

$$\mathrm{rk}(N + L) - \mathrm{rk}(L) = \mathrm{rk}(N) - \mathrm{rk}(N \cap L).$$

□

Chapter 7. The Structure of a Linear Operator

Theorem 117. *Let V be a vector space over a field F and let $\tau \in \mathcal{L}(V)$. If $f(x), g(x) \in F[x]$ are coprime polynomials such that $f(x)g(x)$ annihilates V_τ, then*

$$V = \ker(f(\tau)) \oplus \ker(g(\tau)).$$

Proof. Write

$$a(x)f(x) + b(x)g(x) = 1$$

for some $a(x), b(x) \in F[x]$. If $v \in V$ then

$$v = b(\tau)g(\tau)v + a(\tau)f(\tau)v \in \ker(f(\tau)) + \ker(g(\tau))$$

since

$$f(\tau)b(\tau)g(\tau)v = 0 = g(\tau)a(\tau)f(\tau)v.$$

If $v \in \ker(f(\tau)) \cap \ker(g(\tau))$ then

$$v = b(\tau)g(\tau)v + a(\tau)f(\tau)v = 0,$$

which shows that the sum is direct. □

Remark 118. [Exercise 7.1] We have seen that any $\tau \in \mathcal{L}(V)$ can be used to make V into an $F[x]$-module. Does every module V over $F[x]$ come from some $\tau \in \mathcal{L}(V)$? Explain.

Given an $F[x]$-module V, define $\tau : V \to V$ by $v \mapsto xv$; τ is easily seen to be a linear map. If

$$p(x) = \sum_i a_i x^i \in F[x]$$

and $v \in V$, then

$$\begin{aligned} p(x)v &= \sum_i a_i x^i v \\ &= \sum_i a_i \tau^i v \\ &= p(\tau)v. \end{aligned}$$

Thus the $F[x]$-module V does come from the linear operator τ.

Theorem 119. *[Exercise 7.2] Let $\tau \in \mathcal{L}(V)$ have minimal polynomial*

$$m_\tau(x) = p_1^{e_1}(x) \cdots p_n^{e_n}(x)$$

where $p_i(x)$ are distinct monic primes. The following are equivalent:

(1) *V_τ is cyclic.*
(2) *$\deg(m_\tau(x)) = \dim(V)$, i.e. τ is nonderogatory.*

(3) *The elementary divisors of* τ *are the prime power factors* $p_i^{e_i}(x)$ *and so*

$$V_\tau = \langle v_1 \rangle \oplus \cdots \oplus \langle v_k \rangle$$

is a direct sum of τ*-cyclic submodules* $\langle v_i \rangle$ *of order* $p_i^{e_i}(x)$*.*

Proof. If V_τ is cyclic, then V is τ-cyclic of dimension $\deg(m_\tau(x))$ by Theorem 7.5, which shows (1) $\Rightarrow$ (2). If $\deg(m_\tau(x)) = \dim(V)$, then $m_\tau(x) = c_\tau(x)$ and

$$p_1^{e_1}(x), \ldots, p_n^{e_n}(x)$$

must be the complete list of elementary divisors. This proves (2) $\Rightarrow$ (3). Finally, if (3) holds then V_τ is cyclic by Theorem 6.4, proving (3) $\Rightarrow$ (1). $\square$

Theorem 120. *[Exercise 7.3] A matrix* $A \in \mathcal{M}_n(F)$ *is nonderogatory if and only if it is similar to a companion matrix.*

Proof. If A is nonderogatory, then Theorem 7.9 shows that V_A is cyclic and Theorem 7.12 shows that the multiplication operator τ_A can be represented by a companion matrix. This means that A is similar to a companion matrix. Conversely, if $A = B(C[p(x)])B^{-1}$ then $[\tau_A]_B = C[p(x)]$, and V_A is cyclic by Theorem 7.12. By Theorem 7.9, A is nonderogatory. $\square$

Theorem 121. *[Exercise 7.4] If* A *and* B *are block diagonal matrices with the same blocks, but in possibly different order, then* A *and* B *are similar.*

Proof. One simple way to see this is to explicitly produce a permutation matrix P such that $A = PBP^{-1}$. We will use another approach here. Let A' be the matrix formed by taking the blocks of A into the elementary divisor version of the rational canonical form; then A is similar to A'. Define B' similarly so that B is similar to B'. The resulting companion blocks in A' must be the same as those in B' up to reordering, so A' is similar to B' by Theorem 7.14. $\square$

Remark 122. [Exercise 7.5] Let $A \in \mathcal{M}_n(F)$. Justify the statement that the entries of any invariant factor version of a rational canonical form for A are "rational" expressions in the entries of A, hence the origin of the term *rational canonical form*. Is the same true for the elementary divisor version?

If $F = \mathbb{C}$ and A contains only rational entries, then any invariant factor version of a rational canonical form contains only polynomials with rational coefficients. This is despite the fact that the invariant factors will split in $\mathbb{C}[x]$. The elementary divisors may not be polynomials with rational coefficients, so the elementary divisor version is not "rational" as in the invariant factor version.

Theorem 123. *[Exercise 7.6] Let* $\tau \in \mathcal{L}(V)$ *where* V *is finite-dimensional. If* $p(x) \in F[x]$ *is irreducible and if* $p(\tau)$ *is not one-to-one, then* $p(x)$ *divides the minimal polynomial of* τ*.*

Proof. Let $K = \ker(p(\tau))$. If $p(\tau)$ is not one-to-one, then $K \neq \{0\}$ and $\text{ann}(K)$ is not generated by a unit in $F[x]$. Since $p(x)$ annihilates K and is irreducible, we must have $\text{ann}(K) = \langle p(x) \rangle$. But the minimal polynomial of τ also annihilates K, so it must be divisible by $p(x)$. □

Theorem 124. *[Exercise 7.7] The minimal polynomial of $\tau \in \mathcal{L}(V)$ is the least common multiple of its elementary divisors.*

Proof. Follows from Theorem 7.7. □

Remark 125. [Exercise 7.8] Let $\tau \in \mathcal{L}(V)$ where V is a finite-dimensional vector space over a field F. Describe conditions on the minimal polynomial of τ that are equivalent to the fact that the elementary divisor version of the rational canonical form of τ is diagonal. What can you say about the elementary divisors?

If the elementary divisor version of the rational canonical form is diagonal, then each elementary divisor must be linear. Equivalently, the minimal polynomial of τ splits into linear factors over F.

Theorem 126. *[Exercise 7.10] Given any multiset of monic prime power polynomials*

$$M = \left\{ p_1^{e_{1,1}}(x), \cdots, p_1^{e_{1,k_1}}(x), \cdots, \cdots, p_n^{e_{n,1}}(x), \cdots, p_n^{e_{n,k_n}}(x) \right\}$$

and given any vector space V of dimension equal to the sum of the degrees of these polynomials, there exists an operator $\tau \in \mathcal{L}(V)$ whose multiset of elementary divisors is M.

Proof. Define the matrix

$$A = \text{diag}\left(C[p_1^{e_{1,1}}(x)], \cdots, C[p_1^{e_{1,k_1}}(x)], \cdots, \cdots, C[p_n^{e_{n,1}}(x)], \cdots, C[p_n^{e_{n,k_n}}(x)] \right).$$

Then by Theorem 7.12, the operator $\tau_A(v) = Av$ has elementary divisors equal to M. □

Example 127. [Exercise 7.11] Find all rational canonical forms (up to the order of the blocks on the diagonal) for a linear operator on $\mathbb{R}^6$ having minimal polynomial $(x-1)^2(x+1)^2$.

The possible elementary divisor multisets and their rational canonical forms for both the elementary divisor and invariant factor versions are:

- $\{(x-1)^2, (x+1)^2, (x-1)^2\}$ with matrices

$$\begin{bmatrix} 0 & -1 & & & & \\ 1 & 2 & & & & \\ & & 0 & -1 & & \\ & & 1 & -2 & & \\ & & & & 0 & -1 \\ & & & & 1 & 2 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 & 0 & 0 & -1 & & \\ 1 & 0 & 0 & 0 & & \\ 0 & 1 & 0 & 2 & & \\ 0 & 0 & 1 & 0 & & \\ & & & & 0 & -1 \\ & & & & 1 & 2 \end{bmatrix}$$

- $\{(x-1)^2, (x+1)^2, (x+1)^2\}$ with matrices

$$\begin{bmatrix} 0 & -1 & & & & \\ 1 & 2 & & & & \\ & & 0 & -1 & & \\ & & 1 & -2 & & \\ & & & & 0 & -1 \\ & & & & 1 & -2 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 & 0 & 0 & -1 & & \\ 1 & 0 & 0 & 0 & & \\ 0 & 1 & 0 & 2 & & \\ 0 & 0 & 1 & 0 & & \\ & & & & 0 & -1 \\ & & & & 1 & -2 \end{bmatrix}$$

- $\{(x-1)^2, (x+1)^2, x-1, x-1\}$ with matrices

$$\begin{bmatrix} 0 & -1 & & & & \\ 1 & 2 & & & & \\ & & 0 & -1 & & \\ & & 1 & -2 & & \\ & & & & 1 & \\ & & & & & 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 & 0 & 0 & -1 & & \\ 1 & 0 & 0 & 0 & & \\ 0 & 1 & 0 & 2 & & \\ 0 & 0 & 1 & 0 & & \\ & & & & 1 & \\ & & & & & 1 \end{bmatrix}$$

- $\{(x-1)^2, (x+1)^2, x+1, x+1\}$ with matrices

$$\begin{bmatrix} 0 & -1 & & & & \\ 1 & 2 & & & & \\ & & 0 & -1 & & \\ & & 1 & -2 & & \\ & & & & -1 & \\ & & & & & -1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 & 0 & 0 & -1 & & \\ 1 & 0 & 0 & 0 & & \\ 0 & 1 & 0 & 2 & & \\ 0 & 0 & 1 & 0 & & \\ & & & & -1 & \\ & & & & & -1 \end{bmatrix}$$

- $\{(x-1)^2, (x+1)^2, x-1, x+1\}$ with matrices

$$\begin{bmatrix} 0 & -1 & & & & \\ 1 & 2 & & & & \\ & & 0 & -1 & & \\ & & 1 & -2 & & \\ & & & & 1 & \\ & & & & & -1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 & 0 & 0 & -1 & & \\ 1 & 0 & 0 & 0 & & \\ 0 & 1 & 0 & 2 & & \\ 0 & 0 & 1 & 0 & & \\ & & & & 0 & 1 \\ & & & & 1 & 0 \end{bmatrix}$$

Example 128. [Exercise 7.12] How many possible rational canonical forms (up to order of blocks) are there for linear operators on $\mathbb{R}^6$ with minimal polynomial $(x-1)(x+1)^2$?

We give the general method for computation here, even if it is unnecessary for this example. The largest invariant factor must be $(x-1)(x+1)^2$, so the remaining elementary divisors must be a combination of $x-1$, $x+1$ and $(x+1)^2$. The number of different elementary divisor multisets is given by the coefficient of x^3 in the generating function

$$f(x) = (1 + x + x^2 + \cdots)^2(1 + x^2 + x^4 + \cdots).$$

We can expand f:

$$\begin{aligned} f(x) &= \frac{1}{(1-x)^2(1-x^2)} \\ &= \frac{1}{8}\left(\frac{1}{1+x}\right) + \frac{1}{8}\left(\frac{1}{1-x}\right) + \frac{1}{4}\left(\frac{1}{(1-x)^2}\right) + \frac{1}{2}\left(\frac{1}{(1-x)^3}\right) \\ &= \sum_{n\geq 0}\left(\frac{1}{8}(-1)^n + \frac{1}{8} + \frac{1}{4}(n+1) + \frac{1}{2}\binom{n+2}{2}\right)x^n. \end{aligned}$$

The coefficient of x^3 is then equal to

$$1 + \frac{1}{2}\binom{5}{2} = 6.$$

Remark 129. [Exercise 7.13]

(1) Show that if A and B are $n \times n$ matrices, at least one of which is invertible, then AB and BA are similar.
(2) What do the matrices

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

have to do with this issue?
(3) Show that even without the assumption on invertibility the matrices AB and BA have the same characteristic polynomial.

Assume without loss of generality that A is invertible. Then

$$AB = A(BA)A^{-1},$$

so AB and BA are similar. For (2), we can compute

$$AB = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad BA = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

These two matrices are not similar, so (1) may not hold if both A and B are singular.

For (3), write (by row and column operations)

$$A = PI_{n,r}Q$$

where P and Q are invertible and $I_{n,r}$ is an $n \times n$ matrix that has the $r \times r$ identity in the upper left-hand corner and 0s elsewhere. Let $B' = QBP$. Then $AB = PI_{n,r}B'P^{-1}$ and $BA = Q^{-1}B'I_{n,r}Q$, so it remains to show that $\det(xI - I_{n,r}B') = \det(xI - B'I_{n,r})$, for AB is similar to $I_{n,r}B'$ and BA is similar to $B'I_{n,r}$. Write

$$B' = \begin{bmatrix} B'_{1,1} & B'_{2,1} \\ B'_{2,1} & B'_{2,2} \end{bmatrix}$$

where $B'_{1,1}$ is an $r \times r$ matrix. Then

$$\begin{aligned} \det(xI - I_{n,r}B') &= \det\left(xI - \begin{bmatrix} B'_{1,1} & B'_{2,1} \\ 0 & 0 \end{bmatrix}\right) \\ &= \begin{vmatrix} xI - B'_{1,1} & -B'_{2,1} \\ 0 & xI \end{vmatrix} \\ &= x^{n-r}\det(xI - B'_{1,1}) \end{aligned}$$

and

$$\begin{aligned} \det(xI - B'I_{n,r}) &= \det\left(xI - \begin{bmatrix} B'_{1,1} & 0 \\ B'_{2,1} & 0 \end{bmatrix}\right) \\ &= \begin{vmatrix} xI - B'_{1,1} & 0 \\ -B'_{2,1} & xI \end{vmatrix} \\ &= x^{n-r}\det(xI - B'_{1,1}), \end{aligned}$$

which completes the proof.

Example 130. [Exercise 7.14] Let τ be a linear operator on F^4 with minimal polynomial $m_\tau(x) = (x^2+1)(x^2-2)$. Find the elementary divisor version of the rational canonical form for τ if $F = \mathbb{Q}$, $F = \mathbb{R}$ or $F = \mathbb{C}$.

If $F = \mathbb{Q}$, the matrix is

$$\begin{bmatrix} 0 & -1 & & \\ 1 & 0 & & \\ & & 0 & 2 \\ & & 1 & 0 \end{bmatrix}.$$

If $F = \mathbb{R}$, we can write $m_\tau(x) = (x^2+1)(x+\sqrt{2})(x-\sqrt{2})$ and the matrix is

$$\begin{bmatrix} 0 & -1 & & \\ 1 & 0 & & \\ & & -\sqrt{2} & \\ & & & \sqrt{2} \end{bmatrix}.$$

If $F = \mathbb{C}$, we can write $m_\tau(x) = (x+i)(x-i)(x+\sqrt{2})(x-\sqrt{2})$ and the matrix is

$$\begin{bmatrix} -i & & & \\ & i & & \\ & & -\sqrt{2} & \\ & & & \sqrt{2} \end{bmatrix}.$$

Remark 131. [Exercise 7.15] Suppose that the minimal polynomial of $\tau \in \mathcal{L}(V)$ is irreducible. What can you say about the dimension of V?

If the minimal polynomial $m_\tau(x)$ is irreducible, then any other elementary divisors must be equal to $m_\tau(x)$. Therefore the dimension of V is a multiple of $\deg(m_\tau(x))$.

Theorem 132. *[Exercise 7.16] Let $\tau \in \mathcal{L}(V)$ where V is finite-dimensional. Suppose that $p(x)$ is an irreducible factor of the minimal polynomial $m(x)$ of τ. Suppose further that $u, v \in V$ have the property that $o(u) = o(v) = p(x)$. Then $u = f(\tau)v$ for some polynomial $f(x)$ if and only if $v = g(\tau)u$ for some polynomial $g(x)$.*

Proof. Suppose that $u = f(\tau)v$ for some polynomial $f(x)$. Since $f(x) \notin \operatorname{ann}(v) = \langle p(x) \rangle$, it must be relatively prime to $p(x)$. Write

$$a(x)f(x) + b(x)p(x) = 1$$

for some polynomials $a(x), b(x)$. Then

$$\begin{aligned} v &= a(x)f(x)v + b(x)p(x)v \\ &= a(x)u \end{aligned}$$

since $p(x)$ annihilates v. The other direction follows by symmetry. □

Chapter 8. Eigenvalues and Eigenvectors

Theorem 133. *[Exercise 8.1] Let J be the $n \times n$ matrix whose entries are all equal to 1. Find the minimal polynomial and characteristic polynomial of J and the eigenvalues.*

Proof. We can calculate $J^2 = nJ$, so the minimal polynomial is

$$m_J(x) = x^2 - nx = x(x-n).$$

Since J has rank 1, we can find $n-1$ eigenvectors for the eigenvalue 0. Also, $(1, \ldots, 1)$ is an eigenvector with eigenvalue n. Therefore the characteristic polynomial of J is $x^{n-1}(x-n)$. □

Example 134. [Exercise 8.2] Prove that the eigenvalues of a matrix do not form a complete set of invariants under similarity.

Consider the following matrices:

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 0 & -1 \\ 1 & 2 \end{bmatrix}.$$

Both have characteristic polynomial $(x-1)^2$, but $m_A(x) = x-1$ while $m_B(x) = (x-1)^2$.

Theorem 135. *[Exercise 8.3] A linear operator $\tau \in \mathcal{L}(V)$ is invertible if and only if 0 is not an eigenvalue of τ.*

Proof. As a simple consequence of the definition, 0 is not an eigenvalue of τ if and only if $\tau - 0\iota = \tau$ is invertible. □

Theorem 136. *[Exercise 8.4] Let A be an $n \times n$ matrix over a field F that contains all roots of the characteristic polynomial of A. Then $\det(A)$ is the product of the eigenvalues of A, counting multiplicity.*

Proof. We have $c_A(x) = \det(xI - A) = (x - \lambda_1) \cdots (x - \lambda_n)$. Setting $x = 0$,

$$\det(-A) = (-1)^n \det(A) = (-1)^n \lambda_1 \cdots \lambda_n$$

so $\det(A) = \lambda_1 \cdots \lambda_n$. □

Theorem 137. *[Exercise 8.5] If λ is an eigenvalue of τ, then $p(\lambda)$ is an eigenvalue of $p(\tau)$, for any polynomial $p(x)$. Also, if $\lambda \neq 0$, then λ^{-1} is an eigenvalue for τ^{-1}.*

Proof. Let $p(x) = \sum_k a_k x^k$ be any polynomial. If $\tau v = \lambda v$ for some nonzero v, then

$$p(\tau)v = \sum_k a_k \tau^k v = \sum_k a_k \lambda^k v = p(\lambda)v.$$

If $\lambda \neq 0$, then

$$\tau^{-1}v = \lambda^{-1}(\tau^{-1}\lambda v) = \lambda^{-1}\tau^{-1}\tau v = \lambda^{-1}v.$$

□

Theorem 138. *[Exercise 8.6] An operator $\tau \in \mathcal{L}(V)$ is **nilpotent** if $\tau^n = 0$ for some positive $n \in \mathbb{N}$.*

(1) *If τ is nilpotent, then the (point) spectrum of τ is $\{0\}$.*
(2) *There exists a nonnilpotent operator with (point) spectrum $\{0\}$.*

Proof. The following proof for (1) is only valid if V is finite-dimensional: Let n denote the dimension of V and let $\tau \in \mathcal{L}(V)$ be a nilpotent operator with $\tau^k = 0$ for some $k > 0$. Since x^k annihilates V_τ we have $m_\tau(x) = x^j$ for some $j \leq k$. But the characteristic polynomial has the same roots as $m_\tau(x)$, so $c_\tau(x) = x^n$. We now give a more general proof.

We can assume that $V \neq \{0\}$, for otherwise (1) is trivial. If τ is injective, then $\tau v \neq 0$ for every $v \neq 0$. If we choose any nonzero $v \in V$, then $\tau^n v$ must be nonzero, which contradicts the fact that $\tau^n = 0$. Therefore $\ker(\tau) \neq \{0\}$ and $0 \in \operatorname{Spec}(\tau)$. If $\tau v = \lambda v$ for some nonzero $v \in V$, then

$$\lambda^n v = \tau^n v = 0$$

and $\lambda^n = 0 = \lambda$ since $v \neq 0$. Therefore $\operatorname{Spec}(\tau) = \{0\}$, which proves (1).

(2) is not true if V is finite-dimensional, since the Jordan normal form shows that operators with spectrum $\{0\}$ are similar to a matrix with only subdiagonal entries. However, we can show (2) by choosing an infinite-dimensional vector space. Let $V = \mathbb{R}[x]$ and consider the differential operator $D : V \to V$. To see that $\operatorname{Spec}(\tau) = \{0\}$, we have $D(1) = 0$ which shows that $0 \in \operatorname{Spec}(\tau)$, and if $Df(x) = \lambda f(x)$ and $f(x) \neq 0$ then $\lambda = 0$ for otherwise the degree of $Df(x)$ is strictly less than the degree of $\lambda f(x)$. But τ is not nilpotent, since $D^n(x^n) \neq 0$ for every n. □

Theorem 139. *[Exercise 8.7] If $\sigma, \tau \in \mathcal{L}(V)$ and one of σ and τ is invertible, then $\sigma\tau \sim \tau\sigma$ and so $\sigma\tau$ and $\tau\sigma$ have the same eigenvalues, counting multiplicity.*

Proof. Assume without loss of generality that σ is invertible. Then $\tau\sigma = \sigma^{-1}(\sigma\tau)\sigma$. □

Example 140. [Exercise 8.8]

(1) Find a linear operator τ that is not idempotent but for which $\tau^2(\iota - \tau) = 0$.
(2) Find a linear operator τ that is not idempotent but for which $\tau(\iota - \tau)^2 = 0$.
(3) Prove that if $\tau^2(\iota - \tau) = \tau(\iota - \tau)^2 = 0$, then τ is idempotent.

For (1) and (2), we can choose operators on $\mathbb{R}^4$ defined by the matrices

$$A = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & 0 & 0 \\ & & 1 & 0 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 0 & & \\ 1 & 1 & & \\ & & 0 & \\ & & & 0 \end{bmatrix},$$

where $A^2(I - A) = 0$ and $B(I - B)^2 = 0$. The condition in (3) implies that

$$\tau^2 - \tau^3 = 0 = \tau(\iota - 2\tau + \tau^2) = \tau - 2\tau^2 + \tau^3,$$

so $\tau^3 = \tau^2$ and

$$\tau - \tau^2 = \tau - 2\tau^2 + \tau^2 = \tau - 2\tau^2 + \tau^3 = 0.$$

Remark 141. [Exercise 8.9] An **involution** is a linear operator θ for which $\theta^2 = \iota$. If τ is idempotent, what can you say about $2\tau - \iota$? Construct a one-to-one correspondence between the set of idempotents on V and the set of involutions.

If τ is idempotent, then $\tau^2 = \tau$ and

$$(2\tau - \iota)^2 = 4\tau^2 - 4\tau + \iota = \iota.$$

Let V be a vector space over a field F with characteristic other than 2. Let I be the set of idempotent operators on V and let J be the set of involutions on V. Then the map $f : I \to J$ defined by

$$\tau \mapsto 2\tau - \iota$$

is a bijection, for if $\tau \in J$ then

$$f\left(\frac{\tau + \iota}{2}\right) = \tau$$

and if $f(\tau) = f(\tau')$ then $2\tau - \iota = 2\tau' - \iota$ which implies that $\tau = \tau'$.

Lemma 142. *Let V be a vector space over a field F. Let $\tau \in \mathcal{L}(V)$ and suppose that*

$$B = (v, \tau v, \ldots, \tau^{n-1} v)$$

is a basis for an n-dimensional subspace S. Let $\{p_0(x), \ldots, p_{n-1}(x)\}$ be a set of polynomials where each $p_k(x)$ has degree k. Then

$$B' = (p_0(\tau)v, p_1(\tau)v, \ldots, p_{n-1}(\tau)v)$$

is also a basis for S.

Proof. For each k, write $p_k(x) = \sum_{j=0}^{k} a_j x^j$. Define a linear operator $\sigma : V \to V$ by

$$[\sigma]_B = \begin{bmatrix} a_{0,0} & & & \\ a_{1,0} & a_{1,1} & & \\ \vdots & \vdots & \ddots & \\ a_{n-1,0} & a_{n-1,1} & \cdots & a_{n-1,n-1} \end{bmatrix}.$$

Since each $p_k(x)$ has degree k, the coefficient $a_{k,k}$ must be nonzero. This implies that σ is invertible and therefore an isomorphism; it carries the basis B to B'. By Theorem 16, B' is also a basis. □

Theorem 143. *[Exercise 8.11] Let V be a vector space over a field F. Let $\tau \in \mathcal{L}(V)$ and let*

$$S = \langle v, \tau v, \ldots, \tau^{de-1} v \rangle$$

be a τ-cyclic subspace of V with minimal polynomial $p(x)^e$ where $p(x)$ is prime of degree d. Let $\sigma = p(\tau)$ restricted to $\langle v \rangle = Fv$. Then S is the direct sum of d σ-cyclic subspaces each of dimension e, that is,

$$S = T_1 \oplus \cdots \oplus T_d.$$

Proof. For each $0 \leq i < d$, let

$$B_{i+1} = \left\{\tau^i v, p(\tau)\tau^i v, \ldots, p(\tau)^{e-1}\tau^i v\right\}.$$

Then $B_1 \cup \cdots \cup B_d$ is a set that contains exactly de polynomials, with degrees $0, 1, \ldots, de-1$. Applying Lemma 142 completes the proof. □

Theorem 144. *[Exercise 8.12] Fix $\varepsilon > 0$. Define the "almost" Jordan block to be the matrix*

$$\tilde{\mathcal{J}}(\lambda, n) = \begin{bmatrix} \lambda & & & \\ \varepsilon & \lambda & & \\ & \ddots & \ddots & \\ & & \varepsilon & \lambda \end{bmatrix}.$$

Every complex square matrix is similar to a matrix composed of "almost" Jordan blocks (for the given ε).

Proof. Let A be a complex square matrix. By Theorem 8.6, A is similar to a matrix of the form

$$\operatorname{diag}(\mathcal{J}(a_1, e_1), \ldots, \mathcal{J}(a_k, e_k)).$$

For each $1 \leq i \leq k$, define the similarity block $B_i = \operatorname{diag}(1, \varepsilon, \ldots, \varepsilon^{e_k})$ so that $B_i \mathcal{J}(a_i, e_i) B_i^{-1} = \tilde{\mathcal{J}}(a_i, e_i)$. Then A is similar to

$$\operatorname{diag}(\tilde{\mathcal{J}}(a_1, e_1), \ldots, \tilde{\mathcal{J}}(a_k, e_k))$$

with similarity matrix $\operatorname{diag}(B_1, \ldots, B_k)$. □

Remark 145. [Exercise 8.13] Show that the Jordan canonical form is not very robust in the sense that a small change in the entries of a matrix A may result in a large jump in the entries of the Jordan form J. Consider the matrix

$$A_\varepsilon = \begin{bmatrix} \varepsilon & 0 \\ 1 & 0 \end{bmatrix}.$$

What happens to the Jordan form of A_ε as $\varepsilon \to 0$?

If $\varepsilon \neq 0$, then

$$J = \begin{bmatrix} 0 & 0 \\ 0 & \varepsilon \end{bmatrix}$$

with eigenvectors $(0, 1)$ and $(\varepsilon, 1)$. But if $\varepsilon = 0$,

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

since the eigenvalue $\lambda = 0$ now has algebraic multiplicity 2.

Example 146. [Exercise 8.14] Give an example of a complex nonreal matrix all of whose eigenvalues are real. Show that any such matrix is similar to a real matrix. What about the type of the invertible matrices that are used to bring the matrix to Jordan form?

The matrix

$$A = \begin{bmatrix} i & 1 \\ 1 & -i \end{bmatrix}$$

has characteristic polynomial $c_A(x) = x^2$, so all eigenvalues are real. The Jordan normal form of any matrix M with eigenvalues all real must also be real, and M is similar to its Jordan normal form. However, if $M = PJP^{-1}$ where M is nonreal and J is real, then P must be complex for otherwise the product PJP^{-1} would be real.

