MAT 442: Advanced Linear Algebra

Arizona State University

Fall 2020

Contents

1 Vector Spaces

1.1 Fields

Definition 1 A *field* is a triple $(\mathbb{F}, +, \cdot)$ with $+ : \mathbb{F} \times \mathbb{F} \to \mathbb{F}$ and $\cdot : \mathbb{F} \times \mathbb{F} \to \mathbb{F}$ that satisfies the following axioms.

(1) For all $a, b, c \in \mathbb{F}$, $(a + b) + c = a + (b + c)$ (Associativity of addition)

(2) For all $a, b, c \in \mathbb{F}$, $(a \cdot b) \cdot c = a \cdot (b \cdot c)$ (Associativity of multiplication)

(3) For all $a, b \in \mathbb{F}$, $a + b = b + a$ (Commutativity of addition)

(4) For all $a, b \in \mathbb{F}$, $a \cdot b = b \cdot a$ (Commutativity of multiplication)

(5) For all $a, b, c \in \mathbb{F}$, $a \cdot (b + c) = a \cdot b + a \cdot c$ (Distributive law)

(6) There exists $0 \in \mathbb{F}$ such that $a + 0 = a$ for all $a \in \mathbb{F}$ (Identity element of addition)

(7) There exists $0 \in \mathbb{F}$ such that $a \cdot 0 = a$ for all $a \in \mathbb{F}$ (Identity element of multiplication)

(8) For all $a \in \mathbb{F}$, there exists $(-a) \in \mathbb{F}$ such that $a + (-a) = 0$ (Additive inverse)

(9) For all $a \in \mathbb{F}$, there exists $a^{-1} \in \mathbb{F}$ such that $a \cdot a^{-1} = 0$ (Multiplicative inverse)

Examples of Fields

(1) $(\mathbb{R}, +, \cdot)$

(2) $(\mathbb{C}, +, \cdot)$

(3) $(\mathbb{Q}, +, \cdot)$

(4) $(\{a + b\sqrt{2} : a, b \in \mathbb{Q}\}, +, \cdot)$

(5) $(\mathbb{Z}_2, +, \cdot)$

(6) $(\mathbb{Z}_p, +, \cdot)$, where p prime

1.2 Vector Spaces

Definition 1 A *vector space* V over a field F is a triple $(V, +, \cdot)$ where $+ : V \times V \to V$ (addition), $\cdot : F \times V \to V$ (scalar multiplication) and the following conditions hold.

(VS 1) For $x, y \in V$, $x + y = y + x$ (Commutativity of addition)

(VS 2) For $x, y, z \in V$, $(x + y) + z = x + (y + z)$ (Associativity of addition)

(VS 3) There exists an element $0_V \in V$ such that $0_V + x = x$ for all $x \in V$ (Identity element of addition)

(VS 4) For $x \in V$, there is $y_x \in V$ such that $x + y_x = 0_V$ (Inverse elements of addition)

(VS 5) For $x \in V$, $1 \cdot x = x$ (Identity element of scalar multiplication)

(VS 6) For $a, b \in \mathbb{F}$, $x \in V$, $(ab)x = a(bx)$ (Compatibility of scalar multiplication)

(VS 7) For $a \in \mathbb{F}$, $x, y \in V$, $a(x + y) = ax + ay$ (Distributivity of scalar multiplication with respect to vector addition)

(VS 8) For $a, b \in \mathbb{F}$, $x \in V$, $(a + b)x = ax + bx$ (Distributivity of scalar multiplication with respect to field addition)

Conventions

- We will often identify the set V with the vector space V.
- We will write av instead of $a \times v$.
- We will often write 0 for 0_V when V is clear from the context.

An $m \times n$ matrix with entries from a field F is a function $A : \{1, \ldots, m\} \times \{1, \ldots, n\} \to \mathbb{F}$. We often organize the values in a rectangular array with m rows and n columns and use A_{ij} for $A(i, j)$.

Examples of Vector Fields

(1) $\mathbb{F}^n = \{(a_1, \ldots, a_n) : a_i \in \mathbb{F}\}$

(2) $\{f : S \to \mathbb{F} : S \neq \emptyset\}$. Also notated $\mathcal{F}(S, F)$.

(3) Set of all sequences over $\mathbb{F}$

(4) $P(\mathbb{F})$. Set of all polynomials with coefficients in $\mathbb{F}$.

(5) $M_{m \times n}(\mathbb{F})$

Example: Is the empty set a vector space?

Answer. No. By rule (VS 3), there must exist $0_V \in V$, such that $0_V + x = x$ for all $x \in V$, but since $V = \emptyset$, $\nexists 0_V \in V$.

Example: Decide if V is a vector space.

(1) $V = \{(a_1, a_2) : a_1, a_2 \in \mathbb{R}\}$, $(a_1, a_2) + (b_1, b_2) = (a_1 + b_1, a_2 - b_2)$, $c(a_1, a_2) = (ca_1, ca_2)$

Answer. No. V violates axiom (VS 1).

(2) $V = \{(a_1, a_2) : a_1, a_2 \in \mathbb{R}\}$, $(a_1, a_2) + (b_1, b_2) = (a_1 + b_1, 0)$, $c(a_1, a_2) = (ca_1, 0)$

Answer. No. V violates axiom (VS 5).

(3) $V = \{(a_1, a_2) : a_1, a_2 \in \mathbb{R}\}$, $(a_1, a_2) + (b_1, b_2) = (a_1 + b_1, a_2 b_2)$, $c(a_1, a_2) = (ca_1, a_2)$

Answer. No. V violates axiom (VS 4).

Theorem 1.1. (Cancellation Law for Vector Addition) Let $x, y, z \in V$. If $x + z = y + z$, then $x = y$.

Proof. Suppose $x + z = y + z$. Let w be such that $z + w = 0_V$. Then

$$\begin{aligned} x &= x + 0_V \\ &= x + (z + w) \\ &= (x + z) + w \\ &= (y + z) + w \\ &= y + (z + w) \\ &= y + 0_V \\ &= y. \end{aligned}$$

□

Corollary 2

- There is unique $0_V \in V$ such that $0_V + x = x$ for all $x \in V$.

  Proof. Let $v \in V$. Suppose that 0_V and $0'_V$ are zero elements.

  $$v + 0_V = v = v + 0'_V.$$

  Thus, by cancellation law, $0_V = 0'_V$. □

- For every $x \in V$, there is unique $y_x \in V$ such that $x + y_x = 0_V$. *Proof.* Let $x \in V$. Suppose that y_x and y'_x are inverses of x. Then

$$x + y_x = 0 = x + y'_x.$$

Thus, by cancellation law, $y_x = y'_x$. □

Theorem 1.2. Let $x \in V$, $a \in \mathbb{F}$. Then

- $0x = 0_V$;

Proof.

$$\begin{aligned} 0x + 0x &= (0+0)x \\ &= 0x \\ &= 0x + 0_V. \end{aligned}$$

Thus, by cancellation law, $0x = 0_V$. □

- $(-a)x = -(ax) = a(-x)$;

Proof. We can show

$$\begin{aligned} (-a)x + (ax) &= (-a+a)x \\ &= 0x \\ &= 0_V. \end{aligned}$$

Thus, the inverse of ax is unique, which implies $-ax = (-a)x$. We have $(-x) = (-1)x$, which implies $a(-x) = (a(-1))x = (-a)x$. □

- $a0_V = 0_V$.

Proof.

$$a0_V = a(0_V - 0_V) = a0_V - a0_V = 0_V.$$

□

1.3 Subspaces

Definition 2 Let $(V, +, \cdot)$ be a vector space and let $W \subseteq V$. Then W is called a *subspace* if $(W, +, \cdot)$ is a vector space.

Note: W is a subspace if

- $x + y \in W$ for $x, y \in W$

- $cx \in W$ for $x \in W$, $c \in \mathbb{F}$
- W has a zero vector 0_W
- For $x \in W$ there is $y_x \in W$ such that $x + y_x = 0_W$

Theorem 1.3. Let W be a subset of V a vector space $(V, +, \cdot)$. Then W is a subspace of V if and only if the following hold.

(1) $0_V \in W$

(2) if $x, y \in W$, then $x + y \in W$

(3) if $c \in \mathbb{F}$ and $x \in W$, then $cx \in W$

Proof. ($\Longrightarrow$): Suppose W is a subspace. Then (2) and (3) are satisfied. Then, $0_W + 0_W = 0_W = 0_W + 0_V$. Thus, by cancellation law, $0_W = 0_V$.

($\Longleftarrow$): Suppose $x \in W$. We have $-x = (-1)x$ and $-1 \in \mathbb{F}, x \in W$. Thus, $-x \in W$. $\square$

The *transpose* of an $m \times n$ matrix A is the $n \times m$ matrix A^t such that $(A^t)_{ij} = A_{ji}$. Note, $(A + B)^t = A^t + B^t$. Also, $(cA^t) = cA^t$. This is proven by,

$$\begin{aligned} ((A+B)^t)_{ij} &= (A+B)_{ji} \\ &= A_{ji} + B_{ji} \\ &= (A^t)_{ij} + (B^t)_{ij}. \end{aligned}$$

A is called *symmetric* if $A^t = A$.

An $n \times n$ matrix A is called *diagonal* if $A_{ji} = 0$ whenever $j \neq i$.

Theorem 1.4. Any intersection of subspaces of a vector space V is a subspace of V.

Definition 3

- Let $S_1, S_2 \subseteq V$. Then $S_1 + S_2$ is the set $\{x + y : x \in S_1, y \in S_2\}$
- V is called a direct sum of W_1 and W_2 if W_1, W_2 are subspaces of V such that $W_1 \cap W_2 = \{0\}$ and $W_1 + W_2 = V$. We then write $V = W_1 \oplus W_2$.

  Example: Let $V = \mathbb{R}^2$, $S_1 = \{(a, 0) : a \in \mathbb{R}\}$, $S_2 = \{(0, b) : b \in \mathbb{R}\}$. Then $S_1 + S_2$ generates V.

Examples of subspaces

(1) $\mathcal{C} = \{f \in \mathcal{F}(\mathbb{R}, \mathbb{R}) : f \text{ is continuous}\}$ is a subspace of $\mathcal{F}(\mathbb{R}, \mathbb{R})$.

(2) $P_n(\mathbb{R}) = \{f : f \text{ is a polynomial and } deg(f) \leq n\}$ is a subspace of $P(\mathbb{R})$.

(3) $D = \{A \in M_{n\times n}(\mathbb{R}) : A \text{ is diagonal}\}$ is a subspace of $M_{n\times n}(\mathbb{R})$.

(4) $W = \{A \in M_{n\times n}(\mathbb{R}) : tr(A) = 0\}$ is a subspace of $M_{n\times n}(\mathbb{R})$.

1.4 Linear combinations and systems of linear equations

Definition 4 Let S be a subset of a vector space V and let $v \in V$. Then v is called a linear combination of vectors of S if there exists $u_1, \dots, u_n \in S$ and $a_1, \dots, a_n \in \mathbb{F}$ such that

$$v = \sum_{i=1}^{n} a_i v_i.$$

When solving a system of linear equations we reduced to what is called the reduction echelon form by performing the following three operations.

- Interchange two equations
- Multiply an equation by a non-zero element from F
- Add a scalar multiple of one equation to another

Example: $v = (2, 6, 8) \in \mathbb{R}^3$ is a linear combination of $v_1 = (1, 2, 1)$, $v_2 = (-2, -4, -2)$, $v_3 = (0, 2, 3)$, $v_4 = (2, 0, -3)$, $v_5 = (-3, 8, 16)$. Specifically, $v = a_1 v_1 + \cdots + a_5 v_5$ where $(a_1, a_2, a_3, a_4, a_5) = (-4, 0, 7, 3, 0)$ is one solution.

Definition 5 Let S be a subset of a vector space V. If S is non-empty, we let $span(S)$ be the set of all linear combinations of vectors from S, and we set $span(\emptyset) = \{0_V\}$.

Theorem 1.5. Let S be a subset of a vector space V. Then $span(S)$ is a subspace of V. Moreover, if W is a subspace of V such that $S \subseteq W$, then $span(S) \subseteq W$.

Proof.

(1) (Show $span(S)$ is a subspace): If $S = \emptyset$, then $span(S) = \{0_V\}$, which is a subspace. Assume $S \neq \emptyset$. Let $v \in S$. Then,

   (a) $0v = 0_V \in span(S)$

   (b) Suppose $x, y \in span(S)$. Then $x = \sum_{i=1}^{n} a_i v_i$, $y = \sum_{i=1}^{m} b_i w_i$ where $v_1, \dots v_n, w_1, \dots, w_m \in S$ and $a_1, \dots, b_1, \dots, b_m \in \mathbb{F}$. Let

$$u_i = \begin{cases} v_i & i \leq n \\ w_{i-n} & i > n \end{cases}, \qquad c_i = \begin{cases} a_i & i \leq n \\ b_{i-n} & i > n. \end{cases}$$

   Then $x + y = \sum_{i=1}^{n+m} c_i u_i$ and $u_i \in S$, $c_i \in \mathbb{F}$ for $i = 1, \dots, n+m$. Thus, $x + y \in span(S)$.

   (c) Let $x \in span(S)$ and let $c \in \mathbb{F}$. Then $x = \sum_{i=1}^{n} a_i v_i$ where $v_1, \dots, v_n \in S$ and $a_1, \dots, a_n \in \mathbb{F}$. Then $cx = \sum_{i=1}^{n} (ca_i) v_i$ and $ca_i \in \mathbb{F}$.

Therefore $span(S)$ is a subspace.

(2) Let $x \in span(S)$. Then, $x = \sum_{i=1}^{n} a_i v_i$, where $a_i \in \mathbb{F}$ and $v_i \in S$. We will use induction on n.

- If $n = 1$, $x = a_1 v_1$, and $v_1 \in W$. Thus, $x \in W$.
- Suppose $x = \sum_{i=1}^{n} a_i v_i$. Then $x = \sum_{i=1}^{n-1}(a_i v_i) + a_n v_n$. Then $\sum_{i=1}^{n-1}(a_i v_i) \in W$ by inductive hypothesis and $a_n v_n \in W$. Thus $x \in W$.

Therefore, $span(S) \subseteq W$.

$\square$

Note: In particular, $span(span(S)) = span(S)$.
Proof.

(a) $span(S) \subseteq span(span(S))$

(b) Let $x = span(S)$. Then $span(S) \subseteq span(S)$ implies $span(S) \subseteq x$, which gives us $span(span(S)) \subseteq span(S)$. Therefore, $(span(span(S)) = span(S)$.

$\square$

Definition 6 A subset S of V *generates* (spans) V if $span(S) = V$.

Example: Let $V = M_{2\times 2}(\mathbb{R})$ and $S = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \right\}$. Then $span(S) = V$.

1.5 Linear dependence and linear independence

Definition 7 A subset S of a vector space V is called *linearly dependent* if there exists distinct vectors $u_1, \ldots, u_n \in S$ and $a_1, \ldots, a_n \in \mathbb{F}$ not all zero such that $a_1 u_1 + \cdots + a_n u_n = 0$.

Notes:

- The empty set is linearly dependent.
- $S = \{u\}$ is linearly dependent if and only if $u \neq 0$.
- A set is called linearly independent if and only if the only representation of 0 as linear combinations of its vectors are trivial.

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

Example: Let $S = \{(1,0,0,-1),(0,1,0,-1),(0,0,1,-1),(0,0,0,1)\} \subseteq \mathbb{R}^4$. We'll show S is linearly independent. Suppose

$$a_1(1,0,0,-1) + a_2(0,1,0,-1) + a_3(0,0,1,-1) + a_4(0,0,0,1) = (0,0,0,0).$$

We get $a_1 = 0, a_2 = 0, a_3 = 0, -a_1 - a_2 - a_3 + a_4 = 0$, which implies $a_1 = a_2 = a_3 = a_4 = 0$.

Example: $P_k(x) = x^k + \cdots + x^n$ where $1 \leq k \leq n$. We'll show that $\{P_0(x), \ldots, P_n(x)\}$ is linearly independent. Suppose $a_0P_0(x) + \cdots + a_nP_n(x) = 0$. The coefficient of x^i on the left hand side is $a_0 + \cdots + a_i$. Then we have

$$\begin{aligned} a_0 + \cdots + a_n &= 0 \\ a_0 + \cdots + a_{n-1} &= 0 \\ &\vdots \\ a_0 &= 0. \end{aligned}$$

Thus, $a_0 = \cdots = a_n = 0$.

Corollary 8 If $S_1 \subseteq S_2 \subseteq V$ and S_2 is linearly independent, then so is S_1.

Theorem 1.7. Let S be a linearly independent subset of V and let $v \in V \setminus S$. Then $S \cup \{v\}$ is linearly dependent if and only if $v \in span(S)$.

Proof. Let S be linearly independent, $v \notin S$. We'll show $S \cup \{v\}$ is linearly dependent if and only if $v \in span(S)$.

($\Longrightarrow$): Suppose $T = S \cup \{v\}$ is linearly dependent. Then, there exists $a_1, \ldots, a_n$ not all zero and there exists $u_1, \ldots, u_n \in T$ such that $a_1u_1 + \cdots + a_nu_n = 0_V$. Since S is linearly independent, $u_i = v$ for some $1 \leq i \leq n$. Say $u_1 = v$ and $a_i \neq 0$. Then $a_1u_1 = -a_2u_2 - \cdots - a_nu_n$. Thus $u_1 = -\frac{a_2}{a_1}u_2 - \cdots - \frac{a_n}{a_1}u_n$. Also $u_2, \ldots, u_n \in S$. Therefore $u_1 \in span(S)$.

($\Longleftarrow$): Let $v \in span(S)$. Then there exists $a_1, \ldots, a_n$ not all zero and there exists $v_1, \ldots, v_n \in S$ such that $v = a_1v_1 + \cdots + a_nv_n$. Then $v - a_1v_1 - \cdots - a_nv_n = 0_V$. The coefficient on v is $1 \neq 0$. Thus, $\{v\} \cup S$ is linearly dependent. $\square$

1.6 Bases and dimension

Definition 8 A *basis* β for a vector space V is a linearly independent subset of V that generates V.

Theorem 1.8. Let V be a vector space and let $\beta = \{u_1, \ldots, u_n\}$. Then β is a basis of V if and only if each v can be uniquely written as

$$v = a_1u_1 + \cdots + a_nu_n$$

where $a_1, \ldots, a_n \in \mathbb{F}$.

Proof. ($\Longrightarrow$): Suppose $\beta = \{u_1, \dots, u_n\}$ is a basis for a vector space V. Then $span(\beta) = V$ and every $v \in V$ can be written as a linear combination of $u_1, \dots, u_n$. Suppose $v = \sum_{i=1}^n a_i u_i$ and $v = \sum_{i=1}^n b_i u_i$. Then $0_V = \sum_{i=1}^n (a_i - b_i) u_i$. Thus $a_i = b_i = 0$ for every $i = 1, \dots, n$.

($\Longleftarrow$): Suppose every $v \in V$ can be written uniquely as $v = \sum_{i=1}^n a_i u_i$. Then $span(\{u_1, \dots, u_n\}) = V$. Suppose $a_1 u_1 + \cdots + a_n u_n = 0_V$. Note, $0_V = u_1 \cdot 0 + \cdots + u_n \cdot 0$. Since every vector has a unique representation as a linear vector, then $a_1 = \cdots = a_n = 0$. Thus, $\{u_1, \dots, u_n\}$ is linearly independent and a basis for V. $\square$

Theorem 1.9. Let S be a finite set that generates V. Then there is a subset of S which is a basis for V.

Proof. If $S = \emptyset$ or $S = \{0_V\}$, then $V = \{0_V\}$ and $\emptyset$ is a basis. Assume S contains a finite set of at least one nonzero vector v, which generates V. Then $\{v\}$ is linearly independent. Let $\{u_1, \dots, u_k\}$ is a maximal linearly independent subset of S. Then $S \subseteq span(\{u_1, \dots, u_k\})$. Then $span(S) \subseteq span(span(\{u_1, \dots, u_k\}))$. So, $span(S) = V \subseteq span(\{u_1, \dots, u_k\})$. Therefore $\{u_1, \dots, u_k\}$ is a basis for V. $\square$

Theorem 1.10. Let $G \subset V$, $|G| = n$ and $V = span(G)$. Suppose further that $L \subseteq V$, $|L| = m$ and L is linearly independent. Then $m \leq n$ and there exists a subset H of G such that $|H| = n - m$ and $L \cup H$ generates V.

Proof. (Outline)

- Induction on m. For the inductive step consider $L = \{v_1, \dots, v_{m+1}\}$
- Apply induction to $\{v_1, \dots, v_m\}$. Thus $m \leq n$ and there is a subset $H' \dots$
- H' can't be empty, say $H' = \{u_1, \dots, u_{n-m}\}$. Thus $m \leq n$.
- To find H for L, use the fact $H' \cup \{v_1, \dots, v_m\}$ generates V and substitute one of vectors from H' with v_{m+1}.

Proof. Let $|G| = n$, $span(G) = V$, $|L| = m$, L is linearly independent. Then, using induction on m,

(Base Case): If $m = 0$, then $L = \emptyset$, $|L| = 0$ and $H = G$ and $|H| = n = n - 0$.

(Inductive Step): Suppose $L = \{v_1, \dots, v_{m+1}\}$ is linearly independent. Let $L' = \{v_1, \dots, v_m\}$. By inductive hypothesis, there exists $m \leq n$ and there exists $H' \subseteq G$, such that $|H'| = n - m$ and $span(L' \cup H') = V$. Note, $H' \neq \emptyset$. Otherwise, $span(L') = V$. In particular, $v_{m+1} \in span(\{v_1, \dots, v_m\})$, but L is linearly independent, contradiction. Therefore, $H' \neq \emptyset$ and so $n - m > 0$. Thus, $m + 1 \leq n$. Say $H' = \{u_1, \dots, u_{n-m}\}$. Since $span(L' \cup H') = V$,

$$v_{m+1} = \sum_{i=1}^{m} a_i v_i + \sum_{i=1}^{n-m} b_i u_i.$$

Now, at least one of scalars b_i is nonzero, say $b_1 \neq 0$. Then, $b_1 u_1 = v_{m+1} - \sum_{i=1}^{m} a_i v_i - \sum_{i=2}^{n-m} b_i u_i$. So,

$$u_1 = \frac{1}{b_1} v_{m+1} - \sum_{i=1}^{m} \frac{a_i}{b_1} v_i - \sum_{i=2}^{n-m} \frac{b_i}{b_1} u_i,$$

which implies $u_1 \in span(\{v_1, \ldots, v_{m+1}\} \cup \{u_2, \ldots, u_{n-m}\})$. Thus, $u_1 \in span(L \cup \{u_2, \ldots, u_{n-m}\})$. Let $H = \{u_2, \ldots, u_{n-m}\}$. Then $|H| = n - m - 1 = n - (m+1)$ and $H \subseteq G$. We have $u_i \in span(L \cup H)$ for $i = 1, \ldots, n-m$ and $L \subseteq span(L \cup H)$. Thus, $V \subseteq span(L \cup \{u_1, \ldots, u_{n-m}\}) \subseteq span(L \cup H)$. □

Corollary 13 Suppose V has finite basis. Then every basis for V has the same cardinality.

Proof. Let β and γ be bases for V.

- β is linearly independent, $span(\gamma) = V$, $|\beta| \leq |\gamma|$ by Theorem 1.10.
- So, γ is linearly independent and $span(\beta) = V$. Again, by Theorem 1.10, $|\gamma| \leq |\beta|$.

□

Definition 9 A vector space is called *finite-dimensional* if it has a finite basis. The number of vectors in a basis, is called the *dimension* of V, notated $dim(V)$. A vector space which is not finite-dimensional is called *infinite-dimensional*.

Corollary 14 Suppose $dim(V) = n$.

(a) If S is finite and $span(S) = V$, then $n \leq |V|$. If $|S| = n$, then S is a basis.

Proof. Let S be finite and $span(S) = V$. Let β be a basis for V. Then, β is linearly independent and $|\beta| = n$. By Theorem 1.10, $|S| \geq |\beta| = n$. Suppose $|S| = n$. Then S contains a subset T such that T is linearly independent and $span(T) = V$. Consequently, T is a basis fr V. Thus, $|T| = dim(V) = n$, $T \subseteq S$, $|T| = n$. Therefore, $S = T$, which implies S is a basis. □

(b) If $|S| = n$ and S is linearly independent, then S is a basis.

Proof. Suppose L is linearly independent and $|L| = dim(V)$. Let β be a basis for V. Then $|L| \leq |\beta|$, in addition, there exists $H \subseteq \beta$ such that $span(L \cup H) = V$ and $|H| = |\beta| - |L| = dim(V) - dim(V) = 0$. Thus, $H = \emptyset$. Thus $span(L) = V$. □

(c) Every linearly independent set can be extended to a basis.

Proof. Let L be linearly independent. Suppose $|L| < n$. Let β be a basis. Then there exists $H \subseteq \beta$ such that $span(L \cup H) = V$ and $|H| = n - |L|$. So, $|L \cup H| \leq n$. As before, $L \cup H$ contains an independent subset T such that $span(T) = V$. Then T is a basis and so $|T| = n$. Therefore, $T = L \cup H$. Thus $L \cup H$ is a basis. □

Example:

(1) $\mathbb{F}^n$, $dim(\mathbb{F}^n) = n$.

(2) $V = M_{m\times n}(\mathbb{F})$, $dim(V) = mn$.

(3) $V = \{A \in M_{n\times n} : A \text{ is symmetric}\}$, $dim(V) = \frac{n(n+1)}{2}$.

Theorem 1.11. Let W be a subspace of a finite-dimensional space V. Then W is finite-dimensional and $dim(W) \leq dim(V)$. Moreover, if $dim(W) = dim(V)$, then $W = V$.

1.7 Maximal linearly independent subsets

Definition 10 A collection $\mathcal{C}$ of sets is called a *chain* if for every $A, B \in \mathcal{C}$, $A \subseteq B$ or $B \subseteq A$.

Maximal Principle: Let $\mathcal{F}$ be a family of sets. If for every chain $\mathcal{C}$ in $\mathcal{F}$ there is a set in $\mathcal{F}$ which contains all members of $\mathcal{C}$, then $\mathcal{F}$ contains a maximal element.

Theorem 1.12. Every vector space has a basis.

Proof. (Outline)

- Start with an arbitrary (finite) linearly independent set S in V and consider the family $\mathcal{F}$ of all independent sets contains S. Argue that Maximal Principle applies and take the element β in $\mathcal{F}$.
- β generates V.

Cauchy's functional equation:

$$f(x + y) = f(x) + f(y)$$

What type of functions can f be?

(1) $f : \mathbb{Q} \to \mathbb{Q}$ where $f(x) = \alpha \cdot x$ and $\alpha = f(1)$.

(2) $f : \mathbb{R} \to \mathbb{R}$ allows for other fairly exotic functions.

For example, $\mathbb{R}$ is a vector space over $\mathbb{Q}$. Let H be a basis of this vector space, commonly called Hamel basis. Then for every element $x \in \mathbb{R}$ there exists unique $h_1, \dots, h_n \in H$ and unique scalars $c_1, \dots, c_n \in \mathbb{Q}$ with $c_1, \dots, c_n \neq 0$ such that $x = \sum_{i=1}^n c_i h_i$. Then, for any $g : H \to \mathbb{R}$, we can extend g to $\overline{g} : \mathbb{R} \to \mathbb{R}$ defined by

$$\overline{g}(x) = \sum_{i=1}^n c_i g(h_i).$$

So,

$$x + y = \sum_{i=1}^n d_i h_i + \sum_{i=1}^n a_i h_i = \sum_{i=1}^n c_i h_i.$$

2 Linear Transformations and Matrices

2.1 Linear transformations, null spaces, and ranges

Definition 1 A function $T : V \to W$ is called a *linear transformation* from V to W if for all $x, y \in V$ and $c \in \mathbb{F}$ the following hold.

- $T(x + y) = T(x) + T(y)$
- $T(cx) = cT(x)$

Observations:

- T is linear if and only if $x_1, \ldots, x_n \in V$, $a_1, \ldots, a_n \in \mathbb{F}$

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

- T is linear if and only if $T(cx + y) = cT(x) + T(y)$ for $x, y \in V$, $c \in \mathbb{F}$.

Example:

(1) Define $T : \mathbb{R}^2 \to \mathbb{R}^2$ to be $T(a_1, a_2) = (2a_1 + a_2, a_1)$. Then

$$\begin{aligned} T((a_1, a_2) + (b_1, b_2)) &= T(a_1 + b_1, a_2 + b_2) \\ &= (2(a_1 + b_1) + a_2 + b_2, a_1 + b_1) \\ &= (2a_1 + a_2, a_1) + (2b_1 + b_2, b_1) \\ &= T(a_1, a_2) + T(b_1, b_2). \end{aligned}$$

It's also easy to show,

$$T(c(a_1, a_2)) = cT(a_1, a_2).$$

(2) Let $T_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ where $T_\theta(a_1, a_2)$ is the vector obtained by (a_1, a_2) by the angle θ.

(3) Let $V = \mathcal{C}(\mathbb{R})$, and $a, b \in \mathbb{R}$, with $a < b$. Let $T : V \to \mathbb{R}$, where

$$T(f) = \int_a^b f(t)dt.$$

Then, $T(f + g) = T(f) + T(g)$. Also, $T(cf) = cT(f)$.

Definition 2 Let $T : V \to W$ be linear.

- The *null space* (kernel) $N(T) = \{x \in V : T(x) = 0\}$

- The *range* (image) $R(T) = \{T(x) : x \in V\}$

Theorem 2.1. Let $T : V \to W$ be linear. Then $N(T)$ and $R(T)$ are subspaces of V and W.

Proof.

(1) We have $0_V \in N(T)$ because $T(0_V) = 0_W$.

(2) Suppose $x, y \in N(T)$. Then $T(x) = 0_W$, $T(y) = 0_W$. Thus, $T(x + y) = T(x) + T(y) = 0_W + 0_W = 0_W$. Thus $x + y \in N(T)$.

(3) Suppose $x \in N(T)$ and $c \in \mathbb{F}$. then $T(x) = 0_W$ and so $T(cx) = cT(x) = 0_W$. Thus, $cx \in N(T)$.

Therefore $N(T)$ is a subspace of V. $R(T)$ can be shown to be a subspace of V in a similar manner. □

Theorem 2.2. Let $T : V \to W$ be linear and let $\beta = \{v_1, \ldots, v_n\}$ be a basis for V. Then $R(T) = span(T(\beta)) = span(\{T(v_1), \ldots, T(v_n)\})$.

Proof.

(1) Let $w \in R(T)$. Then $w = T(v)$ for some $v \in V$. Then $v = \sum_{i=1}^n c_i v_i$ for some $c_1, \ldots, c_n \in \mathbb{F}$. Thus, $T(v) = \sum_{i=1}^n c_i T(v_i)$. Therefore, $w = T(v) \in span(T(\beta))$.

(2) Let $w \in span(T(\beta))$. Then $w = \sum_{i=1}^n c_i T(v_i)$ for some $c_1, \ldots, c_n \in \mathbb{F}$. Thus $w = T\left(\sum_{i=1}^n c_i v_i\right)$ and $\sum_{i=1}^n c_i v_i \in V$. Thus $w \in R(T)$.

□

- $nullity(T) = dim(N(T))$
- $rank(T) = dim(R(T))$

Theorem 2.3. (Dimension Theorem) Let $T : V \to W$ be linear. If V is finite-dimensional, then

$$nullity(T) + rank(T) = dim(V).$$

Proof. (Outline)

- Start with a basis for $N(T)$, $\{v_1, \ldots, v_k\}$ and extend it to a basis of V, $\{v_1, \ldots, v_k, v_{k+1}, \ldots, v_n\}$.
- Prove that $\{T(v_{k+1}), \ldots, T(v_n)\}$ is a basis for $R(T)$.

Proof. Let $\{v_1, \ldots, v_k\}$ be a basis for $N(T)$. Extend this basis to a basis β for V, say $\beta = \{v_1, \ldots, v_k, v_{k+1}, \ldots, v_n\}$. Note, $k = dim(N(T))$ and $n = dim(V)$. We claim that $\{T(v_{k+1}), \ldots, T(v_n)\}$ is a basis for $R(T)$.

(1) Clearly, $span(\{T(v_{k+1}), \ldots, T(v_n)\}) \subseteq R(T)$. Let $w \in R(T)$. Then $w = T(v)$ for some $v \in V$ and $v = \sum_{i=1}^{n} c_i v_i$ for some $c_1, \ldots, c_n \in \mathbb{F}$. Then $w = T(v) = \sum_{i=1}^{n} c_i T(v_i) = \sum_{i=k+1}^{n} c_i T(v_i)$ since $T(v_i) = 0$ for $i \leq k$. Therefore, $w \in span(\{T(v_{k+1}), \ldots, T(v_n)\})$. So, $span(\{T(v_{k+1}), \ldots, T(v_n)\}) = R(T)$.

(2) We'll show $\{T(v_{k+1}), \ldots, T(v_n)\}$. is linearly independent. Suppose $\sum_{i=k+1}^{n} c_i T(v_i) = 0$. Then, $T\left(\sum_{k+1}^{n} c_i v_i\right) = 0$. Thus, $\sum_{k+1}^{n} c_i v_i \in N(T)$. So, $\sum_{k+1}^{n} c_i v_i = \sum_{i=1}^{k} d_i v_i$ for some $d_1, \ldots, d_k \in \mathbb{F}$. Therefore, $\sum_{k+1}^{n} c_i v_i - \sum_{i=1}^{k} d_i v_i = 0$. Since β is linearly independent, $c_{k+1} = \cdots = c_n = 0$. In addition, $d_1 = \cdots = d_k = 0$. Thus, $\{T(v_{k+1}), \ldots, T(v_n)\}$ is linearly independent.

$\square$

Theorem 2.4. Let $T : V \to W$ be linear. Then T is injective if and only if $N(T) = \{0\}$.

Proof.
($\Longrightarrow$): Suppose $T : V \to W$ is injective. If $T(x) = 0$ then since $T(0) = 0$, we have $T(x) = T(0)$ and so $x = 0$.
($\Longleftarrow$): Suppose $N(T) = \{0\}$. If $T(x) = T(y)$, then $T(x) - T(y) = 0$. Thus, $T(x - y) = 0$. Thus, $x - y \in N(T)$, which implies $x - y = 0$ and thus $x = y$. $\square$

Theorem 2.5. Let V, W be finite-dimensional vector spaces such that $dim(V) = dim(W)$ and let $T : V \to W$ be linear. Then the following are equivalent:

(a) T is injective;

(b) T is surjective;

(c) $rank(T) = dim(V) = dim(W)$.

Proof. Note, T being injective is equivalent to $N(T) = \{0\}$, which is equivalent to $nullity(T) = 0$. By the dimension theorem, $rank(T) = dim(V)$. So, $rank(T) = dim(W)$, which is equivalent to $dim(R(T)) = dim(W)$. This is equivalent to $R(T) = W$ since $R(T) \subseteq W$. Thus, T is surjective. $\square$

Theorem 2.6. Suppose $\{v_1, \ldots, v_n\}$ is a basis for V. For $w_1, \ldots, w_n \in W$ there exists exactly one linear transformation $T : V \to W$ such that $T(v_i) = w_i$ for every i.

Proof. (Outline)

- For $x \in V$ we can write $x = \sum_{i=1}^{n} a_i v_i$ uniquely.
- Let $T : V \to W$ be $T(x) = \sum_{i=1}^{n} a_i w_i$.

Proof. Let $\{v_1, \ldots, v_n\}$ be a basis for V and $w_1, \ldots, w_n \in W$. Then there exists unique $T : V \to W$ defined by $T(v_i) = w_i$.

(1) Let $x \in V$. Then x can be uniquely written as $x = \sum_{i=1}^{n} a_i v_i$ where $a_i \in \mathbb{F}$. Let $T(x) \coloneqq \sum_{i=1}^{n} a_i w_i$. $T : V \to W$ is well-defined, and thus a function.

(2) Let T be linear and $u, v \in V$. Let $c \in \mathbb{F}$. We will show $T(cu + v) = cT(u) + T(v)$. Then, $u = \sum_{i=1}^{n} a_i v_i$ and $v = \sum_{i=1}^{n} b_i v_i$ for unique $a_1, \ldots, a_n, b_1, \ldots, b_n \in \mathbb{F}$. Then $cu + v = \sum_{i=1}^{n} (ca_i + b_i) v_i$. So,

$$\begin{aligned} T(cu + v) &= \sum_{i=1}^{n} (ca_i + b_i) w_i \\ &= c \sum_{i=1}^{n} a_i w_i + \sum_{i=1}^{n} b_i w_i \\ &= cT(u) + T(v). \end{aligned}$$

(3) Suppose $U : V \to W$ is linear and $U(v_i) = w_i$. Let $x \in V$. Then $x = \sum_{i=1}^{n} a_i v_i$. Then

$$\begin{aligned} U(x) &= U\left(\sum_{i=1}^{n} a_i v_i\right) \\ &= \sum_{i=1}^{n} a_i U(v_i) \\ &= \sum_{i=1}^{n} a_i w_i \\ &= T(x). \end{aligned}$$

(4) Note, $T(v_i) = w_i$ because $v_i = 1 \cdot v_i$. Thus, $T(v_i) = 1 \cdot w_i = w_i$.

$\square$

Corollary 7 Let $\{v_1, \ldots, v_n\}$ be a basis of V. If $U, T : V \to W$ are linear and $T(v_i) = U(v_i)$, then $T = U$.

2.2 The matrix representation of a linear transformation

Definition 3 Let V be a finite-dimensional vector space. An *ordered basis* for V is a basis equipped with an ordering.

Example: Let $\beta = \left\{ e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, e_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}$. Let $x = \begin{pmatrix} 5 \\ 6 \\ 7 \end{pmatrix} = 5e_1 + 6e_2 + 7e_3$.
Then $\left[x\right]_\beta = \begin{pmatrix} 5 \\ 6 \\ 7 \end{pmatrix}$.

Definition 4 Let $\beta = \{u_1, \ldots, u_n\}$ be an ordered basis for V. Then for every $x \in V$, $x = \sum a_i u_i$ and $a_1, \ldots, a_n$ are unique. The coordinate vector of x relative to β, $\left[x\right]_\beta = (a_1, \ldots, a_n)^T$.

Let $T : V \to W$, $\beta = \{v_1, \ldots, v_n\}$, $\gamma = \{w_1, \ldots, w_m\}$ be ordered bases for V and W. We have

$$T(v_j) = \sum_{i=1}^{m} a_{ij} w_i.$$

Then $A = \left[a_{ij}\right]$ is called the matrix representation of T in the ordered bases β and γ and we write

$$A = \left[T\right]_\beta^\gamma.$$

If $V = W$ and $\beta = \gamma$, we say $A = \left[T\right]_\beta$.

Observations:

- The jth column of A is $\left[T(v_j)\right]_\gamma$.
- If $T, U : V \to W$ and $\left[U\right]_\beta^\gamma = \left[T\right]_\beta^\gamma$, then $T = U$.

Example: Let $T : \mathbb{R}^2 \to \mathbb{R}^2$, $\beta = \{(1,0), (0,1)\}$ and $\gamma = \{(1,1), (1,-1)\}$. Let $T((1,0)) = (1,0)$ and $T((0,1)) = (1,1)$. Find $\left[T\right]_\beta^\gamma$.

We can see $T((1,0)) = (1,0) = \frac{1}{2}(1,1) + \frac{1}{2}(1,-1)$. Also, $T((0,1)) = (1,1) = 1(1,1) + 0(1,-1)$. Therefore, $\left[T\right]_\beta^\gamma = \begin{pmatrix} \frac{1}{2} & 1 \\ \frac{1}{2} & 0 \end{pmatrix}$.

Theorem 2.7. Let $T, U : V \to W$ are linear and let $a \in \mathbb{F}$. Then,

- $aT + U$ is linear.

- The set of all linear transformations from V to W (with addition of functions and scalar multiplication) forms a vector space over F.

We notate $\mathcal{L}(V, W)$ as the vector space of all linear transformations from V to W. Also, $\mathcal{L}(V) = \mathcal{L}(V, V)$.

Theorem 2.8. Let $T, U : V \to W$ be linear. Then,

- $\left[T + U\right]_\beta^\gamma = \left[T\right]_\beta^\gamma + \left[U\right]_\beta^\gamma$
- $\left[aT\right]_\beta^\gamma = a\left[T\right]_\beta^\gamma$.

2.3 Composition of linear transformations and matrix multiplication

Theorem 2.9. Let $T : V \to W$, $U : W \to Z$ be linear. Then $UT : V \to Z$ is linear.

Proof. Let $T : V \to W$, $U : W \to Z$ be linear. Then,

$$\begin{aligned} UT(au + v) &= U(T(au + v)) \\ &= U(aT(u) + T(v)) \\ &= aU(T(u)) + U(T(v)) \\ &= aUT(u) + UT(v). \end{aligned}$$

□

Theorem 2.10. Let $T, U_1, U_2 \in \mathcal{L}(V)$. Then

- $T(U_1 + U_2) = TU_1 + TU_2$ and $(U_1 + U_2)T = U_1T + U_2T$
- $T(U_1U_2) = (TU_1)U_2$
- $TI = IT = T$
- $a(U_1U_2) = (aU_1)U_2 = U_1(aU_2)$

Definition 5 Let A be an $m \times n$ matrix, B be an $n \times p$ matrix. We define AB to be the $m \times p$ matrix such that

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

Note: $(AB)^t = B^tA^t$.

Theorem 2.11. Let $T : V \to W$. $U : W \to Z$ be linear and let α, β, γ be ordered bases in V, W, Z. Then

$$\left[UT\right]_\alpha^\gamma = \left[U\right]_\beta^\gamma \left[T\right]_\alpha^\beta .$$

Proof. (Outline)

- $\alpha = \{v_1, \ldots, v_m\}$, $\beta = \{w_1, \ldots, w_n\}$, $\gamma = \{z_1, \ldots, z_p\}$.
- $A = \left[U\right]_\beta^\gamma$, $B = \left[T\right]_\alpha^\beta$
- $U(T(v_j)) = U\left(\sum_{k=1}^n B_{kj} w_k\right) = \sum_{k=1}^n B_{kj} U(w_k) = \sum_{k=1}^n B_{kj} \sum_{i=1}^p A_{ik} z_i = \sum_{i=1}^p C_{ij} z_i$
- Thus, by definition, the i, jth entry in $\left[UT\right]_\alpha^\gamma$ is $C_{ij} = \sum_{k=1}^n A_{ik} B_{kj}$.

Proof. Let $\alpha = \{v_1, \ldots, v_n\}, \beta = \{w_1, \ldots, w_m\}$ and $\gamma = \{z_1, \ldots, z_p\}$. Let $A = \left[U\right]_\beta^\gamma$ and $B = \left[T\right]_\alpha^\beta$. Consider $\left[UT\right]_\alpha^\gamma$. Then

$$\begin{aligned}
(UT)(v_j) &= U(T(v_j)) \\
&= U\left(\sum_{i=1}^m B_{ij} w_i\right) \\
&= \sum_{i=1}^m B_{ij} U(w_i) \\
&= \sum_{i=1}^m B_{ij} \cdot \sum_{k=1}^p A_{ki} z_k \\
&= \sum_{k=1}^p \left(\sum_{i=1}^m A_{ki} B_{ij}\right) z_k.
\end{aligned}$$

If $C = \left[UT\right]_\alpha^\gamma$, then

$$C_{kj} = \sum_{i=1}^m A_{ki} B_{ij}.$$

□

Example: Let $\alpha = \{1, x, x^2\}$ and $\beta = \{1, x, x^2, x^3\}$ be standard bases for $P_2(\mathbb{R})$ and $P_3(\mathbb{R})$. Let $U : P_3(\mathbb{R}) \to P_2(\mathbb{R})$ and $T : P_2(\mathbb{R}) \to P_3(\mathbb{R})$ defined by $U(f(x)) = f'(x)$ and $T(f(x)) = \int_0^x f(t)dt$. From calculus, $UT = I$. We can see

$$\begin{aligned}
U(1) &= 0 = 0 \cdot 1 + 0 \cdot x + 0 \cdot x^2 \\
U(x) &= 1 = 1 \cdot 1 + 0 \cdot x + 0 \cdot x^2 \\
U(x^2) &= 2x = 0 \cdot 1 + 2 \cdot x + 0 \cdot x^2 \\
U(x^3) &= 3x^2 = 0 \cdot 1 + 0 \cdot x + 3 \cdot x^2.
\end{aligned}$$

Also,

$$\begin{aligned} T(1) &= x = 0 \cdot 1 + 1 \cdot x + 0 \cdot x^2 + 0 \cdot x^3 \\ T(x) &= \frac{1}{2}x^2 = 0 \cdot 1 + 0 \cdot x + \frac{1}{2}x^2 + 0 \cdot x^3 \\ T(x^2) &= \frac{1}{3}x^3 = 0 \cdot 1 + 0 \cdot x + 0 \cdot x^2 + \frac{1}{3}x^3. \end{aligned}$$

Thus,

$$\begin{aligned} \left[UT\right]_\alpha^\alpha &= \left[U\right]_\beta^\alpha \left[T\right]_\alpha^\beta \\ &= \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{3} \end{pmatrix} \\ &= I. \end{aligned}$$

Theorem 2.12. Let A be an $m \times n$ matrix, B and C be $n \times p$ matrices, D, E be $q \times m$ matrices. Then

- $A(B + C) = AB + AC$, $(D + E)A = DA + EA$
- $a(AB) = (aA)B = A(aB)$
- $I_m A = A = AI_n$
- If V is n-dimensional with ordered basis β, then $\left[I_V\right]_\beta = I_n$

Theorem 2.13. Let A be an $m \times n$ matrix, B be an $n \times p$ matrix. Let u_j, v_j denote the jth columns of AB and B. Then

(a) $u_j = Av_j$

(b) $v_j = Be_j$

Theorem 2.14. Let V, W be finite-dimensional vector spaces with ordered bases β and γ and let $T : V \to W$ be linear. Then for $u \in V$

$$\left[T(u)\right]_\gamma = \left[T\right]_\beta^\gamma \left[u\right]_\beta .$$

Proof. Let V, W be finite-dimensional vector spaces with ordered bases β and γ and let $T : V \to W$ be linear. Fix $u \in V$. Let $f : \mathbb{R} \to V$ and $g : \mathbb{R} \to W$ where $f(a) = au$

and $g(a) = aT(u)$. Let $\alpha = \{1\}$ be the standard basis for $\mathbb{R}$. Note that, g and f are linear transformations. Then,

$$\begin{aligned} \left[T(u)\right]_\gamma &= \left[g(1)\right]_\gamma \\ &= \left[g\right]_\alpha^\gamma \\ &= \left[Tf\right]_\alpha^\gamma \\ &= \left[T\right]_\beta^\gamma \left[f\right]_\alpha^\beta \\ &= \left[T\right]_\beta^\gamma \left[f(1)\right]_\beta \\ &= \left[T\right]_\beta^\gamma \left[u\right]_\beta . \end{aligned}$$

□

Let $L_A : \mathbb{F}^n \to \mathbb{F}^m$ be given by $L_A(x) = Ax$ where A is an $m \times n$ matrix.

Theorem 2.15. Let A be an $m \times n$ matrix. Then L_A is linear. Moreover, if $B \in M_{m\times n}(\mathbb{F})$ and β and γ are the standard ordered bases of $\mathbb{F}^n$ and $\mathbb{F}^m$, then

(a) $\left[L_A\right]_\beta^\gamma = A$

(b) $L_A = L_B$ if and only if $A = B$

(c) $L_{A+B} = L_A + L_B$, $L_{aA} = aL_A$

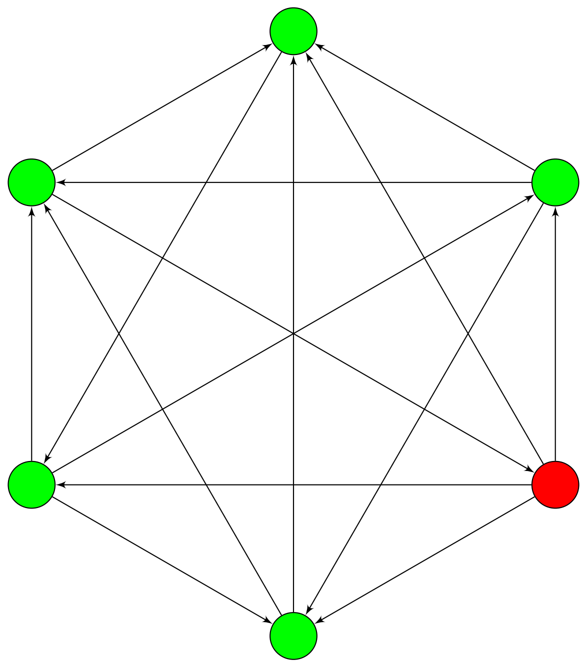
(d) If $T : \mathbb{F}^n \to \mathbb{F}^m$ is linear, then there exists unique matrix C (namely $C = \left[T\right]_\beta^\gamma$) such that $T = L_C$.

(e) If E is an $n \times p$ matrix, then $L_{AE} = L_A L_E$

(f) If $m = n$, $L_{I_n} = I_{\mathbb{F}^n}$.

Theorem 2.16. Let A, B, C be such that $A(BC)$ is defined then $(AB)C = A(BC)$.

Tournaments: There are n players $\{1, \ldots, n\}$ for every $i \neq j$ there is exactly one game between i and j which results in i winning or j winning. Let A be the incidence matrix of a tournament where we put $A_{ij} = 1$ if i wins with j and 0 otherwise. Show that $A^2 + A$ contains a column such that each entry but the diagonal is at least one.



2.4 Invertibility and isomorphisms

Definition 6 Let $T : V \to W$ be linear. A function $U : W \to V$ is called an *inverse* of T if $UT = I_V$, $TU = I_W$. If T has an inverse, then it's called *invertible*.

Note:

- If T is invertible, then the inverse is unique, denoted by T^{-1}.
- $(UT)^{-1} = T^{-1}U^{-1}$
- $(T^{-1})^{-1} = T$
- T is invertible if and only if T is a bijection.
- $T : V \to W$, V, W are finite-dimensional of equal dimensions.
- T is invertible if and only if $rank(T) = dim(v)$.

Theorem 2.17. Let $T : V \to W$ be linear and invertible. Then $T^{-1} : W \to V$ is linear.