Theorem 147. *Let $\tau \in \mathcal{L}(V)$ and suppose that τ is represented by a Jordan block $J(\lambda, e)$. Then $m_\tau(x) = (x-\lambda)^e$ and*

$$J(\lambda, e) \sim C[(x-\lambda)^e].$$

Proof. Since

$$J(\lambda, e) - \lambda I = \begin{bmatrix} 0 & & & \\ 1 & 0 & & \\ & \ddots & \ddots & \\ & & 1 & 0 \end{bmatrix},$$

a simple computation shows that $(J(\lambda, e) - \lambda I)^e = 0$ and $(x-\lambda)^e$ annihilates V_τ. Then $m_\tau(x) = (x-\lambda)^k$ for some $k \leq e$, but $(J(\lambda, e) - \lambda I)^k \neq 0$ if $k < e$. Therefore $m_\tau(x) = (x-\lambda)^e$ and τ is nonderogatory. Similarity to the companion matrix now follows from Theorem 7.12. □

Theorem 148. *The Jordan normal form, up to the order of the Jordan blocks, is a complete invariant for similarity.*

Proof. If $\tau, \sigma \in \mathcal{L}(V)$ and $\tau \sim \sigma$, then their Jordan normal forms are identical up to order of Jordan blocks since their elementary divisors are the same. Conversely, if

$$J = \operatorname{diag}(J(\lambda_1, e_1), \ldots, J(\lambda_k, e_k))$$

is a Jordan form matrix, then

$$J \sim \operatorname{diag}(C[(x-\lambda_1)^{e_1}], \ldots, C[(x-\lambda_k)^{e_k}])$$

by Theorem 147. Applying Theorem 7.14 completely determines the elementary divisors of J. □

Theorem 149. *[Exercise 8.15] Let $J = [\tau]_B$ be the Jordan form of a linear operator $\tau \in \mathcal{L}(V)$. For a given Jordan block of $J(\lambda, e)$ let U be the subspace of V spanned by the basis vectors of B associated with that block.*

(1) *$\tau|_U$ has a single eigenvalue λ with geometric multiplicity 1. In other words, there is essentially only one eigenvector (up to scalar multiple) associated with each Jordan block. Hence, the geometric multiplicity of λ for τ is the number of Jordan blocks for λ.*
(2) *The algebraic multiplicity of an eigenvalue λ is the sum of the dimensions of the Jordan blocks associated with λ.*
(3) *The number of Jordan blocks in J is the maximum number of linearly independent eigenvectors of τ.*
(4) *If the algebraic multiplicity of every eigenvalue is equal to its geometric multiplicity, then each Jordan block is a 1×1 matrix.*

Proof. For (1), it suffices to show that the matrix $J(\lambda, e)$ satisfies the required properties. We can compute

$$J(\lambda, e) - \lambda I = \begin{bmatrix} 0 & & & \\ 1 & 0 & & \\ & \ddots & \ddots & \\ & & 1 & 0 \end{bmatrix},$$

which has a one-dimensional kernel:

$$\ker(J(\lambda, e) - \lambda I) = \left\langle \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} \right\rangle.$$

Since the Jordan blocks are taken over subspaces with pairwise trivial intersections, the geometric multiplicity of λ is the number of Jordan blocks. This proves (1). Each $k \times k$ Jordan block for an eigenvalue λ contributes k to the total algebraic multiplicity of λ, so (2) follows. (3) is immediate from (1) and the definition of geometric multiplicity. Finally, if the algebraic multiplicity of every eigenvalue is equal to its geometric multiplicity, then for each λ the sum of the dimensions of the Jordan blocks for λ is equal to the number of Jordan blocks for λ. In other words, each Jordan block has size 1. This proves (4). □

Theorem 150. *[Exercise 8.16] Assume that the base field F is algebraically closed. Then assuming that the eigenvalues of a matrix A are known, it is possible to determine the Jordan form J of A by looking at the rank of various matrix powers. A matrix B is **nilpotent** if $B^n = 0$ for some $n > 0$. The smallest such exponent is called the **index of nilpotence**.*

(1) *Let $J = J(\lambda, n)$ be a single Jordan block of size $n \times n$. Then $J - \lambda I$ is nilpotent of index n. Thus, n is the smallest integer for which $\operatorname{rk}(J - \lambda I)^n = 0$.*

Now let J be a matrix in Jordan form but possessing only one eigenvalue λ.

(2) *$J - \lambda I$ is nilpotent. If m is its index of nilpotence, then m is the maximum size of the Jordan blocks of J and $\mathrm{rk}(J - \lambda I)^{m-1}$ is the number of Jordan blocks in J of maximum size.*
(3) *$\mathrm{rk}(J - \lambda I)^{m-2}$ is equal to 2 times the number of Jordan blocks of maximum size plus the number of Jordan blocks of size one less than the maximum.*
(4) *The sequence $\mathrm{rk}(J - \lambda I)^k$ for $k = 1, \ldots, m$ uniquely determines the number and size of all of the Jordan blocks in J, that is, it uniquely determines J up to the order of the blocks.*
(5) *Now let J be an arbitrary Jordan matrix. If λ is an eigenvalue for J, then the sequence $\mathrm{rk}(J - \lambda I)^k$ for $k = 1, \ldots, m$, where m is the first integer for which $\mathrm{rk}(J - \lambda I)^m = \mathrm{rk}(J - \lambda I)^{m+1}$, uniquely determines J up to the order of the blocks.*
(6) *For any matrix A with spectrum $\{\lambda_1, \ldots, \lambda_s\}$ the sequence $\mathrm{rk}(A - \lambda_i I)^k$ for $i = 1, \ldots, s$ and $k = 1, \ldots, m$, where m is the first integer for which $\mathrm{rk}(A - \lambda_i I)^m = \mathrm{rk}(A - \lambda_i I)^{m+1}$, uniquely determines the Jordan matrix J for A up to the order of the blocks.*

This provides an alternative proof of the uniqueness of the Jordan normal form.

Proof. (1) follows from Theorem 147. In (2), since J has only one eigenvalue λ, the matrix $J - \lambda I$ is lower triangular with zero diagonal entries; this is clearly nilpotent. If

$$J = \mathrm{diag}(J(\lambda, e_1), \ldots, J(\lambda, e_k))$$

then

$$\begin{aligned}(J - \lambda I)^n &= \mathrm{diag}((J(\lambda, e_1) - \lambda I)^n, \ldots, (J(\lambda, e_k) - \lambda I)^n). \\ &= \mathrm{diag}(B_1^n, \ldots, B_k^n)\end{aligned}$$

where we write $B_i = J(\lambda, e_i) - \lambda I$ for convenience. If m is the index of nilpotence of $J - \lambda I$, then m must be the index of nilpotence for at least one of B_i, and no other block has a higher index of nilpotence. By (1), the maximum size of any Jordan block in J must be m. Also, for each B_i with maximum size, B_i^{m-1} consists of a single 1 in the bottom-left corner, so its rank is 1. This proves (2). In general, if B_i has size e_i, then B_i^k contains exactly $e_i - k$ nonzero entries, and has rank $e_i - k - 1$. From this, (3) and(4) are clear. If J is an arbitrary Jordan matrix, we can reorder the blocks to group blocks with the same eigenvalue together. Applying (4) then proves (5) and (6). □

Theorem 151. *[Exercise 8.17] Let $A \in \mathcal{M}_n(F)$.*

(1) *If all the roots of the characteristic polynomial of A lie in F, then A is similar to its transpose A^T.*
(2) *Any matrix is similar to its transpose.*

Proof. If all the roots of $c_A(x)$ lie in F, then $A \sim J$ where J is a matrix in Jordan form. But every Jordan block is similar to its transpose, since

$$\begin{bmatrix} \lambda & & & \\ 1 & \lambda & & \\ & \ddots & \ddots & \\ & & 1 & \lambda \end{bmatrix} = \begin{bmatrix} & & & 1 \\ & & \iddots & \\ & 1 & & \\ 1 & & & \end{bmatrix} \begin{bmatrix} \lambda & 1 & & \\ & \lambda & \ddots & \\ & & \ddots & 1 \\ & & & \lambda \end{bmatrix} \begin{bmatrix} & & & 1 \\ & & \iddots & \\ & 1 & & \\ 1 & & & \end{bmatrix}.$$

Therefore $A \sim J \sim J^T \sim A^T$, which proves (1). If $c_A(x)$ does not split into linear factors in F, then let $K \supseteq F$ be the splitting field of $c_A(x)$. Now $A \sim A^T$ in K, so $A \sim A^T$ in F by Theorem 7.20. □

The Trace of a Matrix.

Theorem 152. *[Exercise 8.18] Let $A \in \mathcal{M}_n(F)$.*

(1) $\operatorname{tr}(rA) = r \operatorname{tr}(A)$ *for* $r \in F$.
(2) $\operatorname{tr}(A + B) = \operatorname{tr}(A) + \operatorname{tr}(B)$.
(3) $\operatorname{tr}(AB) = \operatorname{tr}(BA)$.
(4) $\operatorname{tr}(ABC) = \operatorname{tr}(CAB) = \operatorname{tr}(BCA)$, *but* $\operatorname{tr}(ABC)$ *does not necessarily equal* $\operatorname{tr}(ACB)$.
(5) *The trace is an invariant under similarity.*
(6) *If F is algebraically closed, then the trace of A is the sum of the eigenvalues of A, including multiplicity.*

Proof. (1) and (2) are obvious. For (3), we have

$$\operatorname{tr}(AB) = \sum_{i=1}^{n} (AB)_{i,i} = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{i,j} B_{j,i} = \sum_{j=1}^{n} \sum_{i=1}^{n} B_{j,i} A_{i,j} = \operatorname{tr}(BA).$$

(4) follows immediately from (3), noting that

$$\operatorname{tr}\left(\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \right) = \operatorname{tr}\left(\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \right) = 1$$

but

$$\operatorname{tr}\left(\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \right) = \operatorname{tr}\left(\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \right) = 0.$$

If $A = PBP^{-1}$ then

$$\operatorname{tr}(A) = \operatorname{tr}(PBP^{-1}) = \operatorname{tr}(BP^{-1}P) = \operatorname{tr}(B),$$

which proves (5). If F is algebraically closed, then A is similar to its Jordan normal form, which has the eigenvalues of A as its diagonal entries. This proves (6). □

Example 153. [Exercise 8.19] Use the concept of the trace of a matrix, as defined in the previous exercise, to prove that there are no matrices $A, B \in \mathcal{M}_n(\mathbb{C})$ for which $AB - BA = I$.

If $AB - BA = I$ then $n = \operatorname{tr}(I) = \operatorname{tr}(AB - BA) = \operatorname{tr}(AB) - \operatorname{tr}(BA) = 0$, which is a contradiction.

Theorem 154. *[Exercise 8.20] Let $T : \mathcal{M}_n(F) \to F$ be a function with the following properties. For all matrices $A, B \in \mathcal{M}_n(F)$ and $r \in F$,*

(1) $T(rA) = rT(A)$.
(2) $T(A + B) = T(A) + T(B)$.
(3) $T(AB) = T(BA)$.

Then there exists $s \in F$ for which $T(A) = s \operatorname{tr}(A)$, for all $A \in \mathcal{M}_n(F)$. That is, up to a constant factor, tr *is the unique function that satisfies the above three properties.*

Proof. If $A = PBP^{-1}$ then

$$T(A) = T(PBP^{-1}) = T(BP^{-1}P) = \operatorname{tr}(B),$$

which shows that T is a similarity invariant. We also have $T(0) = T(0 \cdot I) = 0T(I) = 0$. Write $E_{i,j}$ for an $n \times n$ matrix with 1 in the (i, j) position and 0 everywhere else. Let $s = T(E_{1,1})$; we have $T(E_{i,i}) = s$ for every i since $E_{i,i} \sim E_{1,1}$. Now consider $E_{i,j}$ where $i \neq j$. We have $E_{i,j}E_{j,j} = E_{i,j}$ but $E_{j,j}E_{i,j} = 0$, so

$$T(E_{i,j}) = T(E_{i,j}E_{j,j}) = T(E_{j,j}E_{i,j}) = 0.$$

Finally, let $A \in \mathcal{M}_n(F)$ and write

$$A = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{i,j} E_{i,j}$$

so that

$$T(A) = T\left(\sum_{i=1}^{n} \sum_{j=1}^{n} a_{i,j} E_{i,j} \right) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{i,j} T(E_{i,j}) = s \sum_{i=1}^{n} a_{i,i} = s \operatorname{tr}(A).$$

□

Commuting Operators.

Lemma 155. *If $\tau \in \mathcal{L}(V)$ is diagonalizable and S is a τ-invariant subspace of V, then $\tau|_S$ is also diagonalizable.*

Proof. Since τ is diagonalizable, its minimal polynomial $m_\tau(x)$ has no multiple roots. But S is τ-invariant, so $m_\tau(\tau|_S) = 0$ and $m_{\tau|_S}(x)$ divides $m_\tau(x)$. Then $m_{\tau|_S}(x)$ has no multiple roots, so $\tau|_S$ is diagonalizable. □

Theorem 156. *[Exercise 8.21] A pair of linear operators $\sigma, \tau \in \mathcal{L}(V)$ is **simultaneously diagonalizable** if there is an ordered basis B of V for which $[\tau]_B$ and $[\sigma]_B$ are both diagonal, that is, B is an ordered basis of eigenvectors for both τ and σ. Two diagonalizable operators σ and τ are simultaneously diagonalizable if and only if they commute, that is, $\sigma\tau = \tau\sigma$.*

Proof. If τ and σ are simultaneously diagonalizable, then $[\tau]_B = D_1$ and $[\sigma]_B = D_2$ where D_1 and D_2 are diagonal matrices. We have

$$[\tau\sigma]_B = D_1 D_2 = D_2 D_1 = [\sigma\tau]_B ,$$

which shows that τ and σ commute. Conversely, suppose that $\sigma\tau = \tau\sigma$ and write

$$V = \mathcal{E}_{\lambda_1} \oplus \cdots \oplus \mathcal{E}_{\lambda_k}$$

where $\lambda_1, \ldots, \lambda_k$ are the distinct eigenvalues of τ. If $v \in \mathcal{E}_{\lambda_i}$,

$$\tau(\sigma v) = \sigma\tau v = \lambda_i(\sigma v)$$

so $\sigma v \in \mathcal{E}_{\lambda_i}$. This shows that $(\mathcal{E}_{\lambda_1}, \ldots, \mathcal{E}_{\lambda_k})$ reduces σ. But each $\sigma|_{\mathcal{E}_{\lambda_i}}$ is diagonalizable by Lemma 155, so there exists a basis B_i of $\mathcal{E}_{\lambda_i}$ such that $\left[\sigma|_{\mathcal{E}_{\lambda_i}}\right]_{B_i}$ is diagonal. Let $B = B_1 \cup \cdots \cup B_k$; then $[\sigma]_B$ is diagonal and $[\tau]_B$ is also diagonal. □

Theorem 157. *[Exercise 8.22] Let $\sigma, \tau \in \mathcal{L}(V)$. If σ and τ commute, then every eigenspace of σ is τ-invariant. Thus, if $\mathcal{F}$ is a commuting family, then every eigenspace of any member of $\mathcal{F}$ is $\mathcal{F}$-invariant.*

Proof. As in Theorem 156. □

Theorem 158. *[Exercise 8.23] Let $\mathcal{F}$ be a finite family of operators in $\mathcal{L}(V)$ with the property that each operator in $\mathcal{F}$ has a full set of eigenvalues in the base field F, that is, the characteristic polynomial splits over F. If $\mathcal{F}$ is a commuting family, then $\mathcal{F}$ has a common eigenvector $v \in V$.*

Proof. Write $\mathcal{F} = \{\tau_1, \ldots, \tau_n\}$. If $n = 1$, then we are done. Otherwise, assume that any commuting family of $n-1$ operators has a common eigenvector. Choose an eigenspace $\mathcal{E}_\lambda$ of τ_n; then $\mathcal{E}_\lambda$ is $\mathcal{F}$-invariant by Theorem 157. By the induction hypothesis, $\{\tau_1|_{\mathcal{E}_\lambda}, \ldots, \tau_{n-1}|_{\mathcal{E}_\lambda}\}$ has a common eigenvector $v \in \mathcal{E}_\lambda$. But v is also an eigenvector of τ_n, and this completes the proof. □

Geršgorin Disks.

Example 159. [Exercise 8.25] Find and sketch the Geršgorin region and the eigenvalues for the matrix

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}.$$

Write $\overline{D}_r(z_0)$ for the closed disc $\{z \in \mathbb{C} \mid |z - z_0| \leq r\}$. The row region is

$$\mathrm{GR}(A) = D_5(1) \cup D_{10}(5) \cup D_{15}(9)$$

and the column region is

$$\mathrm{GC}(A) = D_{11}(1) \cup D_{10}(5) \cup D_9(9).$$

Therefore the Geršgorin region is

$$G(A) = D_5(1) \cup D_{10}(5) \cup D_9(9).$$

Definition 160. A matrix $A \in \mathcal{M}_n(\mathbb{C})$ is **diagonally dominant** if for each $k = 1, \ldots, n$,

$$|A_{kk}| \geq R_k(A),$$

and it is **strictly diagonally dominant** if strict inequality holds.

Theorem 161. *[Exercise 8.26] If A is strictly diagonally dominant, then it is invertible.*

Proof. Any eigenvalue λ must lie in the Geršgorin row region

$$\bigcup_{k=1}^{n} \{z \in \mathbb{C} \mid |z - A_{kk}| \leq R_k(A)\}.$$

Since A is strictly diagonally dominant, we have

$$\begin{aligned} R_k(A) &\geq |\lambda - A_{kk}| \\ &\geq |A_{kk}| - |\lambda| \\ &> R_k(A) - |\lambda| \end{aligned}$$

for all k, which implies that $|\lambda| > 0$. By Theorem 135, A is invertible. □

Example 162. [Exercise 8.27,8.28] Find a matrix $A \in \mathcal{M}_n(\mathbb{C})$ that is diagonally dominant but not invertible, and find a matrix $B \in \mathcal{M}_n(\mathbb{C})$ that is invertible but not strictly diagonally dominant.

Take

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

Semisimple Operators.

Definition 163. A linear operator $\tau \in \mathcal{L}(V)$ on a finite-dimensional vector space is **semisimple** if every τ-invariant subspace has a complement that is also τ-invariant.

Theorem 164. *Let $\tau \in \mathcal{L}(V)$ be a semisimple operator. If S is a τ-invariant subspace of V, then $\tau|_S$ is also semisimple.*

Proof. Let T be a $\tau|_S$-invariant subspace of S. Since T is τ-invariant, there exists a τ-invariant subspace U of V such that $V = T \oplus U$. Then $S = T \oplus (U \cap S)$ by Theorem 5, so $U \cap S$ is a $\tau|_S$-invariant complement of T in S. □

Theorem 165. *Let V be a finite-dimensional vector space over an algebraically closed field F. Then τ is semisimple if and only if it is diagonalizable.*

Proof. Suppose that τ is diagonalizable and write

$$V = \mathcal{E}_{\lambda_1} \oplus \cdots \oplus \mathcal{E}_{\lambda_k}$$

where $\lambda_1, \ldots, \lambda_k$ are distinct eigenvalues. If S is a τ-invariant subspace, then $\tau|_S$ is diagonalizable by Theorem 155 and we can write

$$S = \widetilde{\mathcal{E}}_{\lambda_1} \oplus \cdots \oplus \widetilde{\mathcal{E}}_{\lambda_k}$$

where each $\widetilde{\mathcal{E}}_{\lambda_i} \subseteq \mathcal{E}_{\lambda_i}$ may be trivial. For each i, choose a subspace $\mathcal{D}_{\lambda_i}$ such that $\mathcal{E}_{\lambda_i} = \widetilde{\mathcal{E}}_{\lambda_i} \oplus \mathcal{D}_{\lambda_i}$; then

$$T = \mathcal{D}_{\lambda_1} \oplus \cdots \oplus \mathcal{D}_{\lambda_k}$$

is a τ-invariant complement of S (any subspace of an eigenspace is τ-invariant). To prove the converse, we use induction on the dimension n of V. If $n = 0$ then the statement is trivial. Otherwise, assume that every semisimple linear operator on a vector space of dimension $n-1$ over F is diagonalizable. Since F is algebraically closed, τ has a some eigenvector v. Choose a τ-invariant subspace T of V such that $T = \langle v \rangle \oplus T$; then $\tau|_T$ is semisimple, and by the induction hypothesis, also diagonalizable. Therefore

$$T = \langle v \rangle \oplus \langle v_1 \rangle \oplus \cdots \oplus \langle v_k \rangle$$

where $v_1, \ldots, v_k$ are eigenvectors of τ, so τ is diagonalizable. □

Chapter 9. Real and Complex Inner Product Spaces

Theorem 166. *[Exercise 9.1] If a matrix M is unitary, upper triangular and has positive entries on the main diagonal, then it must be the identity matrix.*

Proof. We use induction on the size of M. If M is a 1×1 matrix, the result is clear. Otherwise, write

$$M = \begin{bmatrix} m_{1,1} & * \\ 0 & M_1 \end{bmatrix} = \begin{bmatrix} m_{1,1} & m_{1,2} & \cdots & m_{1,n} \\ & m_{2,2} & \ddots & \vdots \\ & & \ddots & m_{n-1,n} \\ & & & m_{n,n} \end{bmatrix},$$

noting that the columns $m_1, \ldots, m_n$ of M form a orthonormal set. For each $i > 1$ we have $\langle m_1, m_i \rangle = m_{1,1}\overline{m_{1,i}} = 0$ and $m_{1,1} > 0$, so $m_{1,i} = 0$. Also $\langle m_1, m_1 \rangle = m_{1,1}^2 = 1$ and $m_{1,1} > 0$, so $m_{1,1} = 1$. Since M is unitary and

$$M^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & M_1^{-1} \end{bmatrix},$$

M_1 must also be unitary and upper triangular with positive entries on its main diagonal. By the induction hypothesis, $M_1 = I$ and therefore $M = I$. □

Theorem 167. *[Exercise 9.2] Any triangularizable matrix is unitarily (orthogonally) triagularizable.*

Proof. Let A be a triangularizable matrix with $A = PBP^{-1}$ where B is upper triangular. Write $P = QR$ where Q is unitary and R is upper triangular; then

$$A = (QR)B(QR)^{-1} = Q(RBR^{-1})Q^{-1}$$

where RBR^{-1} is upper triangular. □

Theorem 168. *[Exercise 9.3] In the triangle inequality*

$$\|u + v\| \leq \|u\| + \|v\|,$$

equality holds if and only if one of u and v is a nonnegative scalar multiple of the other.

Proof. If $u = av$ where $a \geq 0$, then

$$\|u + v\| = (a + 1)\|v\| = \|u\| + \|v\|;$$

the same holds when $v = au$. Conversely, if $\|u + v\| = \|u\| + \|v\|$ then

$$\|u\|^2 + \langle u, v \rangle + \langle v, u \rangle + \|v\|^2 = \|u\|^2 + 2\|u\|\|v\| + \|v\|^2.$$

But

$$|\langle u, v \rangle| \leq \|u\|\|v\| = \mathrm{Re}(\langle u, v \rangle) \leq |\langle u, v \rangle|,$$

so $\langle u, v \rangle$ is equal to $\|u\|\|v\|$. Equality in the Cauchy-Schwarz inequality holds only if one of u and v is a scalar multiple of the other; assume without loss of generality that $u = av$ for some $a \in F$. In this case,

$$\|u\|\|v\| = \langle u, v \rangle = \langle av, v \rangle = a\|v\|^2$$

so a must be nonnegative. □

Lemma 169. *[Exercise 9.4] If $u, v \in V$ then*

$$\|u+v\|^2 + \|u-v\|^2 = 2(\|u\|^2 + \|v\|^2).$$

Proof. We have

$$\|u+v\|^2 = \|u\|^2 + \langle u, v\rangle + \langle v, u\rangle + \|v\|^2$$

and

$$\|u-v\|^2 = \|u\|^2 - \langle u, v\rangle - \langle v, u\rangle + \|v\|^2 .$$

□

Theorem 170. *[Exercise 9.5] If $u, v, w \in V$, then*

$$\|w-u\|^2 + \|w-v\|^2 = \frac{1}{2}\|u-v\|^2 + 2\left\|w - \frac{1}{2}(u+v)\right\|^2 .$$

Proof. We have

(*) $\|w-u\|^2 + \|w-v\|^2 = 2\|w\|^2 + \|u\|^2 + \|v\|^2 - \langle w, u\rangle - \langle u, w\rangle - \langle w, v\rangle - \langle v, w\rangle$

while the right hand side is

$$\begin{aligned}&\frac{1}{2}\left(\|u-v\|^2 + \|u+v\|^2\right) + 2\|w\|^2 - \langle w, u+v\rangle - \langle u+v, w\rangle \\ &= 2\|w\|^2 + \|u\|^2 + \|v\|^2 - \langle w, u+v\rangle - \langle u+v, w\rangle ,\end{aligned}$$

which equals (*). □

Theorem 171. *[Exercise 9.6] Let V be an inner product space with basis B. The inner product is uniquely defined by the values $\langle u, v\rangle$ for all $u, v \in B$.*

Proof. If $u, v \in V$, write

$$\begin{aligned}u &= a_1 b_1 + \cdots + a_m b_m, \\ v &= a'_1 b'_1 + \cdots + a'_n b'_n\end{aligned}$$

where $a_i, a'_i \in F$ and $b_i, b'_i \in B$. Then

$$\begin{aligned}\langle u, v\rangle &= \left\langle \sum_{i=1}^{m} a_i b_i, \sum_{j=1}^{n} a'_j b'_j \right\rangle \\ &= \sum_{i=1}^{m} a_i \left\langle b_i, \sum_{j=1}^{n} a'_j b'_j \right\rangle \\ &= \sum_{i=1}^{m}\sum_{j=1}^{n} a_i \overline{a'_i} \langle b_i, b'_i\rangle ,\end{aligned}$$

which depends only on the values of $\langle \cdot, \cdot \rangle$ for vectors in B. □

Theorem 172. *[Exercise 9.7] Two vectors u and v in a real inner product space V are orthogonal if and only if*

$$\|u+v\|^2 = \|u\|^2 + \|v\|^2 .$$

Proof. This is clear from the expansion

$$\|u+v\|^2 = \|u\|^2 + 2\langle u, v\rangle + \|v\|^2 .$$

□

Theorem 173. *[Exercise 9.8] Any isometry is injective.*

Proof. Let $\tau \in \mathcal{L}(V, W)$ be an isometry. If $\tau v = 0$ then

$$\langle v, v\rangle = \langle \tau v, \tau v\rangle = 0$$

and $v = 0$. □

Theorem 174. *[Exercise 9.9] Let $V \neq \{0\}$ be an inner product space. Then V has a Hilbert basis.*

Proof. Let $\mathcal{A}$ be the collection of all orthonormal sets in V, partially ordered by set inclusion; $\mathcal{A}$ is not empty since we can choose a nonzero $v \in V$ and take $\{v/\|v\|\} \in \mathcal{A}$. Suppose that $\mathcal{C} = \{\mathcal{O}_i \mid i \in I\} \subseteq \mathcal{S}$ is a nonempty chain in $\mathcal{S}$. The union

$$U = \bigcup_{i \in I} \mathcal{O}_i$$

is orthonormal and $U \subseteq \mathcal{A}$. By Zorn's lemma, $\mathcal{A}$ has a maximal element $\mathcal{O}$. This is a Hilbert basis. □

Theorem 175. *[Exercise 9.10] Let V be a finite-dimensional inner product space, let $v \in V$ and let $\widehat{v}$ be the orthogonal projection of v onto a subspace S of V. Then*

$$\|\widehat{v}\| \leq \|v\| .$$

Proof. Since $v - \widehat{v} \perp \widehat{v}$,

$$\|v\|^2 = \|v - \widehat{v} + \widehat{v}\|^2 = \|v - \widehat{v}\|^2 + \|\widehat{v}\|^2 \geq \|\widehat{v}\|^2 .$$

□

Theorem 176. *[Exercise 9.11-9.13] Let $\mathcal{O} = \{u_1, \ldots, u_k\}$ be an orthonormal subset of an inner product space V and let $S = \langle \mathcal{O}\rangle$. The following are equivalent:*

(1) *$\mathcal{O}$ is an orthonormal basis for V.*
(2) *$\langle \mathcal{O}\rangle^{\perp} = \{0\}$.*
(3) *Every vector is equal to its Fourier expansion, that is, $\widehat{v} = v$ for all $v \in V$.*

(4) ***Bessel's identity*** *holds for all $v \in V$, that is, $\|\widehat{v}\| = \|v\|$.*
(5) ***Parseval's identity*** *holds for all $v, w \in V$, that is,*

$$\langle v, w\rangle = [\widehat{v}]_{\mathcal{O}} \cdot [\widehat{w}]_{\mathcal{O}}$$

where $\cdot$ is the standard inner product in F^k.