Proof. Suppose $T : V \to W$ is linear and invertible. Then $T^{-1} : W \to V$. Let $y_1, y_2 \in W$. Since T is bijective, there exists unique $x_1, x_2 \in V$ such that $T(x_1) = y_1$ and

$T(x_2) = y_2$. Let $c \in \mathbb{F}$. Then

$$\begin{aligned} T^{-1}(cy_1 + y_2) &= T^{-1}(cT(x_1) + T(x_2)) \\ &= T^{-1}(T(cx_1 + x_2)) \\ &= cx_1 + x_2 \\ &= cT^{-1}(y_1) + T^{-1}(y_2). \end{aligned}$$

$\square$

Definition 7 Let A be an $n \times n$ matrix. Then A is *invertible* if there exists an $n \times n$ matrix B such that $AB = BA = I$. Then B is called the inverse of A denoted by A^{-1}.

Lemma 19. Let T be an invertible transformation from V to W. Then V is finite-dimensional if and only if W is. In this case $dim(V) = dim(W)$.

Proof. Let T be an invertible transformation from V to W. Then T must be bijective. Suppose V is finite-dimensional. Let $\beta = \{v_1, \ldots, v_n\}$ be a basis for V. Then $dim(V) = n$. Then $T(\beta) = \{T(v_1), \ldots, T(v_n)\}$ generates W because T is surjective. Let $w \in W$. Then there exists $v \in V$ such that $T(v) = w$. Then $v = \sum_{i=1}^{n} a_i v_i$. Thus, $T(v) = \sum_{i=1}^{n} a_i T(v_i)$. As a result, $dim(W) \leq n = dim(V)$. However, $T^{-1} : W \to V$ is also linear. Then the same argument implies $dim(V) \leq dim(W)$. Therefore, $dim(V) = dim(W)$. $\square$

Theorem 2.18. Let V, W be finite-dimensional with ordered bases β and γ. Let $T : V \to W$ be linear. Then T is invertible if and only if $\left[T\right]_\beta^\gamma$ is invertible. Furthermore $\left[T^{-1}\right]_\gamma^\beta = (\left[T\right]_\beta^\gamma)^{-1}$.

Proof. (Outline)

- If T is invertible, then by the lemma $dim(V) = dim(W)$ and the fact that $\left[T\right]_\beta^\gamma$ follows from previous facts and $T^{-1}T = I_V$.
- If $\left[T\right]_\beta^\gamma$ is invertible then it has an inverse B and so there is a transformation $U : W \to V$ with $B = \left[U\right]_\gamma^\beta$. Now check that $UT = I_V$.

Proof. Suppose T is invertible. Then by the lemma $dim(V) = dim(W) = n$, and in addition, there exists $U : W \to V$ such that $TU = I_W$ and $UT = I_V$. Let β, γ be bases for V, W respectively. Then

$$I = \left[I_V\right]_\beta = \left[UT\right]_\beta = \left[U\right]_\gamma^\beta \left[T\right]_\beta^\gamma.$$

In the same way,

$$\left[I_W\right]_\gamma = \left[TU\right]_\gamma = \left[T\right]_\beta^\gamma \left[U\right]_\gamma^\beta.$$

Thus, $\left[U\right]_{\gamma}^{\beta} = \left(\left[T\right]_{\beta}^{\gamma}\right)^{-1}$.

Now, assume $\left[T\right]_{\beta}^{\gamma}$ is invertible. Let $A = \left[T\right]_{\beta}^{\gamma}$. Then A is an $n \times n$ matrix. Let $\beta = \{v_1, \ldots, v_n\}$ and $\gamma = \{w_1, \ldots, w_n\}$. Since A is invertible, there exists B such that $AB = I_n = BA$. Then there exists unique linear transformation $U : W \to V$ such that $U(w_j) = \sum_{i=1}^{n} B_{ij} v_i$. Then $\left[U\right]_{\gamma}^{\beta} = B$. Thus $\left[TU\right]_{\gamma} = \left[T\right]_{\beta}^{\gamma} \left[U\right]_{\gamma}^{\beta} = A \cdot B = I_n = \left[I_W\right]_{\gamma}$. Also, $\left[UT\right]_{\beta} = B \cdot A = I_n = \left[I_V\right]_{\beta}$. Thus $TU = I_W$ and $UT = I_V$. $\square$

Definition 8 Vector space V is isomorphic to W if there is an invertible linear transformation $T : V \to W$.

Example: Let $T : P_3(\mathbb{R}) \to M_{2\times 2}(\mathbb{R})$ be defined by

$$T(f) = \begin{pmatrix} f(1) & f(2) \\ f(3) & f(4) \end{pmatrix}.$$

Then

(1) T is linear.

$$T(f+g) = T(f) + T(g)$$
$$T(cf) = cT(f)$$

(2) T is injective. (This is because $N(T) = \{0\}$.)

(3) $dim(P_3(\mathbb{R})) = 4$ and $dim(M_{2\times 2}(\mathbb{R})) = 4$

(4) T is surjective

(5) T is bijective

(6) $P_3(\mathbb{R}) \cong M_{2\times 2}(\mathbb{R})$

Theorem 2.19. Let V, W be finite-dimensional over F. Then V is isomorphic to W if and only if $dim(V) = dim(W)$.

Proof.
($\Longrightarrow$): Suppose $V \cong W$. By the lemma, $dim(V) = dim(W)$.
($\Longleftarrow$): Suppose $dim(V) = dim(W)$. Let $\beta = \{v_1, \ldots, v_n\}$ be a basis for V and let $\gamma = \{w_1, \ldots, w_n\}$ be a basis for W. Then there exists a unique linear transformation $T : V \to W$ such that $T(v_i) = w_i$. Then T is surjective. Since $dim(V) = dim(W)$, T is injective. Thus T is invertible. Therefore, $V \cong W$. $\square$

Theorem 2.20. Let V, W be finite-dimensional over F of dimensions n and m. Let β, γ be ordered bases for V and W. Then $\Phi : \mathcal{L}(V,W) \to M_{m\times n}(\mathbb{F})$ given by $\Phi(T) = \left[T\right]_\beta^\gamma$ for $T \in \mathcal{L}(V,W)$ is an isomorphism.

Proof. Let $\Phi(T) = \left[T\right]_\beta^\gamma$.

(1) Let $\phi : \mathcal{L}(V,W) \to M_{m\times n}(\mathbb{F})$. Then Φ is linear.

$$\begin{aligned}\Phi(aT+U) &= \left[aT+U\right]_\beta^\gamma \\ &= \left[aT\right]_\beta^\gamma + \left[U\right]_\beta^\gamma \\ &= a\left[T\right]_\beta^\gamma + \left[U\right]_\beta^\gamma\end{aligned}$$

(2) For every $A \in M_{m\times n}(\mathbb{F})$ there exists unique transformation T such that $\Phi(T) = A = \left[A_{ij}\right]$. Let $\beta = \{v_1, \ldots, v_n\}$ and $\gamma = \{w_1, \ldots, w_m\}$. Then there exists unique transformation $T : V \to W$ such that

$$T(v_j) = \sum_{i=1}^{n} A_{ij} w_i,$$

for $j = 1, \ldots, n$ and $i = 1, \ldots, m$. Thus

$$\left[T\right]_\beta^\gamma = A.$$

Thus $\Phi(T) = \left[T\right]_\beta^\gamma = A$.

□

Definition 9 Let β be an ordered basis for an n-dimensional vector space V over F. The *standard representation of V with respect to β* is the function $\phi_\beta : V \to \mathbb{F}^n$ defined as $\phi_\beta(x) = \left[x\right]_\beta$.

Theorem 2.21. Let V be a finite-dimensional vector space with ordered basis β. Then ϕ_β is an isomorphism.

Note: For vector fields V and W with basis β and γ, respectively, we have the following schematic representation:

$$\begin{tikzcd} V \arrow[r, "T"] \arrow[d, "\phi_\beta"'] & W \arrow[d, "\phi_\gamma"] \\ \mathbb{F}^n \arrow[r, "L_A"] & \mathbb{F}^m \end{tikzcd}$$

Where $L_A(x) = Ax$ and $A = \left[T\right]_\beta^\gamma$, we have

$$L_A \phi_\beta = \phi_\gamma T.$$

That is,

$$\left[T\right]_\beta^\gamma \left[u\right]_\beta = \left[T(u)\right]_\gamma.$$

2.5 The change of coordinate matrix

Theorem 2.22. Let β, β' be two ordered bases for a finite-dimensional vector space V and let $Q = \left[I_V\right]_{\beta'}^{\beta}$. Then

(a) Q is invertible.

(b) For any $v \in V$, $\left[v\right]_\beta = Q\left[v\right]_{\beta'}$

Proof.

(a) We know $\left[I_V\right]_{\beta'}^{\beta}$ is invertible because $I_V : V \to V$ is invertible. So, $I_V^{-1} = I_V$.

(b) We have $\left[u\right]_\beta = Q \cdot \left[u\right]_{\beta'}$. So, $\left[u\right]_\beta = \left[I_V(u)\right]_\beta = \left[I_V\right]_{\beta'}^{\beta}\left[u\right]_{\beta'}$.

∎

Example: Let $\beta = \left\{\begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}1\\-1\end{pmatrix}\right\}$ and $\beta' = \left\{\begin{pmatrix}2\\4\end{pmatrix}, \begin{pmatrix}3\\1\end{pmatrix}\right\}$. Then

$$\begin{aligned}
\begin{pmatrix}2\\4\end{pmatrix} &= 3\begin{pmatrix}1\\1\end{pmatrix} - \begin{pmatrix}1\\-1\end{pmatrix}, \\
\begin{pmatrix}3\\1\end{pmatrix} &= 2\begin{pmatrix}1\\1\end{pmatrix} + \begin{pmatrix}1\\-1\end{pmatrix}, \\
\left[I_V\right]_{\beta'}^{\beta} &= \begin{pmatrix}3 & 2\\-1 & 1\end{pmatrix}, \\
\left[\begin{pmatrix}2\\4\end{pmatrix}\right]_\beta &= Q\left[\begin{pmatrix}2\\4\end{pmatrix}\right]_{\beta'} \\
&= Q\begin{pmatrix}1\\0\end{pmatrix} \\
&= \begin{pmatrix}3\\-1\end{pmatrix}.
\end{aligned}$$

Q is called the change of coordinate matrix. A linear operator on V is the linear transformation from V to V.

Theorem 2.23. Let T be a linear operator on a finite-dimensional vector space V. Let β, β' be ordered bases for V. Suppose $Q = \left[I_V\right]_{\beta'}^{\beta}$. Then

$$\left[T\right]_{\beta'} = Q^{-1}\left[T\right]_\beta Q.$$

Proof. Suppose $Q = \left[I_V\right]_{\beta'}^{\beta}$. Note that $I_V Q = Q = Q I_V$. Then,

$$\begin{aligned} Q\left[T\right]_{\beta'} &= \left[I_V\right]_{\beta'}^{\beta}\left[T\right]_{\beta'} \\ &= \left[I_V T\right]_{\beta'}^{\beta} \\ &= \left[T I_V\right]_{\beta'}^{\beta} \\ &= \left[T\right]_{\beta}\left[I_V\right]_{\beta'}^{\beta} \\ &= \left[T\right]_{\beta}\cdot Q. \end{aligned}$$

So, $Q^{-1}Q\left[T\right]_{\beta'} = Q^{-1}\left[T\right]_{\beta} Q$. Therefore, $\left[T\right]_{\beta'} = Q^{-1}\left[T\right]_{\beta} Q$. □

Corollary 26 Let $A \in M_{n\times n}(\mathbb{F})$ and let $\gamma = \{u_1, \ldots, u_n\}$ be an ordered basis for $\mathbb{F}^n$. Then $\left[L_A\right]_{\gamma} = Q^{-1}AQ$ where Q is the matrix with the jth column equal to u_j.

Note: We say B **is similar to** A if there exists an invertible matrix C such that

$$B = C^{-1}AC.$$

Example: Let $A = \begin{pmatrix} 2 & 0 & 0 \\ 1 & 2 & 1 \\ -1 & 0 & 1 \end{pmatrix}$. Find A^{100}.

Solution. Let $\gamma = \left\{ \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right\}$.

(1) Find $\left[L_A\right]_{\gamma}$.

$$\begin{aligned} L_A\left(\begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}\right) &= A\begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 & 0 & 0 \\ 1 & 2 & 1 \\ -1 & 0 & 1 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}, \\ L_A\left(\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}\right) &= A\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 & 0 & 0 \\ 1 & 2 & 1 \\ -1 & 0 & 1 \end{pmatrix}\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -2 \\ 0 \\ 2 \end{pmatrix}, \\ L_A\left(\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}\right) &= A\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 & 0 & 0 \\ 1 & 2 & 1 \\ -1 & 0 & 1 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix}. \end{aligned}$$

Thus

$$\left[L_A\right]_{\gamma} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

(2) Note, $Q = \begin{pmatrix} 0 & -1 & 0 \\ -1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$ and $Q^{-1} = \begin{pmatrix} 1 & 0 & 1 \\ -1 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix}$. Then $\left[L_A\right]_\gamma = Q^{-1}AQ$. So, $A = Q\left[L_A\right]_\gamma Q^{-1}$. Then we have

$$\begin{aligned} A^{100} &= \left(Q\left[L_A\right]_\gamma Q^{-1}\right)^{100} \\ &= Q\left[L_A\right]_\gamma Q^{-1}Q\left[L_A\right]_\gamma Q^{-1}\dots Q\left[L_A\right]_\gamma Q^{-1} \\ &= Q\left[L_A\right]_\gamma^{100} Q^{-1} \\ &= Q\begin{pmatrix} 1^{100} & 0 & 0 \\ 0 & 2^{100} & 0 \\ 0 & 0 & 2^{100} \end{pmatrix} Q^{-1}. \end{aligned}$$

2.6 Dual spaces

- Linear functional on V – linear transformation from V to F.
- $V^* = \mathcal{L}(V, F)$
- $V^{**} = (V^*)^*$

$$dim(V^*) = dim(V)$$

Let $\beta = \{x_1, \dots, x_n\}$ for $v \in V$ let $\left[v\right]_\beta = (a_1, \dots, a_n)^T$. Define

$$f_i(v) = a_i.$$

Example:

(1) Let V = continuous functions $f : [0, 2\pi] \to \mathbb{R}$. Let $g \in V$. Define $h : V \to \mathbb{R}$ by

$$h(x) = \frac{1}{2\pi}\int_0^{2\pi} x(t)g(t)dt.$$

Note, h is linear. If $g(t) = \sin(nt)$ or $g(t) = \cos(nt)$, then $h(x)$ is called the nth Fourier coefficient of x.

(2) Let $V = M_{n\times n}(\mathbb{F})$ and $f : V \to \mathbb{F}$ where $f(A) = tr(A)$.

Theorem 2.24. Let V be a finite-dimensional vector space with ordered basis $\beta = \{x_1, \dots, x_n\}$ and let $\beta^* = \{f_1, \dots, f_n\}$. Then β^* is an ordered basis for V^* and for $f \in V^*$ where

$$f = \sum_{i=1}^{n} f(x_i)f_i.$$

Note, $f_i : V \to \mathbb{F}$ is defined as $f_i(x_j) = \delta_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{else} \end{cases}.$

Proof. (Outline)

- Enough to show $f = \sum f(x_i) f_i$.
- Let $F \coloneqq \sum f(x_i) f_i$. Then $F(x_j) = f(x_j)$ for every j.

Proof. Let V be a finite-dimensional vector space with ordered basis $\beta = \{x_1, \ldots, x_n\}$ and let $\beta^* = \{f_1, \ldots, f_n\} \subseteq V^*$ where $n = dim(V) = dim(V^*)$. Since $|\beta^*| = n = dim(V^*)$, it is enough to show that β^* generates V^*. To that end, we will argue that for $f \in V^*$,

$$f = \sum_{i=1}^{n} f(x_i) f_i.$$

Let $g = \sum_{i=1}^{n} f(x_i) f_i$. Then

$$\begin{aligned} g(x_j) &= \left(\sum_{i=1}^{n} f(x_i) f_i \right)(x_j) \\ &= \sum_{i=1}^{n} f(x_i) f_i(x_j) \\ &= \sum_{i=1}^{n} f(x_i) \delta_{ij} \\ &= f(x_j). \end{aligned}$$

Thus, $g(x_j) = f(x_j)$ for every $x_j \in \beta$. Therefore $g = f$. $\square$

Definition 10 An ordered basis $\beta^* = \{f_1, \ldots, f_n\}$ for V^* such that $f_i(x_j) = \delta_{ij}$ is called the *dual basis* of $\beta = \{x_1, \ldots, x_n\}$.

Theorem 2.25. Let V, W be finite-dimensional vector spaces over $\mathbb{F}$ with ordered bases β and γ. Let $T : V \to W$ be linear. Then $T^t : W^* \to V^*$ given by $T^t(g) = gT$ is linear and

$$\left[T^t\right]_{\gamma^*}^{\beta^*} = (\left[T\right]_{\beta}^{\gamma})^t.$$

Proof. (Outline)

- It's easy to see that T^t is a linear transformation from W^* to V^*.
- Let $\beta = \{x_1, \ldots, x_n\}$, $\gamma = \{y_1, \ldots, y_m\}$, $\beta^* = \{f_1, \ldots, f_n\}$, $\gamma^* = \{g_1, \ldots, g_m\}$, $A = \left[T\right]_{\beta}^{\gamma}$.

- The jth column of $\left[T^t\right]_{\gamma^*}^{\beta^*}$ is $T^t(g_j)$ which is

$$\sum_{k=1}^{n}(g_jY)(x_k)f_k.$$

- Thus the i,j-th entry is $(T^t(g_j))(x_i)$ which is A_{ji}.

Proof. Let $T^t : W^* \to V^*$, $T^t(g) = gT$.

(1) Note that $gT : V \to \mathbb{F}$ and gT is a linear transformation. Thus $T^t(g) \in V^*$.

(2) We have T^t is linear. So,

$$\begin{aligned} T^t(cg+h) &= (cg+h)T \\ &= cgT + hT \\ &= cT^t(g) + T^t(h). \end{aligned}$$

(3) Lastly, $\left[T^t\right]_{\gamma^*}^{\beta^*} = \left(\left[T\right]_{\beta}^{\gamma}\right)^t$. Let $\beta = \{x_1, \ldots, x_n\}$, $\gamma = \{y_1, \ldots y_m\}$, $\beta^* = \{f_1, \ldots, f_n\}$, and $\gamma^* = \{g_1, \ldots, g_m\}$. Let $A = \left[T\right]_{\beta}^{\gamma}$. To obtain the jth column of $\left[T\right]_{\gamma^*}^{\beta^*}$,

$$\begin{aligned} T^t(g_j) &= g_jT \\ &= \sum_{k=1}^{n}(g_jT)x_kf_k. \end{aligned}$$

Thus, the i,jth entry of $\left[T\right]_{\gamma^*}^{\beta^*}$ is

$$\begin{aligned} (g_jT)(x_i) &= g_j(T(x_i)) \\ &= g_j\left(\sum_{k=1}^{m}A_{ki}y_i\right) \\ &= \sum_{k=1}^{m}A_{ki}g_j(y_i) && (g_j(y_i) = \delta_{ij}) \\ &= A_{ji}. \end{aligned}$$

□

For $x \in V$ let $\hat{x} : V^* \to \mathbb{F}$ given by $\hat{x}(f) = f(x)$.

Theorem 2.26. Let V be finite-dimensional and let $\psi : V \to V^{**}$ be given by $\psi(x) = \hat{(x)}$. Then ψ is an isomorphism.

3 Elementary matrix operations and systems of linear equations

3.1 Elementary matrix operations and elementary matrices

Definition 1 Let A be a matrix. Elementary row operations:

- Interchange any two rows of A
- Add a scalar multiple of a row of A to another row
- Multiply any row of A by a non-zero scalar

Note: The same can be done for columns.

Definition 2 An $n \times n$ elementary matrix is a matrix obtained from I_n by an elementary operation. Its type is the type of the operation performed.

Theorem 3.1. Let $A \in M_{m\times n}(\mathbb{F})$ and suppose B is obtained by performing an elementary row (column) operation. Then there exists an $m \times m(n \times n)$ elementary matrix E such that $B = EA(B = AE)$.

Example: Note, $\begin{pmatrix} 1 & -7 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ is an elementary matrix.

$$\begin{pmatrix} a-7x & b-7y & c-7z \\ x & y & z \\ u & v & w \end{pmatrix} = \begin{pmatrix} 1 & -7 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Theorem 3.2. Elementary matrices are invertible and the inverse of an elementary matrix is an elementary matrix of the same type.

Example: Let $A = \begin{pmatrix} 1 & -7 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. Then

$$\begin{aligned} AA^{-1} &= \begin{pmatrix} 1 & -7 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 7 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \end{aligned}$$

3.2 The rank of a matrix and matrix inverse

Definition 3 Let $A \in M_{m\times n}(\mathbb{F})$. The $rank(A)$ is the rank of $L_A : \mathbb{F}^n \to \mathbb{F}^m$. Also $rank(L_A) = dim(R(L_A))$.

Theorem 3.3. Let $T : V \to W$ be linear and let β, γ be ordered bases for V and W. Then

$$rank(T) = rank\left(\left[T\right]_\beta^\gamma\right).$$

Recall, where V and W are vector spaces with bases β and γ, respectively, and $A = \left[T\right]_\beta^\gamma$, we have

$$\begin{array}{ccc} V & \xrightarrow{T} & W \\ \updownarrow & & \updownarrow \\ \mathbb{F}^n & \xrightarrow{L_A} & \mathbb{F}^m \end{array}$$

Corollary 3.4.1 Elementary operations are rank-preserving.

Theorem 3.4. Let A be an $m \times n$ matrix. If P and Q are invertible $m \times m$ and $n \times n$ matrices, then

- $rank(AQ) = rank(A)$
- $rank(PA) = rank(A)$

and so $rank(PAQ) = rank(A)$.

Proof. (Outline)

- $R(L_{AQ}) = R(L_A)$ because $L_Q(\mathbb{F}^n) = \mathbb{F}^n$.
- $dim(L_P(L_A(\mathbb{F}^n))) = dim(L_A(\mathbb{F}^n))$ because $L_P : \mathbb{F}^n \to \mathbb{F}^n$ is an isomorphism.

Proof.

(1) Note,

$$\begin{aligned} rank(AQ) &= rank(L_{AQ}) \\ &= dim(R(L_{AQ})). \end{aligned}$$

So,

$$\begin{aligned} R(L_{AQ}) &= R(L_A L_Q) \\ &= L_A L_Q(\mathbb{F}^n) \\ &= L_A(L_Q(\mathbb{F}^n)) \\ &= L_A(\mathbb{F}^n) && \text{(Since } L_Q(\mathbb{F}^n) = \mathbb{F}^n) \\ &= R(L_A). \end{aligned}$$

Therefore, $dim(R(L_{AQ})) = dim(R(L_A)) = rank(A)$.

(2) We have to show $rank(PA) = rank(A)$. So,

$$\begin{aligned} rank(PA) &= dim(R(L_{PA})) \\ &= dim(L_{PA}(\mathbb{F}^n)) \\ &= dim(L_P(L_A(\mathbb{F}^n))). \end{aligned}$$

Since $L_P : L_A(\mathbb{F}^n) \to L_P(L_A(\mathbb{F}^n))$, then L_P is an isomorphism. Thus, $dim(L_P(L_A(\mathbb{F}^n))) = dim(L_A(\mathbb{F}^n)) = rank(A)$. Therefore, $rank(PA) = rank(A)$.

□

Theorem 3.5. The rank of a matrix equals the maximum number of its linearly independent columns.

Proof. (Outline)

- $R(L_A) = span(L_A(\{e_1, \ldots, e_n\}))$ and $L_A(e_j)$ is the ith column of A.

Proof. Let $A \in M_{m\times n}(\mathbb{F}^n)$. Then

$$\begin{aligned} rank(A) &= dim(R(L_A)) \\ &= dim(L_A(\mathbb{F}^n). \end{aligned}$$

Let β be the standard basis for $\mathbb{F}^n$. We have $span(\beta) = \mathbb{F}^n$. Therefore,

$$\begin{aligned} R(L_A) &= span(L_A(\beta)) \\ &= span(\{L_A(e_1), \ldots, L_A(e_n)\}. \end{aligned}$$

We have $L_A(e_i) = a_i$ where a_i is the ith column of A. Thus, $R(L_A) = span(\{a_1, \ldots, a_n\}$. Thus, $dim(R(L_A))$ is the maximum number of linearly independent columns. □

Example: Find the rank of

$$A = \begin{pmatrix} 1 & 2 & 1 \\ 1 & 0 & 3 \\ 1 & 1 & 1 \end{pmatrix}.$$

Solution.

$$
\begin{aligned}
A &\xrightarrow{\text{Row Op.}} \begin{pmatrix} 1 & 2 & 1 \\ 0 & -2 & 2 \\ 0 & -1 & 0 \end{pmatrix} \\
&\xrightarrow{\text{Col. Op.}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & -2 & 2 \\ 0 & -1 & 0 \end{pmatrix} \\
&\xrightarrow{\text{Row Op.}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 2 & 2 \end{pmatrix} \\
&\xrightarrow{\text{Row Op.}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.
\end{aligned}
$$

Therefore, $rank(A) = 3$.