Proof. We first prove (1) $\Leftrightarrow$ (2). Suppose that $\mathcal{O}$ is an orthonormal basis and there exists a nonzero $v \in \langle\mathcal{O}\rangle^{\perp}$. Then $\mathcal{O} \cup \{v/\|v\|\}$ is orthonormal, so $\mathcal{O}$ is not maximal. Conversely, if $\mathcal{O}$ is not maximal then $\mathcal{O} \subset \mathcal{P}$ where $\mathcal{P}$ is orthonormal; any $v \in \mathcal{P} \setminus \mathcal{O}$ is nonzero and a member of $\langle\mathcal{O}\rangle^{\perp}$. Now suppose that (1) holds. By Theorem 9.14, $v - \widehat{v} \in \langle\mathcal{O}\rangle^{\perp} = \{0\}$ and $v = \widehat{v}$. This implies (3), while (3) implies (4). If (4) holds and $v \perp \langle\mathcal{O}\rangle$, then

$$\|v\|^2 = \|\widehat{v}\|^2 = \|\langle v, u_1\rangle\, u_1 + \cdots + \langle v, u_k\rangle\, u_k\|^2 = 0,$$

so $v = 0$. This implies (2). It now remains to show (1) $\Leftrightarrow$ (5). If (1) holds then (3) also holds. Therefore

$$\langle v, w\rangle = \langle\widehat{v}, \widehat{w}\rangle = [\widehat{v}]_{\mathcal{O}} \cdot [\widehat{w}]_{\mathcal{O}}$$

since $\mathcal{O}$ is an isometry. Suppose that (5) holds and $v \perp \langle\mathcal{O}\rangle$. Then $\widehat{v} = 0$, so

$$\langle v, v\rangle = [\widehat{v}]_{\mathcal{O}} \cdot [\widehat{v}]_{\mathcal{O}} = 0$$

and $v = 0$, which implies (2). □

Theorem 177. *[Exercise 9.14] Let $u = (r_1, \ldots, r_n)$ and $v = (s_1, \ldots, s_n)$ be in $\mathbb{R}^n$. Then*

$$(|r_1 s_1| + \cdots + |r_n s_n|)^2 \leq (r_1^2 + \cdots + r_n^2)(s_1^2 + \cdots + s_n^2).$$

Proof. Assume each r_k, s_k is nonnegative; we can compute

$$\begin{aligned}
\left(\sum_k r_k^2\right)\left(\sum_k s_k^2\right) - \left(\sum_k r_k s_k\right)^2 &= \sum_{j,k}(r_j^2 s_k^2 - r_j r_k s_j s_k) \\
&= \sum_{j \neq k}(r_j^2 s_k^2 - r_j r_k s_j s_k) \\
&= \sum_{j<k}(r_j^2 s_k^2 - r_j r_k s_j s_k) + \sum_{k<j}(r_j^2 s_k^2 - r_j r_k s_j s_k) \\
&= \sum_{j<k}(r_j^2 s_k^2 - r_j r_k s_j s_k) + \sum_{j<k}(r_k^2 s_j^2 - r_k r_j s_k s_j) \\
&= \sum_{j<k}(r_j s_k - r_k s_j)^2 \\
&\geq 0.
\end{aligned}$$

□

Theorem 178. *Let V be an inner product space and let $X, Y \subseteq V$ (cf. Theorem 62).*

(1) *$X^\perp$ is a subspace of V.*
(2) *If $X \subseteq Y$ then $Y^\perp \subseteq X^\perp$.*
(3) *$X \subseteq \operatorname{span}(X) \subseteq X^{\perp\perp}$.*

Proof. Let $v, w \in X^\perp$, $a \in F$ and $x \in X$. Clearly $0 \in X^\perp$; we also have $\langle v, x\rangle = \langle w, x\rangle = 0$ so $\langle v + w, x\rangle = 0$ and $\langle av, x\rangle = 0$. Therefore $v + w, av \in X^\perp$, which proves (1). For (2), let $v \in Y^\perp$ so that $v \perp y$ for every $y \in Y$. In particular, $v \perp x$ for every $x \in X$, so $v \in X^\perp$.

If $v \in \operatorname{span}(X)$ then $v = a_1 v_1 + \cdots + a_n v_n$ where $v_1, \ldots, v_n \in X$. If $w \in X^\perp$ then

$$\langle v, w\rangle = a_1 \langle v_1, w\rangle + \cdots + a_n \langle v_n, w\rangle = 0,$$

so $v \in X^{\perp\perp}$. This proves (3). □

Theorem 179. *[Exercise 9.15] Let V be a finite-dimensional inner product space. For any subset X of V, we have $X^{\perp\perp} = \operatorname{span}(X)$. In particular, if X is a subspace of V then $X^{\perp\perp} = X$.*

Proof. We know that $\operatorname{span}(X) \subseteq X^{\perp\perp}$ from Theorem 178. If $v \in X^{\perp\perp}$ then $v = s + s'$ where $s \in \operatorname{span}(X)$ and $s' \in \operatorname{span}(X)^\perp$, by Theorem 9.15. Since $X \subseteq \operatorname{span}(X)$ we have $\operatorname{span}(X)^\perp \subseteq X^\perp$ by Theorem 178, so $s' \in X^\perp$ and $s' \perp v$. But $s' \perp s$, so s' is orthogonal to itself and $s' = 0$. This implies that $v = s \in \operatorname{span}(X)$. □

Example 180. [Exercise 9.16] Let $\mathcal{P}_3$ be the inner product space of all polynomials of degree at most 3, under the inner product

$$\langle p(x), q(x)\rangle = \int_{-\infty}^{\infty} p(x)q(x)e^{-x^2}\,dx.$$

Apply the Gram-Schmidt process to the basis $\{1, x, x^2, x^3\}$, thereby computing the first four **Hermite polynomials** (at least up to a multiplicative constant).

We have

$$h_0(x) = 1$$

$$h_1(x) = x - \frac{\int_{-\infty}^{\infty} xe^{-x^2}\,dx}{\int_{-\infty}^{\infty} e^{-x^2}\,dx} = x$$

$$h_2(x) = x^2 - \frac{\int_{-\infty}^{\infty} x^2e^{-x^2}\,dx}{\int_{-\infty}^{\infty} e^{-x^2}\,dx} - \frac{\int_{-\infty}^{\infty} x^3e^{-x^2}\,dx}{\int_{-\infty}^{\infty} x^2e^{-x^2}\,dx}x = x^2 - \frac{1}{2}$$

$$h_3(x) = x^3 - \frac{\int_{-\infty}^{\infty} x^3 e^{-x^2}\,dx}{\int_{-\infty}^{\infty} e^{-x^2}\,dx} - \frac{\int_{-\infty}^{\infty} x^4 e^{-x^2}\,dx}{\int_{-\infty}^{\infty} x^2 e^{-x^2}\,dx}x - \frac{\int_{-\infty}^{\infty} (x^5 - \frac{1}{2}x^3) e^{-x^2}\,dx}{\int_{-\infty}^{\infty} (x^2 - \frac{1}{2})^2 e^{-x^2}\,dx}x^2$$
$$= x^3 - \frac{3}{2}x.$$

Theorem 181. *[Exercise 9.17] The linear map $\tau : V \to V^*$ defined by $\tau x = \langle \cdot, x \rangle$ is injective.*

Proof. If $\langle \cdot, x \rangle = 0$ then $\langle x, x \rangle = 0$, so $x = 0$. □

Theorem 182. *[Exercise 9.18] Let V be a complex inner product space and let S be a subspace of V. Suppose that $v \in V$ is a vector for which $\langle v, s \rangle + \langle s, v \rangle \leq \langle s, s \rangle$ for all $s \in S$. Then $v \in S^\perp$.*

Proof. Let $s \in S$ be a nonzero vector. For $z \in \mathbb{C}$,

$$\langle v, zs \rangle + \langle zs, v \rangle \leq \langle zs, zs \rangle,$$
$$\overline{z}\,\langle v, s \rangle + z\overline{\langle v, s \rangle} \leq |z|^2 \|s\|^2.$$

For all $r \in \mathbb{R}$, putting $z = r$ gives $2r \operatorname{Re}(\langle v, s \rangle) \leq r^2 \|s\|^2$ and putting $z = ir$ gives $2r \operatorname{Im}(\langle v, s \rangle) \leq r^2 \|s\|^2$. Therefore

$$\frac{2\,|\operatorname{Re}(\langle v, s \rangle)|}{\|s\|^2} \leq |r| \quad \text{and} \quad \frac{2\,|\operatorname{Im}(\langle v, s \rangle)|}{\|s\|^2} \leq |r|\,;$$

taking $r \to 0$ shows that $|\operatorname{Re}(\langle v, s \rangle)| = |\operatorname{Im}(\langle v, s \rangle)| = 0$. This completes the proof.

If S is assumed to be finite-dimensional, we sketch an alternative proof. The given condition implies that $\|v\| \leq \|v - s\|$ for all $s \in S$; if $\widehat{v} \neq 0$ then Theorem 9.14 implies that $\|v\| \leq \|v - \widehat{v}\| < \|v\|$, which is a contradiction. Therefore $\widehat{v} = 0$ and $v = v - \widehat{v} \in S^\perp$. □

Example 183. [Exercise 9.19] If V and W are inner product spaces, consider the function on $V \boxplus W$ defined by

$$\langle (v_1, w_1), (v_2, w_2) \rangle = \langle v_1, v_2 \rangle + \langle w_1, w_2 \rangle.$$

Is this an inner product on $V \boxplus W$?

Yes. Clearly $\langle (v, w), (v, w) \rangle \geq 0$, and

$$\langle (v, w), (v, w) \rangle = \|v\|^2 + \|w\|^2 = 0$$

implies that $\|v\| = \|w\| = 0$. The (conjugate) symmetry and linearity conditions are similarly satisfied.

Theorem 184. *[Exercise 9.20] If V is a real normed space and if the norm satisfies the parallelogram law*

$$\|u+v\|^2 + \|u-v\|^2 = 2\|u\|^2 + 2\|v\|^2,$$

then the polarization identity

$$\langle u, v\rangle = \frac{1}{4}(\|u+v\|^2 - \|u-v\|^2)$$

defines an inner product on V.

Proof. Since

$$0 = \|v-v\| \le \|-v\| + \|v\| = 2\|v\|,$$

we have $\|v\| \ge 0$ for all $v \in V$. Therefore $\langle v, v\rangle = \|v\|^2$ is positive-definite. Also, $\langle\cdot,\cdot\rangle$ is clearly symmetric, and

$$\begin{aligned} 8\langle u, x\rangle + 8\langle v, x\rangle &= 2(\|u+x\|^2 + \|v+x\|^2) - 2(\|u-x\|^2 + \|v-x\|^2) \\ &= \|u+v+2x\|^2 + \|u-v\|^2 - \|u+v-2x\|^2 - \|u-v\|^2 \\ &= 4\langle u+v, 2x\rangle \end{aligned}$$

and $2(\langle u, x\rangle + \langle v, x\rangle) = \langle u+v, 2x\rangle$. Setting $v = 0$ gives $\langle v, 2x\rangle = 2\langle v, x\rangle$ and thus

$$\langle u+v, x\rangle = \frac{1}{2}\langle u+v, 2x\rangle = \langle u, x\rangle + \langle v, x\rangle.$$

For fixed nonzero $u, v \in V$, define $f : \mathbb{R} \to \mathbb{R}$ by $r \mapsto \langle u, rv\rangle$. Let $a \in \mathbb{R}$, $\varepsilon > 0$ and choose $\delta = \min(\varepsilon/(2\|u\|\|v\|), \|u\|/\|v\|)$. Then for all $|b-a| < \delta$,

$$\begin{aligned} |f(b) - f(a)| &= |\langle u, (b-a)v\rangle| \\ &= \frac{1}{4}(\|u+(b-a)v\| + \|u-(b-a)v\|)\,|\|u+(b-a)v\| - \|u-(b-a)v\||| \\ &\le \frac{1}{4}(2\|u\| + 2\|(b-a)v\|)\,\|2(b-a)v\| \\ &< \delta\|v\|(\|u\| + \delta\|v\|) \\ &\le \varepsilon \end{aligned}$$

where the triangle inequality is used in the third line. This shows that f is continuous. We have

$$f(2^k) = \langle u, 2^k v\rangle = 2^k\langle u, v\rangle = 2^k f(1)$$

for all $k \in \mathbb{Z}$ and $f(a+b) = f(a) + f(b)$ for all $a, b \in \mathbb{R}$. Let $r \in \mathbb{R}$ and let

$$r = \sum_{k=-\infty}^{n} b_k 2^k$$

be a binary expansion of r, where $b_k \in \{0, 1\}$. Then for all m,

$$f\left(\sum_{k=-m}^{n} b_k 2^k\right) = \sum_{k=-m}^{n} b_k 2^k f(1),$$

and since f is continuous, letting $m \to \infty$ gives $f(r) = rf(1)$, i.e. $\langle u, rv\rangle = r\langle u, v\rangle$. □

Theorem 185. *[Exercise 9.21] Let S be a subspace of a finite-dimensional inner product space V. Then each coset in V/S contains exactly one vector that is orthogonal to S (cf. Theorem 59).*

Proof. From Theorem 9.15, $V = S \odot S^\perp$. Since $S^\perp$ is a complement of S, by Theorem 3.6 we have $S^\perp \cong V/S$. The result follows immediately. □

Extensions of Linear Functionals.

Theorem 186. *[Exercise 9.22] Let f be a linear functional on a subspace S of a finite-dimensional inner product space V. Let $f(v) = \langle v, R_f\rangle$. Suppose that $g \in V^*$ is an extension of f, that is, $g|_S = f$. Then $\widehat{R_g} = R_f$, where $\widehat{R_g}$ is the orthogonal projection of R_g onto S.*

Proof. Write $R_g = \widehat{R_g} + x$ where $\widehat{R_g} \in S$ and $x \in S^\perp$. We have

$$\begin{aligned}\langle \widehat{R_g} - R_f, \widehat{R_g} - R_f\rangle &= \langle \widehat{R_g} - R_f, R_g - R_f - x\rangle \\ &= \langle \widehat{R_g} - R_f, R_g\rangle - \langle \widehat{R_g} - R_f, R_f\rangle \\ &= 0\end{aligned}$$

since $x \perp \widehat{R_g} - R_f$ and $\langle v, R_g\rangle = \langle v, R_f\rangle$ for all $v \in S$. Therefore $\widehat{R_g} = R_f$. □

Theorem 187. *[Exercise 9.23] Let f be a nonzero linear functional on a subspace S of a finite-dimensional inner product space V and let $K = \ker(f)$. If $g \in V^*$ is an extension of f, then $R_g \in K^\perp \setminus S^\perp$. Moreover, for each vector $u \in K^\perp \setminus S^\perp$ there is exactly one scalar λ for which the linear functional $g(x) = \langle x, \lambda u\rangle$ is an extension of f.*

Proof. Let $\widehat{R_g}$ be the orthogonal projection of R_g onto S. By Theorem 186, $\widehat{R_g} = R_f$ where R_f is the Riesz vector for f. If $v \in K$ then $\langle v, R_g\rangle = \langle v, R_f\rangle = f(v) = 0$, but $\langle R_f, R_g\rangle = \langle R_f, R_f\rangle > 0$ since $f \neq 0$. Therefore $R_g \in K^\perp \setminus S^\perp$. Write $V = S \odot S^\perp = \langle w\rangle \odot K \odot S^\perp$ where $w \in K^\perp$ and w is nonzero. By Theorem 9.18, R_f is given by

$$R_f = \frac{\overline{f(w)}}{\|w\|^2} w.$$

If $u \in K^\perp \setminus S^\perp$ then $u = aw + z \in \langle w \rangle$ for some nonzero $a \in F$ and some $z \in S^\perp$; if $g(x) = \langle x, \lambda u \rangle$ is an extension of f, then

$$g(R_f) = \langle R_f, \lambda a w \rangle = \overline{\lambda a} \langle R_f, \frac{\|w\|^2}{\overline{f(w)}} R_f \rangle = \frac{\overline{\lambda a}\, \|w\|^2}{f(w)} \|R_f\|^2$$

and $g(R_f) = f(R_f) = \|R_f\|^2$ so that

$$\lambda = \frac{\overline{f(w)}}{a \|w\|^2}.$$

Conversely, this value of λ does define an extension of f, for $\langle x, \lambda u \rangle = \langle x, R_f \rangle = f(x)$. □

Lemma 188. *Let V be a finite-dimensional inner product space and let $f : V \to F$ be a nonzero linear functional. Write $V = \langle w \rangle \odot \ker(f)$ where $w \in \ker(f)^\perp$ is nonzero. Then $R_f \in \langle w \rangle$.*

Proof. By definition, for any $v \in \ker(f)$ we have $\langle v, R_f \rangle = f(v) = 0$. So $R_f \in \ker(f)^\perp = \langle w \rangle$. □

Positive Linear Functionals on $\mathbb{R}^n$.

Theorem 189. *[Exercise 9.24] A linear functional f on $\mathbb{R}^n$ is positive if and only if $R_f \geq 0$ and strictly positive if and only if $R_f \gg 0$.*

Proof. Write $R_f = (r_1, \ldots, r_n)$ and $v = (v_1, \ldots, v_n)$. If $R_f \geq 0$ and $v > 0$, then

$$R_f \cdot v = r_1 v_1 + \cdots + r_n v_n \geq 0$$

since $r_i, v_i \geq 0$. Conversely, if some $r_k < 0$, then put $v_k = 1$ and $v_i = 0$ for $i \neq k$ so that $R_f \cdot v = r_k v_k$ is not nonnegative. If $R_f \gg 0$ and $v > 0$, then $R_f \cdot v > 0$ since $r_1, \ldots, r_n > 0$ at least one of v_i is positive. Conversely, if some $r_k \leq 0$, then put $v_k = 1$ and $v_i = 0$ for $i \neq k$ so that $R_f \cdot v = r_k v_k$ is not strictly positive. □

Lemma 190. *Let $p, s \in \mathbb{R}^n$ where p is strongly positive. Then there exists a $C > 0$ such that*

$$\left| \frac{\langle v, s \rangle}{\langle v, p \rangle} \right| < C$$

for all positive vectors $v \in \mathbb{R}^n$.

Proof. Let $E = \{v \in \mathbb{R}^n \mid \|v\| = 1 \text{ and } v > 0\}$, which is a compact set. By Theorem 41, the maps $v \mapsto \langle v, s \rangle$ and $v \mapsto \langle v, p \rangle$ are continuous. Additionally, $\langle v, p \rangle \neq 0$ for any $v \in E$ since $v > 0$ and $p \gg 0$. Therefore $f(v) = \langle v, s \rangle / \langle v, p \rangle$ is a continuous on E, and

since E is compact, $f(E)$ is compact. Choose a $C > 0$ such that $|f(v)| < C$ whenever $v \in E$. Then

$$|f(v)| = \left| \frac{\langle v/\|v\|, s \rangle}{\langle v/\|v\|, p \rangle} \right| < C$$

for all $v > 0$. □

Theorem 191. *[Exercise 9.25] Let $f : S \to \mathbb{R}$ be a strictly positive linear functional on a subspace S of $\mathbb{R}^n$. Then f has a strictly positive extension to $\mathbb{R}^n$.*

Proof. We split the proof into two cases. We write V for $\mathbb{R}^n$. Suppose that S contains a strictly positive vector v_1 and let $K = \ker(f)$. By Theorem 188, we can write $V = S^\perp \odot S = S^\perp \odot \langle R_f \rangle \odot K$ where $R_f \in K^\perp$ is nonzero. If $v \in \ker(f)$ then $f(v) = 0$, so K does not contain any strictly positive vectors. By the fact given in the text, $K^\perp = S^\perp \odot \langle R_f \rangle$ contains some strongly positive vector $z = s + aR_f$. Furthermore, since $a \langle R_f, v_1 \rangle = \langle z, v_1 \rangle > 0$, we must have $a \neq 0$. So $z \in K^\perp \setminus S^\perp$, and by Theorem 187 there exists a λ such that the linear functional $g(x) = \langle x, \lambda z \rangle$ extends f. But $v_1 > 0$ and $z \gg 0$, so

$$\lambda \langle v_1, z \rangle = g(v_1) = f(v_1) > 0$$

implies that $\lambda > 0$. Finally, applying Theorem 189 shows that g is strictly positive.

Now suppose that S does not contain any strictly positive vectors. Then $S^\perp$ contains a strongly positive vector v_1 by the fact given in the text. Choose a subspace T of $S^\perp$ such that $S^\perp = \langle v_1 \rangle \odot T$, and write $V = S^\perp \odot S = S^\perp \odot \langle w \rangle \odot K$ where $K = \ker(f)$ and $w \in K^\perp$. By Lemma 190 there exists a $C > 0$ such that $|\langle v, w \rangle / \langle v, v_1 \rangle| < C$ for all $v > 0$. Let $M = C\,|f(w)|$ and define $g : \mathbb{R}^n \to \mathbb{R}$ by

$$a_1 v_1 + t + aw + k \mapsto a_1 M + a f(w)$$

where $a_1, a \in \mathbb{R}$, $t \in T$ and $k \in K$. If $v = a_1 v_1 + t + aw + k$ is a positive vector then $a_1 = \langle v, v_1 \rangle > 0$ and $a = \langle v, w \rangle$, so

$$\left| \frac{a}{a_1} f(w) \right| < C\,|f(w)| = M.$$

Therefore

$$\begin{aligned} g(v) &= a_1 M + a f(w) \\ &= a_1 \left(M + \frac{a}{a_1} f(w) \right) \\ &> 0. \end{aligned}$$

This shows that g is a strictly positive extension of f. □

Theorem 192. *[Exercise 9.26] If V is a real inner product space, then we can define an inner product on its complexification $V^{\mathbb{C}}$ as follows:*

$$\langle u + vi, x + yi\rangle = \langle u, x\rangle + \langle v, y\rangle + (\langle v, x\rangle - \langle u, y\rangle)i.$$

Then

$$\|(u + vi)\|^2 = \|u\|^2 + \|v\|^2.$$

Proof. We can compute

$$\begin{aligned}\langle u + vi, u + vi\rangle &= \langle u, u\rangle + \langle v, v\rangle + (\langle v, u\rangle - \langle u, v\rangle)i \\ &= \|u\|^2 + \|v\|^2\end{aligned}$$

since $\langle v, u\rangle = \langle u, v\rangle$. □

Conjugate-linear Maps.

Definition 193. A function $\sigma : V \to W$ on complex vector spaces is **conjugate-linear** or **antilinear** if

$$\sigma(v + v') = \sigma v + \sigma v'$$

and

$$\sigma(rv) = \overline{r}\sigma v$$

for all $v, v' \in V$ and $r \in \mathbb{C}$.

Definition 194. Let V be a complex vector space. The **complex conjugate** $\overline{V}$ of V is the complex vector space of all formal complex conjugates of V, that is,

$$\overline{V} = \{\overline{v} \mid v \in V\},$$

equipped with addition and scalar multiplication defined by $\overline{v}+\overline{w} = \overline{v + w}$ and $r\overline{v} = \overline{\overline{r}v}$. If B is a basis for V, then $\overline{B} = \left\{\overline{b} \mid b \in B\right\}$ is easily seen to be a basis for $\overline{V}$.

Theorem 195. *If $\sigma : V \to W$ is a conjugate-linear map, then the corresponding map $\widetilde{\sigma} : \overline{V} \to W$ defined by $\overline{v} \mapsto \sigma(v)$ is a linear map.*

Proof. For all $\overline{v}, \overline{v'} \in \overline{V}$ and $r \in \mathbb{C}$, we have

$$\widetilde{\sigma}(\overline{v} + \overline{v'}) = \widetilde{\sigma}(\overline{v + v'}) = \sigma(v) + \sigma(v') = \widetilde{\sigma}(\overline{v}) + \widetilde{\sigma}(\overline{v'})$$

and

$$\widetilde{\sigma}(r\overline{v}) = \widetilde{\sigma}(\overline{\overline{r}v}) = \sigma(\overline{r}v) = r\sigma(v) = r\widetilde{\sigma}(v).$$

□

Theorem 196. *Let $\sigma : U \to V$ and $\tau : V \to W$ be conjugate-linear maps.*

(1) *$\tau\sigma : U \to W$ is linear.*
(2) *If $\ker(\tau) = \{0\}$ then τ is injective.*

(3) *If τ is a (conjugate) isomorphism and V is finite-dimensional, then $V \cong W$ in the usual sense.*

Proof. For (3), Theorem 195 shows that $\widetilde{\tau} : \overline{V} \to W$ is a linear isomorphism, and since $\dim(V) = \dim(\overline{V})$, we have $V \cong \overline{V} \cong W$. □

Chapter 10. Structure Theory for Normal Operators

Theorem 197. *Let V and W be inner product spaces.*

(1) *$\iota^* = \iota$, where $\iota : V \to V$ is the identity.*
(2) *$0^* = 0$, where 0 is the zero operator.*

Proof. For all $v \in V$ and $w \in W$,

$$\langle \iota v, w \rangle = \langle v, w \rangle = \langle v, \iota w \rangle$$

and

$$\langle 0v, w \rangle = 0 = \langle v, 0w \rangle .$$

□

Theorem 198. *Let V and W be finite-dimensional inner product spaces. For every $\sigma, \tau \in \mathcal{L}(V, W)$ and $r \in F$,*

(1) *$(\sigma + \tau)^* = \sigma^* + \tau^*$.*
(2) *$(r\tau)^* = \overline{r}\tau^*$.*
(3) *$\tau^{**} = \tau$.*
(4) *If $V = W$, then $(\sigma\tau)^* = \tau^*\sigma^*$.*
(5) *If τ is invertible, then $(\tau^{-1})^* = (\tau^*)^{-1}$.*
(6) *If $V = W$ and $p(x) \in \mathbb{R}[x]$, then $p(\tau)^* = p(\tau^*)$.*

Proof. For (1) and (2), we have for all $v \in V$, $w \in W$ and $r \in \mathbb{C}$,

$$\langle (\sigma + \tau)v, w \rangle = \langle \sigma v, w \rangle + \langle \tau v, w \rangle = \langle v, \sigma^* w \rangle + \langle v, \tau^* w \rangle = \langle v, (\sigma^* + \tau^*)w \rangle$$

and

$$\langle r\tau v, w \rangle = r \langle v, \tau^* w \rangle = \langle v, \overline{r}\tau^* w \rangle .$$

Part (3) follows from the conjugate symmetry of the inner product:

$$\langle \tau^* w, v \rangle = \overline{\langle v, \tau^* w \rangle} = \overline{\langle \tau v, w \rangle} = \langle w, \tau v \rangle .$$

For (4),

$$\langle \sigma\tau v, w \rangle = \langle \tau v, \sigma^* w \rangle = \langle v, \tau^*\sigma^* w \rangle$$

and (5) follows from (4) by setting $\sigma = \tau^{-1}$. Part (6) is clear from (1), (2) and (4). □

Theorem 199. *Let V be a finite-dimensional inner product space and let S and T be subspaces of V. If ρ is a projection onto S along T, then $(\rho_{S,T})^* = \rho_{T^\perp,S^\perp}$.*

Proof. We have $\rho^2 = \rho$ since ρ is a projection; by Theorem 198, $(\rho^*)^2 = \rho^*$ and ρ^* is a projection. But

$$\operatorname{im}(\rho^*) = \ker(\rho)^\perp = T^\perp \quad \text{and} \quad \ker(\rho^*) = \operatorname{im}(\rho)^\perp = S^\perp$$

by Theorem 10.3, so ρ^* is a projection onto $T^\perp$ along $S^\perp$. □

Remark 200. [Exercise 10.1] Let $\tau \in \mathcal{L}(U, V)$. If τ is surjective, find a formula for the right inverse of τ in terms of τ^*. If τ is injective, find a formula for the left inverse of τ in terms of τ^*.