Theorem 3.6. Let A be an $m \times n$ matrix of rank r. Then $r \leq min\{m, n\}$ and A can be transformed to

$$D = \begin{pmatrix} I_r & 0_1 \\ 0_2 & 0_3 \end{pmatrix}$$

using a finite number of elementary row and column operations.

Proof. (Outline) Induction on m. In the inductive step use row and column operations to reduce to

$$\left(\begin{array}{c|ccc} 1 & 0 & \dots & 0 \\ \hline 0 & & & \\ & & B' & \\ 0 & & & \end{array}\right).$$

Proof. If A is the zero matrix, then $rank(A) = 0$ and $D = A$, thus $r = 0$. Suppose A is non-zero. We will use induction on m.

(Base Step) Let $m = 1$. Then A has one row. Then by applying elementary column operations, we can transform A to $\begin{pmatrix} 1 & 0 & \dots & 0 \end{pmatrix}$ and so $r = 1$ and $rank(A) = 1$.

(Induction Step) Suppose $n \geq 2$. If $n = 1$, then A can be transformed into $\begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$.

So, $r = 1 = rank(A)$. Let $n \geq 2$. So, there exists A_{ij} such that $A_{ij} \neq 0$ and we can transform A so that A_{ij} is in position (1,1). Therefore, we can transform A to

$$\left(\begin{array}{c|ccc} 1 & 0 & \dots & 0 \\ \hline 0 & & & \\ & & B & \\ 0 & & & \end{array}\right)$$

and $rank(B) = rank(A) - 1 = r - 1$. By the inductive hypothesis, $r - 1 \leq m - 1$ and $r - 1 \leq n - 1$ and B can be transformed to

$$\begin{pmatrix} I_{r-1} & 0_4 \\ 0_5 & 0_6 \end{pmatrix}.$$

Thus, A can be transformed to

$$\begin{pmatrix} I_r & 0_1 \\ 0_2 & 0_3 \end{pmatrix}$$

for some 0_i.

□

Note, if $m = n$ and $rank(A) = n$, then $B = I_n$. The converse of this statement is also true.

Corollary 7 Let A be an $m \times n$ matrix of rank r. Then there exist invertible matrices B and C of sizes $m \times m$ and $n \times n$ such that $D = BAC$.

As a consequence of the above corollary, A is invertible if and only if $rank(A) = n$.

Corollary 8 Let A be an $m \times n$ matrix. Then

- $rank(A^t) = rank(A)$
- $rank(A)$ is equal to the dimension of the row space of A
- dimension of the row space is equal to the dimension of the column space.

Proof. We will show $rank(A^t) = rank(A)$. By the Theorem 3.6, there exits invertible matrices B and C such that $D = BAC$. Then $D^t = (BAC)^t = C^t A^t B^t$. Also, B^t and C^t are invertible. Recall, $(B^t)^{-1} = (B^{-1})^t$. We have $rank(D^t) = r = rank(D)$ and $rank(A) = rank(D) = rank(D^t) = rank(A^t)$. □

Corollary 9 Every invertible matrix is a product of elementary matrices.

Proof. Suppose A is invertible. Then there exists invertible matrices B and C such that $D = BAC$. Thus, D is invertible, which implies $D = I_n$. Also, $B = E_1 \dots E_p$ and $C = G_1 \dots G_q$, where E_i and G_j are elementary. Therefore, $BAC = I_n$ and $A = B^{-1}C^{-1}$. Thus,

$$\begin{aligned} A &= (E_1 \dots E_p)^{-1}(G_1 \dots G_q)^{-1} \\ &= E_p^{-1} \dots E_1^{-1} G_q^{-1} \dots G_1^{-1}. \end{aligned}$$

□

Theorem 3.7. Let $T : V \to W$ and $U : W \to Z$ be linear transformations on finite-dimensional vector spaces. Let A, B be matrices such that AB is defined. Then

(a) $rank(UT) \leq rank(U)$

(b) $rank(UT) \leq rank(T)$

(c) $rank(AB) \leq rank(A)$

(d) $rank(AB) \leq rank(B)$

Proof. (Outline)

- For (a), $R(UT) = U(R(T)) \subseteq U(W) = R(U)$
- (c) and (d) follow from (a) and discussion of the transpose
- (b) follows from the previous by considering matrix representations.

Proof. Let $T : V \to W$ and $U : W \to Z$ be linear transformations on finite-dimensional vector spaces. Let A, B be matrices such that AB is defined. For (a), we have $R(T) = T(V) \subseteq W$. Then

$$\begin{aligned} R(UT) &= (UT)(V) \\ &= U(T(V)) \\ &\subseteq U(W) \\ &= R(U). \end{aligned}$$

Thus, $rank(UT) = dim(R(UT)) \leq dim(R(U)) = rank(U)$.

Now, for (c),

$$\begin{aligned} rank(AB) &= rank(L_{AB}) \\ &= rank(L_A L_B) \\ &\leq rank(L_A) && \text{(By (a))} \\ &= rank(A). \end{aligned}$$

Now, for (d),

$$\begin{aligned} rank(AB) &= rank((AB)^t) \\ &= rank(B^tA^t) \\ &\leq rank(B^t) && \text{(By (c))} \\ &= rank(B). \end{aligned}$$

Now, for (b), let α, β, γ be ordered bases in V, W, and Z. Let $A = \left[U\right]_\beta^\gamma$ and $B = \left[T\right]_\alpha^\beta$. Then $AB = \left[UT\right]_\alpha^\gamma$. Thus,

$$\begin{aligned} rank(UT) &= rank\left(\left[UT\right]_\alpha^\gamma\right) \\ &= rank(AB) \\ &\leq rank(B) && \text{(By (b))} \\ &= rank\left(\left[T\right]_\alpha^\beta\right) \\ &= rank(T). \end{aligned}$$

□

Observations: A is an invertible $n \times n$ matrix if and only if $(A|I_n)$ can be transformed into $(I_n|B)$ by elementary row operations, in this case $B = A^{-1}$.

So, $C = (A|I_n)$. Then $A^{-1}C = (A^{-1}A|A^{-1}) = (I_n|A^{-1})$. Consequently, $A^{-1} = E_1 \ldots E_p$ where E_i is elementary. Thus, $(E_1 \ldots E_p)C = (I_n|A^{-1})$. So, C can be converted to $(I_n|A^{-1})$ by elementary row operations.

Further, suppose we can transform C to $(I_n|B)$ by using elementary row operations. So, $E_1 \ldots E_p(A|I_n) = (I_n|B)$. Let $M = E_1 \ldots E_p$. Then, $MA = I_n$ and $M = B$. So, $MA = I_n$ implies $M = A^{-1} = B$. Finally, if A is not invertible, then by Theorem 3.6, $r < n$.

Example: Determine if $\begin{pmatrix} 4 & 0 & 1 \\ 2 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$ is invertible and find its inverse.

Solution.

$$\left(\begin{array}{ccc|ccc} 4 & 0 & 1 & 1 & 0 & 0 \\ 2 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 \end{array}\right) \xrightarrow{\text{Row Op.}} \left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 0 & 0 & 1 \\ 2 & 1 & 1 & 0 & 1 & 0 \\ 4 & 0 & 1 & 1 & 0 & 0 \end{array}\right)$$
$$\xrightarrow{\text{Row Op.}} \left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 0 & 0 & 1 \\ 0 & -1 & -1 & 0 & 1 & -2 \\ 0 & -4 & -3 & 1 & 0 & -4 \end{array}\right)$$
$$\xrightarrow{\text{Row Op.}} \left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & -1 & 2 \\ 0 & 4 & 3 & -1 & 0 & 4 \end{array}\right)$$
$$\vdots$$
$$\xrightarrow{\text{Row Op.}} \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 0 & 1 & -1 \\ 0 & 1 & 0 & -1 & 3 & -2 \\ 0 & 0 & 1 & 1 & -4 & 4 \end{array}\right)$$

Example: Let $T : P_2(\mathbb{R}) \to P_2(\mathbb{R})$ be defined by $T(f) = f + f' + f''$. Find T^{-1}.

Solution. Let $\beta = \{1, x, x^2\}$ be the standard ordered basis for $P_2(\mathbb{R})$. Then

$$\begin{aligned} T(1) &= 1 + 0 + 0 = 1 \\ T(x) &= x + 1 = 1 + x \\ T(x^2) &= x^2 + 2x + 2 = 2 + 2x + x^2. \end{aligned}$$

We have

$$\left[T^{-1}\right]_\beta = \left[T\right]_\beta^{-1} = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{pmatrix}.$$

Therefore,

$$\begin{aligned} T^{-1}(a_0 + a_1 x + a_2 x^2) &= \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} \\ &= \begin{pmatrix} a_0 - a_1 \\ a_1 - 2a_2 \\ a_2 \end{pmatrix} \\ &= (a_0 - a_1) + (a_1 - 2a_2)x + a_2 x^2. \end{aligned}$$

3.3 Systems of linear equations (theoretical aspect)

- $Ax = b$ is consistent if it has at least one solution and inconsistent otherwise.

- $Ax = 0$ is called a homogeneous system and $Ax = b$ for $b \neq 0$ a nonhomogeneous system.

Theorem 3.8. Let $Ax = 0$ be a homogeneous system over F in n unknowns and let K be the set of solutions. Then $K = N(L_A)$ and so K is a subspace of $\mathbb{F}^n$ and $dim(K) = n - rank(A)$.

Corollary 12 If $m < n$, then $Ax = 0$ has a non-zero solution.

Proof. We have $dim(K) = dim(N(L_A)) = n - rank(A)$. We know $rank(A) \leq m$. So, $dim(K) = n - rank(A) \geq n - m > 0$. □

Theorem 3.9. Let K be the solution set to $Ax = b$ and K_H be the solution set to $Ax = 0$. Then for any $s \in K$, $K = \{s\} + K_H = \{s + k : k \in K_H\}$.

Proof. Let $s \in K$. Then $K = s + K_H$.

(1) Let $w \in K$. Then $Aw = b$. So, $A(w - s) = Aw - As = b - b = 0$. Thus, $w - s \in K_H$ and we have $w = s + (w - s)$. So, $w \in \{s\} + K_H$.

(2) Let $w \in \{s\} + K_H$. Then $w = s + k$ for some $k \in K_H$, and so $Aw = A(s + k) = As + Ak = b + 0 = b$. Thus, $w \in K$.

□

Theorem 3.10. Let $Ax = b$ be a system of n linear equations in n unknowns. Then A is invertible if and only if the system has exactly one solution. Namely, $x = A^{-1}b$.

Proof. Let $Ax = b$ be a system of n linear equations in n unknowns.

($\Longrightarrow$): Suppose A is invertible. Then $x = A^{-1}b$ is a solution to $Ax = b$. Clearly, $AA^{-1}b = b$. If s is a solution to $Ax = b$, then $As = b$. So, $A^{-1}As = A^{-1}b$ implies $s = A^{-1}b$.

($\Longleftarrow$): Suppose $Ax = b$ has exactly one solution. Let s be this solution. Let K_H be the set of solutions to $Ax = 0$. Then, by theorem 3.9, $\{s\} = \{s\} + K_H$. Thus, $K_H = \{0\}$. Therefore, $N(L_A) = K_H = \{0\}$. Therefore, L_A is injective and surjective. Thus, L_A has an inverse. So, A has an inverse.

□

Theorem 3.11. The system $Ax = b$ is consistent if and only if $rank(A) = rank(A|b)$.

Proof. Note, $R(L_A) = span(\{a_1, \ldots, a_n\})$ where a_i is the ith column of A.

$$\begin{aligned}
Ax = b \text{ is consistent } &\Longleftrightarrow b \in R(L_A) \\
&\Longleftrightarrow b \in span(\{a_1, \ldots, a_n\}) \\
&\Longleftrightarrow span(\{a_1, \ldots, a_n, b\}) = span(\{a_1, \ldots, a_n\}) \\
&\Longleftrightarrow dim(span(\{a_1, \ldots, a_n, b\})) = dim(span(\{a_1, \ldots, a_n\})) \\
&\Longleftrightarrow rank(A|b) = rank(A).
\end{aligned}$$

□

4 Determinants

4.1 Determinants of order 2

Definition 1 If $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, then the determinant of A is $ad - bc$.

Notation: The determinant of A will be denoted by $det(A)$ or $|A|$.

Theorem 4.1. For $u, v, w \in \mathbb{F}^2$ and $k \in \mathbb{F}$

$$det \begin{pmatrix} u + kv \\ w \end{pmatrix} = det \begin{pmatrix} u \\ w \end{pmatrix} + k det \begin{pmatrix} v \\ w \end{pmatrix}$$
$$det \begin{pmatrix} w \\ u + kv \end{pmatrix} = det \begin{pmatrix} w \\ u \end{pmatrix} + k det \begin{pmatrix} w \\ v \end{pmatrix}.$$

Theorem 4.2. Let $A \in M_{2\times 2}(\mathbb{F})$. Then the determinant of A is non-zero if and only if A is invertible. If A is invertible, then $A^{-1} = \frac{1}{det(A)} \begin{pmatrix} A_{22} & -A_{12} \\ -A_{21} & A_{22} \end{pmatrix}$.

Proof. Let $A \in M_{2\times 2}(\mathbb{F})$.

($\Longrightarrow$): Let $M = \frac{1}{det(A)} \begin{pmatrix} A_{22} & -A_{12} \\ -A_{21} & A_{11} \end{pmatrix}$. Then,

$$\begin{aligned} AM &= \frac{1}{det(A)} \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} A_{22} & -A_{12} \\ -A_{21} & A_{11} \end{pmatrix} \\ &= I_2 \\ &= MA. \end{aligned}$$

($\Longleftarrow$): Suppose A is invertible. Then $rank(A) = 2$. If $A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$, then $A_{11} \neq 0$ or $A_{12} \neq 0$.

(a) Suppose $A_{11} \neq 0$. Then we can transform A into

$$\begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} - \frac{A_{21}A_{12}}{A_{11}} \end{pmatrix},$$

which has a *rank* of 2. Therefore, $A_{22} - \frac{A_{21}A_{12}}{A_{11}} \neq 0$. Thus, $det(A) \neq 0$.

(b) Suppose $A_{11} = 0$. Then $A_{12} \neq 0$ and $A_{21} \neq 0$. So, $det(A) = -A_{12} - A_{21} \neq 0$.

$\square$

Observation: Let $\delta : M_{2\times 2}(\mathbb{F}) \to \mathbb{F}$ be such that

(1) $\delta \begin{pmatrix} u+kv \\ w \end{pmatrix} = \delta \begin{pmatrix} u \\ w \end{pmatrix} + k\delta \begin{pmatrix} v \\ w \end{pmatrix}$

$\delta \begin{pmatrix} w \\ u+kv \end{pmatrix} = \delta \begin{pmatrix} w \\ u \end{pmatrix} + k\delta \begin{pmatrix} w \\ v \end{pmatrix}$;

(2) $\delta \begin{pmatrix} u \\ u \end{pmatrix} = 0$;

(3) $\delta(I_2) = 1$.

Then $\delta = det$.

Proof. We have

$$\delta\left(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right) + \delta\left(\begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}\right) = \delta\left(\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\right)$$
$$\delta\left(\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\right) + \delta\left(\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}\right) = \delta\left(\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}\right)$$

and

$$\delta\left(\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\right) + \delta\left(\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}\right) = \delta\left(\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\right) = 0.$$

Thus,

$$\delta\left(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right) + \delta\left(\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\right) = 0$$

and

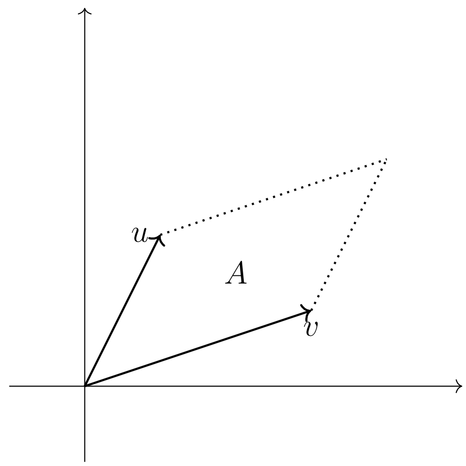
$$\delta\left(\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\right) = -1.$$

Then,

$$\begin{aligned}
\delta\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right) &= \delta\left(\begin{pmatrix} a & 0 \\ c & d \end{pmatrix}\right) + \delta\left(\begin{pmatrix} 0 & b \\ c & d \end{pmatrix}\right) \\
&= a\delta\left(\begin{pmatrix} 1 & 0 \\ c & d \end{pmatrix}\right) + b\delta\left(\begin{pmatrix} 0 & 1 \\ c & d \end{pmatrix}\right) \\
&= a\left(c\delta\left(\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}\right) + d\delta\left(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)\right) + b\left(c\delta\left(\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\right) + d\delta\left(\begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}\right)\right) \\
&= ad - bc \\
&= det\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right).
\end{aligned}$$

□

Observation: When $\mathbb{F} = \mathbb{R}$ and $M_{2\times 2}(\mathbb{R})$ and for $\begin{pmatrix} u \\ v \end{pmatrix}$, then we have the following diagram.



We will define the area of the parallelogram to be $A\begin{pmatrix} u \\ v \end{pmatrix}$. If u and v are linearly dependent, then $A\begin{pmatrix} u \\ v \end{pmatrix} = 0$. We can show $A\begin{pmatrix} u \\ v \end{pmatrix} = \text{sign}\left(det\begin{pmatrix} u \\ v \end{pmatrix}\right) det\begin{pmatrix} u \\ v \end{pmatrix}$.

Proof. Define $O : M_{2\times 2}(\mathbb{R}) \to \{-1, 1\}$ by

$$O\begin{pmatrix} u \\ v \end{pmatrix} = \begin{cases} \frac{det\begin{pmatrix} u \\ v \end{pmatrix}}{\left|det\begin{pmatrix} u \\ v \end{pmatrix}\right|} & \text{if } u, v \text{ are linearly independent} \\ 1 & \text{if } u, v \text{ are linearly dependent.} \end{cases}$$

Thus, we can show $det\begin{pmatrix} u \\ v \end{pmatrix} = O\begin{pmatrix} u \\ v \end{pmatrix} A\begin{pmatrix} u \\ v \end{pmatrix}$. We will show $\delta\begin{pmatrix} u \\ v \end{pmatrix} = O\begin{pmatrix} u \\ v \end{pmatrix} A\begin{pmatrix} u \\ v \end{pmatrix}$ satisfies the previous observation. Note,

$$\delta\begin{pmatrix} u \\ u \end{pmatrix} = O\begin{pmatrix} u \\ u \end{pmatrix} A\begin{pmatrix} u \\ u \end{pmatrix} = 0 \quad \text{and} \quad \delta\left(I_2\right) = \delta\begin{pmatrix} e_1 \\ e_2 \end{pmatrix} A\begin{pmatrix} e_1 \\ e_2 \end{pmatrix} = 1.$$

We now have three steps.

(a) If $c = 0$, then $\delta\begin{pmatrix} u \\ 0 \end{pmatrix} = 0$. If $c \neq 0$, then $\delta\begin{pmatrix} u \\ cv \end{pmatrix} = O\begin{pmatrix} u \\ cv \end{pmatrix} A\begin{pmatrix} u \\ cv \end{pmatrix}$ and $A\begin{pmatrix} u \\ cv \end{pmatrix} = |c|A\begin{pmatrix} u \\ v \end{pmatrix}$ and $O\begin{pmatrix} u \\ cv \end{pmatrix} = \frac{det\begin{pmatrix} u \\ cv \end{pmatrix}}{\left|det\begin{pmatrix} u \\ cv \end{pmatrix}\right|} = \frac{c}{|c|}$. So, $\delta\begin{pmatrix} u \\ cv \end{pmatrix} = c\delta\begin{pmatrix} u \\ v \end{pmatrix}$.

(b) Note, $\delta\begin{pmatrix} u \\ u+v \end{pmatrix} = \delta\begin{pmatrix} u \\ v \end{pmatrix}$. If $a=0$, then $\delta\begin{pmatrix} u \\ au+bv \end{pmatrix} = \delta\begin{pmatrix} u \\ bv \end{pmatrix} = b\delta\begin{pmatrix} u \\ v \end{pmatrix}$. If $a \neq 0$, then

$$\begin{aligned}\delta\begin{pmatrix} u \\ au+bv \end{pmatrix} &= a\delta\begin{pmatrix} u \\ u+\frac{b}{a}v \end{pmatrix} \\ &= a\delta\begin{pmatrix} u \\ \frac{b}{a}v \end{pmatrix} \\ &= b\delta\begin{pmatrix} u \\ v \end{pmatrix}.\end{aligned}$$

(c) Assume $u \neq 0$. Let $w \in \mathbb{R}^2$ such that uw is linearly independent. Then $v_1 = a_1u + b_1w$ and $v_2 = a_2u + b_2w$. So,

$$\begin{aligned}\delta\begin{pmatrix} u \\ v_1+v_2 \end{pmatrix} &= \delta\begin{pmatrix} u \\ (a_1+a_2)u + (b_1+b_2)w \end{pmatrix} \\ &= (b_1+b_2)\begin{pmatrix} u \\ w \end{pmatrix}. && \text{(by (b))}\end{aligned}$$

Also, by (b), $\delta\begin{pmatrix} u \\ v_1 \end{pmatrix} + \delta\begin{pmatrix} u \\ v_2 \end{pmatrix} = b_1\begin{pmatrix} u \\ w \end{pmatrix} + b_2\delta\begin{pmatrix} u \\ w \end{pmatrix}$. Therefore,

$$\delta\begin{pmatrix} u \\ v_1+v_2 \end{pmatrix} = \delta\begin{pmatrix} u \\ v_1 \end{pmatrix} + \delta\begin{pmatrix} u \\ v_2 \end{pmatrix}.$$

□

4.2 Determinants of order n

Let $\tilde{A}_{ij}$ be the matrix obtained from A by deleting the ith row and the jth column.

Definition 2 Let $A \in M_{n\times n}(\mathbb{F})$.

- If $n = 1$, then $det(A) = A_{11}$.
- If $n \geq 2$, then $det(A) = \sum_{j=1}^{n}(-1)^{1+j}A_{1j}det(\tilde{A}_{1j})$.

The cofactor of the i,jth entry of A,

$$c_{ij} = (-1)^{i+j}det(\tilde{A}_{ij}).$$

Example:

$$\begin{vmatrix} 0 & 1 & 3 \\ -2 & -3 & -5 \\ 4 & -4 & 4 \end{vmatrix} = 0 \cdot (-1)^2 \cdot \begin{vmatrix} -3 & -5 \\ -4 & 4 \end{vmatrix} + 1 \cdot (-1)^3 \cdot \begin{vmatrix} -2 & -5 \\ 4 & 4 \end{vmatrix} + 3 \cdot (-1)^4 \cdot \begin{vmatrix} -2 & -3 \\ 4 & -4 \end{vmatrix}$$
$$= 0 - 12 + 60$$
$$= 48.$$

Theorem 4.3. Let $a_1, \ldots, a_n \in \mathbb{F}^n$, let $k \in \mathbb{F}$ and suppose $a_r = u + kv$ for some $u, v \in \mathbb{F}^n$. Then

$$\begin{vmatrix} a_1 \\ a_{r-1} \\ a_r \\ a_{r+1} \\ a_n \end{vmatrix} = \begin{vmatrix} a_1 \\ a_{r-1} \\ u \\ a_{r+1} \\ a_n \end{vmatrix} + k \begin{vmatrix} a_1 \\ a_{r-1} \\ v \\ a_{r+1} \\ a_n \end{vmatrix}.$$

Proof. Let $a_r = u + kv$.

(Base Case): Let $n = 1$. Then clearly, $det(A) = A_{11}$.

(Inductive Step): Let $n \geq 2$.

- If $r = 1$, then $A_{1j} = u_j + kv_j$. Let B and C be matrices obtained from A by replacing row r by u and v. Then

$$det(\tilde{A}_{1j}) = det(\tilde{B}_{1j}) = det(\tilde{C}_{1j}).$$

Thus,

$$\begin{aligned} det(A) &= \sum_{j=1}^{n} (-1)^{1+j} A_{1j} det(\tilde{A}_{1j}) \\ &= \sum_{j=1}^{n} (-1)^{1+j} (u_j + kv_j) det(\tilde{A}_{1j}) \\ &= \sum_{j=1}^{n} (-1)^{1+j} u_j det(\tilde{B}_{1j}) + \sum_{j=1}^{n} (-1)^{1+j} kv_j det(\tilde{C}_{1j}) \\ &= \sum_{j=1}^{n} (-1)^{1+j} B_{1j} det(\tilde{B}_{1j}) + k \sum_{j=1}^{n} (-1)^{1+j} C_{1j} det(\tilde{C}_{1j}) \\ &= det(B) + kdet(C). \end{aligned}$$

– If $r > 1$, then $\tilde{A}_{1j} = \tilde{B}_{1j} = \tilde{C}_{1j}$ are the same except row $r-1$, which is $(u_1 + kv_1, \ldots, u_{j-1} + kv_{j-1}, u_{j+1} + kv_{j+1}, \ldots, u_n + kv_n)$ in A, $(u_1, \ldots, u_{j-1}, u_{j+1}, \ldots, u_n)$ in B, and $(v_1, \ldots, v_{j-1}, v_{j+1}, \ldots, v_n)$ in C. Therefore, by the inductive hypothesis

$$det(\tilde{A}_{1j}) = det(\tilde{B}_{1j}) + k \cdot det(\tilde{C}_{1j}).$$

Thus,

$$\begin{aligned} det(A) &= \sum_{j=1}^{n} (-1)^{1+j} A_{1j} det(\tilde{A}_{1j}) \\ &= \sum_{j=1}^{n} (-1)^{1+j} B_{1j} det(\tilde{B}_{1j}) + k \cdot \sum_{j=1}^{n} (-1)^{1+j} C_{1j} det(\tilde{C}_{1j}) \\ &= det(A) + k \cdot det(C). \end{aligned}$$

Therefore,

$$det(A) = det(B) + k \cdot det(C).$$

□

Corollary 4 If A has a row consisting of zeroes, then $det(A) = 0$.

Lemma 5. Let $B \in M_{n \times n}(\mathbb{F})$ and $n \geq 2$. Suppose that the ith row of B is e_k for some $1 \leq k \leq n$. Then $det(B) = (-1)^{i+k} det(\tilde{B}_{ik})$.

Theorem 4.4. For $A \in M_{n \times n}(\mathbb{F})$ and $i \in \{1, \ldots, n\}$ and $n \geq 2$

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

Proof. Let B_j denote the matrix obtained from A by replacing the ith row with e_j. Then

$$A = \sum_{j=1}^{n} A_{ij} B_j.$$

Thus,

$$\begin{aligned} det(A) &= \sum_{j=1}^{n} A_{ij} det(B_j) && \text{(By Theorem 4.3)} \\ &= \sum_{j=1}^{n} A_{ij} (-1)^{i+j} det(\tilde{A}_{ij}). && \text{(By Lemma 5)} \end{aligned}$$

□

Corollary 7 If $A \in M_{n\times n}(\mathbb{F})$ has two identical rows, then $det(A) = 0$.

Proof. Let $A \in M_{n\times n}(\mathbb{F})$.

(Base Case): Let $n = 2$. Then $\begin{pmatrix} a & b \\ a & b \end{pmatrix} = ab - ab = 0$.

(Inductive Step): Let $n \geq 3$. Suppose rows r and s are identical. Let $i \in \{1, \ldots, n\}$ such that $i \neq r$ and $i \neq s$. By Theorem 4.4,

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

and $\tilde{A}_{ij}$ has two identical rows. So, by the inductive hypothesis, $det(\tilde{A}_{ij}) = 0$. Therefore, $det(A) = 0$.

□

Theorem 4.5. If $A \in M_{n\times n}(\mathbb{F})$ and B is obtained from A by interchanging two rows, then $det(B) = -det(A)$.

Proof. Let $A \in M_{n\times n}(\mathbb{F})$. Suppose B is obtained from A by interchanging rows r and s

with $r < s$, without loss of generality. Let a_i denote the ith row of A. Then

$$\begin{aligned}
0 = det\begin{pmatrix} a_1 \\ \vdots \\ a_r + a_s \\ \vdots \\ a_r + a_s \\ \vdots \\ a_n \end{pmatrix}
&= det\begin{pmatrix} a_1 \\ \vdots \\ a_r \\ \vdots \\ a_r + a_s \\ \vdots \\ a_n \end{pmatrix} + det\begin{pmatrix} a_1 \\ \vdots \\ a_s \\ \vdots \\ a_r + a_s \\ \vdots \\ a_n \end{pmatrix} \\
&= det\begin{pmatrix} a_1 \\ \vdots \\ a_r \\ \vdots \\ a_r \\ \vdots \\ a_n \end{pmatrix} + det\begin{pmatrix} a_1 \\ \vdots \\ a_r \\ \vdots \\ a_s \\ \vdots \\ a_n \end{pmatrix} + det\begin{pmatrix} a_1 \\ \vdots \\ a_s \\ \vdots \\ a_r \\ \vdots \\ a_n \end{pmatrix} + det\begin{pmatrix} a_1 \\ \vdots \\ a_s \\ \vdots \\ a_s \\ \vdots \\ a_n \end{pmatrix} \\
&= 0 + det(A) + det(B) + 0.
\end{aligned}$$

Therefore, $det(B) = -det(A)$. □

Theorem 4.6. If $A \in M_{n\times n}(\mathbb{F})$ and B is obtained from A by adding a multiple of one row to another, then $det(B) = det(A)$.

Proof. Let $A \in M_{n\times n}(\mathbb{F})$. Let B be obtained from A by adding ka_r to a_s. Let C be obtained from A by replacing row s with a_r. Then $C = \begin{pmatrix} a_1 \\ \vdots \\ a_r \\ \vdots \\ a_r \\ \vdots \\ a_n \end{pmatrix}$ and so, $det(C) = 0$. In addition, row s of B is equal to row s of A plus k times row s of C. Thus,

$$\begin{aligned} det(B) &= det(A) + kdet(C) \\ &= det(A). \end{aligned}$$

□

Corollary 10 If $A \in M_{n\times n}(\mathbb{F})$ and $rank(A) < n$, then $det(A) = 0$.

Proof. Let $A \in M_{n\times n}(\mathbb{F})$ and $rank(A) < n$. Then the dimension of the row space of A is less than n. Thus, the rows of A are linearly dependent. Therefore, for some $r \in \{1, \dots, n\}$, $a_r = \sum_{j\neq r} c_j a_j$ for some $c_j \in \mathbb{F}$. For every $j \neq r$, if $c_j \neq 0$, then adding $-c_j a_j$ to row r results in the zero vector. Therefore, $det(A) = 0$. □

Example:

$$\begin{aligned} \begin{vmatrix} 0 & 2 & 1 & 5 \\ 1 & 5 & 0 & 6 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & -2 \end{vmatrix} \xrightarrow{\text{row op.}} - &\begin{vmatrix} 1 & 5 & 0 & 6 \\ 0 & 2 & 1 & 5 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & -2 \end{vmatrix} \\ &= -1 \cdot 2 \cdot 4 \cdot (-2) \\ &= 16. \end{aligned}$$

4.3 Properties

Summary of properties

- Let $A, B \in M_{n\times n}(\mathbb{F})$. Then $det(AB) = det(A)det(B)$.

  Proof. Let $A, B \in M_{n\times n}(\mathbb{F})$.

  (1) We will show this is the case when A is an elementary matrix.

  - If A is of Type I (interchange), then $det(A) = -det(I) = -1$ and by Theorem 4.5, $det(AB) = -det(B)$.
  - If A is Type II (multiply by k), then $det(A) = k$ and $det(AB) = kdet(B)$.
  - If A is of type III (adding), then $det(A) = 1$ and $det(AB) = det(B)$.

  So, the fact holds when A is an elementary matrix.

  (2) If $rank(A) < n$, then by Corollary 10, $det(A) = 0$. In addition, $rank(AB) \leq rank(A) < n$. Thus, $det(A) = 0$.

(3) If $rank(A) = n$, then A is invertible and a product of elementary matrices. Say $A = E_1 E_2 \dots E_m$ where E_i are elementary matrices. Then

$$\begin{aligned} det(AB) &= det(E_1 \dots E_m \cdot B) \\ &= det(E_1)det(E_2 \dots E_m \cdot B) \\ &\vdots \\ &= det(E_1)det(E_2)\dots det(E_m)det(B) \\ &= det(E_1 \dots E_m)det(B) \\ &= det(A)det(B). \end{aligned}$$

□

- $A \in M_{n\times n}(\mathbb{F})$ is invertible if and only if $det(A) \neq 0$ and $det(A^{-1}) = \frac{1}{det(A)}$.

  Proof.

  $\Longrightarrow$ If A is not invertible, then $rank(A) < n$ and so $det(A) = 0$.

  $\Longleftarrow$ If A is invertible, then $A \cdot A^{-1} = I_n$. Thus, by the fact above, $1 = det(I_n) = det(A)det(A^{-1})$. So, $det(A) \neq 0$.

  □

- Let $A \in M_{n\times n}(\mathbb{F})$. Then $det(A^t) = det(A)$.

  Proof.

  (1) If A is not invertible, then $rank(A) < n$ and $det(A) = 0$. Also, $rank(A^t) = rank(A) < n$. Thus, $det(A^t) = 0$.

  (2) If A is invertible, then A is a product of elementary matrices, say $A = E_1 E_2 \dots E_m$. Then $A^t = (E_1 \dots E_m)^t = E_m^t \dots E_1^t$. Thus,

$$\begin{aligned} det(A^t) &= det(E_m^t)\dots det(E_1^t) \\ &= det(E_m)\dots det(E_1) \\ &= det(E_1)\dots det(E_m) \\ &= det(A). \end{aligned}$$

  □

- (Cramer's Rule) Let $Ax = b$ where $A \in M_{n\times n}(\mathbb{F})$ and let M_k be the $n \times n$ matrix obtained from A by replacing column k with b. Then $x_k = \frac{det(M_k)}{det(A)}$.

  Proof. Let $A \in M_{n\times n}(\mathbb{F})$ be nonzero. Thus, A is invertible, and so, $Ax = b$ has a unique solution. Fix k. Let X be obtained from I_n by replacing the kth column with

x. Then the ith column of AX is $Ae_i = a_i$ if $i \neq k$ and $Ax = b$ if $i = k$. Therefore $AX = M_k$ and so

$$\begin{aligned} det(A)det(X) &= det(AX) \\ & det(M_k). \end{aligned}$$

So, $det(X) = x_k$. Therefore, $x_k = \frac{det(M_k)}{det(A)}$. $\square$

5 Diagonalization

5.1 Eigenvalues and eigenvectors

Definition 1 Let V be finite-dimensional and let $T : V \to V$ be linear. T is called *diagonalizable* if there is an ordered basis β such that $\left[T\right]_\beta$ is a diagonal matrix.

Matrix A is diagonalizable if L_A is.

Definition 2 A non-zero vector $v \in V$ such that $T(v) = \lambda v$ for some $\lambda \in \mathbb{F}$ is called an *eigenvector*. The scalar λ is called an *eigenvalue* of T.

Theorem 5.1. Let V be finite-dimensional. A linear operator on V is diagonalizable if and only if there is an ordered basis β for V consisting of eigenvectors of T.

Proof.

$\Longrightarrow$ Suppose T is diagonalizable. Then there exists an ordered basis $\beta = \{v_1, \ldots, v_n\}$ for V such that $\left[T\right]_\beta = D$ where $D = \left[D_{ij}\right]$ and $D_{ij} = 0$ if $i \neq j$. Then

$$T(v_j) = \sum_{i=1}^{n} D_{ij} v_i = D_{ij} v_i.$$

Thus, v_j is an eigenvector of T since $v_j \neq 0$.

$\Longleftarrow$ Suppose $\beta = \{v_1, \ldots, v_n\}$ is a basis for V consisting of eigenvectors of T, say $T(v_j) = \lambda_j v_j$. Then $\left[T\right]_\beta = D = [D_{ij}]$ where

$$D_{ij} = \begin{cases} \lambda_j & i = j \\ 0 & i \neq j. \end{cases}$$

Thus, D is diagonal.

$\square$

Theorem 5.2. Let $A \in M_{n\times n}(\mathbb{F})$. Then $\lambda \in \mathbb{F}$ is an eigenvalue of A if and only if $det(A - \lambda I_n) = 0$.

Proof.

$$\begin{aligned}
\lambda \text{ is an eigenvalue of } A &\iff \exists v \neq 0 \text{ such that } Av - \lambda v = 0 \\
&\iff \exists v \neq 0 \text{ such that } (A - \lambda I_n)v = 0 \\
&\iff \exists v \neq 0 \text{ such that } v \in N(A - \lambda I_n) \\
&\iff A - \lambda I_n \text{ is not invertible} \\
&\iff det(A - \lambda I_n) = 0
\end{aligned}$$

□

Definition 3

- Let $A \in M_{n\times n}(\mathbb{F})$. Then $f(t) = det(A - tI_n)$ is called the *characteristic polynomial* of A.
- If T is a linear operator V, then the characteristic polynomial of T is the characteristic polynomial of $A = \left[T\right]_\beta$ where β is an ordered basis for V.

Note: The characteristic polynomial of T is well-defined.

Proof. Let β and α be ordered bases for V. Then $f_T(t) = det(\left[T\right]_\beta - tI) = det(\left[T\right]_\alpha - tI)$. Let $B = \left[T\right]_\beta$ and $A = \left[T\right]_\alpha$. Then there exists Q invertible such that $B = Q^{-1}AQ$. So,

$$\begin{aligned}
det(B - tI) &= det(Q^{-1}AQ - tI) \\
&= det(Q^{-1}(A - tI)Q) \\
&= det(Q^{-1})det(A - tI)det(Q) \\
&= det(A - tI).
\end{aligned}$$

□

Theorem 5.3. Let $A \in M_{n\times n}(\mathbb{F})$.

- $f_A(t)$ is a polynomial of degree n with the leading coefficient $(-1)^n$.
- A has at most n distinct eigenvalues.

Theorem 5.4. Let T be a linear operator on V and let λ be an eigenvalue of T. A vector $v \in V$ is an eigenvector of T corresponding to λ if and only if $v \neq 0$ and $v \in N(T - \lambda I)$.

Example: Let $C^\infty(\mathbb{R}) = \{f : \mathbb{R} \to \mathbb{R} : f^{(n)}$ exists for every $n \in \mathbb{Z}^+\}$. Then $C^\infty(\mathbb{R})$ is a subspace of $\mathcal{F}(\mathbb{R}, \mathbb{R})$. Let $T : C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R})$, $T(f) = f'$. Find all eigen values of T.

Solution. We know f is an eigenvector if $f \neq 0$ and $Tf = \lambda f$ for some $\lambda \in \mathbb{R}$. So, $f' = \lambda f$ implies $\frac{df}{dx} = \lambda f$ implies $\frac{df}{f} = \lambda dx$ implies $ln(f) = \lambda x + C$ implies $|f| = e^{\lambda x + C}$ implies $f(x) = Le^{\lambda x}$ for some $L \in \mathbb{R}$ and $L \neq 0$. All $\lambda \in \mathbb{R}$ are eigenvalues.

Example: Let $T : M_{2\times 2}(\mathbb{R}) \to M_{2\times 2}(\mathbb{R})$ be defined by $T(A) = A^T$. Is T diagonalizable?

Solution. Let $\alpha = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \right\} = \{e_1, e_2, e_3, e_4\}$. So,

$$[T]_\alpha = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

is not diagonalizable. We need to find the eigenvalues of T. If λ is an eigenvalue, then $T(A) = \lambda A$ for some $A \neq 0$. Then, $T^2(A) = T(T(A)) = \lambda^2 A$. Thus, $\lambda^2 A = A \neq 0$. So, $\lambda = \pm 1$.

($\lambda = 1$): Then $T(A) = A$ and $T(A) = A^T$. Thus, $A = A^T$ is a symmetric matrix. Let $\left\{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \right\} = \{f_1, f_2, f_3\}$.

($\lambda = -1$): Then $T(A) = -A$ and $T(A) = A^T$. So, $A^T = -A$ is a skew-symmetric matrix. Let $\left\{ \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \right\} = \{f_4\}$.

Then $\beta = \{f_1, f_2, f_3, f_4\}$ is a basis of eigenvectors of T. Then $T(f_1) = f_1,$, $T(f_2) = f_2$, $T(f_3) = f_3$, and $T(f_4) = f_4$ implies

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Example: Let $A \in M_{2\times 2}(\mathbb{F})$ be defined as $\begin{pmatrix} 1 & 1 \\ 4 & 1 \end{pmatrix}$. Find the eigenvectors of A.

Solution.

- Finding the eigenvalues of A, we have

$$\begin{aligned} f_A(t) &= \begin{vmatrix} 1-t & 1 \\ 4 & 1-t \end{vmatrix} \\ &= (1-t)^2 - 4 \\ &= (t-3)(t+1). \end{aligned}$$

  So, the eigenvalues of A are $\lambda_1 = 3$ and $\lambda_2 = -1$.

- Finding the eigenvectors of A, we have,

  λ_1: We have $A-3I = \begin{pmatrix} -2 & 1 \\ 4 & -2 \end{pmatrix}$. So, $(A-3I)\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 0$ implies $\begin{pmatrix} -2 & 1 \\ 4 & -2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} =$ 0 implies $x_1 = t$ and $x_2 = 2t$. Thus, $N(A-\lambda_1 I) = \left\{ t\begin{pmatrix} 1 \\ 2 \end{pmatrix} : t \in \mathbb{R} \right\}$.

  λ_2: We have $A+I = \begin{pmatrix} 2 & 1 \\ 4 & 2 \end{pmatrix}$. So, $(A+I)\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 0$ implies $\begin{pmatrix} 2 & 1 \\ 4 & 2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 0$ implies $x_1 = t$ and $x_2 = -2t$. Thus, $N(A-\lambda_1 I) = \left\{ t\begin{pmatrix} 1 \\ -2 \end{pmatrix} : t \in \mathbb{R} \right\}$.

Now, is A diagonalizable? Yes. For $\beta = \left\{ \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ -2 \end{pmatrix} \right\}$, we have $[L_A]_\beta = \begin{pmatrix} 3 & 0 \\ 0 & -1 \end{pmatrix}$.

5.2 Diagonalizability

Theorem 5.5. Let T be a linear operator with distinct eigenvalues $\lambda_1, \ldots, \lambda_k$. If $v_1, \ldots, v_k$ are eigenvectors of T such that v_i corresponds to λ_i, then $\{v_1, \ldots, v_k\}$ is linearly independent.

Proof. Let $\lambda_1, \ldots, \lambda_k$ be distinct eigenvalues.

Base Step: If $k = 1$, then $v_1 \neq 0$ implies v_1 is linearly independent.

Induction Step: Suppose $v_1, \ldots, v_{k-1}$ is linearly independent. Suppose $\sum_{i=1}^k c_i v_i = 0$ for some $c_i \in \mathbb{F}$. Then

$$(T-\lambda_k I)\left(\sum_{i=1}^k c_i v_i\right) = (T-\lambda_k I)(0) = 0.$$

Thus,

$$\begin{aligned} 0 &= \sum_{i=1}^k c_i (T-\lambda_k I)(v_i) \\ &= \sum_{i=1}^k c_i (\lambda_i - \lambda_k)(v_i) && \text{(Since } T(v_i) = \lambda_i v_i\text{)} \\ &= \sum_{i=1}^{k-1} c_i (T-\lambda_k I)(v_i). \end{aligned}$$

Since $v_1, \ldots, v_{k-1}$ is linearly independent, then $c_i(\lambda_i - \lambda_k) = 0$. Thus, $c_i = 0$ for $i = 1, \ldots, k-1$ since $\lambda_i \neq \lambda_k$ if $i < k$. Thus, $c_k v_k = 0$ and $v_k \neq 0$ implies $c_k = 0$ and $v_1, \ldots, v_k$ is linearly independent. $\square$

Corollary 6 If T has n distinct eigenvalues, then T is diagonalizable.

Definition 4 A polynomial $f(t) \in P(\mathbb{F})$ splits over $\mathbb{F}$ if $f(t) = c(t-a_1)\dots(t-a_n)$ for some $c, a_1, \dots, a_n \in \mathbb{F}$.

Theorem 5.6. Let T be a linear operator. If T is diagonalizable, then its characteristic polynomial splits.

Proof. Suppose T is diagonalizable. Then there exists $\beta = \{v_1, \dots, v_n\}$ such that v_i is an eigenvector of T, say $T(v_j) = \lambda_j v_j$. Consequently, $A = \left[T\right]_\beta = I\lambda_i$ and so

$$\begin{aligned} f_T(t) &= f_A(t) \\ &= (\lambda_1 - t)\dots(\lambda_n - t) \\ &= (-1)^{n+1}(t-\lambda_1)\dots(t-\lambda_n). \end{aligned}$$

Therefore, $f_T(t)$ splits. □

Lemma 8. Let $X \in M_{n\times n}(\mathbb{F})$ such that $X = \begin{pmatrix} A & B \\ 0 & C \end{pmatrix}$ for $A \in M_{m\times m}(\mathbb{F})$, $B \in M_{m\times p}(\mathbb{F})$, and $C \in M_{p\times p}(\mathbb{F})$. Then $det(X) = det(A)det(C)$.

Proof.

- Suppose $C = I_p$. We will use induction on p. The base case is clearly true. Suppose $p \geq 2$. Use the expansion of the last row. Then $det(X) = (-1)^{n+n} \cdot 1 \cdot det\left(\begin{pmatrix} A & B' \\ 0 & I_{p-1} \end{pmatrix}\right)$ where B' is obtained from B by deleting the last column. Then, by the inductive hypothesis, $det(X) = det(A)det(I_{p-1}) = det(A)$. Thus, $det(X) = det(A)det(C)$.
- Suppose $A = I_m$, then use the first column. Thus, $det(X) = det(A)det(C)$.
- For the general case,
  - If $det(A) = 0$, then the columns of A are linearly dependent. Then the first m columns of X are linearly dependent. Thus, $det(X) = 0 = det(A)det(C)$.
  - If $det(A) \neq 0$, then A is invertible. Then,

    $$X = \begin{pmatrix} A & B \\ 0 & C \end{pmatrix} = \begin{pmatrix} A & 0 \\ 0 & I_p \end{pmatrix} \begin{pmatrix} I_m & A^{-1}B \\ 0 & C \end{pmatrix}.$$

    Therefore, $det(X) = det(A)det(C)$.

□

Definition 5 The (algebraic) *multiplicity* of an eigenvalue λ is the largest positive integer k such that $(t-\lambda)^k | f(t)$.

The eigenspace of T with respect to λ is

$$E_\lambda = N(T - \lambda I_V) = \{x \in V | T(x) = \lambda x\}.$$

Theorem 5.7. Let λ be an eigenvalue of T of multiplicity m. Then

$$1 \le dim(E_\lambda) \le m.$$

Proof. (Outline)

- Start with an ordered basis $\{v_1, \dots, v_p\}$ for E_λ and extend it to an ordered basis β for V.
- Then $A = \big[T\big]_\beta$ will have the following form

$$\begin{pmatrix} \lambda I_p & B \\ 0 & C \end{pmatrix}.$$

Proof. Let $k = dim(E_\lambda)$ and let $\{v_1, \dots, v_k\}$ be a basis for E_λ. Extend the basis to a basis $\beta = \{v_1, \dots, v_k, v_{k-1}, \dots, v_n\}$ for V. Then $T(v_j) = \lambda v_j$ for $j = 1, \dots, k$. So,

$$A = \big[T\big]_\beta = \begin{pmatrix} \lambda I_k & B \\ 0 & C \end{pmatrix}.$$

Therefore,

$$\begin{aligned} f_T(t) &= f_A(t) \\ &= \begin{vmatrix} (\lambda - t)I_k & B \\ 0 & C - tI_{n-k} \end{vmatrix} \\ &= det((\lambda - t)I_k) det(C - tI_{n-k}) && \text{(By Lemma 8)} \\ &= (-1)^k (t-\lambda)^k g(t). \end{aligned}$$

Thus, $(t-\lambda)^k | f_T(t)$ and $k \le m$. □

Example: Let $T : P_3(\mathbb{R}) \to P_3(\mathbb{R})$ be defined as $T(f) = f'$. Is T diagonalizable?

Solution. Let $\beta = \{1, x, x^2, x^3\}$ be a basis for $P_3(\mathbb{R})$. Then

$$\big[T\big]_\beta = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Thus,

$$f_T(t) = \begin{vmatrix} -t & 1 & 0 & 0 \\ 0 & -t & 2 & 0 \\ 0 & 0 & -t & 3 \\ 0 & 0 & 0 & -t \end{vmatrix} = t^4.$$

So, there is only one eigenvalue, $\lambda = 0$. Thus,

$$\begin{aligned} E_0 &= N(T - 0I) \\ &= \{f : Tf = 0\} \\ &= \{f : f' = 0\} \\ &= P_0(\mathbb{R}). \end{aligned}$$

Thus, $dim(E_0) = 1$. Therefore, it is not possible to find a basis consisting of eigenvectors.

Lemma 9. Let $\lambda_1, \dots, \lambda_k$ be distinct eigenvalues of T and let $v_i \in E_{\lambda_i}$. If $v_1 + \cdots + v_k = 0$, then $v_i = 0$ for every i.

Proof. Renumerate $v_1, \dots, v_k$ so that for some m, $v_i = 0$ for $i > m$ and $v \neq 0$ for $i \leq m$. Then $v_1 + \cdots + v_m = 0$ but, $v_1, \dots, v_m$ are linearly independent. Therefore, $m = 0$. □

Theorem 5.8. Let $\lambda_1, \dots, \lambda_k$ be distinct eigenvalues of T and let $S_i \subseteq E_{\lambda_i}$ be finite and linearly independent. Then $S_1 \cup \cdots \cup S_k$ is linearly independent.

Proof. (Outline) Say $S_i = \{v_{i_1}, \dots, v_{i_{n_i}}\}$ and suppose $\sum_{i=1}^{k} \sum_{j=1}^{n_i} a_{ij} v_{i_j} = 0$. Consider $w_i = \sum_{j=1}^{n_i} a_{ij} v_{i_j}$.