If $\operatorname{im}(\tau) = V$, then $\operatorname{im}(\tau\tau^*) = V$ and $\ker(\tau\tau^*) = V^\perp = \{0\}$ so that $\tau\tau^*$ is invertible. Therefore

$$(\tau\tau^*)(\tau\tau^*)^{-1} = \iota$$

and $\tau^*(\tau\tau^*)^{-1}$ is a right inverse for τ. Similarly, if $\ker(\tau) = \{0\}$ then $\operatorname{im}(\tau^*\tau) = \{0\}^\perp = V$ and $\ker(\tau^*\tau) = \{0\}$ so that $\tau^*\tau$ is invertible. Therefore

$$(\tau^*\tau)^{-1}(\tau^*\tau) = \iota$$

and $(\tau^*\tau)^{-1}\tau^*$ is a left inverse for τ.

Theorem 201. *[Exercise 10.2] Let $\tau \in \mathcal{L}(V)$ where V is a complex vector space and let*

$$\tau_1 = \frac{1}{2}(\tau + \tau^*) \quad \text{and} \quad \tau_2 = \frac{1}{2i}(\tau - \tau^*).$$

Then τ_1 and τ_2 are self-adjoint, and

$$\tau = \tau_1 + i\tau_2 \quad \text{and} \quad \tau^* = \tau_1 - i\tau_2.$$

Furthermore, if $\tau = \sigma_1 + i\sigma_2$ where σ_1 and σ_2 are self-adjoint, then $\sigma_1 = \tau_1$ and $\sigma_2 = \tau_2$.

Proof. Computing adjoints gives

$$\tau_1^* = \frac{1}{2}(\tau^* + \tau) = \tau_1 \quad \text{and} \quad \tau_2^* = -\frac{1}{2i}(\tau^* - \tau) = \tau_2,$$

and the last statement follows. If $\tau = \sigma_1 + i\sigma_2$ where σ_1 and σ_2 are self-adjoint, then $\tau^* = \sigma_1 - i\sigma_2$ and we have

$$\sigma_1 + i\sigma_2 = \tau_1 + i\tau_2 \quad \text{and} \quad \sigma_1 - i\sigma_2 = \tau_1 - i\tau_2.$$

Adding the equations gives $\sigma_1 = \tau_1$ and $\sigma_2 = \tau_2$. □

Theorem 202. *[Exercise 10.3] All of the roots of the characteristic polynomial of a skew-Hermitian matrix are pure imaginary.*

Proof. If λ is an eigenvalue of a skew-Hermitian matrix A, then $Av = \lambda v$ implies that $-Av = A^*v = \overline{\lambda}v$ and $(\lambda + \overline{\lambda})v = 0$. Since $v \neq 0$, we have $\lambda + \overline{\lambda} = 2\,\mathrm{Re}(\lambda) = 0$, i.e. λ is pure imaginary. □

Example 203. [Exercise 10.4] Give an example of a normal operator that is neither self-adjoint nor unitary.

The matrix

$$A = \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix}$$

is skew-Hermitian since

$$A^* = \begin{bmatrix} 0 & -i \\ -i & 0 \end{bmatrix} = -A.$$

Theorem 204. *[Exercise 10.5] Let V be a complex inner product space. If $\|\tau v\| = \|\tau^* v\|$ for all $v \in V$, then τ is normal.*

Proof. For all $v \in V$,

$$\langle \tau^*\tau v, v\rangle = \langle \tau v, \tau v\rangle = \langle \tau^* v, \tau^* v\rangle = \langle \tau\tau^* v, v\rangle$$

and $\tau^*\tau = \tau\tau^*$ by Theorem 9.2. □

Theorem 205. *[Exercise 10.6] Let τ be a normal operator on a complex finite-dimensional inner product space V or a self-adjoint operator on a real finite-dimensional inner product space.*

(1) *$\tau^* = p(\tau)$ for some polynomial $p(x) \in \mathbb{C}[x]$.*
(2) *For any $\sigma \in \mathcal{L}(V)$, $\sigma\tau = \tau\sigma$ implies $\sigma\tau^* = \tau^*\sigma$. In other words, τ^* commutes with all operators that commute with τ.*

Proof. We can assume that V is a complex inner product space, for if τ is self-adjoint then the statements follow trivially. Part (1) follows from the spectral resolution

$$\tau = \lambda_1\rho_1 + \cdots + \lambda_k p_k$$

since

$$\tau^* = \overline{\lambda_1}\rho_1 + \cdots + \overline{\lambda_k}\rho_k = p(\tau)$$

where $p(x) \in \mathbb{C}[x]$ is a polynomial such that $p(\lambda_i) = \overline{\lambda_i}$ for all i. For (2), if $\sigma\tau = \tau\sigma$ then σ commutes with any polynomial in τ; since (1) shows that $\tau^* = p(\tau)$ for some $p(x) \in \mathbb{C}[x]$, σ commutes with τ^*. □

Theorem 206. *A pair of linear operators $\sigma, \tau \in \mathcal{L}(V)$ is **simultaneously unitarily diagonalizable** if there is an orthonormal basis $\mathcal{O}$ of V for which $[\tau]_{\mathcal{O}}$ and $[\sigma]_{\mathcal{O}}$ are both diagonal, that is, $\mathcal{O}$ is an basis of orthonormal eigenvectors for both τ and σ. Two normal operators σ and τ are simultaneously unitarily diagonalizable if and only if they commute, that is, $\sigma\tau = \tau\sigma$ (cf. Theorem 156).*

Proof. If τ and σ are simultaneously unitarily diagonalizable, then $[\tau]_{\mathcal{O}} = D_1$ and $[\sigma]_{\mathcal{O}} = D_2$ where D_1 and D_2 are diagonal matrices. We have

$$[\tau\sigma]_{\mathcal{O}} = D_1 D_2 = D_2 D_1 = [\sigma\tau]_{\mathcal{O}},$$

which shows that τ and σ commute. Conversely, suppose that $\sigma\tau = \tau\sigma$ and write

$$V = \mathcal{E}_{\lambda_1} \odot \cdots \odot \mathcal{E}_{\lambda_k}$$

where $\lambda_1, \ldots, \lambda_k$ are the distinct eigenvalues of τ. If $v \in \mathcal{E}_{\lambda_i}$,

$$\tau(\sigma v) = \sigma\tau v = \lambda_i(\sigma v)$$

so $\sigma v \in \mathcal{E}_{\lambda_i}$. This shows that $(\mathcal{E}_{\lambda_1}, \ldots, \mathcal{E}_{\lambda_k})$ reduces σ. But each $\sigma|_{\mathcal{E}_{\lambda_i}}$ is diagonalizable by Lemma 155, so there exists an orthonormal basis $\mathcal{O}_i$ of $\mathcal{E}_{\lambda_i}$ such that $\left[\sigma|_{\mathcal{E}_{\lambda_i}}\right]_{\mathcal{O}_i}$ is diagonal. Let $\mathcal{O} = \mathcal{O}_1 \cup \cdots \cup \mathcal{O}_k$; then $[\sigma]_{\mathcal{O}}$ is diagonal and $[\tau]_{\mathcal{O}}$ is also diagonal. □

Theorem 207. *[Exercise 10.7] A linear operator τ on a finite-dimensional complex inner product space V is normal if and only if whenever S is an invariant subspace under τ, so is $S^{\perp}$ (cf. Theorem 165).*

Proof. Suppose that τ is normal and write

$$V = \mathcal{E}_{\lambda_1} \odot \cdots \odot \mathcal{E}_{\lambda_k}$$

where $\lambda_1, \ldots, \lambda_k$ are distinct eigenvalues. If S is a τ-invariant subspace, then $\tau|_S$ is diagonalizable by Theorem 155 and we can write

$$S = \widetilde{\mathcal{E}}_{\lambda_1} \odot \cdots \odot \widetilde{\mathcal{E}}_{\lambda_k}$$

where each $\widetilde{\mathcal{E}}_{\lambda_i} \subseteq \mathcal{E}_{\lambda_i}$ may be trivial and the sum is orthogonal by Theorem 10.8. For each i, choose a subspace $\mathcal{D}_{\lambda_i}$ such that $\mathcal{E}_{\lambda_i} = \widetilde{\mathcal{E}}_{\lambda_i} \odot \mathcal{D}_{\lambda_i}$; then

$$S^{\perp} = \mathcal{D}_{\lambda_1} \odot \cdots \odot \mathcal{D}_{\lambda_k}$$

is τ-invariant since any subspace of an eigenspace is τ-invariant. To prove the converse, let v_1 be an eigenvector of τ (which exists since $\mathbb{C}$ is algebraically closed). Since $\langle v_1 \rangle$ is τ-invariant, the orthogonal complement $\langle v \rangle^{\perp}$ is also τ-invariant. Choosing an eigenvector for $\tau|_{\langle v \rangle^{\perp}}$ and continuing the process gives a decomposition

$$V = \langle v_1 \rangle \odot \cdots \odot \langle v_n \rangle$$

where each $\langle v_i \rangle$ is an invariant subspace, and therefore each v_i is an eigenvector. By Theorem 10.9, τ is normal. □

Corollary 208. *Let $\tau \in \mathcal{L}(V)$ be a normal operator. If S is τ-invariant then $(S, S^{\perp})$ reduces τ, and $\tau|_S$ is normal.*

Theorem 209. *[Exercise 10.8] Let V be a finite-dimensional inner product space and let τ be a normal operator on V.*

(1) *If τ is idempotent, then it is also self-adjoint.*
(2) *If τ is nilpotent, then $\tau = 0$.*
(3) *If $\tau^2 = \tau^3$, then τ is idempotent.*

Proof. If $\tau^2 = \tau$ then τ is a projection, and since τ is normal, $\ker(\tau) = \ker(\tau^*) = \mathrm{im}(\tau)^\perp$. Therefore τ is an orthogonal projection and is self-adjoint by Theorem 10.5, which proves (1). Part (2) follows from the fact that a diagonalizable nilpotent matrix must be zero (see Theorem 138). For (3), if $\tau^2 = \tau^3$ then the minimal polynomial of τ divides $x^3 - x^2 = x^2(x-1)$. Since τ is normal, $m_\tau(x)$ must divide $x(x-1)$. If $m_\tau(x) = x$ then $\tau = 0$, if $m_\tau(x) = x - 1$ then $\tau = \iota$, and if $m_\tau(x) = x(x-1)$ then $\tau^2 = \tau$. In all three cases, τ is idempotent. □

Theorem 210. *[Exercise 10.9] If τ is a normal operator on a finite-dimensional complex inner product space, then the algebraic multiplicity is equal to the geometric multiplicity for all eigenvalues of τ.*

Proof. This is true for any diagonalizable operator. □

Theorem 211. *[Exercise 10.10] Two orthogonal projections $\sigma, \rho \in \mathcal{L}(V)$ are orthogonal to each other if and only if* $\mathrm{im}(\sigma) \perp \mathrm{im}(\rho)$.

Proof. Since σ and ρ are orthogonal projections, they are self-adjoint. Suppose that $\sigma\rho = \rho\sigma = 0$. If $v \in \mathrm{im}(\sigma)$ and $w \in \mathrm{im}(\rho)$, then $v = \sigma x$ and $w = \rho y$ for some $x, y \in V$. We have

$$\langle v, w\rangle = \langle \sigma x, \rho y\rangle = \langle x, \sigma\rho y\rangle = 0,$$

which shows that $\mathrm{im}(\sigma) \perp \mathrm{im}(\rho)$. Conversely, if $\mathrm{im}(\sigma) \perp \mathrm{im}(\rho)$ and $v \in V$,

$$\langle \sigma\rho v, \sigma\rho v\rangle = \left\langle \rho v, \sigma^2 \rho v\right\rangle = 0$$

so $\sigma\rho = 0$. Similarly, $\rho\sigma = 0$. □

Theorem 212. *[Exercise 10.11] Let τ be a normal operator and let σ be any operator on V. If the eigenspaces of τ are σ-invariant, then τ and σ commute.*

Proof. Write $\tau = \lambda_1\rho_1 + \cdots + \lambda_k\rho_k$ where each ρ_i is the orthogonal projection onto the eigenspace $\mathcal{E}_{\lambda_i}$. Since each eigenspace is σ-invariant, σ commutes with every ρ_i by Theorem 2.24. Therefore σ commutes with τ. □

Theorem 213. *[Exercise 10.12] If τ and σ are normal operators on a finite-dimensional complex inner product space and if $\tau\theta = \theta\sigma$ for some operator θ then $\tau^*\theta = \theta\sigma^*$.*

Proof. Write

$$V = \mathcal{E}_{\lambda_1} \odot \cdots \odot \mathcal{E}_{\lambda_k}$$

where $\lambda_1, \ldots, \lambda_k$ are the distinct eigenvalues of σ. If $v \in \mathcal{E}_{\lambda_i}$ then $\tau\theta v = \theta\sigma v = \lambda_i \theta v$, so $\theta(\mathcal{E}_{\lambda_i})$ is an eigenspace of τ with eigenvalue λ_i. By Theorem 10.8,

$$\sigma v = \lambda_i v \Leftrightarrow \sigma^* v = \overline{\lambda_i} v \quad \text{and} \quad \tau\theta v = \lambda_i \theta v \Leftrightarrow \tau^* \theta v = \overline{\lambda_i}\theta v.$$

Then for all $v = v_1 + \cdots + v_k$ where $v_i \in \mathcal{E}_{\lambda_i}$ we have

$$\begin{aligned} \tau^*\theta v &= \tau^*\theta v_1 + \cdots + \tau^*\theta v_k \\ &= \overline{\lambda_i}\theta v_1 + \cdots + \overline{\lambda_k}\theta v_k \\ &= \theta\sigma^* v_1 + \cdots + \theta\sigma^* v_k \\ &= \theta\sigma^* v. \end{aligned}$$

□

Theorem 214. *[Exercise 10.13] If two normal $n \times n$ complex matrices are similar, then they are unitarily similar, that is, similar via a unitary matrix.*

Proof. If A and B are similar normal matrices, then $A = PDP^*$ and $B = QDQ^*$ where P and Q are unitary and D is diagonal. Therefore $A = PQ^*DQP^*$ where PQ^* is unitary. □

Theorem 215. *[Exercise 10.14] If ν is a unitary operator on a complex inner product space, then there exists a self-adjoint operator σ for which $\nu = e^{i\sigma}$.*

Proof. Since ν is unitary, it has a spectral resolution $\nu = \lambda_1\rho_1 + \cdots + \lambda_k\rho_k$ where $|\lambda_j| = 1$ for all j. Write $\lambda_j = e^{i\theta_j}$; then

$$\sigma = \theta_1\rho_1 + \cdots + \theta_k\rho_k$$

is self-adjoint since every θ_j is real, and $\nu = e^{i\sigma}$. □

Theorem 216. *[Exercise 10.15] A positive operator has a unique positive square root.*

Proof. Let τ be a positive operator; we already know that τ has a positive square root. If $\sigma = \lambda_1\rho_1 + \cdots + \lambda_k\rho_k$ is a positive square root of τ, then

$$\tau = \sigma^2 = \lambda_1^2\rho_1 + \cdots + \lambda_k^2\rho_k$$

is the spectral resolution of τ. Since this expression is unique and all eigenvalues of σ are nonnegative, the λ_i are uniquely determined. □

Theorem 217. *[Exercise 10.16] If τ has a square root, that is, if $\tau = \sigma^2$ for some positive operator σ, then τ is positive.*

Proof. Since σ is normal with nonnegative eigenvalues, τ is also normal with nonnegative eigenvalues. Theorem 10.22 then shows that τ is positive. □

Theorem 218. *[Exercise 10.17] If $\sigma \leq \tau$ (that is, $\tau - \sigma$ is positive) and if θ is a positive operator that commutes with both σ and τ, then $\sigma\theta \leq \tau\theta$.*

Proof. This is equivalent to proving that the product of two commuting positive operators σ, τ is positive. By Theorem 206, there exists an orthonormal basis $\mathcal{O}$ such that $[\sigma]_{\mathcal{O}}$ and $[\tau]_{\mathcal{O}}$ are both diagonal. Furthermore, since σ and τ are positive, both matrices have nonnegative diagonal entries. Therefore the product $[\sigma\tau]_{\mathcal{O}}$ also has nonnegative diagonal entries, and Theorem 10.22 shows that $\sigma\tau$ is positive. □

Theorem 219. *[Exercise 10.18] An invertible linear operator $\tau \in \mathcal{L}(V)$ is positive if and only if it has the form $\tau = \rho^*\rho$ where ρ is upper triangularizable. Moreover, ρ can be chosen with positive eigenvalues, in which case the factorization is unique.*

Proof. If $\tau = \rho^*\rho$ then τ is positive by Theorem 10.23. Conversely, if τ is positive and invertible then it must be positive-definite. Therefore $\tau = (\sqrt{\tau})^*\sqrt{\tau}$ is a decomposition where $\sqrt{\tau}$ is positive-definite (and has positive eigenvalues). □

Remark 220. [Exercise 10.19] Does every self-adjoint operator on a finite-dimensional real inner product space have a square root?

No, take the linear operator $\tau : \mathbb{R} \to \mathbb{R}$ given by $x \mapsto -x$.

Theorem 221. *Let A be an $m \times n$ complex matrix. Let $a_1, \ldots, a_n$ be the columns of A and let $\widetilde{a}_1, \ldots, \widetilde{a}_m$ be the rows of A. Then the diagonal entries of A^*A are $\|a_1\|^2, \ldots, \|a_n\|^2$ and the diagonal entries of AA^* are $\|\widetilde{a}_1\|^2, \ldots, \|\widetilde{a}_m\|^2$.*

Proof. For every k,

$$(A^*A)_{k,k} = \sum_{i=1}^{n} (A^*)_{k,i} A_{i,k} = \sum_{i=1}^{n} \overline{A_{i,k}} A_{i,k} = \|a_k\|^2$$

and

$$(AA^*)_{k,k} = \sum_{i=1}^{m} A_{k,i} (A^*)_{i,k} = \sum_{i=1}^{m} A_{k,i} \overline{A_{k,i}} = \|\widetilde{a}_k\|^2 .$$

□

Theorem 222. *Let V be a finite-dimensional inner product space and let $\tau \in \mathcal{L}(V)$. If $\operatorname{tr}(\tau^*\tau) = \operatorname{tr}(\tau\tau^*) = 0$, then $\tau = 0$.*

Proof. This follows by choosing an orthonormal basis for V and applying Theorem 221. Alternatively, $\tau^*\tau$ is a positive operator by Theorem 10.23, so all eigenvalues are nonnegative. Since $\operatorname{tr}(\tau^*\tau) = 0$, all eigenvalues are 0 and $\tau^*\tau = 0$. But $\ker(\tau) = \ker(\tau^*\tau) = V$, so $\tau = 0$. □

Theorem 223. *[Exercise 10.20] Let τ be a linear operator on $\mathbb{C}^n$ and let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of τ, each one written a number of times equal to its algebraic multiplicity. Then*

$$\sum_i |\lambda_i|^2 \leq \operatorname{tr}(\tau^*\tau).$$

Equality holds if and only if τ is normal.

Proof. We can unitarily triangularize τ to prove the first statement. Let $\mathcal{O}$ be an orthonormal basis such that $T = [\tau]_{\mathcal{O}}$ is upper triangular with columns $t_1, \ldots, t_n$. Then by Theorem 221,

$$\operatorname{tr}(\tau^*\tau) = \operatorname{tr}([\tau^*]_{\mathcal{O}}\,[\tau]_{\mathcal{O}}) = \sum_i \|t_i\|^2 \geq \sum_i |\lambda_i|^2$$

since the diagonal entries of $[\tau]_{\mathcal{O}}$ are the eigenvalues of τ. If equality holds then

$$\sum_i |\lambda_i|^2 = \sum_i \|t_i\|^2 = \sum_{i,j} |T_{i,j}|^2 = \sum_i |\lambda_i|^2 + \sum_{i \neq j} |T_{i,j}|^2,$$

so $T_{i,j} = 0$ for all $i \neq j$. Therefore T is diagonal and τ is normal. Conversely, if τ is normal then equality follows by applying Theorem 221 to any unitary diagonalization of τ. □

Theorem 224. *[Exercise 10.21] Let $\tau \in \mathcal{L}(V)$ where V is a real inner product space. Then the Hilbert space adjoint satisfies $(\tau^*)^{\mathbb{C}} = (\tau^{\mathbb{C}})^*$.*

Proof. For all $a, b, c, d \in \mathbb{R}$,

$$\begin{aligned}
\left\langle \tau^{\mathbb{C}}(a+ib), c+id \right\rangle &= \langle \tau a + i\tau b, c+id\rangle \\
&= \langle \tau a, c\rangle + \langle \tau b, d\rangle + (\langle \tau b, c\rangle - \langle \tau a, d\rangle)i \\
&= \langle a, \tau^* c\rangle + \langle b, \tau^* d\rangle + (\langle b, \tau^* c\rangle - \langle a, \tau^* d\rangle)i \\
&= \langle a+ib, \tau^* c\rangle + \langle b - ia, \tau^* d\rangle \\
&= \langle a+ib, \tau^* c\rangle + \langle a+ib, i\tau^* d\rangle \\
&= \left\langle a+ib, \tau^*(c+id)\right\rangle.
\end{aligned}$$

□

Chapter 11. Metric Vector Spaces: The Theory of Bilinear Forms

Note: this section is incomplete.

Theorem 225. *A metric vector space V is nonsingular if and only if all representing matrices M_B are nonsingular.*

Proof. Suppose that V is nonsingular. If $M_B[v]_B = 0$ for some v then $\langle x, v\rangle = [x]_B^T M_B[v]_B = 0$ for all $x \in V$, so $v = 0$ since $\operatorname{rad}(V) = \{0\}$. Therefore M_B is invertible. Conversely, suppose that all representing matrices M_B are invertible and let $v \in \operatorname{rad}(V)$. Then $\langle x, v\rangle = [x]_B^T M_B[v]_B = 0$ for all $x \in V$, and in particular, $[e_i]_B^T M_B[v]_B$ for all i. Therefore $M_B[v]_B = 0$ and $v = 0$ since M_B is invertible. □

Theorem 226. *Let $\tau \in \mathcal{L}(V, W)$ be a linear map between finite-dimensional metric vector spaces V and W.*

(1) *Let $B = \{v_1, \ldots, v_n\}$ be a basis for V. Then τ is an isometry if and only if τ is bijective and*
$$\langle \tau v_i, \tau v_j\rangle = \langle v_i, v_j\rangle$$
for all i, j.

(2) *If V is orthogonal and $\operatorname{char}(F) \neq 2$, then τ is an isometry if and only if it is bijective and*
$$\langle \tau v, \tau v\rangle = \langle v, v\rangle$$
for all $v \in V$.

(3) *Suppose that $\tau : V \to W$ is an isometry and*
$$V = S \odot S^\perp \quad \text{and} \quad W = T \odot T^\perp.$$
If $\tau S = T$, then $\tau(S^\perp) = T^\perp$.

Proof. Let $u, v \in V$ and write $u = a_1 v_1 + \cdots + a_n v_n$ and $v = b_1 v_1 + \cdots + b_n v_n$. Then
$$\begin{aligned}
\langle \tau u, \tau v\rangle &= \left\langle \tau \sum_{i=1}^n a_i v_i, \tau \sum_{j=1}^n b_j v_j \right\rangle \\
&= \sum_{i=1}^n \sum_{j=1}^n a_i b_j \langle \tau v_i, \tau v_j\rangle \\
&= \sum_{i=1}^n \sum_{j=1}^n a_i b_j \langle v_i, v_j\rangle \\
&= \langle u, v\rangle ,
\end{aligned}$$
which proves (1). Part (2) follows from the identity
$$\langle u, v\rangle = \frac{1}{2}\langle u + v, u + v\rangle - \frac{1}{2}\langle u, u\rangle - \frac{1}{2}\langle v, v\rangle .$$
For (3), let $v \in \tau(S^\perp)$ and $t \in T$. Then $v = \tau x$ for some $x \in S^\perp$ and $t = \tau y$ for some $y \in S$ since $\tau S = T$. Therefore
$$\langle v, t\rangle = \langle \tau x, \tau y\rangle = \langle x, y\rangle = 0$$
and $\tau(S^\perp) \subseteq T^\perp$. But $\dim(\tau(S^\perp)) = \dim(S^\perp) = \dim(T^\perp)$, so $\tau(S^\perp) = T^\perp$. □

Chapter 12. Metric Spaces

Theorem 227. *[Exercise 12.1,12.2]*

(1) *If* $x_1, \ldots, x_n \in M$ *then* $d(x_1, x_n) \leq d(x_1, x_2) + \cdots + d(x_{n-1}, x_n)$.
(2) *If* $x, y, a, b \in M$ *then* $|d(x, y) - d(a, b)| \leq d(x, a) + d(y, b)$.
(3) *If* $x, y, z \in M$ *then* $|d(x, z) - d(y, z)| \leq d(x, y)$.

Proof. We use induction on n to prove (1). If $n = 2$ then the result is clear. Otherwise, assume that the result holds for $n - 1$ elements of M. Then

$$\begin{aligned} d(x_1, x_n) &\leq d(x_1, x_{n-1}) + d(x_{n-1}, x_n) \\ &\leq d(x_1, x_2) + \cdots + d(x_{n-1}, x_n). \end{aligned}$$

For (2),

$$\begin{aligned} d(x, y) &\leq d(x, a) + d(a, b) + d(y, b), \\ d(a, b) &\leq d(x, a) + d(x, y) + d(y, b). \end{aligned}$$

Similarly, for (3),

$$\begin{aligned} d(x, z) &\leq d(x, y) + d(y, z), \\ d(y, z) &\leq d(x, y) + d(x, z). \end{aligned}$$

□

Example 228. [Exercise 12.3] Let $S \subseteq \ell^\infty$ be the subspace of all binary sequences (sequences of 0s and 1s). Describe the metric on S.

The ℓ^∞ metric on S is identical to the discrete metric.

Theorem 229. *[Exercise 12.5] Let* $1 \leq p < \infty$.

(1) *If* $x = (x_n) \in \ell^p$ *then* $x_n \to 0$.
(2) *There exists a sequence that converges to* 0 *but is not an element of any* ℓ^p *for* $1 \leq p < \infty$.