Proof. Suppose $S_i = \{v_{i1}, v_{i2}, \dots, v_{in_i}\}$ and $|S_i| = n_i$. Suppose $\sum_{i=1}^{k} \sum_{j=1}^{n_i} a_{ij} v_{ij} = 0$. Let $w_i = \sum_{j=1}^{n_i} a_{ij} v_{ij}$. Then $w_i \in E_{\lambda_i}$ and we have $w_1 + \cdots + w_k = 0$. By Lemma 9, $w_i = 0$ for every $i = 1, \dots, k$. Thus, $\sum_{j=1}^{n_i} a_{ij} v_{ij} = 0$. Since S_i is linearly independent, $a_{ij} = 0$ for all i and j. □

Theorem 5.9. Let T be a linear operator on a finite-dimensional vector space V such that the characteristic polynomial of T splits. Let $\lambda_1, \dots, \lambda_k$ be distinct eigenvalues of T. Then

(a) T is diagonalizable if and only if the multiplicity of each λ_i equals $dim(E_{\lambda_i})$.

(b) If T is diagonalizable and β_i is an ordered basis for E_{λ_i}, then $\beta = \beta_1 \cup \cdots \cup \beta_k$ is an ordered basis for V.

Proof. (Outline)

- Part (b) follows from the proof.

- ($\Longrightarrow$): m_i- the multiplicity of λ_i, $d_i = dim(E_{\lambda_i})$, $\beta_i = \beta \cap E_{\lambda_i}$, $n_i = |\beta_i|$. We have $n_i \leq d_i \leq m_i$ and $\sum n_i = n = \sum m_i$. It follows that $d_i = m_i$ for every i.
- ($\Longleftarrow$): Let β_i be an ordered basis for E_{λ_i}; let $\beta = \beta_1 \cup \cdots \cup \beta_k$. Then β is a basis for V consisting of eigenvectors.

Proof. Let $\lambda_1, \ldots, \lambda_k$ be distince eigenvalues. Let m_i denote the multiplicity of λ_i. Let $n = dim(V)$. Let $d_i = dim(E_{\lambda_i})$. Then $d_i \leq m_i$ by Theorem 5.7. Note, $m_1 + \cdots + m_k = n$.

$\Longrightarrow$ Suppose T is diagonalizable. Then there exists a basis β consisting of eigenvectors. Let $\beta_i = \beta \cap E_{\lambda_i}$. Then β_i is linearly independent and so $|\beta_i| \leq d_i$. In addition,

$$\begin{aligned} n &= |\beta| \\ &= \sum_{i=1}^{k} |\beta_i| \\ &\leq \sum_{i=1}^{k} d_i \\ &\leq \sum_{i=1}^{k} m_i. && \text{(By Theorem 5.7)} \end{aligned}$$

Thus, $m_i = d_i$ for all i.

$\Longleftarrow$ Suppose $m_i = d_i$ for all i. Let β_i be a basis for E_{λ_i}. Thus, $|\beta_i| = d_i$. Let $\beta = \cup_{i=1}^{k} \beta_i$. Thus, β is linearly independent. In addition,

$$\begin{aligned} |\beta| &= \sum_{i=1}^{k} |\beta_i| \\ &= \sum_{i=1}^{k} d_i \\ &= \sum_{i=1}^{k} m_i \\ &= n \\ &= dim(V). \end{aligned}$$

Thus, β is a basis for V.

□

Note: T is diagonalizable if and only if

- The characteristic polynomial splits and
- the multiplicity of λ_i is $nullity(T - \lambda_i I) = n - rank(T - \lambda_i I)$.

Example: Test if $A = \begin{pmatrix} 3 & 1 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 4 \end{pmatrix}$ is diagonalizable over $\mathbb{R}$.

Solution.

- We know $f_A(t) = \begin{vmatrix} 3-t & 1 & 0 \\ 0 & 3-t & 0 \\ 0 & 0 & 4-t \end{vmatrix} = (3-t)^2(4-t)$ splits.
- For $\lambda_1 = 4$, we have $rank(A - \lambda_1 I) = rank\left(\begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}\right) = 2$ and multiplicity of $\lambda_1 = 1$ and $3 - rank(A - \lambda_1 I) = 3 - 2 = 1$.
- For $\lambda_2 = 3$, we have $rank(A - \lambda_2 I) = rank\left(\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}\right) = 2$ and multiplicity of $\lambda_2 = 2$ and $3 - 2 = 1$.

So, A is not diagonalizable.

5.4 Invariant subspaces and the Cayley-Hamilton theorem

Definition 6

- Let $T : V \to V$. A subspace $W \subseteq V$ is called a T-invariant subspace of V if $T(W) \subseteq W$.

  Examples:

  - $\{0\}$
  - $N(T)$: if $x \in N(T)$, then $T(x) = 0$.
  - $R(T)$: if $y \in R(T)$, then $y \in V$ and so, $T(y) \in R(T)$.
  - E_λ: Let $v \in E_\lambda$. Then $T(v) = \lambda v$. Thus, $T(T(v)) = T(\lambda v) = \lambda T(v)$. Thus $T(v) \in E_\lambda$.

- The T-cyclic subspace of V generated by x is

$$span(\{x, T(x), T^2(x), \dots\}).$$

Note: T-cyclic subspace is a minimal subspaces which is T-invariant and contains x.

Note: If W is $T-invariant$ and $v \in W$, then $T^i(v) \in W$ for all $i \in \mathbb{Z}^*$ and W contains a T-cyclic subspace generated by x.

Notation: If W is T-invariant, then $T_W = T\big|_W : W \to W$.

Theorem 5.21. Let T be a linear operator on a finite-dimensional vector space V and let W be a T-invariant subspace of V. Then the characteristic polynomial of the restriction of T to W, T_W, divides the characteristic polynomial of T.

Proof. Let $\alpha = \{v_1, \ldots, v_k\}$ be a basis for W and extend it to a basis $\beta = \{v_1, \ldots, v_n\}$ for V. Then

$$\big[T\big]_\beta = \begin{pmatrix} B_1 & B_2 \\ 0 & B_3 \end{pmatrix}$$

for some $B_1 \in M_{k\times k}(\mathbb{F})$ and $\big[T_W\big] = B_1$. Therefore, $f_T(t) = f_{T_W}(t)g(t)$ and so, $f_{T_W}(t)|f_T(t)$. $\square$

Example: Let $T : \mathbb{R}^4 \to \mathbb{R}^4$ be defined as $T(a, b, c, d) = (a + b + 5c - d, b + d, c - d, c + d)$. Let $W = \{x, y, 0, 0 : x, y \in \mathbb{R}\}$. Thus, $T(W) \subseteq W$. So, W is T-invariant. Let $\beta = \{e_1, e_2\}$ be a basis for W. Then $T(e_1) = e_1$ and $T(e_2) = (1, 1, 0, 0) = e_1 + e_2$. Thus,

$$\big[T_W\big]_\beta = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

Let $\alpha = \{e_1, e_2, e_3, e_4\}$ be a basis for V. So, $T(e_3) = (5, 0, 1, 1)$ and $T(e_4) = (-1, 1, -1, 1)$. Thus,

$$\big[T\big]_\alpha = \left(\begin{array}{cc|cc} 1 & 1 & 5 & -1 \\ 0 & 1 & 0 & 1 \\ \hline 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & 1 \end{array}\right).$$

Therefore,

$$\begin{aligned} f_T(t) &= \begin{vmatrix} 1-t & 1 \\ 0 & 1-t \end{vmatrix} \begin{vmatrix} 1-t & -1 \\ 1 & 1-t \end{vmatrix} \\ &= (1-t)^2((1-t)^2 + 1) \\ &= f_{T_W}(t)g(t). \end{aligned}$$

So, W is T-invariant. Now, is W T-cyclic? We know $T(e_1) = e_1$ and $T(e_2) = e_1 + e_2$. Thus, $e_1 \in span(\{e_2, T(e_2)\})$. Therefore, W is T-cyclic.

Example: Is it possible to have T and W such that W is T-invariant, but is not T-cyclic?

Solution. Yes. Consider $N(T)$ such that $dim(N(T)) > 1$. Then is is easy to show that W is T-invariant, but not T-cyclic.

Theorem 5.22. Let T be a linear operator on a vector space V of dimension n and let W be a T-cyclic subspace of V generated by a non-zero vector v of dimension k. Then

(a) $\{v, T(v), \dots, T^{k-1}(v)\}$ is a basis for W.

(b) If $a_0 v + a_1 T(v) + \cdots + a_{k-1}T^{k-1}(v) + T^k(v) = 0$, then the characteristic polynomial of T_W is $f(t) = (-1)^k(a_0 + a_1 t + \cdots + a_{k-1}t^{k-1} + t^k)$.

Proof. (Outline)

(a) Let j be the largest integer such that $\beta = \{v, T(v), \dots, T^{j-1}(v)\}$ is linearly independent. Then $span(\beta)$ is T-invariant and it follows that $j = k$.

(b) Let $\beta = v, T(v), \dots, T^{k-1}(v)$ and suppose $a_0 v + a_1 T(v) + \cdots + a_{k-1}T^{k-1} + T^k(v) = 0$. Use the form of $\left[T_W\right]_\beta$ to notice that $f_{T_W}(t) = (-1)^k(a_0 + a_1 t + \cdots + a_{k-1}t^{k-1} + t^k)$.

Proof. Let W be T-cyclic generated by v and $dim(W) = k$. Since $v \neq 0$, $\{v\}$ is linearly independent. Let j be the largest positive integer such that $\{v, T(v), \dots, T^{j-1}(v)\}$ is linearly independent. Then $j \leq k$. We will show that $j = k$. Let $Z = span(\{v, T(v), \dots, T^{j-1}(v)\})$. We claim that Z is T-invariant. Let $w \in Z$. Then, $w = a_0 v + a_1 T(v) + \cdots + a_{j-1}T^{j-1}(v)$. So, $T(w) = a_0 T(v) + \cdots + a_{j-1}T^j(v)$. Thus, $T^j(v)$ is linearly dependent on $\{v, T(v), \dots, T^{j-1}(v)\}$. Thus, $T^j(v) \in Z$ and $T(v), \dots, T^{j-1}(v) \in Z$ and so, $T(w) \in Z$. Therefore, Z is T-invariant. Since Z is T-invariant and $v \in Z$, then $W \subseteq Z$. So, $W = Z$ and $j = k$. Thus, $\{v, T(v), \dots, T^{k-1}(v)\}$ is a basis for W. Then

$$\left[T_W\right]_\beta = \begin{pmatrix} 0 & 0 & \dots & \dots & 0 & -a_0 \\ 1 & 0 & \ddots & \ddots & \vdots & -a_1 \\ 0 & 1 & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & 1 & 0 & \vdots \\ 0 & 0 & \dots & 0 & 1 & -a_{k-1} \end{pmatrix}.$$

Thus, $f_{T_W}(t) = (-1)^k(a_0 + a_1 t + \cdots + a_{k-1}t^{k-1} + t^k)$. ☐

Theorem 5.23, Cayley-Hamilton. Let V be finite-dimensional and let $T : V \to V$ be linear with the characteristic polynomial $f(t)$. Then $f(T) = T_0$.

Proof. (Outline) Note that $f(T)$ is a linear operator. We will show $f(T)(v) = 0$ for every v.

- Take $v \neq 0$ and consider the T-cycle subspace generated by v, W.

- Let $k = dim(W)$. Then $T^k \in span(\{v, T(v), \ldots, T^{k-1}(v)\})$ by 5.22(a) and we are done by 5.22(b).

Proof. We will show $f(T)v = 0_V$ for all $v \in V$. Note, $f(T)$ is a linear operator. Let $v = 0$. Then $f(T)v = 0$. Let $v \neq 0$. Consider the T-cyclic subspace generated by v, say W. Suppose that $dim(W) = k$. Then by Theorem 5.22, $\{v, T(v), \ldots, T^{k-1}(v)\}$ is a basis for W. So, $T^k(v) \in span(\{v, \ldots, T^{k-1}(v)\})$. So, $a_0v + a_1T(v) + \cdots + a_{k-1}T^{k-1}(v) + T^k(v) = 0$. Again, by Theorem 5.22, the characteristic polynomial of T_W satisfies $f_{T_W}(t) = (-1)^k(a_0 + \cdots + a_{k-1}t^{k-1} + t^k)$. Therefore, $f_{T_W}(T) = (-1)^k(a_0I + \cdots + a_{k-1}T^{k-1} + T^k)$. Thus, $f_{T_W} = (-1)^k(a_0v + \cdots + a_{k-1}T^{k-1}(v) + T^k(v)) = 0$. By Theorem 5.21, $F_T(t) = g(t)f_{T_W}(t)$. Thus, $f_T(T) = g(T)f_{T_W}(T)$. Therefore,

$$\begin{aligned} f_T(T)v &= g(T)f_{T_W}(T)v \\ &= g(T)(0_V) \\ &= 0_V. \end{aligned}$$

$\square$

Corollary 15 Let A be an $n \times n$ matrix and let $f(t)$ be its characteristic polynomial. Then $f(A) = 0_n$, the $n \times n$ zero matrix.

Example: Let $A = \begin{pmatrix} 2 & 1 \\ 0 & 1 \end{pmatrix}$. Then $f_A(t) = \begin{vmatrix} 2-t & 1 \\ 0 & 1-t \end{vmatrix} = (t-1)(t-2) = t^2 - 3t + 2$. Thus, $f(A) = \begin{pmatrix} 2 & 1 \\ 0 & 1 \end{pmatrix}^2 - 3\begin{pmatrix} 2 & 1 \\ 0 & 1 \end{pmatrix} + 2\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

6 Inner product spaces

6.1 Inner products and norms

Definition 1 Let V be a vector space over $\mathbb{F}$. An *inner product* on V is a function $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{F}$ such that the following conditions hold.

- $\langle x + z, y \rangle = \langle x, y \rangle + \langle z, y \rangle$
- $\langle cx, y \rangle = c\langle x, y \rangle$
- $\overline{\langle x, y \rangle} = \langle y, x \rangle$
- $\langle x, x \rangle > 0$ if $x \neq 0$ and $\langle 0, 0 \rangle = 0$.

Examples:

- Let $x = (a_1, \ldots, a_n)$, $y = (b_1, \ldots, b_n) \in \mathbb{F}^n$. Then $\langle x, y \rangle = \sum_{i=1}^n a_i \overline{b_i}$ is called the standard inner product.
- For $V = C[0,1] = \{f : [0,1] \to \mathbb{R} : f \text{ is continuous}\}$, $\langle f, g \rangle = \int_0^1 f(t)g(t)dt$ is an inner product on V.
- (Frobenius inner product)

  Let $V = M_{n\times n}(\mathbb{F})$. For $A, B \in V$, let $\langle A, B \rangle = tr(B^*A)$ where B^* is the adjoint of B. The four properties of an inner product are satisfied on V as follows:

  (1)

$$\begin{aligned}\langle A_1 + A_2, B \rangle &= tr(B^*(A_1 + A_2)) \\ &= tr(B^*A_1 + B^*A_2) \\ &= tr(B^*A_1) + tr(B^*A_2) \\ &= \langle A_1, B \rangle + \langle A_2, B \rangle\end{aligned}$$

  (2)

$$\begin{aligned}\langle cA, B \rangle &= tr(B^*cA) \\ &= c \cdot tr(B^*A) \\ &= c\langle A, B \rangle\end{aligned}$$

  (3)

$$\begin{aligned}\overline{\langle A, B \rangle} &= \overline{tr(B^*A)} \\ &= tr(A^*) \\ &= \langle B, A \rangle\end{aligned}$$

  (4)

$$\begin{aligned}\langle A, A \rangle &= tr(A^*A) \\ &= \sum_{i=1}^n \sum_{j=1}^n \overline{A_{ji}} A_{ji} \\ &= \sum_{i,j=1}^n |A_{ji}|^2 \\ &> 0 && \text{(When } A \neq 0\text{)}\end{aligned}$$

Definition 2 The *adjoint* of an $m \times n$ matrix A is the $n \times m$ matrix A^* such that $(A^*)_{ij} = \overline{A_{ji}}$.

Theorem 6.1.

- $\langle x, y+z\rangle = \langle x, y\rangle + \langle x, z\rangle$
- $\langle x, cy\rangle = \bar{c}\langle x, y\rangle$
- $\langle x, 0\rangle = \langle 0, x\rangle = 0$
- $\langle x, x\rangle = 0$ if and only if $x = 0$
- If $\langle x, y\rangle = \langle x, z\rangle$ for every $x \in V$, then $y = z$.

Definition 3 Let V be an inner product space. For $x \in V$, the *norm* of x is $\|x\| = \sqrt{\langle x, x\rangle}$.

Theorem 6.2. Let V be an inner product space over F. For $x, y \in V$ and $c \in \mathbb{F}$.

(a) $\|cx\| = |c| \cdot \|x\|$

(b) $\|x\| = 0$ if and only if $x = 0$ and $\|x\| \geq 0$ for any x.

(c) (Cauchy-Schwarz Inequality) $|\langle x, y\rangle| \leq \|c\|\|y\|$

(d) (Triangle Inequality) $\|x + y\| \leq \|x\| + \|y\|$

Proof. (Outline)

(c) Expand $\|x - cy\|^2$ and apply with $c = \frac{\langle x,y\rangle}{\langle y,y\rangle}$

(d) Note that $\langle x, y\rangle + \langle y, x\rangle = 2Re\langle x, y\rangle$ and $Re\langle x, y\rangle \leq |\langle x, y\rangle|$.

Proof.

(c) Note, the inequality holds if $y = 0$. Suppose $y \neq 0$. Let $c \in \mathbb{F}$. Then

$$0 \leq \langle x - cy, x - cy\rangle = \langle x, x\rangle - c\langle y, x\rangle - \bar{c}\langle x, y\rangle + c\bar{c}\langle y, y\rangle.$$

Let $c = \frac{\langle x,y\rangle}{\langle y,y\rangle}$. Then

$$\begin{aligned} 0 &\leq \langle x, x\rangle - \frac{\langle x, y\rangle\langle y, x\rangle}{\langle y, y\rangle} \\ &= \|x\|^2 - \frac{|\langle x, y\rangle|^2}{\|y\|^2}. \end{aligned}$$

Thus, $|\langle x, y\rangle|^2 \leq \|x\|^2\|y\|^2$.

(d)

$$\begin{aligned}
\|x+y\|^2 &= \langle x+y, x+y\rangle \\
&= \langle x,x\rangle + \langle y,x\rangle + \langle x,y\rangle + \langle y,y\rangle \\
&= \langle x,x\rangle + 2Re\langle x,y\rangle + \langle y,y\rangle \\
&\leq \|x\|^2 + 2\,|\langle x,y\rangle| + \|y\|^2 \\
&= \|x\|^2 + 2\|x\|\|y\| + \|y\|^2 && \text{(By Cauchy-Schwartz)} \\
&= (\|x\|+\|y\|)^2.
\end{aligned}$$

□

Definition 4

- Vectors $x, y \in V$ are called *orthogonal* if $\langle x, y\rangle = 0$.
- A set of vectors $S \subseteq V$ is called *orthogonal* if any two distinct vectors are orthogonal.
- S is called *orthonormal* if it is orthogonal and $\|x\| = 1$ for every $x \in S$.

Example: Let $H = \{f : [0, 2\pi] \to \mathbb{C}\}$. Let $\langle f, g\rangle = \frac{1}{2\pi}\int_0^{2\pi} f(t)\overline{g(t)}dt$. Let $f_n(t) = e^{int} = \cos(nt) + i\sin(nt)$. Then $S = \{f_n(t) : n \in \mathbb{N}\}$ is orthonormal.

Proof. If $n \neq m$, then

$$\begin{aligned}
\langle f_n, f_m\rangle &= \frac{1}{2\pi}\int_0^{2\pi} e^{int}\overline{e^{imt}}dt \\
&= \frac{1}{2\pi}\int_0^{2\pi} e^{int}e^{-imt}dt \\
&= \frac{1}{2\pi}\int_0^{2\pi} e^{it(n-m)}dt \\
&= \frac{1}{2\pi}\int_0^{2\pi} \cos(t(n-m)) + i\sin(t(n-m))dt \\
&= \frac{1}{2\pi}\big[\sin(t(n-m)) - \cos(t(n-m))\big]_0^{2\pi} \\
&= 0.
\end{aligned}$$

If $n = m$, then

$$\langle f_n, f_n\rangle = \frac{1}{2\pi}\int_0^{2\pi} e^{int}e^{-int}dt = 1.$$

□

Odd Town Problem: There are n people living in an odd town and they form clubs. A club must contain an odd number of members and for any two distinct clubs there must be an even (possibly zero) number of people in both of them. What is the maximum number of clubs that can be formed?

Proof. Let $S_1, \ldots, S_m$ denote these clubs. Let $v_i = \begin{pmatrix} v_{i1} \\ \vdots \\ v_{in} \end{pmatrix}$ when $v_{ij} = \begin{cases} 1 & j \in S_i \\ 0 & \text{else} \end{cases}$.
Then $v_i \in \mathbb{F}_2^n$. Then $v_1, \ldots, v_m$ are linearly independent. Suppose $\sum_{i=1}^m c_i v_i = 0$. Then

$$0 = \left\langle \sum_{i=1}^m c_i v_i, v_j \right\rangle = \sum_{i=1}^m c_i \langle v_i, v_j \rangle = c_i.$$

Then, $m \leq dim(\mathbb{F}^n) = n$. Therefore, the number of clubs is at most n. □

6.2 The Gram-Schmidt orthogonalization

Definition 5 An ordered basis which is orthonormal is called an *orthonormal basis.*

Theorem 6.3. Suppose $S = \{v_1, \ldots, v_k\}$ is an orthogonal subset of V such that $v_i \neq 0$. For $y \in span(S)$,

$$y = \sum_{i=1}^k \frac{\langle y, v_i \rangle}{\|v_i\|^2} v_i.$$

Proof. Let $y \in span(S)$. Then $y = \sum_{i=1}^k a_i v_i$ and so $\langle y, v_j \rangle = \sum_{i=1}^k a_k \langle v_i, v_j \rangle = a_j \langle v_j, v_j \rangle$. Thus, $a_j = \frac{\langle y, v_i \rangle}{\langle v_j, v_j \rangle} = \frac{\langle y, v_j \rangle}{\|v_j\|^2}$. □

Corollary 4 Any orthogonal set of non-zero vectors is linearly independent.

Proof. Suppose $\sum_{i=1}^k a_i v_i = 0$. Then $a_i = \frac{\langle 0, v_i \rangle}{\|v_i\|^2} = 0$ by Theorem 6.3. □

Theorem 6.4 (Graham-Schmidt algorithm). Let $S = \{w_1, \ldots, w_n\}$ be a linearly independent subset. Define $S' = \{v_1, \ldots, v_n\}$ as follows, $v_1 := w_1$ and for $k \geq 2$

$$v_k = w_k - \sum_{j=1}^{k-1} \frac{\langle w_k, v_j \rangle}{\|v_j\|^2} v_j.$$

Then S' is orthogonal and $span(S') = span(S)$.

Proof. Let $S_k = \{w_1, \ldots, w_k\}$. We will use induction on k.

- Base Case: For $k = 1$, clearly $span(S_1') = span(S_1)$.

- Induction Step: Suppose $S'_{k-1} = \{v_1, \ldots, v_{k-1}\}$ has been constructed so that S'_{k-1} is orthogonal and $span(S'_{k-1}) = span(S_k)$. Let $v_k = w_k - \sum_{i=1}^{k-1} \frac{\langle w_k, v_i \rangle}{\|v_i\|^2} v_i$. Then we have the following properties:
  - We know $v_k \neq 0$ because otherwise, $w_k - \sum_{i=1}^{k-1} \frac{\langle w_k, v_i \rangle}{\|v_i\|^2} v_i = 0$ and $v_k \in span(S'_{k-1}) = span(S_{k-1})$, but S is linearly independent.
  - For $j \leq k-1$,

$$\begin{aligned}
\langle v_k, v_j \rangle &= \langle w_k, v_j \rangle - \sum_{i=1}^{k-1} \frac{\langle w_k, v_i \rangle}{\|v_i\|^2} \langle v_i, v_j \rangle && \text{(Note, } \langle v_i, v_j \rangle = 0 \text{ if } i \neq j) \\
&= \langle w_k, v_j \rangle - \frac{\langle w_k, v_j \rangle}{\|v_j|^2} \langle v_j, v_j \rangle && \text{(By inductive hypothesis)} \\
&= 0.
\end{aligned}$$

  Thus, $S_k = \{v_1, \ldots, v_k\}$ is orthogonal. From $v_k = w_k - \sum_{i=1}^{k-1} \frac{\langle w_k, v_i \rangle}{\|v_i\|^2} v_i$, we have $v_k \in span(w_k, v_1, \ldots, v_{k-1}) \in span(w_1, \ldots, w_k) = span(S_k)$. Thus, $span(S'_k) \subseteq span(S_k)$. Also, $w_k \in span(v_1, \ldots, v_k)$ and so $span(S_k) \subseteq span(S'_k)$.

$\square$

Example: $\left(\text{Legendre Polynomials of the form } \frac{v_k(x)}{v_k(1)}\right)$

Let $P_2(\mathbb{R})$ with the standard basis $\beta = \{1, x, x^2\}$. Consider $\langle f, g \rangle = \int_{-1}^{1} f(x)g(x)dx$. Replace β with an orthogonal basis.

Proof. Let $\beta = \{w_1, w_2, w_3\}$. Then

- $v_1 = w_1 = 1$, $\|v_1\|^2 = \langle v_1, v_1 \rangle = 2$
- $v_2 = w_2 - \frac{\langle w_2, v_1 \rangle}{\|v_1\|^2} v_1 = x$
- $v_3 = w_3 - \frac{\langle w_3, v_2 \rangle}{\|v_2\|^2} v_2 - \frac{\langle w_3, v_1 \rangle}{\|v_1\|^2} v_1 = x^2 - \frac{1}{3}$.