Proof. If $(x_n) \in \ell^p$ then

$$\sum_{n=1}^{\infty} |x_n|^p < \infty$$

by definition. Therefore $|x_n|^p \to 0$ and $x_n \to 0$. For (2), choose the sequence $x = (1/\log n)_{n \geq 2}$. Then

$$\sum_{n=2}^{\infty} \frac{1}{(\log n)^p}$$

never converges for any $1 \leq p < \infty$, so $x \notin \ell^p$. □

Theorem 230. *[Exercise 12.6]*

(1) *If $x = (x_n) \in \ell^p$, then $x \in \ell^q$ for all $q > p$.*
(2) *There exists a sequence $x = (x_n)$ that is in ℓ^p for $p > 1$, but is not in ℓ^1.*

Proof. If $(x_n) \in \ell^p$ then

$$\sum_{n=1}^{\infty} |x_n|^p < \infty.$$

For some integer N, we have $|x_n|^p < 1$ for all $n \geq N$. For these n we have $|x_n|^q \leq |x_n|^p$ since $q > p$, and therefore

$$\sum_{n=1}^{\infty} |x_n|^q < \infty$$

by the comparison test. For (2), take $x_n = 1/n$. □

Theorem 231. *[Exercise 12.7] A subset S of a metric space M is open if and only if S contains an open neighborhood of each of its points.*

Proof. If S is open, then S clearly contains an open neighborhood (in fact an open ball) around each of its points. Conversely, if every point $x \in S$ has an open neighborhood N in S, then we may choose an open ball around x since N is open. □

Theorem 232. *The union of any collection of open sets in a metric space is open.*

Proof. Let $\{E_i\}$ be a collection of open sets and let $E = \bigcup_i E_i$. If $x \in E$ then $x \in E_i$ for some i. Since E_i is open, there exists an open ball $B \subseteq E_i$ around x. But $B \subseteq E$, and this proves that E is open. □

Theorem 233. *[Exercise 12.8] The intersection of any collection of closed sets in a metric space is closed.*

Proof. This follows from Theorem 232 and the fact that

$$\left(\bigcap_i E_i \right)^c = \bigcup_i E_i^c.$$

□

Theorem 234. *[Exercise 12.9] Let (M, d) be a metric space. The **diameter** of a nonempty subset $S \subseteq M$ is*

$$\delta(S) = \sup_{x,y \in S} d(x, y).$$

*A set S is **bounded** if $\delta(S) < \infty$.*

(1) *S is bounded if and only if there is some $x \in M$ and $r \in \mathbb{R}$ for which $S \subseteq B(x, r)$.*

(2) $\delta(S) = 0$ *if and only if* S *consists of a single point.*
(3) $S \subseteq T$ *implies* $\delta(S) \leq \delta(T)$.
(4) *If* S *and* T *are bounded, then* $S \cup T$ *is also bounded.*

Proof. Suppose that S is bounded and choose any $p \in S$. Then $S \subseteq B(p, \delta(S) + 1)$ since any $x \in S$ has $d(p, x) \leq \delta(S)$. Conversely, if $S \subseteq B(p, r)$ for some $p \in S$ then

$$d(x, y) \leq d(x, p) + d(p, y) < 2r$$

for every $x, y \in S$. Therefore $\sup_{x,y \in S} d(x, y) \leq 2r$, and in particular, it is finite. This proves (1). For (2), if S contains exactly one point x then $\delta(S) = d(x, x) = 0$. Conversely, if $x, y \in S$ are distinct points then $0 < d(x, y) \leq \delta(S)$. For (3), we have $d(x, y) \leq \delta(T)$ for every $x, y \in S$, so $\delta(S) \leq \delta(T)$. For (4), suppose that S and T are bounded. Choose any two points $p \in S$ and $q \in T$. Let $x, y \in S \cup T$. If $x, y \in S$ then $d(x, y) \leq \delta(S)$ and if $x, y \in T$ then $d(x, y) \leq \delta(T)$. If $x \in S$ and $y \in T$ then

$$d(x, y) \leq d(x, p) + d(p, q) + d(q, y) \leq \delta(S) + d(p, q) + \delta(T),$$

and the same inequality holds if $x \in T$ and $x \in S$. In all cases the above inequality holds, so $S \cup T$ is bounded. □

Theorem 235. *[Exercise 12.10] Let* (M, d) *be a metric space. Let* d' *be the function defined by*

$$d'(x, y) = \frac{d(x, y)}{1 + d(x, y)}.$$

(1) (M, d') *is a metric space and* M *is bounded under this metric, even if it is not bounded under the metric* d.
(2) (M, d) *and* (M, d') *have the same open sets.*

Proof. We only prove the triangle inequality for d'. Let $x, y, z \in M$ and let $a = d(x, y)$, $b = d(x, z)$, $c = d(z, y)$ for convenience. Then

$$a \leq b + c$$
$$a \leq b + bc + c + cb$$
$$a + ab + ac \leq (b + ab + bc) + (c + ac + cb)$$

Adding abc to both sides and simplifying gives

$$a(1 + c)(1 + b) \leq b(1 + a)(1 + c) + c(1 + a)(1 + b)$$
$$\frac{a}{1 + a} \leq \frac{b}{1 + b} + \frac{c}{1 + c},$$

which proves that $d'(x, y) \leq d'(x, z) + d'(z, y)$. Since

$$\frac{d(x, y)}{1 + d(x, y)} = 1 - \frac{1}{1 + d(x, y)} \leq 1,$$

every subset of M is bounded. This proves (1). The metric d' is monotonically increasing as a function of d, so (2) follows. □

Theorem 236. *[Exercise 12.11] If S and T are subsets of a metric space (M, d), we define the **distance** between S and T by*

$$\rho(S,T) = \inf_{x\in S, t\in T} d(x,y).$$

(1) *$x \in \mathrm{cl}(S)$ if and only if $\rho(\{x\}, S) = 0$.*
(2) *(Is it true that $\rho(S,T) = 0$ if and only if $S = T$? Is ρ a metric?*

Proof. If $x \in \mathrm{cl}(S)$, then every open ball $B(x,\delta)$ contains a point of S. Therefore $\inf_{s\in S} d(x,s) \leq \delta$ for arbitrarily small δ, and $\rho(\{x\}, S) = 0$. Conversely, if $\rho(\{x\}, S) = 0$ then for every $\delta > 0$ there exists a $s \in S$ such that $d(x,s) < \delta$, so $x \in \mathrm{cl}(S)$. This proves (1). For (2), choose $S = \mathbb{Q}$ and $T = \mathbb{R} \setminus \mathbb{Q}$. Then $\rho(S,T) = 0$ but $S \neq T$ (in fact they are disjoint), so ρ cannot be a metric. □

Theorem 237. *[Exercise 12.12,12.13] Let (M, d) be a metric space and let $S \subseteq M$. The following are equivalent:*

(1) *$x \in M$ is a limit point of S.*
(2) *Every neighborhood of x meets S in a point other than x itself.*
(3) *Every open ball $B(x,r)$ contains infinitely many points of S.*

Proof. (1) $\Leftrightarrow$ (2) is obvious. Suppose that $x \in M$ is a limit point of S and let $B(x,r)$ be an open ball. Choose a sequence $\{x_i\}$ of points in $B(x,r) \cap S$ as follows: take x_1 to be any point in $B(x,r) \cap S$ not equal to x, which exists since x is a limit point of S. Having chosen the points $x_1, \ldots, x_k$, let

$$r' = \min_{1\leq i\leq k} d(x, x_i)$$

and choose some x_{k+1} in $B(x,r') \cap S$ not equal to x. Continuing this process produces an infinite sequence of distinct points in $B(x,r) \cap S$. This proves (1) $\Rightarrow$ (3). For the converse, suppose that x is not a limit point of S. Then there exists an open ball $B(x,r)$ such that $B(x,r) \cap S$ is empty or contains only x; this contradicts the condition in (3). Therefore (3) $\Rightarrow$ (1). □

Theorem 238. *[Exercise 12.14] Limits are unique. That is, $(x_n) \to x$ and $(x_n) \to y$ implies that $x = y$.*

Proof. Let $\varepsilon > 0$ and choose M, N such that $d(x, x_n) < \varepsilon$ whenever $n \geq M$ and $d(y, x_n) < \varepsilon$ whenever $n \geq N$. For $n = \max(M, N)$ we have

$$d(x,y) \leq d(x,x_n) + d(x_n,y) < 2\varepsilon;$$

since ε was arbitrary, $x = y$. □

Theorem 239. *[Exercise 12.15] Let S be a subset of a metric space M. Then $x \in \mathrm{cl}(S)$ if and only if there exists a sequence (x_n) in S that converges to x.*

Proof. Let $x \in \mathrm{cl}(S)$. If $x \in S$ then the constant sequence (x) converges to x. Otherwise, x is a limit point of S. For each integer $n \geq 1$, choose a point x_n from $B(x, 1/n) \cap S$. Then $(x_n) \to x$ since $d(x, x_n) < 1/n \to 0$ as $n \to \infty$. Conversely, suppose that there is a sequence (x_n) in S that converges to x. We may assume $x_n \neq x$ for all n, for otherwise $x \in S$ and we are done. If $B(x, r)$ is a neighborhood of x, then there exists an integer N such that $d(x, x_n) < r$ whenever $n \geq N$. In particular, $x_N \in B(x, r) \cap S$ and $x_N \neq x$ by our previous assumption. This proves that x is a limit point of S. □

Theorem 240. *[Exercise 12.16] The closure has the following properties:*

(1) $S \subseteq \mathrm{cl}(S)$.
(2) $\mathrm{cl}(\mathrm{cl}(S)) = \mathrm{cl}(S)$.
(3) $\mathrm{cl}(S \cup T) = \mathrm{cl}(S) \cup \mathrm{cl}(T)$.
(4) $\mathrm{cl}(S \cap T) \subseteq \mathrm{cl}(S) \cap \mathrm{cl}(T)$. *Equality does not necessarily hold.*

Proof. We only prove (3). The inclusion $\mathrm{cl}(S) \cup \mathrm{cl}(T) \subseteq \mathrm{cl}(S \cup T)$ is obvious. We have $S \cup T \subseteq \mathrm{cl}(S) \cup \mathrm{cl}(T)$, and since $\mathrm{cl}(S) \cup \mathrm{cl}(T)$ is closed, $\mathrm{cl}(S \cup T) \subseteq \mathrm{cl}(S) \cup \mathrm{cl}(T)$. This proves (3). Part (4) is similar. If we take $S = \{1/n \mid n \in \mathbb{Z}, n \geq 1\}$ and $T = \{-1/n \mid n \in \mathbb{Z}, n \geq 1\}$, then $\mathrm{cl}(S \cap T)$ is empty but $\mathrm{cl}(S) \cap \mathrm{cl}(T) = \{0\}$. □

Theorem 241. *[Exercise 12.17]*

(1) *The closed ball $\overline{B}(x, r)$ is always a closed subset.*
(2) *There exists a metric space in which the closure of an open ball $B(x, r)$ is not equal to the closed ball $\overline{B}(x, r)$.*

Proof. Let $E = \overline{B}(x, r)$. If $p \in E^c$ then $d(x, p) > r$. Choose some s with $d(x, p) > s > r$; then $B(p, d(x, p) - s)$ is an open ball around p that is contained in E^c. This shows that E is closed. For (2), consider $\mathbb{Z}$ with the discrete metric. We have $B(0, 1) = \{0\}$ and $\overline{B}(0, 1) = \mathbb{Z}$, but $\mathrm{cl}(B(0, 1)) = \{0\}$. □

Theorem 242. *[Exercise 12.20] A discrete metric space is separable if and only if it is countable.*

Proof. Let M be a discrete metric space. If $S \subseteq M$ then $\mathrm{cl}(S) = S$ since every subset of M is closed. Therefore, S is dense in M if and only if $S = \mathrm{cl}(S) = M$. The result follows easily. □

Theorem 243. *[Exercise 12.25] Any convergent sequence is a Cauchy sequence.*

Proof. Let (x_n) be a sequence that converges to some point x and let $\varepsilon > 0$. Choose an integer N such that $d(x, x_n) < \varepsilon/2$ for all $n \geq N$. Then for all $m, n \geq N$ we have

$$d(x_m, x_n) \leq d(x_m, x) + d(x, x_n) < \varepsilon.$$

□

Theorem 244. *[Exercise 12.26] If $(x_n) \to x$ in a metric space M, show that any subsequence (x_{n_k}) of (x_n) also converges to x.*

Proof. Obvious. □

Theorem 245. *[Exercise 12.27] If (x_n) is a Cauchy sequence in a metric space M and some subsequence (x_{n_k}) of (x_n) converges to $x \in M$, then (x_n) converges to x.*

Proof. Let $\varepsilon > 0$. Choose an integer M such that $d(x, x_{n_k}) < \varepsilon/2$ for all k such that $n_k \geq M$. Also choose an integer N such that $d(x_m, x_n) < \varepsilon/2$ whenever $m, n \geq N$. Then for all $n \geq \max(M, N)$, there exists a $n_k \geq n$, and

$$d(x, x_n) \leq d(x, x_{n_k}) + d(x_{n_k}, x_n) < \varepsilon.$$

□

Theorem 246. *[Exercise 12.28] If (x_n) is a Cauchy sequence, then the set $\{x_n\}$ is bounded. However, a bounded sequence is not necessarily a Cauchy sequence.*

Proof. Choose an integer N such that $d(x_m, x_n) < 1$ whenever $m, n \geq N$. In particular, $d(x_N, x_n) < 1$ for all $n > N$, so $\{x_N, x_{N+1}, \dots\}$ is bounded. Since $\{x_1, \dots, x_{N-1}\}$ is finite, it is also bounded. Therefore $\{x_n\}$ is bounded. The converse is not true, since the sequence defined by $x_n = (-1)^n$ is bounded but not Cauchy. □

Theorem 247. *[Exercise 12.29] Let (x_n) and (y_n) be Cauchy sequences in a metric space M. Then the sequence $d_n = d(x_n, y_n)$ converges.*

Proof. Let $\varepsilon > 0$. Choose integers M, N such that $d(x_m, x_n) < \varepsilon/2$ whenever $m, n \geq M$ and $d(y_m, y_n) < \varepsilon/2$ whenever $m, n \geq N$. Then for all $m, n \geq \max(M, N)$,

$$\begin{aligned} d_m - d_n &= d(x_m, y_m) - d(x_n, y_n) \\ &\leq d(x_m, x_n) + d(x_n, y_n) + d(y_n, y_m) - d(x_n, y_n) \\ &= d(x_m, x_n) + d(y_n, y_m) \\ &< \varepsilon, \end{aligned}$$

and a similar argument shows that $d_n - d_m < \varepsilon$, so $|d_m - d_n| < \varepsilon$. This proves that (d_n) is a Cauchy sequence. Since $\mathbb{R}$ is complete, (d_n) converges. □

Theorem 248. *[Exercise 12.36] The metric spaces $C[a, b]$ and $C[c, d]$, under the* sup *metric, are isometric.*

Proof. Take the isometry $\varphi : C[a,b] \to C[c,d]$ given by

$$(\varphi f)(x) = f\left(a + \frac{b-a}{d-c}(x-c)\right).$$

□

Theorem 249. *[Exercise 12.37] Let $p, q \geq 1$ with $p + q = pq$. If $\sum |x_n|^p$ and $\sum |y_n|^q$ converge, then $\sum |x_n y_n|$ converges and*

$$\sum_{n=1}^{\infty} |x_n y_n| \leq \left(\sum_{n=1}^{\infty} |x_n|^p\right)^{1/p} \left(\sum_{n=1}^{\infty} |y_n|^q\right)^{1/q}.$$

Proof. We first establish the inequality

$$(*) \qquad uv \leq \frac{u^p}{p} + \frac{v^q}{q}$$

for positive real numbers u and v. We have

$$\begin{aligned} uv &= (u^p)^{1/p}(v^q)^{1/q} \\ &= \exp\left(\frac{1}{p}\log u^p\right)\exp\left(\frac{1}{q}\log v^q\right) \\ &= \exp\left(\frac{1}{p}\log u^p + \frac{1}{q}\log v^q\right) \\ &\leq \frac{1}{p}\exp(\log u^p) + \frac{1}{q}\exp(\log v^q) \\ &= \frac{u^p}{p} + \frac{v^q}{q} \end{aligned}$$

since exp is convex and $1/p + 1/q = 1$. Now let

$$X = \sum_{n=1}^{\infty} |x_n|^p \quad \text{and} \quad Y = \sum_{n=1}^{\infty} |y_n|^q.$$

We can assume that $X, Y \neq 0$, for otherwise the result follows immediately. For each n, set

$$u = \frac{|x_n|}{X^{1/p}} \quad \text{and} \quad v = \frac{|y_n|}{Y^{1/q}}$$

in (*) to obtain

$$\frac{|x_n|\,|y_n|}{X^{1/p}Y^{1/q}} \leq \frac{1}{pX}|x_n|^p + \frac{1}{qY}|y_n|^q.$$

By the comparison test the series $\sum |x_n|\,|y_n|$ converges, and summing on n gives

$$\begin{aligned}\frac{1}{X^{1/p}Y^{1/q}}\sum_{n=1}^{\infty}|x_n|\,|y_n| &\le \frac{1}{pX}\sum_{n=1}^{\infty}|x_n|^p+\frac{1}{qY}\sum_{n=1}^{\infty}|y_n|^q\\ &=\frac{1}{p}+\frac{1}{q}\\ &=1.\end{aligned}$$

Therefore

$$\begin{aligned}\sum_{n=1}^{\infty}|x_n|\,|y_n| &\le X^{1/p}Y^{1/q}\\ &=\left(\sum_{n=1}^{\infty}|x_n|^p\right)^{1/p}\left(\sum_{n=1}^{\infty}|y_n|^q\right)^{1/q}.\end{aligned}$$

□

Theorem 250. *[Exercise 12.38] Let $p \ge 1$. If $\sum |x_n|^p$ and $\sum |y_n|^p$ converge, then $\sum |x_n + y_n|^p$ converges and*

$$\left(\sum_{n=1}^{\infty}|x_n+y_n|^p\right)^{1/p}\le\left(\sum_{n=1}^{\infty}|x_n|^p\right)^{1/p}+\left(\sum_{n=1}^{\infty}|y_n|^p\right)^{1/p}.$$

Proof. The case $p = 1$ is clear since $|x_n + y_n| \le |x_n| + |y_n|$ for all n. Now assume that $p > 1$. We have

$$\begin{aligned}|x_n+y_n|^p &= |x_n+y_n|\,|x_n+y_n|^{p-1}\\ &\le |x_n|\,|x_n+y_n|^{p-1}+|y_n|\,|x_n+y_n|^{p-1}\end{aligned}$$

for all n; summing from 1 to k gives

$$\sum_{n=1}^{k}|x_n+y_n|^p\le\sum_{n=1}^{k}|x_n|\,|x_n+y_n|^{p-1}+\sum_{n=1}^{k}|y_n|\,|x_n+y_n|^{p-1}.$$

Let $q = p/(p-1)$ so that $1/p + 1/q = 1$. Applying Theorem 249 to the finite sums on the right, we have

$$\begin{aligned}\sum_{n=1}^{k}|x_n|\,|x_n+y_n|^{p-1} &\le \left(\sum_{n=1}^{k}|x_n|^p\right)^{1/p}\left(\sum_{n=1}^{k}|x_n+y_n|^p\right)^{1/q},\\ \sum_{n=1}^{k}|y_n|\,|x_n+y_n|^{p-1} &\le \left(\sum_{n=1}^{k}|y_n|^p\right)^{1/p}\left(\sum_{n=1}^{k}|x_n+y_n|^p\right)^{1/q}\end{aligned}$$

and therefore

$$\left(\sum_{n=1}^{k}|x_n+y_n|^p\right)^{1/p}\leq\left(\sum_{n=1}^{k}|x_n|^p\right)^{1/p}+\left(\sum_{n=1}^{k}|y_n|^p\right)^{1/p}.$$

Since the left hand side is nonnegative and the right hand side converges as $k\to\infty$, the left hand side must also converge. Taking $k\to\infty$ gives

$$\left(\sum_{n=1}^{\infty}|x_n+y_n|^p\right)^{1/p}\leq\left(\sum_{n=1}^{\infty}|x_n|^p\right)^{1/p}+\left(\sum_{n=1}^{\infty}|y_n|^p\right)^{1/p}.$$

□

Chapter 13. Hilbert Spaces

Theorem 251. *Let V be an inner product space.*

(1) *If $(x_n)\to x$ then $\|x_n\|\to\|x\|$.*
(2) *If $(x_n)\to x$ and $(y_n)\to y$ then $\langle x_n,y_n\rangle\to\langle x,y\rangle$.*

Proof. For (1), $|\|x_n\|-\|x\||\leq\|x_n-x\|\to 0$ as $n\to\infty$. Part (2) follows from the polarization identities and (1). □

Theorem 252. *Let $\tau:H_1\to H_2$ be a bounded linear map.*

(1) $\|\tau\|=\sup\limits_{\|x\|=1}\|\tau x\|$.
(2) $\|\tau\|=\sup\limits_{\|x\|\leq 1}\|\tau x\|$.
(3) $\|\tau\|=\inf\{c\in\mathbb{R}\mid\|\tau x\|\leq c\|x\|\text{ for all }x\in H_1\}$.

Proof. By definition, $\|\tau v\|\leq\|\tau\|$ for all $\|v\|=1$, so $\sup_{\|x\|=1}\|\tau x\|\leq\|\tau\|$. For any $\|v\|\neq 0$ we have

$$\frac{\|\tau v\|}{\|v\|}=\left\|\tau\left(\frac{v}{\|v\|}\right)\right\|\leq\sup_{\|x\|=1}\|\tau x\|,$$

so $\|\tau\|\leq\sup_{\|x\|=1}\|\tau x\|$. The proof of (2) is similar. Let

$$M=\inf\{c\in\mathbb{R}\mid\|\tau x\|\leq c\|x\|\text{ for all }x\in H_1\}.$$

Since $\|\tau x\|\leq\|\tau\|\|x\|$ for all $x\in H_1$, we have $M\leq\|\tau\|$. But $\|\tau x\|\leq M\|x\|$ for all $x\neq 0$, so $\|\tau\|\leq M$. □

Theorem 253. *[Exercise 13.1] The* sup *metric on the metric space $C[a,b]$ of continuous functions on $[a,b]$ does not come from an inner product.*

Proof. Let $f(t) = 1$ and $g(t) = (t-a)/(b-a)$. Then

$$\begin{aligned} \|f+g\| + \|f-g\| &= 2+1 \\ &\neq 2(\|f\|^2 + \|g\|^2) = 2, \end{aligned}$$

so the norm does not satisfy the parallelogram law. □

Theorem 254. *[Exercise 13.3] Let V be an inner product space and let A and B be subsets of V.*

(1) $A \subseteq B \Rightarrow B^\perp \subseteq A^\perp$.
(2) $A^\perp$ *is a closed subspace of* V.
(3) $[\text{cspan}(A)]^\perp = A^\perp$.

Proof. (1) is proved in Theorem 178. Let x be a limit point of $A^\perp$; by Theorem 239, there exists a sequence (x_n) in $A^\perp$ that converges to x. If $a \in A$ then $\langle x_n, a\rangle = 0$ for every n, so $\langle x, a\rangle = \lim_{n\to\infty} \langle x_n, a\rangle = 0$. Therefore $x \in A^\perp$, and this proves (2). For (3), we have $[\text{cspan}(A)]^\perp \subseteq A^\perp$ by (1). Let $x \in A^\perp$. If $a \in \text{cspan}(A)$ then there exists a sequence (a_n) in $\text{span}(A)$ that converges to a. Since $\langle a_n, x\rangle = 0$ for every n, we have $\langle a, x\rangle = \lim_{n\to\infty} \langle a_n, x\rangle = 0$. Therefore $x \in [\text{cspan}(A)]^\perp$, which proves (3). □

Remark 255. [Exercise 13.4] Let V be an inner product space and $S \subseteq V$. Under what conditions is $S^{\perp\perp\perp} = S^\perp$?

Theorem 254 shows that $S^\perp$ is closed, so if V is a Hilbert space then $S^{\perp\perp\perp} = \text{cl}(S^\perp) = S^\perp$ by Theorem 13.13.

Theorem 256. *[Exercise 13.5] A subspace S of a Hilbert space H is closed if and only if $S = S^{\perp\perp}$.*

Proof. By Theorem 13.13, $\text{cl}(S) = S^{\perp\perp}$. Therefore $S = S^{\perp\perp}$ if and only if $S = \text{cl}(S)$, i.e. S is closed. □

Theorem 257. *[Exercise 13.7] Let $\mathcal{O} = \{u_1, u_2, \dots\}$ be an orthonormal set in H. If $x = \sum_k r_k u_k$ converges, then*

$$\|x\|^2 = \sum_{k=1}^{\infty} |r_k|^2 .$$

Proof. For all n,

$$\left\| \sum_{k=1}^{n} r_k u_k \right\|^2 = \sum_{k=1}^{n} |r_k|^2 .$$

Since $\|\cdot\|$ is continuous, letting $n \to \infty$ proves the result. □

Theorem 258. *[Exercise 13.8] If an infinite series*

$$\sum_{k=1}^{\infty} x_k$$

converges absolutely in a Hilbert space H, then it also converges in the sense of the "net" definition.

Proof. Suppose that $\sum_{k=1}^{\infty} x_k \to x$ and let $\varepsilon > 0$. Choose an integer N such that $|\sum_{k=1}^{n} x_k - x| < \varepsilon/2$ for every $n \geq N$. Since the sum converges absolutely, we may also choose an integer $N' \geq N$ such that $\sum_{k=N'}^{\infty} \|x_k\| < \varepsilon/2$. Let $K = \{1, 2, \ldots, N'\}$, let T be a finite set with $K \subset T \subset \mathbb{N}$, and let $m = \max T$. Then

$$\begin{aligned}
\left| \sum_{k \in T} x_k - x \right| &= \left| \sum_{k \in K} x_k - x + \sum_{k \in T \setminus K} x_k \right| \\
&\leq \left| \sum_{k \in K} x_k - x \right| + \sum_{k \in T \setminus K} \|x_k\| \\
&< \varepsilon.
\end{aligned}$$

□

Example 259. [Exercise 13.10] Find a countably infinite sum of real numbers that converges in the sense of partial sums, but not in the sense of nets.

Consider the sum

$$\sum_{k=1}^{\infty} \frac{(-1)^k}{k}.$$

If this sum converges in the "net" sense, then by Theorem 13.20 there exists a finite set $I \subseteq K$ such that

$$J \cap I = \emptyset,\ J \text{ finite} \Rightarrow \left| \sum_{k \in J} \frac{(-1)^k}{k} \right| \leq \varepsilon.$$

Let $m = \max I$ and let $J = \{m+1, m+3, \ldots, m+2n+1\}$. As $n \to \infty$ we must have a contradiction since the harmonic series diverges.

Theorem 260. *[Exercise 13.11] If a Hilbert space H has infinite Hilbert dimension, then no Hilbert basis for H is a Hamel basis.*

Proof. Choose some infinite sequence (v_n) from H and consider the element

$$v = \sum_{n=1}^{\infty} \frac{1}{n} v_n.$$

The sum converges since $\sum_{n=1}^{\infty} 1/n^2$ converges. But if H is a Hamel basis, then v can be written as a (finite) linear combination of elements from H, which contradicts the equation above. □

Theorem 261. *[Exercise 13.13] Any linear map between finite-dimensional Hilbert spaces is bounded.*

Proof. Let $\tau \in \mathcal{L}(V, W)$ and choose an orthonormal basis $B = \{v_1, \ldots, v_m\}$ for V. Let $M = \max\{\|\tau v_1\|, \ldots, \|\tau v_m\|\}$. If $\|x\| = 1$ then we can write $x = a_1 v_1 + \cdots + a_m v_m$ where $|a_1|, \ldots, |a_m| \leq 1$, so

$$\begin{aligned} \|\tau x\| &= \|a_1 \tau v_1 + \cdots + a_m \tau v_m\| \\ &\leq |a_1| \|\tau v_1\| + |a_m| \|\tau v_m\| \\ &\leq mM. \end{aligned}$$

This shows that τ is bounded. □

Theorem 262. *[Exercise 13.14] If $f \in H^*$, then* $\ker(f)$ *is a closed subspace of H. (Note that H^* denotes the set of all bounded linear functionals on H.)*

Proof. Since f is bounded, it is continuous. But $\ker(f) = f^{-1}(\{0\})$, which is closed since $\{0\}$ is closed. □

Theorem 263. *[Exercise 13.15] A Hilbert space H is separable if and only if* $\mathrm{hdim}(H) \leq \aleph_0$.

Proof. Let B be a Hilbert basis for H so that $\mathrm{cspan}(B) = H$, i.e. B is dense in H. It follows that H is separable if and only if $\mathrm{hdim}(H) = |B| \leq \aleph_0$. □

Remark 264. [Exercise 13.16] Can a Hilbert space have countably infinite Hamel dimension?

Suppose that a Hilbert space H with infinite Hilbert dimension has a countably infinite Hamel basis $B = \{v_1, v_2, \ldots\}$. By Theorem 9.11, there exists an orthogonal sequence $\mathcal{O} = \{u_1, u_2, \ldots\}$ with the property that $\langle u_1, \ldots, u_k \rangle = \langle v_1, \ldots, v_k \rangle$ for all $k > 0$. In particular, $\mathrm{span}(\mathcal{O}) = H$, so $\mathcal{O}$ is a Hilbert basis. This contradicts Theorem 260.

Theorem 265. *[Exercise 13.18] Let τ and σ be bounded linear operators on H.*

(1) $\|r\tau\| \leq |r| \|\tau\|$.
(2) $\|\tau + \sigma\| \leq \|\tau\| + \|\sigma\|$.
(3) $\|\tau\sigma\| \leq \|\tau\| \|\sigma\|$.

Proof. For all $\|x\| = 1$ we have

$$\|(r\tau)x\| = |r| \, \|\tau x\| \leq |r| \, \|\tau\| \,,$$
$$\|(\tau + \sigma)x\| \leq \|\tau x\| + \|\sigma x\| \leq \|\tau\| + \|\sigma\| \,,$$
$$\|\tau\sigma x\| \leq \|\tau\| \, \|\sigma x\| \leq \|\tau\| \, \|\sigma\| \,.$$

The results follow from the properties of sup and Theorem 252. □

Theorem 266. *[Exercise 13.19] There exists a conjugate isomorphism between H^* and H for any Hilbert space H. In particular, if H is a real Hilbert space then $H \cong H^*$.*

Proof. Let $R : H^* \to H$ be the Riesz map, which is injective by Theorem 13.32. Furthermore, for any $z_0 \in H$ the linear functional $x \mapsto \langle x, z_0 \rangle$ is bounded, so R is a bijection. It is easy to see that R is a conjugate-linear map. □

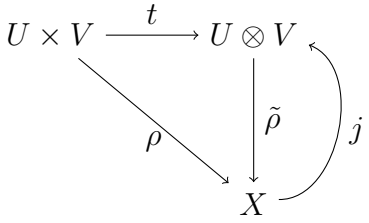
Chapter 14. Tensor Products

Definition 267. If $f : V_1 \times \cdots \times V_n \to W$ is a multilinear map, then we denote the subspace of W generated by $f(V_1 \times \cdots \times V_n)$ by $\mathrm{Im}(f)$.

Theorem 268. *Let U, V be vector spaces. A pair (T, t) is a tensor product of U with V if and only if the following two conditions hold:*

1. $T = \mathrm{Im}(t)$.
2. *If $f : U \times V \to W$ is a bilinear map then there exists a (not necessarily unique) linear map $\tau : T \to W$ such that $f = \tau t$.*

Proof. Suppose that the two conditions hold; we need to prove that if $\tau, \sigma : T \to W$ are two linear maps such that $f = \tau t$ and $f = \sigma t$, then $\tau = \sigma$. But $(\tau - \sigma)t = f - f = 0$, which implies that $\tau = \sigma$ since t is surjective (from the first condition). Now suppose that (T, t) is a tensor product of U with V. Condition (2) follows immediately, so it remains to prove (1). Write $X = \mathrm{Im}(t)$ and $U \otimes V = T$.



Define the bilinear map $\rho : U \times V \to X$ by $(u, v) \mapsto u \otimes v$; there exists a unique $\widetilde{\rho} : U \otimes V \to X$ such that $\widetilde{\rho}t = \rho$. Let $j : X \to U \otimes V$ be the inclusion map. Now $j\rho = t$, so $j\widetilde{\rho}t = j\rho = t$. But $\iota t = t$ where ι is the identity map on $U \otimes V$, so by the universal property (for bilinear maps from $U \times V$ to $U \otimes V$) we have $j\widetilde{\rho} = \iota$. This shows that j is surjective, i.e. $U \otimes V \subseteq X$. □

Theorem 269. *Let $V_1, \ldots, V_n, W_1, \ldots, W_m$ be vector spaces over a field F.*

(1) *There exists a unique isomorphism*

$$\tau : (V_1 \otimes \cdots \otimes V_n) \otimes (W_1 \otimes \cdots \otimes W_m) \to V_1 \otimes \cdots \otimes V_n \otimes W_1 \otimes \cdots \otimes W_m$$

for which

$$\tau[(v_1 \otimes \cdots \otimes v_n) \otimes (w_1 \otimes \cdots \otimes w_m)] = v_1 \otimes \cdots \otimes v_n \otimes w_1 \otimes \cdots \otimes w_m.$$

In particular,

$$(U \otimes V) \otimes W \cong U \otimes (V \otimes W) \cong U \otimes V \otimes W.$$

(2) *If π is a permutation of the indices $\{1, \ldots, n\}$, then there exists a unique isomorphism*

$$\sigma : V_1 \otimes \cdots \otimes V_n \to V_{\pi(1)} \otimes \cdots \otimes V_{\pi(n)}$$

for which

$$\sigma(v_1 \otimes \cdots \otimes v_n) = v_{\pi(1)} \otimes \cdots \otimes v_{\pi(n)}.$$

(3) *There exists a unique isomorphism $\rho : F \otimes V \to V$ for which*

$$\rho(a \otimes v) = av.$$

Similarly, there exists a unique isomorphism $\rho' : V \otimes F \to V$ for which

$$\rho'(v \otimes a) = av.$$

Hence, $F \otimes V \cong V \cong V \otimes F$.

(4) *There exists a unique isomorphism $\varphi : (U \oplus V) \otimes W \to (U \otimes W) \boxplus (V \otimes W)$ for which*

$$\varphi((u + v) \otimes w) = (u \otimes w, v \otimes w),$$

and similarly $W \otimes (U \oplus V) \cong (W \otimes U) \boxplus (W \otimes V)$.

Proof. Define the multilinear maps

$$t : (V_1 \otimes \cdots \otimes V_n) \times (W_1 \otimes \cdots \otimes W_m) \to V_1 \otimes \cdots \otimes V_n \otimes W_1 \otimes \cdots \otimes W_m$$
$$(v_1 \otimes \cdots \otimes v_n, w_1 \otimes \cdots \otimes w_m) \mapsto v_1 \otimes \cdots \otimes v_n \otimes w_1 \otimes \cdots \otimes w_m$$

and

$$s : V_1 \times \cdots \times V_n \to V_{\pi(1)} \otimes \cdots \otimes V_{\pi(n)}$$
$$v_1 \times \cdots \times v_n \mapsto v_{\pi(1)} \otimes \cdots \otimes v_{\pi(n)}$$

so that the linear maps τ and σ exist by the universal property. It is easy to see that these maps are bijections; this proves (1) and (2). For (3), define the bilinear map $r : F \times V \to V$ by $(a, v) \mapsto av$ so that there exists a unique linear map $\rho : F \otimes V \to V$ with $\rho(a \otimes v) = av$. But linear map $v \mapsto 1 \otimes v$ is an inverse to ρ, so ρ is an isomorphism. For (4), define

$$e : (U \oplus V) \times W \to (U \otimes W) \boxplus (V \otimes W)$$
$$(u + v, w) \mapsto (u \otimes w, v \otimes w)$$

so that φ exists by the universal property. Similarly, the bilinear maps $(u, w) \mapsto u \otimes w$ and $(v, w) \mapsto v \otimes w$ induce unique linear maps $f : U \otimes W \to (U \oplus V) \otimes W$ and $g : V \otimes W \to (U \oplus V) \otimes W$. The map $h : (U \otimes W) \boxplus (V \otimes W) \to (U \oplus V) \otimes W$ given by $(a, b) \mapsto f(a) + g(b)$ is an inverse to e, so e is an isomorphism. □

Theorem 270. *Let $z \in U \otimes V$ with*

$$z = \sum_{i=1}^{r} x_i \otimes y_i$$

where the sets $\{x_i\} \subseteq U$ and $\{y_i\} \subseteq V$ contain exactly r elements and are linearly independent. If $\{w_i\} \subseteq U$ and $\{z_i\} \subseteq V$ are sets with exactly r elements and are linearly independent, then the equation

$$z = \sum_{i=1}^{r} w_i \otimes z_i$$

holds if and only if there exist invertible $r \times r$ matrices A, B for which $A^T B = I$ and

$$w_i = \sum_{j=1}^{r} A_{i,j} x_j \quad \text{and} \quad z_i = \sum_{k=1}^{r} B_{i,k} y_k$$

for $i = 1, \ldots, r$.

Proof. If there exist matrices A, B satisfying the given conditions, then

$$\begin{aligned}
\sum_{i=1}^{r} w_i \otimes z_i &= \sum_{i=1}^{r} \sum_{j=1}^{r} \sum_{k=1}^{r} A_{i,j} B_{i,k} (x_j \otimes y_k) \\
&= \sum_{j=1}^{r} \sum_{k=1}^{r} \sum_{i=1}^{r} A^T_{j,i} B_{i,k} (x_j \otimes y_k) \\
&= \sum_{j=1}^{r} \sum_{k=1}^{r} [A^T B]_{j,k} (x_j \otimes y_k)
\end{aligned}$$

$$= \sum_{i=1}^{r} x_i \otimes y_i$$
$$= z$$

since $A^T B = I$. The converse was shown in (14.3) from the text. □

Theorem 271. *[Exercise 14.1] If $\tau : W \to X$ is a linear map and $b : U \times V \to W$ is bilinear, then $\tau \circ b : U \times V \to X$ is bilinear.*

Proof. We have

$$\begin{aligned}(\tau \circ b)(\alpha x + \beta y, z) &= \tau(\alpha b(x, z) + \beta b(y, z)) \\ &= \alpha(\tau \circ b)(x, z) + \beta(\tau \circ b)(y, z).\end{aligned}$$

Similarly, $\tau \circ b$ is linear in the other coordinate. □

Theorem 272. *[Exercise 14.2] The only map that is both linear and n-linear (for $n \geq 2$) is the zero map.*

Proof. Let $\tau : V^n \to W$ be a linear map that is also n-linear. For all $v_1, \ldots, v_n \in V$,

$$\begin{aligned}\tau(v_1, \ldots, v_n) &= \sum_{k=1}^{n} \tau(0, \ldots, v_k, \ldots, 0) \\ &= 0\end{aligned}$$

where the first equality follows from the linearity of τ and the second equality follows from the n-linearity of τ. □

Example 273. [Exercise 14.3] Find an example of a bilinear map $\tau : V \times V \to W$ whose image $\mathrm{im}(\tau) = \{\tau(u, v) \mid u, v \in V\}$ is not a subspace of W.

The tensor map $t : U \times V \to U \otimes V$ is an example, since its image consists of all decomposable tensors. If $u \otimes v, u' \otimes v'$ are decomposable tensors, then $u \otimes v + u' \otimes v'$ is not necessarily a decomposable tensor.

Theorem 274. *[Exercise 14.4] Let $B = \{u_i \mid i \in I\}$ be a basis for U and let $C = \{v_j \mid j \in J\}$ be a basis for V. Then the set*

$$D = \{u_i \otimes v_j \mid i \in I, j \in J\}$$

is a basis for $U \otimes V$.

Proof. We give an argument that shows that D is linearly independent and spans $U \otimes V$. Suppose that

$$\sum_{i,j} r_{i,j} u_i \otimes v_j = \sum_i u_i \otimes \sum_j r_{i,j} v_j = 0.$$

Since B is linearly independent, Theorem 14.5 shows that

$$\sum_j r_{i,j} v_j = 0$$

for each i. But C is linearly independent, so $r_{i,j} = 0$ for all i, j. This shows that D is linearly independent. Let $t : U \times V \to U \otimes V$ be the tensor map. Since B and C span U and V respectively, we have

$$\begin{aligned} \operatorname{span}(\{t(u_i, v_j) \mid i \in I, j \in J\}) &= \operatorname{span}(\{t(u, v) \mid u \in U, v \in V\}) \\ &= \operatorname{Im}(t) \\ &= U \otimes V \end{aligned}$$

by Theorem 268. □

Theorem 275. *[Exercise 14.5] The following property of a pair $(W, g : U \times V \to W)$ with g bilinear characterizes the tensor product $(U \otimes V, t : U \times V \to U \otimes V)$ up to isomorphism: For a pair $(W, g : U \times V \to W)$ with g bilinear, if $\{u_i\}$ is a basis for U and $\{v_i\}$ is a basis for V, then $\{g(u_i, v_j)\}$ is a basis for W.*

Proof. Theorem 274 already shows one direction; we must prove that if $(W, g : U \times V \to W)$ satisfies the property, then it is a tensor product of U with V. But $\{u_i \otimes v_j\}$ is a basis for $U \otimes V$, so we can define a linear map $\tau : U \otimes V \to W$ by setting $\tau(u_i \otimes v_j) = g(u_i, v_j)$. This map carries a basis of $U \otimes V$ to a basis of W, so it is an isomorphism and therefore $W \cong U \otimes V$. □

Theorem 276. *[Exercise 14.7] Let X and Y be nonempty sets. Then $\mathcal{F}_{X \times Y} \cong \mathcal{F}_X \otimes \mathcal{F}_Y$, where $\mathcal{F}_A$ is the free vector space on A.*

Proof. Let $t : \mathcal{F}_X \times \mathcal{F}_Y \to \mathcal{F}_X \otimes \mathcal{F}_Y$ be the tensor map. Define the bilinear map $\tau : \mathcal{F}_X \times \mathcal{F}_Y \to \mathcal{F}_{X \times Y}$ by

$$\tau\left(\sum_i a_i x_i, \sum_j b_j y_j\right) = \sum_{i,j} a_i b_j (x_i, y_j)$$

so that there exists a unique $\widetilde{\tau} : \mathcal{F}_X \otimes \mathcal{F}_Y \to \mathcal{F}_{X \times Y}$ with $\widetilde{\tau} t = \tau$. Clearly $\ker(\widetilde{\tau}) = \{0\}$, so $\widetilde{\tau}$ is injective. If

$$\sum_{i,j} r_{i,j} (x_i, y_j) \in \mathcal{F}_{X \times Y}$$

then

$$\widetilde{\tau}\left(\sum_{i,j} r_{i,j} (x_i \otimes y_j)\right) = \sum_{i,j} r_{i,j} (x_i, y_j),$$

which shows that $\widetilde{\tau}$ is surjective. □

Theorem 277. *[Exercise 14.8] Let $u, u' \in U$ and $v, v' \in V$ with $u \otimes v \neq 0$. Then $u \otimes v = u' \otimes v'$ if and only if $u' = ru$ and $v' = r^{-1}v$ for some $r \neq 0$.*

Proof. One direction is evident. Suppose that $u \otimes v = u' \otimes v'$. For any multilinear map $f : U \times V \to W$ we have $f(u, v) = \tau(u \otimes v)$ and in particular,

$$f(u, v) = \tau(u \otimes v) = \tau(u' \otimes v') = f(u', v').$$

This is true for any multilinear map f, so it is true if $g \in U^*$, $h \in V^*$ and $f : U \times V \to F$ is given by $f(u, v) = g(u)h(v)$:

$$g(u)h(v) = g(u')h(v').$$

Note that $u \otimes v \neq 0$ implies that $u, v \neq 0$; we can choose $\widetilde{h} \in V^*$ with $\widetilde{h}(v) \neq 0$ so that

$$g(u) = g\left(\frac{\widetilde{h}(v')}{\widetilde{h}(v)}u'\right)$$

for all $g \in U^*$. By Corollary 50 we have

$$u = \frac{\widetilde{h}(v')}{\widetilde{h}(v)}u',$$

noting that $\widetilde{h}(v')/\widetilde{h}(v) \neq 0$ since $u \neq 0$. Now choose $\widetilde{g} \in U^*$ with $\widetilde{g}(u) = 1$ so that

$$h(v) = \widetilde{g}(u)h(v) = \widetilde{g}(u')h(v') = h\left(\frac{\widetilde{h}(v)}{\widetilde{h}(v')}v'\right)$$

for all $h \in V^*$. Applying Corollary 50 once more shows that

$$v = \frac{\widetilde{h}(v)}{\widetilde{h}(v')}v'.$$

Note that this result can also be proved by using Theorem 14.6. □

Theorem 278. *[Exercise 14.9] Let $B = \{b_i\}$ be a basis for U and $C = \{c_j\}$ be a basis for V. Any function $f : B \times C \to W$ can be extended uniquely to a linear function $\overline{f} : U \otimes V \to W$. Therefore, the function f can be extended in a unique way to a bilinear map $\widehat{f} : U \times V \to W$. All bilinear maps are in fact obtained in this way.*

Proof. The first statement follows from the fact that $\{b_i \otimes c_j\}$ is a basis for $U \otimes V$ and Theorem 14. Let $t : U \times V \to U \otimes V$ be the tensor map. Since $\overline{f}t : U \times V \to W$ is a bilinear map that extends f, the function $\widehat{f}$ exists. In fact, if $\widehat{f}$ is a bilinear map that extends f then $\widehat{f} = \tau t$ for a unique $\tau : U \otimes V \to W$ such that $\tau(b_i \otimes c_j) = \widehat{f}(b_i, c_j)$ for all i, j. But $\overline{f}$ is already such a linear map, so $\widehat{f} = \tau t = \overline{f}t$. This proves the second statement. If $g : U \times V \to W$ is a bilinear map then we can define $f(b_i, c_j) = g(b_i, c_j)$

for all i, j. The function f can be extended in a unique way to a bilinear map $\widehat{f}$ such that $f(b_i, c_j) = \widehat{f}(b_i, c_j)$ for all i, j. By the uniqueness of $\widehat{f}$, we have $\widehat{f} = g$. □

Theorem 279. *Let U, V be vector spaces, let S be a subspace of U and let T be a subspace of V. If (T, t) is a tensor product of U with V then $(\mathrm{Im}(t|_{S\times T}), t|_{S\times T})$ is a tensor product of S with T.*

Proof. Let $f : S \times T \to W$ be a bilinear map. Choose a bilinear map $g : U \times V \to W$ such that $g|_{S\times T} = f$; there exists a linear map $\tau : T \to W$ with $g = \tau t$. But

$$f = g|_{S\times T} = \tau|_{\mathrm{Im}(t|_{S\times T})} t|_{S\times T},$$

so condition (1) of Theorem 268 is satisfied. Condition (2) is automatically satisfied, so this proves the result. □

Lemma 280. *[Exercise 14.10] Let S_1, S_2 be subspaces of V. Then*

$$(S_1 \otimes V) \cap (S_2 \otimes V) = (S_1 \cap S_2) \otimes V$$

where the tensor products are interpreted as subspaces of $V \otimes V$ (cf. Theorem 279).

Proof. Let $B = \{v_i\}$ be a basis of V. Let $z \in (S_1 \otimes V) \cap (S_2 \otimes V)$ and write

$$z = \sum_{i=1}^{m} x_i \otimes v_{j_i} = \sum_{i=1}^{n} y_i \otimes v_{k_i}$$

where $x_i \in S_1$ and $y_i \in S_2$. After reindexing, we have

$$z = \sum_{i=1}^{r} x_i \otimes v_{j_i} = \sum_{i=1}^{r} y_i \otimes v_{j_i}$$

so that

$$\sum_{i=1}^{r} (x_i - y_i) \otimes v_{j_i} = 0.$$

Since the $\{v_{j_i}\}$ are linearly independent, $x_i = y_i$ for each i and therefore $z \in (S_1 \cap S_2) \otimes V$. The other inclusion is obvious. □

Lemma 281. *[Exercise 14.11] Let $S \subseteq U$ and $T \subseteq V$ be subspaces of vector spaces U and V respectively. Then*

$$(S \otimes V) \cap (U \otimes T) = S \otimes T.$$

Proof. Choose a subspace $T' \subseteq V$ such that $V = T \oplus T'$. Then

$$\begin{aligned}(S \otimes V) \cap (U \otimes T) &= (S \otimes (T + T')) \cap (U \otimes T) \\ &= ((S \otimes T) + (S \otimes T')) \cap (U \otimes T) \\ &= (S \otimes T) + ((S \otimes T') \cap (U \otimes T))\end{aligned}$$

by Theorem 5. But

$$\begin{aligned}(S \otimes T') \cap (U \otimes T) &\subseteq (U \otimes T') \cap (U \otimes T) \\ &= U \times (T' \cap T) \\ &= \{0\}\end{aligned}$$

by Lemma 280 so

$$(S \otimes V) \cap (U \otimes T) = (S \otimes T) + \{0\} = S \otimes T.$$

□

Theorem 282. *[Exercise 14.12] Let $S_1, S_2 \subseteq U$ and $T_1, T_2 \subseteq V$ be subspaces of U and V respectively. Then*

$$(S_1 \otimes T_1) \cap (S_2 \otimes T_2) = (S_1 \cap S_2) \otimes (T_1 \cap T_2).$$

Proof. Applying Lemma 280 twice gives

$$(S_1 \otimes T_1) \cap (S_2 \otimes T_2) \subseteq (S_1 \otimes V) \cap (S_2 \otimes V) = (S_1 \cap S_2) \otimes V$$

and

$$(S_1 \otimes T_1) \cap (S_2 \otimes T_2) \subseteq (U \otimes T_1) \cap (U \otimes T_2) = U \otimes (T_1 \cap T_2).$$

By Lemma 281,

$$\begin{aligned}(S_1 \otimes T_1) \cap (S_2 \otimes T_2) &\subseteq ((S_1 \cap S_2) \otimes V) \cap (U \otimes (T_1 \cap T_2)) \\ &= (S_1 \cap S_2) \otimes (T_1 \cap T_2).\end{aligned}$$

The other inclusion is clear. □

Theorem 283. *[Exercise 14.14] Let ι_X denote the identity operator on a vector space X. For any vector spaces V and W, we have $\iota_V \odot \iota_W = \iota_{V \otimes W}$.*

Proof. By definition, $\iota_V \odot \iota_W : V \otimes W \to V \otimes W$ is the unique linear map such that

$$(\iota_V \odot \iota_W)(v \otimes w) = \iota_V v \otimes \iota_W w = v \otimes w$$

for all $v \in V$ and $w \in W$. But $\iota_{V \otimes W}$ also satisfies the equation above, so $\iota_V \odot \iota_W = \iota_{V \otimes W}$. □

Theorem 284. *[Exercise 14.15] Let $\tau_1 : U \to V$, $\tau_2 : V \to W$ and $\sigma_1 : U' \to V'$, $\sigma_2 : V' \to W'$. Then*

$$(\tau_2 \circ \tau_1) \odot (\sigma_2 \circ \sigma_1) = (\tau_2 \odot \sigma_2) \circ (\tau_1 \odot \sigma_1).$$

Proof. For all $u \in U$ and $u' \in U'$,

$$\begin{aligned}[(\tau_2 \odot \sigma_2) \circ (\tau_1 \odot \sigma_1)](u \otimes u') &= (\tau_2 \odot \sigma_2)(\tau_1 u \otimes \sigma_1 u') \\ &= (\tau_2 \circ \tau_1)u \otimes (\sigma_2 \circ \sigma_1)u'.\end{aligned}$$

By definition, the result follows. □

Theorem 285. *[Exercise 14.16] Let V be a vector space over a field F and let K be a field extension of F. Then $F^n \otimes_F K \cong K^n$ as K-spaces, with scalar multiplication on $F^n \otimes_F K$ defined as $\alpha(v \otimes k) = v \otimes \alpha k$ where $\alpha \in K$.*

Proof. Define the multilinear F-map $f : F^n \times K \to K^n$ by $(v, r) \mapsto rv$. This induces a linear F-map $\varphi : F^n \otimes_F K \to K^n$ satisfying $\varphi(v \otimes r) = rv$. But if $k \in K$ then

$$\varphi(k(v \otimes r)) = \varphi(v \otimes kr) = krv = k\varphi(v \otimes r),$$

so φ is also a linear K-map. Furthermore, φ carries the basis $\{e_i \otimes 1\}$ of the K-space $F^n \otimes_F K$ to the basis $\{e_i\}$ of K^n, so φ is an isomorphism. □

Example 286. [Exercise 14.17] Prove that in a tensor product $U \otimes U$ for which $\dim(U) \geq 2$, not all vectors have the form $u \otimes v$ for some $u, v \in U$.

Let $u, v \in U$ be linearly independent vectors and consider $z = u \otimes v + v \otimes u$. By Theorem 14.6 the rank of z is 2, so it does not have the form $x \otimes y$.

Theorem 287. *[Exercise 14.18] For the block matrix*

$$M = \begin{bmatrix} A & B \\ 0 & C \end{bmatrix}$$

we have $\det(M) = \det(A)\det(C)$.

Proof. Suppose that M is a $n \times n$ matrix and A is a $m \times m$ matrix. Define the map $f : M_m \to F$ given by

$$X \mapsto \det \begin{bmatrix} X & B \\ 0 & I \end{bmatrix},$$

which is multilinear and antisymmetric in the columns of X. Corollary 14.20 shows that

$$(*) \qquad f(X) = \det \begin{bmatrix} I & B \\ 0 & I \end{bmatrix} \det(X) = \det(X)$$

since the last matrix is upper triangular. If we define the map $g : M_{n-m} \to F$ given by

$$X \mapsto \det \begin{bmatrix} A & B \\ 0 & X \end{bmatrix},$$

a similar argument shows that

$$g(X) = \left(\det \begin{bmatrix} A & B \\ 0 & I \end{bmatrix} \right) \det(X) = \det(A)\det(X)$$

by using (*). Therefore

$$\det \begin{bmatrix} A & B \\ 0 & C \end{bmatrix} = g(C) = \det(A)\det(C).$$

□

Theorem 288. *[Exercise 14.19] Let $A, B \in \mathcal{M}_n(F)$. If either A or B is invertible, then the matrices $A + \alpha B$ are invertible except for a finite number of α's.*

Proof. If A is invertible, then

$$\begin{aligned}\det(A + \alpha B) = 0 &\Leftrightarrow \det(A(I + \alpha A^{-1}B)) = 0 \\ &\Leftrightarrow \det(I + \alpha A^{-1}B) = 0 \\ &\Leftrightarrow \det(\alpha^{-1}I + A^{-1}B) = 0.\end{aligned}$$

But $A^{-1}B$ has only a finite number of eigenvalues, so $\det(A + \alpha B) = 0$ for only finitely many values of α. If B is invertible, a similar argument applies. □

The Tensor Product of Matrices.

Theorem 289. *[Exercise 14.20] Let $A = (a_{i,j})$ be the matrix of a linear operator $\tau \in \mathcal{L}(V)$ with respect to the ordered basis $A = (u_1, \ldots, u_n)$. Let $B = (b_{i,j})$ be the matrix of a linear operator $\sigma \in \mathcal{L}(V)$ with respect to the ordered basis $B = (v_1, \ldots, v_n)$. Consider the ordered basis $C = (u_i \otimes v_j)$ ordered lexicographically; that is $u_i \otimes v_j < u_\ell \otimes v_k$ if $i < \ell$ or $i = \ell$ and $j < k$. Then the matrix of $\tau \otimes \sigma$ with respect to C is*

$$A \otimes B = \begin{bmatrix} a_{1,1}B & a_{1,2}B & \cdots & a_{1,n}B \\ a_{2,1}B & a_{2,2}B & \cdots & a_{2,n}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1}B & a_{n,2}B & \cdots & a_{n,n}B \end{bmatrix}.$$

*This matrix is called the **tensor product**, **Kronecker product** or **direct product** of the matrix A with the matrix B.*

Proof. Let $x, y \in V$ and write

$$x = \sum_{i=1}^{n} r_i u_i \quad \text{and} \quad y = \sum_{j=1}^{n} s_j v_j.$$

We have

$$\begin{aligned}(\tau \otimes \sigma)(x \otimes y) &= (\tau \otimes \sigma)\left(\sum_{1 \le i,j \le n} r_i s_j (u_i \otimes v_j)\right) \\ &= \sum_{1 \le i,j \le n} r_i s_j (\tau u_i \otimes \sigma v_j) \\ &= \sum_{1 \le i,j \le n} r_i s_j (\sum_{k=1}^{n} a_{k,i} u_k \otimes \sum_{\ell=1}^{n} b_{\ell,j} v_\ell)\end{aligned}$$

$$= \sum_{1\le i,j,k,\ell\le n} r_i s_j a_{k,i} b_{\ell,j} (u_k \otimes v_\ell)$$

so that

$$\begin{aligned}
([(\tau\otimes\sigma)(x\otimes y)]_C)_k &= \sum_{1\le i,j\le n} r_i s_j a_{\lceil k/n\rceil,i} b_{n+k-n\lceil k/n\rceil,j} \\
&= \sum_{i=1}^{n^2} a_{\lceil k/n\rceil,\lceil i/n\rceil} b_{n+k-n\lceil k/n\rceil,n+i-n\lceil i/n\rceil} r_{\lceil i/n\rceil} s_{n+i-n\lceil i/n\rceil} \\
&= \sum_{i=1}^{n^2} (A\otimes B)_{k,i} r_{\lceil i/n\rceil} s_{n+i-n\lceil i/n\rceil} \\
&= \sum_{i=1}^{n^2} (A\otimes B)_{k,i} ([x\otimes y]_C)_i \\
&= ((A\otimes B)[x\otimes y]_C)_k.
\end{aligned}$$

Therefore $[\tau\otimes\sigma]_C = A\otimes B$. □

Example 290. [Exercise 14.21] Show that the tensor product is not, in general, commutative.