Thus, $\alpha = \{2, x, x^2 - \frac{1}{3}\}$ is an orthonormal basis for $P_2(\mathbb{R})$. $\square$

Theorem 6.5. Let V be a finite-dimensional inner product space and let $\beta = \{v_1, \ldots, v_n\}$ be an orthonormal basis for V. Then for every $x \in V$, $x = \sum_{i=1}^{n} \langle x, v_i \rangle v_i$.

Corollary 7 Let $\beta = \{v_1, \ldots, v_n\}$ be orthonormal, let $T : V \to V$ be linear, and let $A = \big[T\big]_\beta$. Then $A_{ij} = \langle T(v_j), v_i \rangle$.

Fourier coefficient of x relative to β is $\langle x, y\rangle$ where $y \in \beta$.

Example: Let $S = \{e^{int} : n \in \mathbb{Z}\}$ in H with $\langle f, g\rangle = \frac{1}{2\pi}\int_0^{2\pi} f(t)\overline{g(t)}dt$. Find the Fourier coefficients of $f(t) = 1$.

Proof.

- If $n = 0$, $\langle f, 1\rangle = \pi$
- If $n \neq 0$, $\langle f, f_n\rangle = \frac{1}{2\pi}\int_0^{2\pi} te^{-int}dt = -\frac{1}{in}$

□

Definition 6 Let S be a non-empty subset of V. Define $S^\perp = \{x \in V : \langle x, y\rangle = 0 \text{ for every } y \in S\}$.

Note: $S^\perp$ is a subspace of V.

Theorem 6.6. Let W be a finite-dimensional subspace of an inner product space V and let $y \in V$. Then the exist unique $u \in W$ and $z \in W^\perp$ such that $y = u + z$. Furthermore, if $\{v_1, \ldots, v_k\}$ is an orthonormal basis for W, then $u = \sum\langle y, v_i\rangle v_i$.

Proof. (Outline)

- Let $u = \sum\langle y, v_i\rangle v_i$ and $z = y - u$. Check that $z \in W^\perp$.
- For the uniqueness $u - u' \in W$ and $z' - z \in W^\perp$.

Proof. Let $u = \sum\langle y, v_i\rangle v_i$ and $z = y - u$. Then

$$\begin{aligned}\langle z, v_j\rangle &= \langle y - u, v_j\rangle \\ &= \langle y, v_j\rangle - \sum_{i=1}^{k}\langle y, v_i\rangle\langle v_i, v_j\rangle \\ &= \langle y, v_j\rangle.\end{aligned}$$

Then, for uniqueness, $u' + z' = y = u + z$ and $u' + u = z - z' \in W \cap W^\perp = \{0\}$. So, $u' = u$ and $z' = z$. □

Corollary 9 For any $x \in W$, $\|y - x\| \geq \|y - u\|$ and if $\|y - x\| = \|y - u\|$ then $x = u$.

Proof.

$$
\begin{aligned}
\|y-x\|^2 &= \|u+z-x\|^2 \\
&= \|u-x+z\|^2 \\
&= \langle u-z+z, u-x+z\rangle \\
&= \|u-x\|^2 + \|z\|^2 \\
&\geq \|z\|^2 \\
&= \|y-u\|^2
\end{aligned}
$$

□

Theorem 6.7. Suppose $S = \{v_1, \ldots, v_k\}$ is an orthonormal set in an n-dimensional inner product space V. Then

(a) S can be extended to an orthonormal basis $\{v_1, \ldots, v_n\}$ for V.

(b) $\{v_{k+1}, \ldots, v_n\}$ is an orthonormal basis for $(span(S))^\perp$.

(c) If W is a subspace of V, then $dim(W) + dim(W^\perp) = dim(V)$.

6.3 The adjoint of a linear operator

Theorem 6.8. Let V be a finite-dimensional inner product space over $\mathbb{F}$, and let $g : V \to \mathbb{F}$ be a linear transformation. Then there exists a unique vector $y \in V$ such that $g(x) = \langle x, y\rangle$ for every $x \in V$.

Proof. (Outline) Take an orthonormal basis $\beta = \{v_1, \ldots, v_n\}$ and define $y = \sum \overline{g(v_i)} v_i$.

Proof. Let $\{v_1, \ldots, v_n\}$ be an orthonormal basis for V. Let $y = \sum_{i=1}^n \overline{g(v_i)} v_i \in V$. Let $h : V \to \mathbb{F}$ be defined by $h(x) = \langle x, y\rangle$. Then h is a linear transformation. In addition,

$$
\begin{aligned}
h(v_j) &= \langle v_j, y\rangle \\
&= \left\langle v_j, \sum_{i=1}^n \overline{g(v_i)} v_i \right\rangle \\
&= \sum_{i=1}^n g(v_i)\langle v_j, v_i\rangle \\
&= g(v_j).
\end{aligned}
$$

Thus, $h(v_j) = g(v_j)$ for all $j = 1, \ldots, n$. Thus, $h = g$. Suppose $\langle x, y\rangle = \langle x, y'\rangle$ for all $x \in V$. Then $y = y'$. □

Theorem 6.9. Let V be a finite-dimensional inner product space and let $T : V \to V$ be linear. Then there exists a unique function $T^* : V \to V$ such that $\langle T(x), y\rangle = \langle x, T^*(y)\rangle$ for all x, y. Furthermore, T^* is linear.

Proof. (Outline)

- Fix y. Check that $g(x) = \langle T(x), y\rangle$ is linear.
- Theorem 6.8 gives unique y' such that $g(x) = \langle x, y'\rangle$. Define $T^*(y) = y'$.
- Note that T^* is a function and check that it is linear.

Proof. Fix $y \in V$. Consider $g(x) = \langle T(x), y\rangle$. Then $g : V \to \mathbb{F}$ is linear. By Theorem 6.8, there exists a unique y' such that $g(x) = \langle x, y'\rangle$. Let $T^*(y) = y'$. Then $T^* : V \to V$ and $\langle T(x), y\rangle = g(x) = \langle x, T^*(y)\rangle$. In addition, T^* is linear. Let $y_1, y_2 \in V$ and $c \in \mathbb{F}$. Then

$$\begin{aligned}\langle x, T^*(cy_1 + y_2)\rangle &= \langle T(x), cy_1 + y_2\rangle \\ &= \overline{c}\langle T(x), y_1\rangle + \langle T(x), y_2\rangle \\ &= \langle x, cT^*(y_1)\rangle + \langle x, T^*(y_2)\rangle \\ &= \langle x, cT^*(y_1) + T^*(y_2)\rangle.\end{aligned}$$

Thus, $T^*(cy_1 + y_2) = cT^*(y_1) + T^*(y_2)$. Suppose $\langle T(x), y\rangle = \langle x, U(y)\rangle$ for all $x, y \in V$. Then, $\langle T(x), y\rangle = \langle x, T^*(y)\rangle$, which implies $T^* = U$. □

Theorem 6.10. Let V be a finite-dimensional inner product space and let β be an orthonormal ordered basis for V. If T is a linear operator on V, then $\left[T^*\right]_\beta = \left[T\right]_\beta^*$.

Proof. Let $A = \left[T^*\right]_\beta$ and $B = \left[T\right]_\beta$ where $\beta = \{v_1, \dots, v_n\}$. Note, $B_{ij} = \langle T(v_j), v_i\rangle$. Then,

$$\begin{aligned}A_{ij} &= \langle T^*(v_j), v_i\rangle \\ &= \overline{\langle v_i, T^*(v_j)\rangle} \\ &= \overline{\langle T(v_i), v_j\rangle} \\ &= \overline{B_{ji}}.\end{aligned}$$

□

Theorem 6.11. Let V be an inner-product space, and let T, U be linear operators on V. Then

(a) $(T + U)^* = T^* + U^*$

(b) $(cT)^* = \overline{c}T^*$

(c) $(TU)^* = U^*T^*$

Proof.

$$\begin{aligned}\langle x, (TU)^*y\rangle &= \langle (TU)x, y\rangle \\ &= \langle T(Ux), y\rangle \\ &= \langle Ux, T^*y\rangle \\ &= \langle x, U^*T^*y\rangle\end{aligned}$$

□

(d) $T^{**} = T$

(e) $I^* = I$

6.4 Normal and self-adjoint operators

Lemma 15. Let T be a linear operator on a finite-dimensional inner product space V. If T has an eigenvector, then so does T^*.

Theorem 6.14 (Schur). Let T be a linear operator on a finite-dimensional inner product space V and suppose that the characteristic polynomial of T splits. Then there exists an orthonormal basis β for V such that $\left[T\right]_\beta$ is upper triangular.

Proof. (Outline)

- Show that $W^\perp$ is T-invariant and $dim(W^\perp) = n - 1$.
- The characteristic polynomial of $T_{W^\perp}$ divided $f_T(t)$ (so it splits) and by **IH** for some γ, $\left[T_{W^\perp}\right]_\gamma$ is upper-triangular.
- Let $\beta = \gamma \cup \{z\}$ Then $\left[T\right]_\beta$ is upper-triangular.

Proof. Let $n = dim(V)$.

- Base step: Let n = 1. Then clearly the statement holds.
- Induction Step: Let the statement be true for some $n - 1 \geq 1$. Then we can assume T^* has an eigenvector z such that $\|z\| = 1$ and $T^*(z) = \lambda z$ by Lemma 15. Let

$W = span(z)$. We claim $W^\perp$ is T-invariant. Let $y \in W^\perp$. Then $\langle y, z\rangle = 0$. Now, for $w \in W$,

$$\begin{aligned}\langle T(y), w\rangle &= \langle y, T^*(w)\rangle \\ &= \langle y, cT^*(z)\rangle && \text{(For some } c \in \mathbb{F}) \\ &= \langle y, c\lambda z\rangle \\ &= \bar{c}\bar{\lambda}\langle y, z\rangle \\ &= 0.\end{aligned}$$

Thus, $T(y) \in W^\perp$. In addition, $dim(W^\perp) = n - dim(W) = n - 1$ and $f\big|_{T_{W^\perp}}$ divides $f\big|_T$ and so it splits. By inductive hypothesis, there exists an orthonormal basis γ such that $\left[T_{W^\perp}\right]_\gamma$ is upper-triangular. Let $\beta = \gamma \cup \{z\}$. Then $\left[T\right]_\beta$ is upper-triangular because $T(w) \in W^\perp \in span(\gamma)$.

□

Note: If V has an orthonormal basis of eigenvectors of T, then $TT^* = T^*T$.

Definition 7 $T : V \to V$ is *normal* if $TT^* = T^*T$. $A \in M_{n\times n}(\mathbb{F})$ is *normal* if $AA^* = A^*A$.

Note: If there exists an orthonormal basis β consisting of eigenvectors of T, then T is normal.

Proof. Suppose there exists a basis β consisting of eigenvectors of T. Then $\left[T\right]_\beta$ is diagonal and $\left[T^*\right]_\beta = \left(\left[T\right]_\beta\right)^*$ is diagonal. Therefore,

$$\left[TT^*\right]_\beta = \left[T\right]_\beta \left[T^*\right]_\beta = \left[T^*\right]_\beta \left[T\right]_\beta = \left[T^*T\right]_\beta.$$

Thus, $TT^* = T^*T$. □

Examples:

- If $A \in M_{n\times n}(\mathbb{R})$ is symmetric, then A is normal.
- If $A \in M_{n\times n}(\mathbb{R})$ is asymmetric, $A^T = -A$, and so $AA^* = -A^2 = A^*A$.

Theorem 6.15. Let T be a normal operator on V. Then

(a) $\|T(x)\| = \|T^*(x)\|$

Proof.

$$\begin{aligned}\|Tx\|^2 &= \langle T(x), T(x)\rangle \\ &= \langle T^*Tx, x\rangle \\ &= \langle TT^*x, x\rangle \\ &= \langle T^*x, T^*x\rangle \\ &= \|T^*x\|^2\end{aligned}$$

□

(b) $T - cI$ is normal for every $c \in \mathbb{F}$.

Proof.

$$\begin{aligned}(T-cI)(T-cI)^* &= (T-cI)(T^*-\bar{c}I)\\ &= TT^* - cT^* - \bar{c}T - c\bar{c}I\\ &= T^*T - cT^* - \bar{c}T - c\bar{c}I\\ &= (T-cI)^*(T-cI)\end{aligned}$$

□

(c) If x is an eigenvector of T, then x is an eigenvector of T^*.

Proof. We can show

$$\begin{aligned}0 &= \langle (T-\lambda I)x, (T-\lambda I)x\rangle\\ &= \|(T-\lambda I)x\|^2\\ &= \|(T-\lambda I)^*x\|^2 && \text{(From (a) and (b))}\\ &= \langle (T^*-\overline{\lambda}I)x, (T^*-\overline{\lambda}I)x\rangle.\end{aligned}$$

Thus, $T^*x = \overline{\lambda}x$. □

(d) If λ_1, λ_2 are distinct eigenvalues of T with eigenvectors x_1, x_2, then $\langle x_1, x_2\rangle = 0$.

Proof. We can show

$$\begin{aligned}\lambda_1\langle x_1, x_2\rangle &= \langle \lambda_1 x_1, x_2\rangle\\ &= \langle T(x_1), x_2\rangle\\ &= \langle x_1, T^*(x_2)\rangle\\ &= \langle x_1, \overline{\lambda_2}x_2\rangle && \text{(From (c))}\\ &= \lambda_2\langle x_1, x_2\rangle,\end{aligned}$$

which implies $(\lambda_1 - \lambda_2)\langle x_1, x_2\rangle = 0$. Since $\lambda_1 \neq \lambda_2$, then $\langle x_1, x_2\rangle = 0$. □

Theorem 6.16. Let T be a linear operator on a finite-dimensional complex inner-product space V. Then T is normal if and only if there exists an orthonormal basis for V consisting of eigenvectors of T.

Proof. (Outline) Suppose T is normal.

- T splits over $\mathbb{C}$ and so apply Schur's lemma to get an orthonormal basis $\beta = \{v_1, \ldots, v_n\}$.
- $A \coloneqq [T]_\beta$ is upper-triangular and so $T(v_1) = A_{11}v_1$.

- Show that e_k is an eigenvector by induction on k using the fact that $A_{jk} = \langle T(v_k), v_j \rangle$.

Proof. Let T be a linear operator on a finite-dimensional complex inner-product space V.

$\Longrightarrow$ Suppose T is normal. Since we are over $\mathbb{C}$, the characteristic polynomial of T splits. By Schur's Theorem, there exists an orthonormal basis $\beta = \{v_1, \ldots, v_n\}$ such that $A = [T]_\beta$ is upper-triangular. We will show that $v_1, \ldots, v_n$ are in addition eigenvectors of T. We will use induction on n.

- Base Case: Let $n = 1$. We have $T(v_1) = A_{11}v_1$ and so v_1 is an eigenvector.
- Induction Step: Suppose $v_1, \ldots, v_{k-1}$ are eigenvectors of T with eigenvalues $\lambda_1, \ldots, \lambda_{k-1}$. Then $T(v_k) = \sum_{i=1}^{k} A_{ik}v_i$ and in addition, $A_{ik} = \langle T(v_k), v_i \rangle = \langle v_k, T^*(v_i) \rangle$. For $i < k$, we have

$$\begin{aligned} A_{ik} &= \langle v_k, \overline{\lambda_i} v_i \rangle \\ &= \lambda_i \langle v_k, v_i \rangle \\ &= 0. \end{aligned}$$

Thus, $T(v_k) = A_{kk}v_k$. So, v_k is an eigenvector of T.

$\Longleftarrow$ Suppose β is an orthonormal basis consisting of eigenvectors of T. Then $[T]_\beta$ is diagonal, and so $[T^*]_\beta = [T]_\beta^*$ is diagonal. Thus,

$$\begin{aligned} [TT^*]_\beta &= [T]_\beta [T^*]_\beta \\ &= [T^*]_\beta [T]_\beta \\ &= [T^*T]_\beta. \end{aligned}$$

Thus, $TT^* = T^*T$.

$\square$

Definition 8 $T : V \to V$ is *self-adjoint* (Hermitian) if $T = T^*$. $A \in M_{n\times n}(\mathbb{F})$ is *Hermitian* if $A = A^*$.

Lemma 19. Let T be a Hermitian operator on a finite-dimensional inner product space V. Then,

(a) All eigenvalues of T are real.

Proof. Let λ be an eigenvalue correspond with the eigenvector v. Then $\lambda v = T(v) = T^*(v) = \overline{\lambda} v$ and so $\lambda = \overline{\lambda}$. Thus, $\lambda \in \mathbb{R}$. $\square$

(b) If V is a real inner product space, then the characteristic polynomial splits.

Proof. Suppose T is self-adjoint. Let β be an orthonormal basis for T. Then $A = [T]_\beta$ is self-adjoint. Let $L_A : \mathbb{C}^n \to \mathbb{C}^n$ where $n = dim(V)$ defined by $L_A(x) = Ax$. Then $[L_A]_\gamma = A$ where γ is the standard basis. Thus, L_A is self-adjoint. Thus, f_{L_A} splits over $\mathbb{C}$. Then $f_{L_A}(t) = (-1)^n(t-\lambda_1)\dots(t-\lambda_n)$ where λ_i is an eigenvalue of L_A. By (a), $\lambda_i \in \mathbb{R}$. Thus, F_{L_A} splits over $\mathbb{R}$. Therefore, $f_T = f_A = f_{L_A}$ splits over $\mathbb{R}$. □

Theorem 6.17. Let T be a linear operator on a finite-dimensional real inner-product space V. Then T is Hermitian if and only if there exists an orthonormal basis for V consisting of eigenvectors of T.

Proof. (Outline)
The characteristic polynomial splits and so we may apply Schur's lemma. $A := [T]_\beta$ is upper-triangular and so A^*. Thus it must be a diagonal matrix.

Proof. Let T be a linear operator on a finite-dimensional real inner-product space V.

$\implies$ Suppose T is self-adjoint. By Lemma 19, f_T splits. By Schur's Theorem, there exists an orthonormal basis β such that $A = [T]_\beta$ is an upper-triangular. Then,

$$A = [T]_\beta = [T^*]_\beta = [T]_\beta^* = A^*.$$

Thus, A and A^* are upper-triangular. Thus, A is diagonal and so β consists of eigenvectors of T.

$\impliedby$ Suppose there is an orthonormal basis β consisting of eigenvectors of T. Then $[T]_\beta$ is diagonal with entries from $\mathbb{R}$. So, $[T^*]_\beta = [T]_\beta^* = [T]_\beta$. Thus, $T^* = T$. Therefore, T is self-adjoint.

□

6.5 Unitary and orthogonal operators and their matrices

Definition 9 Let V be a finite-dimensional inner-product space over F and let $T : V \to V$ be linear. If $\|T(x)\| = \|x\|$ for every $x \in V$, then T is called *unitary* if $\mathbb{F} = \mathbb{C}$ and orthogonal if $\mathbb{F} = \mathbb{R}$.

Theorem 6.18. Let T be a linear operator on a finite-dimensional inner-product space V. Then the following statements are equivalent.

(a) $TT^* = T^*T = I$

(b) $\langle T(x), T(y)\rangle = \langle x, y\rangle$ for all $x, y \in V$

(c) If β is an orthonormal basis, then so is $T(\beta)$.

(d) There exists an orthonormal basis β such that $T(\beta)$ is orthonormal.

(e) $\|T(x)\|= \|x\|$ for every x.

Proof. (Outline)

- $(d) \implies (e)$ Let $\beta = \{v_1, \dots, v_n\}$ be orthonormal such that $T(\beta)$ is orthonormal. Take $x \in V$. Then $x = \sum a_i v_i$ and

$$\|x\|^2 = \sum |a_i|^2 = \sum \|T(x)\|^2.$$

- $(e) \implies (a)$ $\langle x, x \rangle = \langle x, T^*T(x) \rangle$ be (e). Thus $\langle x, (I - T^*T)(x) \rangle = 0$ for every x. Set $U := I - T^*T$. Then U is self-adjoint and so there is an orthonormal basis consisting of eigenvectors of U. Check that $U(x) = \lambda x$ implies $\lambda = 0$; $U = T_0$; $T^*T = I$; $TT^* = I$ as well because $\left[T\right]_\beta$ is a square matrix.

Proof. Let T be a linear operator on a finite-dimensional inner-product space V. Then the following statements are equivalent.

- (a) $\implies$ (b)

$$\langle x, y \rangle = \langle T^*T(x), y \rangle = \langle Tx, Ty \rangle$$

- (b) $\implies$ (c) Let $\beta = \{v_1, \dots, v_n\}$ be orthonormal. Then $\langle T(v_i), T(v_j) \rangle = \langle v_i, v_j \rangle = \delta_{ij}$.

- (c) $\implies$ (d) This is clearly true.

- (d) $\implies$ (e) Let $\beta = \{v_1, \dots, v_n\}$ be orthonormal such that $T(\beta) = \{T(v_1), \dots, T(v_n)\}$ is also orthonormal. Let $x \in V$. Then $x = \sum_{i=1}^n a_i v_i$. So,

$$\begin{aligned}
\|x\|^2 &= \langle x, x \rangle \\
&= \left\langle \sum_{i=1}^n a_i v_i, \sum_{i=1}^n a_i v_i \right\rangle \\
&= \sum_{i=1}^n \sum_{j=1}^n a_i \overline{a_j} \langle v_i, v_j \rangle \\
&= \sum_{i=1}^n a_i \overline{a_i} \\
&= \sum_{i=1}^n |a_i|^2 .
\end{aligned}$$

Also,

$$\begin{aligned}
\|Tx\|^2 &= \langle Tx, Tx\rangle \\
&= \left\langle \sum_{i=1}^{n} a_i T(v_i), \sum_{i=1}^{n} a_i T(v_i) \right\rangle \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i\overline{a_j}\langle T(v_i), T(v_j)\rangle \\
&= \sum_{i=1}^{n} a_i\overline{a_i} \\
&= \sum_{i=1}^{n} |a_i|^2 .
\end{aligned}$$

Thus, $\|x\| = \|Tx\|$.

- (e) $\implies$ (a) We have

$$\langle x, x\rangle = \|x\|^2 = \|Tx\|^2 = \langle Tx, Tx\rangle = \langle x, T^*Tx\rangle$$

for all x. Thus, for all x, $\langle x, (I - T^*T)x\rangle = 0$. Set $U := I - T^*T$. Then we have $\langle x, Ux\rangle = 0$ for all x and U is self-adjoint. By Theorem 6.17, there exists β orthonormal consisting of eigenvectors of U. Thus, for $v \in \beta$, $Uv = \lambda v$ for some $\lambda \in \mathbb{F}$. So,

$$0 = \langle v, Uv\rangle = \overline{\lambda}\langle v, v\rangle = \overline{\lambda}.$$

Thus, $\overline{\lambda} = 0 = \lambda$. So, $Uw = T_0 w$ for all $x \in V$. Thus $I = T^*T$. Therefore, $TT^* = I = T^*T$.

$\square$

Definition 10 A square matrix A is called *orthogonal* if $A^tA = AA^t = I$ and *unitary* if $A^*A = AA^* = I$.

Note:

- $AA^* = I$ if and only if A are orthonormal.
- $A^*A = I$ if and only if columns of A are orthonormal.
- If λ is an eigenvalue of a unitary (orthogonal) matrix, then $|\lambda| = 1$.

Definition 11 $A, B \in M_{n\times n}(\mathbb{C})$ ($A, B \in M_{n\times n}(\mathbb{R})$) are *unitarily (orthogonally) equivalent* if there exists a unitary (orthogonal) matrix P such that $A = P^{-1}BP$.

Theorem 6.19. Let $A \in M_{n\times n}(\mathbb{C})$. Then A is normal if and only if A is unitarily equivalent to a diagonal matrix.

Theorem 6.20. Let $A \in M_{n\times n}(\mathbb{R})$. Then A is symmetric if and only if A is orthogonally equivalent to a diagonal matrix.

Theorem 6.21 (Schur). Let $A \in M_{n\times n}(\mathbb{F})$ and suppose $f_A(t)$ splits over $\mathbb{F}$.

(a) If $\mathbb{F} = \mathbb{C}$, then A is unitarily equivalent to a complex upper-triangular matrix.

(b) If $\mathbb{F} = \mathbb{R}$, then A is orthogonally equivalent to a real upper-triangular matrix.

7 Homework

Problem 1.2.12. A real-valued function f defined on the real line is called an **even function** if $f(-t) = f(t)$ for each real number t. Prove that the set of even functions defined on the real line with the operations of addition and scalar multiplication defined in Example 3 is a vector space.