If

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

then

$$A\otimes B = \begin{bmatrix} 0 & 1 & & \\ 1 & 0 & & \\ & & 0 & 1 \\ & & 1 & 0 \end{bmatrix} \quad \text{and} \quad B\otimes A = \begin{bmatrix} & & 1 & 0 \\ & & 0 & 1 \\ 1 & 0 & & \\ 0 & 1 & & \end{bmatrix}.$$

Theorem 291. *[Exercise 14.22-14.30] Let A, B be $n\times n$ matrices.*

(1) *The tensor product $A\otimes B$ is bilinear in both A and B.*
(2) *$A\otimes B = 0$ if and only if $A = 0$ or $B = 0$.*
(3) *$(A\otimes B)^T = A^T\otimes B^T$.*
(4) *$(A\otimes B)^* = A^*\otimes B^*$ when $F = \mathbb{C}$.*
(5) *If $u, v\in F^n$, then (as row vectors) $u^T v = u^T\otimes v$.*
(6) *Let $A_{m,n}, B_{p,q}, C_{n,k}, D_{q,r}$ are matrices of the given sizes. Then*

$$(A\otimes B)(C\otimes D) = (AC)\otimes(BD).$$

In particular, if u, v are column vectors of appropriate sizes then

$$(Au)(Bv) = (A\otimes B)(u\otimes v).$$

(7) *If A and B are invertible, then so is $A \otimes B$ and*
$$(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}.$$
(8) $\operatorname{tr}(A \otimes B) = \operatorname{tr}(A)\operatorname{tr}(B)$.
(9) *Suppose that F is algebraically closed. If A has eigenvalues $\lambda_1, \ldots, \lambda_n$ and B has eigenvalues $\mu_1, \ldots, \mu_m$, both lists including multiplicity, then $A \otimes B$ has eigenvalues*
$$\{\lambda_i \mu_j \mid 1 \leq i \leq n, 1 \leq j \leq m\},$$
again counting multiplicity.
(10) $\det(A_{n,n} \otimes B_{m,m}) = \det(A_{n,n})^m \det(B_{m,m})^n$.

Proof. Parts (1) to (5) are clear. Part (6) follows from Theorem 284. If $k = r = 1$ then C and D are column vectors, so we have the result that
$$Au \otimes Bv = (A \otimes B)(u \otimes v) = (A \otimes B) \begin{bmatrix} u_1 v \\ \vdots \\ u_n v \end{bmatrix},$$
which is simply a version of Theorem 289. For (7), put $C = A^{-1}$ and $D = B^{-1}$ in (6). For (8), we have
$$\begin{aligned} \operatorname{tr}(A \otimes B) &= \operatorname{tr}(a_{1,1}B) + \cdots + \operatorname{tr}(a_{n,n}B) \\ &= (a_{1,1} + \cdots + a_{n,n}) \operatorname{tr}(B) \\ &= \operatorname{tr}(A)\operatorname{tr}(B). \end{aligned}$$
For (9), let $u_1, \ldots, u_n$ be eigenvectors for $\lambda_1, \ldots, \lambda_n$ and let $v_1, \ldots, v_m$ be eigenvectors for $\mu_1, \ldots, \mu_m$. Then
$$\begin{aligned} (A \otimes B)(u_i \otimes v_j) &= Au_i \otimes Bv_j \\ &= \lambda_i \mu_j (u_i \otimes v_j) \end{aligned}$$
for each i, j. Part (10) follows from (9), taking the algebraic closure of F if necessary. □

Applications of the Exterior Algebra.

Remark 292. If $u, v \in \bigwedge V$ satisfy $u = \alpha v$ for some scalar α, we write u/v for α.

Theorem 293. *Let V be a finite-dimensional vector space. If $v_1, \ldots, v_k \in V$ are distinct vectors, then $v_1 \wedge \cdots \wedge v_k = 0$ if and only if $v_1, \ldots, v_k$ are linearly dependent.*

Proof. Suppose that $v_1, \ldots, v_k$ are linearly dependent and assume without loss of generality that $a_1 v_1 + \cdots + a_k v_k = 0$ with $a_1 \neq 0$. Then $v_1 = -a_1^{-1}(a_2 v_2 + \cdots + a_k v_k)$, so
$$v_1 \wedge \cdots \wedge v_k = -a_1^{-1}(a_2 v_2 + \cdots + a_k v_k) \wedge v_2 \wedge \cdots \wedge v_k = 0.$$

Conversely, suppose that $v_1, \ldots, v_k$ are linearly independent. We can extend $\{v_1, \ldots, v_k\}$ to a basis $\{v_1, \ldots, v_n\}$ of V, and since $\{v_1 \wedge \cdots \wedge v_n\}$ is a basis of $\bigwedge^n V$ by Theorem 14.17, we must have $v_1 \wedge \cdots \wedge v_n \neq 0$. But $v_1 \wedge \cdots \wedge v_k$ is a factor of $v_1 \wedge \cdots \wedge v_n$, so $v_1 \wedge \cdots \wedge v_n \neq 0$. □

Definition 294. Let $\tau \in \mathcal{L}(V, V)$ where V is finite-dimensional. The alternating multilinear map

$$f : V^n \to \bigwedge^n V$$
$$(v_1, \ldots, v_n) \mapsto \tau v_1 \wedge \cdots \wedge \tau v_n$$

induces a unique linear operator $\varphi : \bigwedge^n V \to \bigwedge^n V$ with $\varphi(v_1 \wedge \cdots \wedge v_n) \mapsto \tau v_1 \wedge \cdots \wedge \tau v_n$. Since $\bigwedge^n V$ is one-dimensional, we can define the **determinant** $\det(\tau)$ to be the unique number satisfying $\varphi = \det(\tau)\iota$ where ι is the identity on $\bigwedge^n V$. Similarly, the alternating multilinear map

$$(v_1, \ldots, v_n) \mapsto \sum_{k=1}^{n} v_1 \wedge \cdots \wedge \tau v_k \wedge \cdots \wedge v_n$$

induces a unique linear operator $\psi : \bigwedge^n V \to \bigwedge^n V$; the **trace** $\operatorname{tr}(\tau)$ is defined as the unique number satisfying $\psi = \operatorname{tr}(\tau)\iota$.

Theorem 295. *Let $\tau, \sigma \in \mathcal{L}(V, V)$ where $\dim(V) = n$. Let $A = \begin{bmatrix} a_1 & \cdots & a_n \end{bmatrix}$ be an $n \times n$ matrix with entries in a field F, and let $\{e_1, \ldots, e_n\}$ be the standard basis of F^n.*

(1) $\det(\iota) = 1$ *and* $\det(0) = 0$.
(2) $\det(\tau\sigma) = \det(\tau)\det(\sigma)$.
(3) $\det(r\tau) = r^n \det(\tau)$.
(4) $a_1 \wedge \cdots \wedge a_n = \det(A) e_1 \wedge \cdots \wedge e_n$.
(5) $\det(A) = \sum_{\rho \in S_n} (-1)^\rho A_{\rho(1),1} \cdots A_{\rho(n),n}$.
(6) $\det(A^T) = \det(A)$.
(7) *If $\varphi : V \to W$ is an isomorphism, then* $\det(\varphi\tau\varphi^{-1}) = \det(\tau)$.
(8) *If $A = [\tau]_B$ where B is a basis of V, then* $\det(A) = \det(\tau)$.

Proof. (1) is obvious. For all $v_1, \ldots, v_n \in V$,

$$\begin{aligned} \det(\tau\sigma) v_1 \wedge \cdots \wedge v_n &= \tau\sigma v_1 \wedge \cdots \wedge \tau\sigma v_n \\ &= \det(\tau)\sigma v_1 \wedge \cdots \wedge \sigma v_n \\ &= \det(\tau)\det(\sigma) v_1 \wedge \cdots \wedge v_n \end{aligned}$$

and

$$\begin{aligned} \det(r\tau) v_1 \wedge \cdots \wedge v_n &= r\tau v_1 \wedge \cdots \wedge r\tau v_n \\ &= r^n \tau v_1 \wedge \cdots \wedge \tau v_n \\ &= r^n \det(\tau) v_1 \wedge \cdots \wedge v_n, \end{aligned}$$

which proves (2) and (3). (4) follows from the computation

$$a_1 \wedge \cdots \wedge a_n = Ae_1 \wedge \cdots \wedge Ae_n = \det(A) e_1 \wedge \cdots \wedge e_n.$$

For (5),

$$\begin{aligned}
\det(A) e_1 \wedge \cdots \wedge e_n &= Ae_1 \wedge \cdots \wedge Ae_n \\
&= \sum_{i_1=1}^{n} A_{i_1,1} e_{i_1} \wedge \cdots \wedge \sum_{i_n=1}^{n} A_{i_n,n} e_{i_n} \\
&= \sum_{i_1=1}^{n} \cdots \sum_{i_n=1}^{n} A_{i_1,1} \cdots A_{i_n,n} e_{i_1} \wedge \cdots \wedge e_{i_n} \\
&= \sum_{\rho \in S_n} A_{\rho(1),1} \cdots A_{\rho(n),n} e_{i_1} \wedge \cdots \wedge e_{i_n} \\
&= \sum_{\rho \in S_n} (-1)^{\rho} A_{\rho(1),1} \cdots A_{\rho(n),n} e_1 \wedge \cdots \wedge e_n.
\end{aligned}$$

For (6),

$$\begin{aligned}
\sum_{\rho \in S_n} (-1)^{\rho} A_{\rho(1),1} \cdots A_{\rho(n),n} &= \sum_{\rho \in S_n} (-1)^{\rho^{-1}} A_{1,\rho^{-1}(1)} \cdots A_{n,\rho^{-1}(n)} \\
&= \sum_{\rho \in S_n} (-1)^{\rho} A_{1,\rho(1)} \cdots A_{n,\rho(n)}.
\end{aligned}$$

Finally, suppose that $\varphi : V \to W$ is an isomorphism. We have an isomorphism $\psi : \bigwedge^n V \to \bigwedge^n W$ induced by the map $(v_1, \ldots, v_n) \mapsto \varphi v_1 \wedge \cdots \wedge \varphi v_n$, so for all $v_1, \ldots, v_n \in V$ we have

$$\begin{aligned}
\det(\tau) v_1 \wedge \cdots \wedge v_n &= \tau v_1 \wedge \cdots \wedge \tau v_n \\
&= \tau \varphi^{-1} \varphi v_1 \wedge \cdots \wedge \tau \varphi^{-1} \varphi v_n \\
&= \psi^{-1}(\varphi \tau \varphi^{-1} \varphi v_1 \wedge \cdots \wedge \varphi \tau \varphi^{-1} \varphi v_n) \\
&= \psi^{-1}(\det(\varphi \tau \varphi^{-1}) \varphi v_1 \wedge \cdots \wedge \varphi v_n) \\
&= \det(\varphi \tau \varphi^{-1}) v_1 \wedge \cdots \wedge v_n.
\end{aligned}$$

This proves (7), and (8) follows by considering the isomorphism $\varphi : V \to F^n$ that takes B to the standard basis of F^n. □

Theorem 296. *Let $A = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}$ be an $n \times n$ matrix with entries in a field F. The following are equivalent:*

(1) *A is invertible.*
(2) $v_1 \wedge \cdots \wedge v_n \neq 0$.
(3) $\det(A) \neq 0$.

Proof. (1) and (2) are equivalent by Theorem 293, and (2) $\Leftrightarrow$ (3) follows from Theorem 295. □

Theorem 297. *[Cramer's rule] Let $A \in \mathcal{M}_n(F)$ be an invertible matrix with columns $v_1, \ldots, v_n$ and let $x = (x_1, \ldots, x_n)$ be a column vector. If $Ax = y$ then*

$$x_k = \frac{v_1 \wedge \cdots \wedge v_{k-1} \wedge y \wedge v_{k+1} \wedge \cdots \wedge v_n}{v_1 \wedge \cdots \wedge v_n} = \frac{\det(A_k)}{\det(A)}$$

for every k, where A_k is the matrix obtained by replacing the kth column of A with y.

Proof. Let $\{e_1, \ldots, e_n\}$ be the standard basis of F^n and write $x = x_1e_1 + \cdots + x_ne_n$. For every k we have

$$\begin{aligned} y \wedge \bigwedge_{i \neq k} v_i &= A(x_1e_1 + \cdots + x_ne_n) \wedge \bigwedge_{i \neq k} v_i \\ &= x_k v_1 \wedge \cdots \wedge v_n, \end{aligned}$$

so the result follows from Theorem 295. □

Chapter 15. Positive Solutions to Linear Systems: Convexity and Separation

Theorem 298. *[Exercise 15.1] Any hyperplane has the form $\mathcal{H}(N, \|N\|^2)$ for an appropriate vector N.*

Proof. Any hyperplane

$$\mathcal{H}(N, b) = \{x \in \mathbb{R}^n \mid \langle N, x \rangle = b\}$$

is equivalent to the hyperplane

$$\mathcal{H}(rN, rb) = \{x \in \mathbb{R}^n \mid \langle rN, x \rangle = rb\}$$

for any nonzero $r \in \mathbb{R}$. In particular, setting

$$r = \frac{b}{\|N\|^2}$$

gives the desired result. □

Theorem 299. *[Exercise 15.2] If A is an $m \times n$ matrix then the set $C = \{Ax \mid x \in \mathbb{R}^n, x > 0\}$ is a convex cone in $\mathbb{R}^m$.*

Proof. It is clear that C is a cone. If $x, y \in C$ then $x = Ax'$ and $y = Ay'$ for some $x', y' \in \mathbb{R}^n_+$. For $0 \leq t \leq 1$,

$$tx + (1-t)y = A(tx' + (1-t)y') \in C$$

since $\mathbb{R}^n_+$ is convex. Therefore C is convex. □

Theorem 300. *[Exercise 15.3] If A and B are strictly separated subsets of $\mathbb{R}^n$ and if A is compact, then A and B are strongly separated as well.*

Proof. Suppose that A and B are strictly separated by the hyperplane $\mathcal{H}(N, b)$, and assume without loss of generality that $\langle N, A\rangle < b < \langle N, B\rangle$. Since A is compact, the continuous function $a \mapsto \langle N, a\rangle$ takes on its maximum $b' < b$ at some $a_0 \in A$. We have

$$\langle N, A\rangle \leq b' < \frac{b+b'}{2} < b < \langle N, B\rangle,$$

so A and B are strongly separated by the hyperplane $\mathcal{H}(N, (b+b')/2)$. □

Theorem 301. *[Exercise 15.4] Let V be a real vector space and let X be a subset of V. The following are equivalent:*

(1) *X is closed under the taking of convex combinations of any two of its points.*
(2) *X is closed under the taking of arbitrary convex combinations, that is, for all $n \geq 1$,*

$$x_1, \ldots, x_n \in X \quad \text{and} \quad \sum_{i=1}^{n} r_i = 1 \quad \text{and} \quad 0 \leq r_i \leq 1$$

implies that

$$\sum_{i=1}^{n} r_i x_i \in X.$$

Proof. The direction (2) ⇒ (1) is clear. Suppose that (1) holds; we use induction on n. If $n = 1, 2$ then (2) clearly holds. Otherwise, suppose that (2) holds for $n-1$ and let $x_1, \ldots, x_n \in X$ with

$$\sum_{i=1}^{n} r_i = 1 \quad \text{and} \quad 0 \leq r_i \leq 1.$$

If $r_1, \ldots, r_{n-1} = 0$ then we are done. Otherwise, let $r = r_1 + \cdots + r_{n-1}$ so that

$$\sum_{i=1}^{n-1} \frac{r_i}{r} = 1 \quad \text{and} \quad 0 \leq \frac{r_i}{r} \leq 1 \text{ for } 1 \leq i \leq n-1$$

and therefore

$$\sum_{i=1}^{n-1} \frac{r_i}{r} x_i \in X$$

by the induction hypothesis. Now $r + r_n = 1$ and $0 \leq r, r_n \leq 1$, so

$$r \sum_{i=1}^{n-1} \frac{r_i}{r} x_i + r_n x_n = \sum_{i=1}^{n} r_i x_i \in X$$

by (1). □

Remark 302. [Exercise 15.5] Explain why an $(n-1)$-dimensional subspace of $\mathbb{R}^n$ is the solution set of a linear equation of the form $a_1x_1 + \cdots + a_nx_n = 0$.

Let $a = (a_1, \ldots, a_n)$ and assume $a \neq 0$; then the solution set is the set of all $x \in \mathbb{R}^n$ such that $\langle a, x\rangle = 0$, i.e. the kernel of the linear functional $x \mapsto \langle a, x\rangle$. It is clear that the kernel of any nonzero linear functional is $(n-1)$-dimensional.

Theorem 303. *[Exercise 15.6] If $N \in \mathbb{R}^n$ is nonzero and $b \in \mathbb{R}$ then*

$$\mathcal{H}_+(N, b) \cap \mathcal{H}_-(N, b) = \mathcal{H}(N, b)$$

and $\mathcal{H}_+^\circ(N, b)$, $\mathcal{H}_-^\circ(N, b)$, $\mathcal{H}(N, b)$ are pairwise disjoint with

$$\mathcal{H}_+^\circ(N, b) \cup \mathcal{H}_-^\circ(N, b) \cup \mathcal{H}(N, b) = \mathbb{R}^n.$$

Proof. Obvious. □

Theorem 304. *[Exercise 15.7] A function $T : \mathbb{R}^n \to \mathbb{R}^m$ is **affine** if it has the form $T(v) = \tau v + b$ for $b \in \mathbb{R}^m$, where $\tau \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$. If $C \subseteq \mathbb{R}^n$ is convex, then so is $T(C) \subseteq \mathbb{R}^m$.*

Proof. If $x, y \in T(C)$ then $x = \tau x' + b$ and $y = \tau y' + b$ for some $x', y' \in \mathbb{R}^n$. For $0 \leq t \leq 1$,

$$\begin{aligned} tx + (1-t)y &= t(\tau x' + b) + (1-t)(\tau y' + b) \\ &= \tau(tx' + (1-t)y') + b \\ &\in T(C) \end{aligned}$$

since $tx' + (1-t)y' \in C$. □

Theorem 305. *[Exercise 15.8]*

(1) *There exist cones in $\mathbb{R}^2$ that are not convex.*
(2) *A subset X of $\mathbb{R}^n$ is a convex cone if and only if $x, y \in X$ implies that $\lambda x + \mu y \in X$ for all $\lambda, \mu \geq 0$.*

Proof. An example for (1) can be constructed by taking

$$\{ax \mid a \geq 0\} \cup \{ay \mid a \geq 0\}$$

for any two nonzero and nonparallel vectors $x, y \in \mathbb{R}^2$. For (2), if X is a convex cone then

$$\lambda x + \mu y = (\lambda + \mu)\left(\frac{\lambda}{\lambda + \mu}x + \frac{\mu}{\lambda + \mu}y\right) \in X$$

if $\lambda + \mu \neq 0$. The converse is clear. □

Theorem 306. *[Exercise 15.9] The convex hull of a set $\{x_1, \ldots, x_n\}$ in $\mathbb{R}^n$ is bounded.*

Proof. This follows easily from Theorem 15.4, but we give another argument. Let $M = \max\{\|x_1\|, \ldots, \|x_n\|\}$. The closed disc

$$D = \{x \in \mathbb{R}^n \mid \|x\| \leq M\}$$

is a convex set that contains $\{x_1, \ldots, x_n\}$, so $\mathcal{C}(S) \subseteq D$ by definition. Since D is bounded, $\mathcal{C}(S)$ is bounded. □

Theorem 307. *[Exercise 15.10] Suppose that a vector $x \in \mathbb{R}^n$ has two distinct representations as convex combinations of the vectors $v_1, \ldots, v_n$. Then the vectors $v_2 - v_1, \ldots, v_n - v_1$ are linearly dependent.*

Proof. Suppose that

$$x = a_1v_1 + \cdots + a_nv_n \quad \text{and} \quad x = b_1v_1 + \cdots + b_nv_n$$

where $(a_1, \ldots, a_n) \neq (b_1, \ldots, b_n)$, $0 \leq a_i, b_i \leq 1$ and $\sum a_i = \sum b_i = 1$. Since $\sum(a_i - b_i) = 0$ we have

$$\sum_{i=1}^{n}(a_i - b_i)v_i = 0 = \sum_{i=1}^{n}(a_i - b_i)v_1,$$

so

$$\sum_{i=2}^{n}(a_i - b_i)(v_i - v_1) = 0.$$

Furthermore, we cannot have $a_i - b_i = 0$ for all $i \geq 2$; otherwise,

$$a_1 = a_n - \cdots - a_2 = b_n - \cdots - b_2 = b_1$$

which contradicts the fact that $(a_1, \ldots, a_n) \neq (b_1, \ldots, b_n)$. This shows that $v_2 - v_1, \ldots, v_n - v_1$ are linearly dependent. □

Theorem 308. *Let $H = \{x \in \mathbb{R}^n \mid \langle N, x\rangle = 0\}$ be a linear hyperplane with $N \neq 0$. If $N' \neq 0$ and $N' \in H^\perp$, then N' is a nonzero scalar multiple of N.*

Proof. We cannot have $\langle N, N'\rangle = 0$ for otherwise $N' \in H$ and $\langle N', N'\rangle = 0$. Now

$$\left\langle N, \frac{N' \|N\|^2}{\langle N, N'\rangle} - N \right\rangle = \frac{\langle N, N'\rangle \|N\|^2}{\langle N, N'\rangle} - \|N\|^2 = 0,$$

so

$$\left\langle N', \frac{N' \|N\|^2}{\langle N, N'\rangle} - N \right\rangle = \frac{\|N'\|^2 \|N\|^2}{\langle N, N'\rangle} - \langle N', N\rangle = 0.$$

Therefore $|\langle N, N'\rangle| = \|N'\| \|N\|$ and the result follows from Theorem 9.3. □

Theorem 309. *[Exercise 15.11] Let C be a nonempty convex subset of $\mathbb{R}^n$ and let $\mathcal{H}(N, b)$ be a hyperplane disjoint from C. Then C lies in one of the open half-spaces determined by $\mathcal{H}(N, b)$.*

Proof. By Theorem 298, we can assume that the hyperplane is of the form $\mathcal{H}(N, \|N\|^2)$. Apply Theorem 15.6 to the convex set $C - N$ and the subspace

$$S = \mathcal{H}(N, \|N\|^2) - N = \{x \in \mathbb{R}^n \mid \langle N, x\rangle = 0\};$$

there exists a nonzero $N' \in S^\perp$such that

$$\langle N', x\rangle \geq \|N'\|^2 > 0$$

for all $x \in C - N$. By Theorem 308, $N' = rN$ for some nonzero $r \in \mathbb{R}$. Assume that $r > 0$. For all $x \in C$ we have

$$\langle N, x\rangle = \langle N, x - N + N\rangle = \langle N, x - N\rangle + \|N\|^2 = \frac{1}{r}\langle N', x - N\rangle + \|N\|^2,$$

and since $\langle N', x - N\rangle \geq \|N'\|^2$,

$$\langle N, x\rangle \geq \frac{1}{r}\|N'\|^2 + \|N\|^2 > \|N\|^2.$$

This shows that C lies in the open half-space $\mathcal{H}_+(N, \|N\|^2)$. Similarly, if $r < 0$ then $C \subseteq \mathcal{H}_-(N, \|N\|^2)$. □

Remark 310. [Exercise 15.12] The conclusion of Theorem 15.6 may fail if we assume only that C is closed and convex.

Let S be the x-axis in $\mathbb{R}^2$ and let C be the set

$$\{(x, y) \mid x \geq 0, f(x) \leq y \leq f(x) + 1\}$$

where f is any nonnegative and monotonically decreasing function continuous on $[0, \infty)$ with $\lim_{x\to\infty} f(x) = 0$. For example, $f(x) = 1/(x + 1)$ is such a function. Then C is closed and convex, but C cannot be strongly separated from S since $\lim_{x\to\infty} f(x) = 0$.

Example 311. [Exercise 15.13] Find two nonempty convex subsets of $\mathbb{R}^2$ that are strictly separated but not strongly separated.

Take $X = \{(x, y) \mid x > 0\}$ and $Y = \{(x, y) \mid x < 0\}$.

Theorem 312. *[Exercise 15.14] Let X and Y be subsets of $\mathbb{R}^n$. Then X and Y are strongly separated by $\mathcal{H}(N, b)$ if and only if*

$$\langle N, x'\rangle > b \text{ for all } x' \in X_\varepsilon \quad \text{and} \quad \langle N, y'\rangle < b \text{ for all } y' \in Y_\varepsilon$$

where $X_\varepsilon = X + \varepsilon B(0, 1)$ and $Y_\varepsilon = Y + \varepsilon B(0, 1)$ and where $B(0, 1)$ is the closed unit ball.

Proof. Obvious. □

Remark 313. [Exercise 15.15] Show that Farkas's lemma cannot be improved by replacing $vA \leq 0$ in statement 2) with $vA \ll 0$.

Let

$$A = \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad b = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

There is no solution to the system $Ax = b$, but there is no vector $v = (v_1, v_2) \in \mathbb{R}^m$ for which $vA \ll 0$, since

$$\begin{bmatrix} v_1 & v_2 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} v_1 & -v_1 \end{bmatrix} \ll 0$$

implies that $0 < v_1 < 0$.

Chapter 16. Affine Geometry

Theorem 314. *Let S and T be subspaces of V and let $X = x + S$ and $Y = y + T$ be flats in V.*

1. *If $X \parallel Y$ then $X \subseteq Y$, $Y \subseteq X$ or $X \cap Y = \emptyset$.*
2. *$X \parallel Y$ if and only if some translation of one of these flats is contained in the other.*

Proof. If $S \subseteq T$ then $w + X \subseteq Y$ for some $w \in V$ by Theorem 16.1. If $X \cap Y \neq \emptyset$ then Theorem 16.1 shows that $X \subseteq Y$. Similarly, $T \subseteq S$ implies that $X \cap Y = \emptyset$ or $Y \subseteq X$. This proves (1). Part (2) follows almost directly from Theorem 16.1. □

Theorem 315. *[Exercise 16.1] If $x_1, \ldots, x_n \in V$, then the set $S = \{\sum r_i x_i \mid \sum r_i = 0\}$ is a subspace of V.*

Proof. The proof is obvious, but the geometric interpretation is not. For example, if $x_1, \ldots, x_n$ are linearly independent then S will be of dimension $n - 1$, and will include every vector of the form $x_i - x_j$. □

Theorem 316. *[Exercise 16.2] If $X \subseteq V$ is nonempty and $x \in X$ then*

$$\operatorname{affhull}(X) = x + \langle X - x \rangle.$$

Proof. It is clear that $x + \langle X - x \rangle$ is a flat that contains X, so $\operatorname{affhull}(X) \subseteq x + \langle X - x \rangle$. Since $x \in \operatorname{affhull}(X)$ we can write $\operatorname{affhull}(X) = x + S$ where S is a subspace of V. For each $y \in X$ we have $y \in x + S$ and $y - x \in S$, so $\langle X - x \rangle \subseteq S$. Therefore

$$x + \langle X - x \rangle \subseteq x + S = \operatorname{affhull}(X).$$

□

Example 317. [Exercise 16.3] The set $X = \{(0,0),(1,0),(0,1)\}$ in $\mathbb{Z}_2^2$ is closed under the formation of lines, but not affine hulls.

If $x, y \in X$ then any point on the line between x and y must simply be either x or y. Therefore X is closed under the formation of lines. However, since $1+1+1=1$ we have the affine combination

$$(0,0)+(1,0)+(0,1)=(1,1) \notin X,$$

so X is not affine closed.

Theorem 318. *[Exercise 16.4,16.5]*

(1) *A flat contains the origin* 0 *if and only if it is a subspace.*
(2) *A flat* X *in* V *is a subspace if and only if for some* $x \in X$ *we have* $rx \in X$ *for some* $1 \neq r \in F$.

Proof. (1) follows from the basic properties of cosets. For (2), write $X = x+S$ where S is a subspace of V. Since $x, rx \in X$ we have $x - rx = (1-r)x \in S$, so $x \in S$. Therefore $x - x = 0 \in S$ and the result follows from (1). □

Theorem 319. *[Exercise 16.6] The join of a collection* $C = \{x_i + S_i \mid i \in K\}$ *of flats in* V *is the intersection of all flats that contain all flats in* C.

Proof. Obvious. □

Theorem 320. *[Exercise 16.8-16.10] Let* V *be an* n*-dimensional vector space over a field* F *with* $n \geq 1$. *Let* $X = x+S$ *and* $Y = y+T$ *be flats in* V.

(1) *If* $\dim(X) = \dim(Y)$ *and* $X \parallel Y$, *then* $S = T$.
(2) *If* X *and* Y *are disjoint hyperplanes, then* $S = T$.
(3) *Let* $v \notin X$. *There is exactly one flat containing* v, *parallel to* X *and having the same dimension as* X.

Proof. For (1) we have $S \subseteq T$ or $T \subseteq S$. In either case, $\dim(S) = \dim(T)$ implies that $S = T$. For (2), suppose that $S \neq T$. There is some $v \in S \setminus T$ or $v \in T \setminus S$; in either case $T + \langle v \rangle$ or $S + \langle v \rangle$ spans V since $\dim(S) = \dim(T) = n-1$, so $S + T = V$. Now $x - y = s + t$ for some $s \in S$ and $t \in T$, so

$$x - s = y + t \in X \cap Y.$$

For (3), write $X = x+S$. The flat $v+S$ satisfies the desired properties. Suppose that $y+T$ contains v, is parallel to $v+S$ and has the same dimension as $v+S$. Then $S = T$ by (1) and $y+T = y+S = v+S$ since $v \in y+T$. □

Theorem 321. *[Exercise 16.11] Let* V *be a finite-dimensional vector space over a field* F *with* $\mathrm{char}(F) \neq 2$.

(1) *The join $X \vee Y$ of two flats may not be the set of all lines connecting all points in the union of these flats.*
(2) *If $X = x + S$ and $Y = y + T$ are flats with $X \cap Y \neq \emptyset$, then $X \vee Y$ is the union of all lines $\overline{xy}$ where $x \in X$ and $y \in Y$.*
(3) *If $\operatorname{char}(F) = 2$ then (2) does not necessarily hold.*

Proof. Let $X \subset \mathbb{R}^2$ be the x-axis and let $Y = X + 1$. We have $X \vee Y = \mathbb{R}^2$, but there is no vertical line of length 2 that joins two points in $X \cup Y$. If X and Y are flats with $X \cap Y \neq \emptyset$, then $X \vee Y = p + (S + T)$ for some $p \in X \cap Y$ by Theorem 16.5. If $z = p + s + t \in X \vee Y$ where $s \in S$ and $t \in T$ then

$$z = \frac{1}{2}(p + 2s) + \frac{1}{2}(p + 2t)$$

is on the line joining the points $p + 2s \in X$ and $p + 2t \in Y$. This proves (2). If $\operatorname{char}(F) = 2$ then the flats

$$X = \{(0,0),(1,0)\} \quad \text{and} \quad Y = \{(0,0),(0,1)\}$$

satisfy $X \cap Y \neq \emptyset$ and $X \vee Y = \mathbb{Z}_2^2$. But Example 317 shows that $X \cup Y$ is two-affine closed, so $(1,1)$ is not on any line $\overline{xy}$ where $x \in X$ and $y \in Y$. □

Theorem 322. *[Exercise 16.12] If $X \parallel Y$ and $X \cap Y = \emptyset$, then*

$$\dim(X \vee Y) = \max\{\dim(X), \dim(Y)\} + 1.$$

Proof. This follows directly from Theorem 16.6 and the fact that $S \subseteq T$ or $T \subseteq S$. □

Theorem 323. *[Exercise 16.13] Let $\dim(V) = 2$.*

(1) *The join of any two distinct points is a line.*
(2) *The intersection of any two nonparallel lines is a point.*

Proof. We apply Theorem 16.6 here. If X and Y are distinct points then $\dim(X \vee Y) = 1+1-0 = 2$. If $X = x+\langle v\rangle$ and $Y = y+\langle w\rangle$ are two nonparallel lines then $\langle v\rangle+\langle w\rangle = V$ since v and w are linearly independent, so $x - y = av + bw$ for some $a, b \in F$ and

$$x - av = y + bw \in X \cap Y.$$

But $\dim(X \cap Y) = 1$, so this proves (2). □

Theorem 324. *[Exercise 16.14] Let $\dim(V) = 3$.*

(1) *The join of any two distinct points is a line.*
(2) *The intersection of any two nonparallel planes is a line.*
(3) *The join of any two lines whose intersection is a point is a plane.*
(4) *The intersection of two coplanar nonparallel lines is a point.*
(5) *The join of any two distinct parallel lines is a plane.*

(6) *The join of a line and a point not on that line is a plane.*
(7) *The intersection of a plane and a line not parallel to that plane is a point.*

Proof. (1) is obvious. Let $X = x + S$ and $Y = y + T$ be two planes as in (2). We have $S + T = V$, so $x - y \in s + t$ and $x - s = y + t \in X \cap Y$. Furthermore,

$$\dim(S \cap T) = \dim(S) + \dim(T) - \dim(S + T) = 2 + 2 - 3 = 1,$$

so $X \cap Y$ is a line. Let X, Y be two lines as in (3); we have

$$\dim(X \vee Y) = 1 + 1 - 0 = 2$$

by Theorem 16.6. Part (4) follows from applying part (2) in Theorem 323. Let X, Y be lines as in (5); we have

$$\dim(X \vee Y) = \dim(S + T) + 1 = 2.$$

For (6), let $X = x + S$ be a line and let $Y = y + \{0\}$ be a point not on that line. Then

$$\dim(X \vee Y) = \dim(S + \{0\}) + 1 = 2.$$

For (7), let $X = x + S$ be a plane and let $Y = y + \langle v \rangle$ be a line not parallel to that plane, where $v \notin S$. Then $S + \langle v \rangle = V$, so $x - y \in s + av$ and $x - s = y + av \in X \cap Y$. We have $\dim(S \cap \langle v \rangle) = 0$ since $v \notin S$. □

Chapter 17. Singular Values and the Moore-Penrose Inverse

Theorem 325. *[Exercise 17.1] For any $\tau \in \mathcal{L}(U)$, the singular values of τ^* are the same as those of τ.*

Proof. Let $U = \langle u_1 \rangle \odot \cdots \odot \langle u_n \rangle$ be a decomposition of U where $u_1, \ldots, u_n$ are eigenvectors of $\tau^*\tau$ with corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$. Then

$$\tau\tau^*(\tau u_i) = \tau \lambda_i u_i = \lambda_i \tau u_i$$

for each i, and furthermore

$$\langle \tau u_i, \tau u_j \rangle = \langle u_i, \tau^*\tau u_j \rangle = \overline{\lambda_j} \langle u_i, u_j \rangle,$$

which shows that $\{\tau u_1, \ldots, \tau u_n\}$ is an orthogonal set. Therefore $\tau u_1, \ldots, \tau u_n$ are linearly independent eigenvectors of $\tau\tau^*$ with eigenvalues $\lambda_1, \ldots, \lambda_n$. Since $\tau^*\tau$ and $\tau\tau^*$ have the same eigenvalues, τ and τ^* have the same singular values.

An alternative proof is to note that if $A = P\Sigma Q^*$ where P, Q are unitary and Σ is diagonal with nonnegative entries then

$$A^* = Q\Sigma^* P^* = Q\Sigma P^*,$$

so the singular values of A^* are the same as those of A by the uniqueness of Σ. □

Example 326. [Exercise 17.2] Find the singular values and the singular value decomposition of the matrix

$$A = \begin{bmatrix} 3 & 1 \\ 6 & 2 \end{bmatrix}.$$

Find A^+.

We compute

$$A^*A = \begin{bmatrix} 3 & 6 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 3 & 1 \\ 6 & 2 \end{bmatrix} = \begin{bmatrix} 45 & 15 \\ 15 & 5 \end{bmatrix},$$

which has orthonormal eigenvectors and eigenvalues

$$u_1 = \begin{bmatrix} 3/\sqrt{10} \\ 1/\sqrt{10} \end{bmatrix} \quad \text{with} \quad \lambda_1 = 50,$$
$$u_2 = \begin{bmatrix} -1/\sqrt{10} \\ 3/\sqrt{10} \end{bmatrix} \quad \text{with} \quad \lambda_2 = 0.$$

The only singular value is $s_1 = 5\sqrt{2}$, with

$$v_1 = \frac{1}{5\sqrt{2}} \begin{bmatrix} 3 & 1 \\ 6 & 2 \end{bmatrix} \begin{bmatrix} 3/\sqrt{10} \\ 1/\sqrt{10} \end{bmatrix} = \begin{bmatrix} 1/\sqrt{5} \\ \sqrt{4}/\sqrt{5} \end{bmatrix},$$
$$v_2 = \begin{bmatrix} -\sqrt{4}/\sqrt{5} \\ 1/\sqrt{5} \end{bmatrix}.$$

Therefore the singular value decomposition of A is

$$\begin{bmatrix} 3 & 1 \\ 6 & 2 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{5} & -\sqrt{4}/\sqrt{5} \\ \sqrt{4}/\sqrt{5} & 1/\sqrt{5} \end{bmatrix} \begin{bmatrix} 5\sqrt{2} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 3/\sqrt{10} & 1/\sqrt{10} \\ -1/\sqrt{10} & 3/\sqrt{10} \end{bmatrix},$$

and the MP inverse is

$$A^+ = Q\Sigma' P^* = \begin{bmatrix} 3/50 & 3/25 \\ 1/50 & 1/25 \end{bmatrix}.$$

Example 327. [Exercise 17.4] Let $X = \begin{bmatrix} x_1 & \cdots & x_m \end{bmatrix}^T$ be a column matrix over $\mathbb{C}$. Find a singular value decomposition of X.

We have

$$X^*X = \begin{bmatrix} |x_1|^2 + \cdots + |x_m|^2 \end{bmatrix} = \begin{bmatrix} \|X\|^2 \end{bmatrix},$$

which has an eigenvector [1] with eigenvalue $\|X\|^2$. The left singular vector is

$$v_1 = \frac{1}{\|X\|} \begin{bmatrix} x_1 & \cdots & x_m \end{bmatrix}^T,$$

and $v_2, \ldots, v_m$ are vectors chosen so that $\{v_1, \ldots, v_m\}$ is an orthonormal basis for $\mathbb{C}^m$. The singular value decomposition of X is then

$$X = \begin{bmatrix} v_1 & \cdots & v_m \end{bmatrix} \begin{bmatrix} \|X\| \\ 0 \\ \vdots \\ 0 \end{bmatrix} [1].$$

Theorem 328. *[Exercise 17.5] Let $A \in \mathcal{M}_{m,n}(F)$ and let $B \in \mathcal{M}_{m+n,m+n}(F)$ be the square matrix*

$$B = \begin{bmatrix} 0 & A \\ A^* & 0 \end{bmatrix}.$$

Counting multiplicity, the nonzero eigenvalues of B are precisely the singular values of A together with their negatives.

Proof. If

$$\begin{bmatrix} 0 & A \\ A^* & 0 \end{bmatrix} \begin{bmatrix} v \\ u \end{bmatrix} = \begin{bmatrix} Au \\ A^*v \end{bmatrix} = \begin{bmatrix} \lambda v \\ \lambda u \end{bmatrix}$$

with $\lambda \neq 0$ then

$$A^*Au = A^*\lambda v = \lambda^2 u,$$

so $|\lambda|$ is a singular value of A with right singular vector u. Conversely, suppose that $s \neq 0$ is a singular value of A with right singular vector u so that $A^*Au = s^2u$. Let $v = (1/s)Au$. Then

$$\begin{bmatrix} 0 & A \\ A^* & 0 \end{bmatrix} \begin{bmatrix} v \\ u \end{bmatrix} = \begin{bmatrix} Au \\ A^*v \end{bmatrix} = \begin{bmatrix} sv \\ su \end{bmatrix}$$

and

$$\begin{bmatrix} 0 & A \\ A^* & 0 \end{bmatrix} \begin{bmatrix} v \\ -u \end{bmatrix} = \begin{bmatrix} -Au \\ A^*v \end{bmatrix} = \begin{bmatrix} -sv \\ su \end{bmatrix},$$

so s and $-s$ are eigenvalues of B with eigenvectors (v, u) and $(v, -u)$. □

Theorem 329. *[Exercise 17.6] For any matrix $A \in \mathcal{M}_{m,n}(F)$, its adjoint A^*, its transpose A^T and its conjugate $\overline{A}$ all have the same singular values. If U and U' are unitary, then A and UAU' have the same singular values.*

Proof. Theorem 325 shows that A^* has the same singular values as A. If $v \in F^n$ is an eigenvector of A^*A with real eigenvalue λ then $(\overline{A}^*\overline{A})\overline{v} = \overline{A^*Av} = \overline{\lambda v} = \lambda\overline{v}$, so λ is also an eigenvalue of $\overline{A}^*\overline{A}$. By symmetry, A has exactly the same singular values as $\overline{A}$. Therefore $A^T = \overline{A^*}$ has the same singular values as A. If U and U' are unitary, then

$$(UAU')^*(UAU') = (U')^*A^*U^*UAU' = (U')^*A^*AU',$$

so A and UAU' have the same singular values. □

The Frobenius Norm.

Definition 330. If $A = (a_{i,j})$ is an $m \times n$ complex matrix, then the **Frobenius norm** of A is

$$\|A\|_F = \left(\sum_{i,j} |a_{i,j}|^2 \right)^{1/2}.$$

Theorem 331. *Let $A \in \mathcal{M}_{m,n}(\mathbb{C})$, $B \in \mathcal{M}_{n,p}(\mathbb{C})$ and $U \in \mathcal{M}_{m,m}(\mathbb{C})$.*

(1) $\|A\|_F^2 = \operatorname{tr}(A^*A) = \operatorname{tr}(AA^*)$.
(2) *If U is unitary then* $\|UA\|_F = \|A\|_F$.
(3) $\|AB\|_F \leq \|A\|_F \|B\|_F$.

Proof. Part (1) follows from Theorem 221. If U is unitary then

$$\|UA\|_F^2 = \operatorname{tr}(A^*U^*UA) = \operatorname{tr}(A^*A) = \|A\|_F^2,$$

which proves (2). For (3), applying the Cauchy-Schwarz inequality gives

$$\begin{aligned}
\|AB\|_F^2 &= \sum_{i=1}^{m} \sum_{j=1}^{p} \left| \sum_{k=1}^{n} A_{i,k} B_{k,j} \right|^2 \\
&\leq \sum_{i=1}^{m} \sum_{j=1}^{p} \left(\sum_{k=1}^{n} |A_{i,k}|^2 \right) \left(\sum_{k=1}^{n} |B_{k,j}|^2 \right) \\
&= \left(\sum_{i=1}^{m} \sum_{k=1}^{n} |A_{i,k}|^2 \right) \left(\sum_{k=1}^{n} \sum_{j=1}^{p} |B_{k,j}|^2 \right) \\
&= \|A\|_F^2 \|B\|_F^2 .
\end{aligned}$$

□

Theorem 332. *[Exercise 17.8] If $s_1, \ldots, s_k$ are the singular values of A, then $\|A\|_F^2 = \sum s_i^2$ (cf. Theorem 223).*

Proof. Let $A = P\Sigma Q^*$ be the singular value decomposition of A. Since P and Q are unitary,

$$\|A\|_F^2 = \|P\Sigma Q^*\|_F^2 = \|\Sigma\|_F^2 = \sum s_i^2$$

by Theorem 331. □

Chapter 18. An Introduction to Algebras

Theorem 333. *Any associative F-algebra A is isomorphic to a subalgebra of the endomorphism algebra $\mathcal{L}_F(A)$. In fact, if μ_a is the left multiplication map defined by*

$$\mu_a x = ax$$

*then the map $\mu : A \to \mathcal{L}_F(A)$ given by $a \mapsto \mu_a$ is an algebra embedding, called the **left regular representation** of A.*

Proof. It is clear that μ is a homomorphism. If $\mu_a = 0$ then $0 = \mu_a 1 = a1 = a$, so μ is injective. □

Theorem 334. *[Exercise 18.1] The subalgebra generated by a nonempty subset X of an algebra A is the subspace spanned by the products of finite subsets of elements of X:*

$$\langle X \rangle_{\text{alg}} = \langle x_1 \cdots x_n \mid x_i \in X \rangle .$$

Proof. Obvious. Note that $1 \in \langle X \rangle_{\text{alg}}$ due to the case $n = 0$ which is an empty product. □

Theorem 335. *[Exercise 18.3] Let A, B be associative unital algebras over a field F. If $\varphi : A \to B$ is an algebra homomorphism, then* $\ker(\varphi)$ *is an ideal of A.*

Proof. Firstly, note that $0 \in \ker(\varphi)$. If $x, y \in \ker(\varphi)$ then $\varphi(x + y) = \varphi x + \varphi y = 0$, so $x + y \in \ker(\varphi)$. Similarly, $x - y \in \ker(\varphi)$. If $a, b \in A$ then

$$\varphi(axb) = (\varphi a)(\varphi x)(\varphi b) = 0.$$

This shows that $\ker(\varphi)$ is an ideal of A. □

Theorem 336. *[Exercise 18.5] If A is an algebra over a field F and $S \subseteq A$ is nonempty, define the **centralizer** $C_A(S)$ of S to be the set of elements of A that commute with all elements of S. The centralizer $C_A(S)$ is a subalgebra of A.*

Proof. Clearly $0, 1 \in C_A(S)$. If $x, y \in C_A(S)$, $r \in F$ and $a \in A$, we have

$$(x + y)a = xa + ya = ax + ay = a(x + y)$$

and

$$xya = xay = axy$$

and

$$rxa = rax = a(rx).$$

□

Example 337. [Exercise 18.6] Show that $\mathbb{Z}_6$ is not an algebra over any field.

The center of $\mathbb{Z}_6$ is simply $\mathbb{Z}_6$, and the only nontrivial subring (with identity) is $\mathbb{Z}_6$. This is not a field, so Theorem 18.1 shows that $\mathbb{Z}_6$ cannot be an algebra over any field.

Remark 338. [Exercise 18.7] Let $A = F[a]$ be the algebra generated over F by a single algebraic element a.

(1) A is isomorphic to the quotient algebra $F[x]/\langle p(x)\rangle$, where $\langle p(x)\rangle$ is the ideal generated by some $p(x) \in F[x]$.
(2) What can you say about $p(x)$?
(3) What is the dimension of A?
(4) What happens if a is not algebraic?

Let $m_a(x)$ be the minimal polynomial of a. Define $\varphi : F[x] \to A$ by $f(x) \mapsto f(a)$. If $f(a) = 0$ then $m_a(x)$ divides $f(x)$, so $\ker(\varphi) \subseteq \langle m_a(x)\rangle$. But $m_a(a) = 0$, so $\langle m_a(x)\rangle \subseteq \ker(\varphi)$. Therefore

$$A = \operatorname{im}(\varphi) \cong F[x]/\langle m_a(x)\rangle$$

since $\operatorname{im}(\varphi) \cong F[x]/\langle m_a(x)\rangle$ is a field and A is by definition the smallest field containing a. The dimension of A is $\dim(m_a(x))$. If a is not algebraic, then there may not be any minimal polynomial and this construction does not work.

Theorem 339. *Let G, H be groups and let F be a field. For any homomorphism $\varphi : G \to H$, there exists a unique homomorphism $\psi : F[G] \to F[H]$ such that*

$$\psi(r_1 g_1 + \cdots + r_n g_n) = r_1\varphi(g_1) + \cdots + r_n\varphi(g_n).$$

Proof. Define ψ as specified. Clearly $\psi(1) = \varphi(1) = 1$. Let $x, y \in F[G]$ and $r \in F$; write

$$x = \sum_{i=1}^{n} r_i g_i \quad \text{and} \quad y = \sum_{i=1}^{n} s_i g_i$$

where $g_1, \ldots, g_n \in G$ (and some r_i's or s_i's may be zero). Then $\psi(rx) = r\psi(x)$,

$$\psi(x + y) = \sum_{i=1}^{n} (r_i + s_i)\varphi(g_i) = \psi(x) + \psi(y)$$

and

$$\begin{aligned}\psi(xy) &= \psi\left(\sum_{i=1}^{n}\sum_{j=1}^{n} r_i s_j g_i g_j\right) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} r_i s_j \varphi(g_i g_j)\end{aligned}$$

$$= \sum_{i=1}^{n} r_i \varphi(g_i) \sum_{j=1}^{n} s_j \varphi(g_j)$$
$$= \psi(x)\psi(y).$$

□

Theorem 340. *[Exercise 18.8] Let $G = \{1 = a_0, \ldots, a_n\}$ be a finite group. For $x \in F[G]$ of the form*

$$x = r_1 a_1 + \cdots + r_n a_n$$

let $T(x) = r_1 + \cdots + r_n$. Then $T : F[G] \to F$ is an algebra homomorphism, where F is an algebra over itself.

Proof. Let $z : G \to \{1\}$ be the trivial homomorphism. By Theorem 339, there exists a unique homomorphism f from $F[G] \to F[\{1\}]$ such that

$$f(r_1 a_1 + \cdots + r_n a_n) = (r_1 + \cdots + r_n)1.$$

The result follows from the fact that $r1 \mapsto r$ is an isomorphism from $F[\{1\}]$ to F. □

Theorem 341. *[Exercise 18.9] A homomorphism $\sigma : A \to B$ of F-algebras induces an isomorphism $\overline{\sigma} : A/\ker(\sigma) \to \mathrm{im}(\sigma)$ defined by $\overline{\sigma}(a + \ker(\sigma)) = \sigma a$.*

Proof. If $a + \ker(\sigma) = b + \ker(\sigma)$ then $a - b \in \ker(\sigma)$, so $\sigma a - \sigma b = \sigma(a - b) = 0$. This shows that $\overline{\sigma}$ is well-defined, and it is now easy to see that $\overline{\sigma}$ is an isomorphism. □

Example 342. [Exercise 18.10] The quaternion field $\mathbb{H}$ is an $\mathbb{R}$-algebra and a field.

We only check that nonzero elements of $\mathbb{H}$ are invertible. Let $q = a + bi + cj + dk \neq 0$, let $\overline{q} = a - bi - cj - dk$, and define $|q|^2 = a^2 + b^2 + c^2 + d^2$. It is easy to check that $q\overline{q} = |q|^2$, so the inverse of q is $\overline{q}/|q|^2$ (noting that $|q|^2 \neq 0$ since $q \neq 0$).

Example 343. [Exercise 18.11] Describe the left regular representation of the quaternions using the ordered basis $B = (1, i, j, k)$.

Let $q = a + bi + cj + dk$. We compute

$$qi = -b + ai + dj - ck,$$
$$qj = -c - di + aj + bk,$$
$$qk = -d + ci - bj + ak,$$

so the matrix representation of q is

$$\begin{bmatrix} a & -b & -c & -d \\ b & a & -d & c \\ c & d & a & -b \\ d & -c & b & a \end{bmatrix}.$$

Theorem 344. *[Exercise 18.12] Let S_n be the group of permutations (bijective functions) of the ordered set $X = (x_1, \ldots, x_n)$ under composition.*

1. *Each $\sigma \in S_n$ defines a linear isomorphism τ_σ on the vector space V with basis X over a field F. This defines an algebra homomorphism $f : F[S_n] \to \mathcal{L}_F(V)$ with the property that $f(\sigma) = \tau_\sigma$.*
2. *The representation f is faithful if and only if $n < 3$.*

Proof. To each $\sigma \in S_n$ there is an associated permutation matrix in $\mathcal{M}_n(F)$. The algebra homomorphism f can then be defined by extending linearly to the group algebra $F[S_n]$. If $n \geq 4$ then $|S_n| = n! > n^2 = \dim(\mathcal{L}_F(V))$ implies that the set $\{\tau_\sigma \mid \sigma \in S_n\}$ is linearly dependent. There exist $r_1, \ldots, r_{n!} \in S_n$ such that

$$r_1\tau_{\sigma_1} + \cdots + r_{n!}\tau_{\sigma_{n!}} = 0,$$

so $f(r_1\sigma_1 + \cdots + r_{n!}\sigma_{n!}) = 0$ and the representation cannot be faithful. If $n = 2$ the only two matrices are

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

which are clearly linearly independent. If $n = 3$ we have the linear combination

$$-\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} - \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} = 0,$$

so the representation cannot be faithful. □

Theorem 345. *[Exercise 18.13,18.14] Let V be a vector space over a field F.*

1. *The algebra $\mathcal{L}_F(V)$ has center*
$$Z = \{r1 \mid r \in F\}$$
and so $\mathcal{L}_F(V)$ is central.
2. *The set I of all elements of $\mathcal{L}_F(V)$ that have finite rank is an ideal of $\mathcal{L}_F(V)$ and is contained in all other nonzero ideals of $\mathcal{L}_F(V)$.*
3. *$\mathcal{L}_F(V)$ is simple if and only if V is finite-dimensional.*

Proof. Theorem 38 proves (1). Let J be an ideal of $\mathcal{L}_F(V)$. Theorem 18.3 proves that any operator of rank 1 is an element of J. Let $\tau \in \mathcal{L}_F(V)$ with rank k and write

$$V = \langle v_1 \rangle \oplus \cdots \oplus \langle v_k \rangle \oplus W$$

where $\text{im}(\tau) = \langle v_1 \rangle \oplus \cdots \oplus \langle v_k \rangle$. By Theorem 2.25, there exists a resolution of the identity

$$\iota = \rho_1 + \cdots + \rho_k + \rho$$

where each ρ_i is a projection onto $\langle v_i \rangle$ and ρ is a projection onto W. Now

$$\tau = \rho_1\tau + \cdots + \rho_k\tau$$

and each $\rho_i\tau \in J$, so $\tau \in J$. This proves (2). If $\mathcal{L}_F(V)$ is simple then $I = \mathcal{L}_F(V)$, where I is defined as in (2). Therefore the identity operator has finite rank and V is finite-dimensional. Conversely, if V is finite-dimensional then $I = \mathcal{L}_F(V)$. If J is a nonzero ideal of $\mathcal{L}_F(V)$ then (2) shows that $I \subseteq J$, which implies that $J = \mathcal{L}_F(V)$. This proves (3). □

Corollary 346. *[Exercise 18.15] The matrix algebras $\mathcal{M}_n(F)$ are central and simple.*

Theorem 347. *[Exercise 18.16] An element $a \in A$ is **left-invertible** if there is a $b \in A$ for which $ba = 1$, in which case b is called the **left inverse** of a. Similarly, $a \in A$ is **right-invertible** if there is a $b \in A$ for which $ab = 1$, in which case b is called a **right inverse** of a. Left and right inverses are called **one-sided inverses** and an ordinary inverse is called a **two-sided inverse**. Let $a \in A$ be algebraic over F.*

(1) *$ab = 0$ for some $b \neq 0$ if and only if $ca = 0$ for some $c \neq 0$. (Does c necessarily equal b?)*
(2) *If a has a one-sided inverse b, then b is a two-sided inverse. (Does this hold if a is not algebraic?)*
(3) *Let $a, b \in A$ be algebraic. Then ab is invertible if and only if a and b are invertible, in which case ba is also invertible.*

Proof. If $ab = 0$ and $b \neq 0$ then $m_a(x) = xp(x)$ for some $p(x) \in F[x]$ by Theorem 18.4. Now $ap(a) = p(a)a = 0$ and $p(a) \neq 0$ since $\deg(p(x)) < \deg(m_a(x))$. By symmetry, $ca = 0$ implies that $ab = 0$. This proves (1). Note that c does not necessarily equal b for otherwise we can multiply c by 2 (if $\mathrm{char}(F) \neq 2$). For (2), suppose that $ba = 1$. If $m_a(x) = xp(x)$ then $0 = ap(a)$, so $0 = bap(a) = p(a)$ gives a contradiction since $\deg(p(x)) < \deg(m_a(x))$. By Theorem 18.4, a has a two-sided inverse and $b = baa^{-1} = a^{-1}$. If a is not algebraic then (2) does not necessarily hold, since there exist injective linear maps on infinite-dimensional vector spaces that are not surjective, i.e. linear maps with left inverses but not right inverses. For (3), if ab is invertible then $abc = cab = 1$ for some $c \in A$, so $a^{-1} = bc$ and $b^{-1} = ca$ by (2). Also, $baa^{-1}b^{-1} = 1$, so $(ba)^{-1} = a^{-1}b^{-1}$. □