Proof. Since $\mathbb{R}$ is a field, then from Example 3, $\mathcal{F}(A, \mathbb{R})$ is a vector space. So, we just need to show that $V = \{f \in \mathcal{F}(A, \mathbb{R}) : f(t) = f(-t)\}$ is a subspace of $\mathcal{F}(A, \mathbb{R})$. Let $f, g \in V$ and $a \in \mathbb{R}$. Then,

$$\begin{aligned}(f+g)(t) &= f(t) + g(t)\\ &= f(-t) + g(-t)\\ &= (f+g)(-t)\end{aligned}$$

implies V is closed under addition. Also,

$$\begin{aligned}(af)(t) &= a[f(t)]\\ &= a[f(-t)]\\ &= (af)(-t)\end{aligned}$$

implies V is closed under scalar multiplication. Suppose $z \in \mathcal{F}(A, \mathbb{R})$ is the zero function. Then, $z(t) = 0 = z(-t)$. Thus $z \in V$ and $(f + z)(t) = f(t) + z(t) = f(t)$ implies that V has a zero vector. Lastly, define $y \in \mathcal{F}(A, \mathbb{R})$ by $y(t) = -f(t) \in \mathbb{R}$, which is possible since $\mathbb{R}$ is a field. Then,

$$\begin{aligned}y(t) &= -f(t)\\ &= -f(-t) && \text{(Since } f \in V\text{)}\\ &= y(-t).\end{aligned}$$

This shows $y \in V$. Then, $(f + y)(t) = f(t) + y(t) = f(t) - f(t) = 0$. Thus, V is closed under additive inverse. Therefore, V is a subspace of $\mathcal{F}(A, \mathbb{R})$. $\square$

Problem 1.2.20. Let V denote the set of all real-valued functions f defined on the real line such that $f(1) = 0$. Prove that V is a vector space with the operations of addition and scalar multiplication defined in Example 3.

Proof. Since $\mathbb{R}$ is a field, then from Example 3, $\mathcal{F}(A, \mathbb{R})$ is a vector space. So, we just need to show that $V = \{f \in \mathcal{F}(A, \mathbb{R}) : f(1) = 0\}$ is a subspace of $\mathcal{F}(A, \mathbb{R})$. Let $f, g \in V$ and $a \in \mathbb{R}$. Then, $(f + g)(1) = f(1) + g(1) = 0$. Thus, V is closed under addition. Further, $(af)(1) = a(f(1)) = a(0) = 0$ implies V is closed under scalar multiplication. Next, let $z \in \mathcal{F}(A, \mathbb{R})$ be the zero function. Clearly $z \in V$. Lastly, define $y \in \mathcal{F}(A, \mathbb{R})$ by $y(t) = -f(t) \in \mathbb{R}$, which is possible since $\mathbb{R}$ is a field. Then, $y(1) = -f(1) = 0$. This shows $y \in V$. Then, $(f + y)(t) = f(t) + y(t) = f(t) - f(t) = 0$. Thus, V is closed under additive inverse. Therefore, V is a subspace of $\mathcal{F}(A, \mathbb{R})$. $\square$

Problem 1.3.18. Prove that a subset W of a vector space V is a subspace of V if and only if $0 \in W$ and $ax + y \in W$ whenever $a \in F$ and $x, y \in W$.

Proof. Suppose $W \subseteq V$, $a \in F$, $x, y \in W$.

($\Longrightarrow$): Suppose W is a subspace of V. Then, the zero vector must be in W by definition of subspace. Then, since W is a vector space, it is closed under addition and scalar multiplication, which implies $ax + y \in W$.

($\Longleftarrow$): Suppose $0 \in W$ and $ax + y \in W$. Then W must be closed under addition and scalar multiplication if $ax + y \in W$. Thus, by Theorem 1.3, W is a subspace of V. $\square$

Problem 1.3.23. Let W_1 and W_2 be subspaces of a vector space V.

(a) Prove that $W_1 + W_2$ is a subspace of V that contains both W_1 and W_2.

Proof. Let W_1 and W_2 be subspaces of a vector space V. Let $W_1 + W_2 = \{x + y : x \in W_1, y \in W_2\}$. Since W_1 and W_2 are subspaces, then they both have the zero vector. Thus, by definition of $W_1 + W_2$, $0 \in W_1 + W_2$. Now, let $x_1, x_2 \in W_1$ and $y_1, y_2 \in W_2$. Let $a = x_1 + y_1 \in W_1 + W_2$ and $b = x_2 + y_2 \in W_1 + W_2$. Then $a + b = (x_1 + y_1) + (x_2 + y_2)$. Since V is a vector space, we can reorder to be $a + b = (x_1 + x_2) + (y_1 + y_2)$. Since W_1 and W_2 are vector spaces, $(x_1 + x_2) \in W_1$ and $(y_1 + y_2) \in W_2$. Thus, $a + b \in W_1 + W_2$. Suppose $c \in F$, where F is a field. Then,

$$\begin{aligned} ca &= c(x_1 + y_1) \\ &= cx_1 + cy_1. && \text{(Since } x_1 + y_1 \in V\text{)} \end{aligned}$$

Then $ca \in W_1 + W_2$ since $cx_1 \in W_1$ and $cy_1 \in W_2$. Therefore, $W_1 + W_2$ is a subspace of V. Clearly, for $x \in W_1$, $x + 0 \in W_1 + W_2$. The same logic applies for $x \in W_2$. Thus, $W_1 \subseteq W_1 + W_2$ and $W_2 \subseteq W_1 + W_2$. $\square$

(b) Prove that any subspace of V that contains both W_1 and W_2 must also contain W_1+W_2.

Proof. Suppose A is a subspace of V such that $W_1 \subseteq A$ and $W_2 \subseteq A$. Then, for $x \in W_1$ and $y \in W_2$, $x + y \in A$ since A is a vector space. Thus, since x and y were arbitrary, $W_! + W_2 \subseteq A$. □

Problem 1.4.12. Show that a subset W of a vector space V is a subspace of V if and only if $span(W) = W$.

Proof. ($\Longrightarrow$): Suppose W is a subspace of V. Note, for any $v \in W$, $1 \cdot v \in W$, which implies $W \subseteq span(W)$. Now, since W is a subspace, then any linear combination of vectors will be in W. That is, $span(W) \subseteq W$. Therefore, $span(W) = W$.

($\Longleftarrow$): Suppose $span(W) = W$. Then by Theorem 1.5, $span(W)$ is a subspace of V, which implies W is a subspace of V. □

Problem 1.4.15. Let S_1 and S_2 be subsets of a vector space V. Prove that $span(S_1 \cap S_2) \subseteq span(S_1) \cap span(S_2)$. Give an example in which $span(S_1 \cap S_2)$ and $span(S_1) \cap span(S_2)$ are equal and one in which they are unequal.

Proof. Let $w \in span(S_1 \cap S_2)$. That is, $w = \sum_{i=1}^{n} c_i v_i$, where $v_i \in S_1$ and $v_i \in S_2$ and $c_i \in F$, where F is a field. Since v_i is in S_1 for all $i = \{1, \ldots, n\}$, then $w \in span(S_1)$. Similarly, since v_i is in S_2 for all $i = \{1, \ldots, n\}$, then $w \in span(S_2)$. Thus, $w \in span(S_1) \cap span(S_2)$. Since w was arbitrary in $span(S_1 \cap S_2)$, then $span(S_1 \cap S_2) \subseteq span(S_1) \cap span(S_2)$. □

In the case of $S_1 = \{2\}$ and $S_2 = \{3\}$ in the real-valued vector space, then $span(S_1 \cap S_2) = \{0\}$, but $span(S_1) \cap span(S_2) = \mathbb{R}$. Now, more simply, if $S_1 = S_2 = \{1\} \subseteq \mathbb{R}$, then $span(S_1 \cap S_2) = span(\{1\}) = \mathbb{R}$ and $span(S_1) \cap span(S_2) = \mathbb{R}$.

Problem 1.5.9. Let u and v be distinct vectors in a vector space V. Show that $\{u, v\}$ is linearly dependent if and only if u or v is a multiple of the other.

Proof. ($\Longrightarrow$): Suppose $\{u, v\}$ are linearly dependent. Then, there is some $c_1, c_2 \in F$, where F is a field, such that $c_1 u + c_2 v = 0$. So, $u = \frac{-c_2}{c_1} v$.

($\Longleftarrow$): Suppose vectors u and v are a multiple of $k \in F$ from each other. That is, $u = ku$. Then $u - ku = 0$. So, there exists $k \in F$, and trivially $1 \in F$, creates a linear combination of u and v which gives the zero vector. Therefore, $\{u, v\}$ are linearly dependent. □

Problem 1.5.15. Let $S = \{u_1, \ldots, u_n\}$ be a finite set of vectors. Prove that S is linearly dependent if and only if $u_1 = 0$ or $u_{k+1} \in span(\{u_1, \ldots, u_k\})$ for some $k \in \{1, \ldots, n-1\}$.

Proof. Let $S = \{u_1, \ldots, u_n\}$ be a finite set of vectors.

($\Longrightarrow$): Suppose S is linearly dependent. Then there exists $\{a_1, \ldots, a_m\}$, not all 0, such that $\sum_{i=1}^{m} a_i u_i = 0$ for $m \in \{1, \ldots, n\}$. If $u_1 = 0$, then $\{1, 0, \ldots, 0\}$ suffices for $\{a_1, \ldots, a_m\}$.

If no vector in S is zero, then $\sum_{i=1}^{m-1} a_i u_i = -a_m u_m$ for $m \in \{1, \dots, n\}$. That is, $u_m \in span(\{u_1, \dots, u_{m-1}\})$ for some $m \in \{1, \dots, n\}$.

($\Longleftarrow$): Suppose $u_1 = 0$. Then $1 \cdot u_1 = 0$ is a nontrivial representation of the zero vector in the set S. So, suppose $u_{k+1} \in span(\{u_1, \dots, u_k\})$ for some $k \in \{1, \dots, n-1\}$. That is,, $a_{k+1}u_{k+1} = \sum_{i=1}^{k} a_k u_k$ for some $k \in \{1, \dots, n-1\}$. That implies there exists $\{a_1, \dots, a_{k+1}\}$ such that $\sum_{i=1}^{k+1} a_i u_i = 0$. So, S is linearly dependent. □

Problem 1.6.16. The set of all upper triangular $n \times n$ matrices is a subspace W of $M_{n\times n}(F)$ (see Exercise 12 of Section 1.3). Find a basis for W. What is the dimension of W?

Proof. Let $W = \{A \in M_{n\times n}(F) : W \text{ is upper triangular}\}$. Let $A \in W$. Then,

$$A = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & \dots & & 0 \end{pmatrix} A_{1,1} + \begin{pmatrix} 0 & 1 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & \dots & & 0 \end{pmatrix} A_{1,2} + \dots + \begin{pmatrix} 0 & 0 & \dots & 1 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & \dots & & 0 \end{pmatrix} A_{1,n}$$

$$+ \begin{pmatrix} 0 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & \dots & & 0 \end{pmatrix} A_{2,2} + \dots + \begin{pmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 1 \\ \vdots & \vdots & \ddots & 0 \\ 0 & \dots & & 0 \end{pmatrix} A_{2,n}$$

$$\ddots$$

$$+ \begin{pmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & \dots & & 1 \end{pmatrix} A_{n,n}$$

is a set of linearly independent vectors, which form a basis for W. This can be simply seen and notated as the set $\beta = \{E_{ij} : i \le j\}$. Then, $dim(\beta) = 1 + 2 + \dots + n$, which is the popular consecutive sum formula $dim(\beta) = \frac{n(n+1)}{2}$. □

Problem 1.6.31. Let W_1 and W_2 be subspaces of a vector space V having dimensions m and n, respectively, where $m \ge n$.

(a) Prove that $dim(W_1 \cap W_2) \le n$.

Proof. Let W_1 and W_2 be subspaces of a vector space V having dimensions m and n, respectively, where $m \ge n$. Let $w \in W_1 \cap W_2$ be arbitrary. Then, since W_2 is a subspace with $dim(W_2) = n \le m$, it must have a basis $\beta = \{u_1, \dots, u_n\}$ that for any $w \in W_2$, $w = \sum_{i=1}^{k} a_i u_i$ and $1 \le k \le n$. Also, $span(\beta) = W$. Since the smaller basis of W_2 can generate at most W_2, that implies β can be used to form a linear combination for all $w \in W_1 \cap W_2$. Therefore, $dim(W_1 \cap W_2) \le n$. □

(b) Prove that $dim(W_1 + W_2) \leq m + n$.

Proof. Let W_1, W_2 be subspaces of V with $dim(W_1) = m$ and $dim(W_2) = n$, with $m \geq n$. Let $\beta = \{u_1, \ldots, u_m\}$ be the basis for W_1 and $\gamma = \{v_1, \ldots, v_n\}$ be the basis for W_2. Then $dim(span(\beta \cup \gamma)) \leq m + n$ since if $W_1 \cap W_2 = \emptyset$, then $dim(span(\beta \cup \gamma)) = dim(W_1) + dim(W_2) = m + n$ at its largest. Let $v \in W_1 + W_2$. Then there exists $x \in W_1$ and $y \in W_2$ such that $v = x + y$. We have that there exists $a_1, \ldots, a_m, b_1, \ldots, b_n \in F$ such that

$$x = \sum_{i=1}^{m} a_i u_i \quad \text{and} \quad y = \sum_{i=1}^{n} b_i v_i.$$

This gives us

$$v = \sum_{i=1}^{m} a_i u_i + \sum_{i=1}^{n} b_i v_i.$$

So, v is a linear combination of elements from $\beta \cup \gamma$. Since v was arbitrary, then $W_1 + W_2 \subseteq span(\beta \cup \gamma)$. Therefore, from Theorem 1.11,

$$\begin{aligned} dim(W_1 + W_2) &\leq dim(span(\beta \cup \gamma)) \\ &\leq m + n. \end{aligned}$$

□

Problem 2.1.14a. Let V and W be vector spaces and $T : V \to W$ be linear. Prove that T is one-to-one if and only if T carries linearly independent subsets of V onto linearly independent subsets of W.

Proof. Let V and W be vector spaces and $T : V \to W$ be linear.

($\Longrightarrow$): Suppose T is injective. Let $V' \subseteq V$ be linearly independent. Suppose for contradiction that $T(V')$ is linearly dependent. Then, for $v_1, \ldots, v_n \in V'$, there exists $c_1, \ldots, c_n \in \mathbb{F}$, not all zero, such that $\sum_{i=1}^{n} c_i T(v_i) = 0_W$. Then, by linearity of T,

$$T\left(\sum_{i=1}^{n} c_i v_i\right) = 0_W.$$

But, $N(T) = \{0_V\}$ by Theorem 2.4, which contradicts the linear independence of V'. Therefore, $T(V') \subseteq W$ is linearly independent.

($\Longleftarrow$): Suppose T maps linearly independent subsets of V onto linearly independent subsets of W. Then $T(v_1) = T(v_2)$ implies $T(v_1) - T(v_2) = 0_W$. By linearity of T, this gives us $T(v_1 - v_2) = 0_W$. Since T maps linearly independent subsets of V onto linearly independent subsets of W, we must have $v_1 - v_2 = 0_V$, which implies $v_1 = v_2$. Therefore, T is injective. □

Problem 2.1.20. Let V and W be vector spaces with subspaces V_1 and W_1, respectively. If $T : V \to W$ is linear, prove that $T(V_1)$ is a subspace of W and that $\{x \in V : T(x) \in W_1\}$ is a subspace of V.

Proof. Let V and W be vector spaces with subspaces V_1 and W_1, respectively. Let $T : V \to W$ be linear and $v_1, v_2 \in V_1$. Then $T(v_1) + T(v_2) = T(v_1 + v_2)$ by linearity of T. Since $v_1 + v_2 \in V_1$ and $T(v_1) + T(v_2) \in W$, then $T(V_1)$ is closed under addition. Also, for $c \in \mathbb{F}$, $cv_1 \in V_1$ and $T(cv_1) = cT(v_1) \in W$, implies $T(V_1)$ is closed under scalar multiplication. Lastly and trivially $0_V \in V_1$ and $T(0_V) = 0_W$. Thus, $T(V_1)$ is a subspace of W.

Now, let $S = \{x \in V : T(x) \in W_1\}$ and $v_1, v_2 \in S$. Then $T(v_1) + T(v_2) \in W_1$ by closure of subspace W_1, which implies $T(v_1) + T(v_2) = T(v_1 + v_2) \in W_1$ by linearity of T. Thus $v_1 + v_2 \in S$. Now $cT(v_1) \in W_1$ by closure of scalar multiplication in subspace W_1. By linearity of T, $cT(v_1) = T(cv_1) \in W_1$. Thus, $cv_1 \in S$. Lastly and trivially, since W_1 is a subspace, $0_W \in W_1$, and linear transformations map zero vectors to zero vectors, then $0_V \in S$, which implies S is a subspace of V. $\square$

Problem 2.2.8. Let V be an n-dimensional vector space with an ordered basis β. Define $T : V \to \mathbb{F}^n$ by $T(x) = \left[x\right]_\beta$. Prove that T is linear.

Proof. Let V be an n-dimensional vector space with an ordered basis $\beta = \{u_1, \dots, u_n\}$. Define $T : V \to \mathbb{F}^n$ by $T(x) = \left[x\right]_\beta$. Let $x_1, x_2 \in V$ be defined by $x_1 = \sum_{i=1}^n a_i u_i$ and $x_2 = \sum_{i=1}^n b_i u_i$ for $a_1, \dots, a_n, b_1, \dots, b_n \in \mathbb{F}$. So,

$$T(x_1) = \left[x_1\right]_\beta = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \qquad \text{and} \qquad T(x_2) = \left[x_2\right]_\beta = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}.$$

So,

$$\begin{aligned} T(x_1) + T(x_2) &= \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} + \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} \\ &= \begin{pmatrix} a_1 + b_1 \\ \vdots \\ a_n + b_n \end{pmatrix} \\ &= T(x_1 + x_2). \end{aligned}$$

For $c \in \mathbb{F}$,

$$\begin{aligned} cT(x_1) &= c \cdot \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \\ &= \begin{pmatrix} c \cdot a_1 \\ \vdots \\ c \cdot a_n \end{pmatrix} \\ &= T(c \cdot x_1). \end{aligned}$$

Therefore T is a linear transformation. $\square$

Problem 2.2.14. Let V and W be vector spaces, and let T and U be nonzero linear transformations from V into W. If $R(T) \cap R(U) = \{0\}$, prove that $\{T, U\}$ is a linearly independent subset of $\mathcal{L}(V, W)$.

Proof. Let V and W be vector spaces with ordered bases $\beta = \{v_1, \ldots, v_n\}$ and $\gamma = \{w_1, \ldots, w_m\}$, respectively. Let T and U be nonzero linear transformations from V into W with $R(T) \cap R(U) = \{0\}$. First, $\{T, U\} \subseteq \mathcal{L}(V, W)$ since T and U are linear transformations. Suppose, for contradiction, that $\{T, U\}$ is linearly dependent. Then, there exists $c_1, c_2 \in \mathbb{F}$, with $c_1 \neq 0$, such that $T = \frac{-c_2}{c_1} U$. This implies that $\left[T(v_j)\right]_\beta^\gamma = \frac{-c_2}{c_1} \left[U(v_j)\right]_\beta^\gamma$, which gives us $\left[T(v_j)\right]_\beta^\gamma \in R(T) \cap R(U)$. This implies that $\left[T(v_j)\right]_\beta^\gamma = 0_W$, which means that T is the zero transformation, contradiction. Therefore, $\{T, U\} \subseteq \mathcal{L}(V, W)$ is linearly independent. $\square$

Problem 2.3.13. Let A and B be $n \times n$ matrices. Recall that the trace of A is defined by

$$tr(A) = \sum_{i=1}^{n} A_{ii}.$$

Prove that $tr(AB) = tr(BA)$ and $tr(A) = tr(A^t)$.

Proof. Let A and B be $n \times n$ matrices of real values. Then

$$\begin{aligned} tr(AB) &= \sum_{i=1}^{n}(AB)_{ii} \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}b_{ji} \\ &= \sum_{j=1}^{n}\sum_{i=1}^{n} b_{ji}a_{ij} \\ &= \sum_{j=1}^{n}(BA)_{jj} \\ &= tr(BA). \end{aligned}$$

Also,

$$\begin{aligned} tr(A) &= \sum_{i,j=1}^{n} a_{ij} \\ &= \sum_{i,j=1}^{n} a_{ji} \\ &= tr(A^t). \end{aligned}$$

□

Problem 2.3.16a. Let V be a finite-dimensional vector space, and let $T : V \to V$ be linear. If $rank(T) = rank(T^2)$, prove that $R(T) \cap N(T) = \{0\}$. Deduce that $V = R(T) \oplus N(T)$ (see the exercises of Section 1.3).

Proof. Let V be a finite-dimensional vector space with $dim(V) = n$, and let $T : V \to V$ be linear. Suppose $rank(T) = rank(T^2)$. We then have,

$$\begin{aligned} nullity(T^2) &= dim(V) - rank(T^2) && \text{(By Theorem 2.3)} \\ &= dim(V) - rank(T) && \text{(Since } rank(T) = rank(T^2)\text{)} \\ &= nullity(T). && \text{(By Theorem 2.3)} \end{aligned}$$

Since T is a linear transformation between the same finite vector space, then by Theorem 2.5, $rank(T) = rank(T^2) = dim(V)$ implies $nullity(T) = nullity(T^2) = 0$. By Theorem 2.4, $nullity(T) = 0$ implies $N(T) = N(T^2) = \{0\}$. Therefore, $R(T) \cap N(T) = \{0\}$. Lastly,

$$dim(R(T) + N(T)) = dim(R(T)) + dim(N(T)) - dim(R(T) \cap N(T)) = dim(V).$$

Since $dim(R(T) \cap N(T)) = 0$, then $R(T) + N(T)$ must be V. Therefore, $V = R(T) \oplus N(T)$.

□

Problem 2.4.7. Let A be an $n \times n$ matrix.

(a) Suppose that $A^2 = 0$. Prove that A is not invertible.

Proof. Suppose that $A^2 = 0$. By Definition 7, A^2 is invertible if there exists $B \in M_{n\times n}(\mathbb{F})$ such that $A^2B = BA^2 = I_n$. We have $(A^2B)_{ij} = \sum_{k=1}^{n} 0 \cdot b_{kj} = 0$ and $(BA^2)_{ij} = \sum_{k=1}^{n} b_{ik} \cdot 0 = 0$ for all $1 \leq i, j \leq n$. Thus there does not exist $B \in M_{n\times n}(\mathbb{F})$ such that $A^2B = BA^2 = I_n$. □

(b) Suppose that $AB = 0$ for some nonzero $n \times n$ matrix B. Could A be invertible? Explain.

Proof. Let $AB = 0$ for some nonzero $n \times n$ matrix B. Suppose A has an inverse. Then $AB = 0$ implies $A^{-1}AB = 0$, which implies $IB = 0$, which implies $B = 0$, contradiction. Therefore, A is not invertible. □

Problem 2.4.15. Let V and W be n-dimensional vector spaces, and let $T : V \to W$ be a linear transformation. Suppose that β is a basis for V. Prove that T is an isomorphism if and only if $T(\beta)$ is a basis for W.

Proof. Let V and W be n-dimensional vector spaces, and let $T : V \to W$ be a linear transformation. Suppose that $\beta = \{v_1, \ldots, v_n\}$ is a basis for V.

($\Longrightarrow$): Suppose T is an isomorphism. From Theorem 2.2,

$$R(T) = span(T(\beta)) = span(\gamma) = W.$$

Since T is injective, T carries linearly independent subsets to linearly independent subsets. Therefore, $T(\beta)$ is a basis for W.

($\Longleftarrow$): Since V and W are both n-dimensional vector spaces with $T : V \to W$ linear, then by Theorem 2.19, $V \cong W$. □

Problem 2.5.9. Prove that "is similar to" is an equivalence relation on $M_{n\times n}(\mathbb{F})$.

Proof. We say "B **is similar to** A", notated $B \sim A$, if there exists $Q \in M_{n\times n}(\mathbb{F})$ invertible such that $B = Q^{-1}AQ$. Let $Q = Q^{-1} = I_n$. Then $A = Q^{-1}AQ$. So $A \sim A$. Thus **is similar to** is reflexive.

Now, suppose $B \sim A$. Then there exists $Q \in M_{n\times n}(\mathbb{F})$ invertible such that $B = Q^{-1}AQ$. By letting $Q^{-1} = Q'$, we have $A = Q'^{-1}BQ'$. That is $A \sim B$. We can similarly show $A \sim B$ implies $B \sim A$. Thus **is similar to** is symmetric.

Now suppose there exists $Q_1, Q_2 \in M_{n\times n}(\mathbb{F})$ invertible such that $B = Q_1^{-1}AQ_1$ and $C = Q_2^{-1}BQ_2$. That is, $B \sim A$ and $C \sim B$. Then, letting $Q_3 = Q_1Q_2$, invertible we get

$$\begin{aligned} C &= Q_2^{-1}BQ_2 \\ &= Q_2^{-1}Q_1^{-1}AQ_1Q_2 \\ &= (Q_1Q_2)^{-1}AQ_1Q_2 \\ &= Q_3^{-1}AQ_3. \end{aligned}$$

Thus $C \sim A$, which implies **is similar to** is transitive. Therefore, **is similar to** is an equivalence relation on $M_{n\times n}(\mathbb{F})$. □

Problem 2.5.10.

(a) Prove that if A and B are similar $n \times n$ matrices, then $tr(A) = tr(B)$. Hint: Use Exercise 13 of Section 2.3.

Proof. Let $A, B \in M_{n\times n}(\mathbb{F})$. Suppose B **is similar to** A. That is, there exists $Q \in M_{n\times n}(\mathbb{F})$ invertible such that $B = Q^{-1}AQ$. Then

$$\begin{aligned} tr(B) &= tr(Q^{-1}AQ) \\ &= tr(Q^{-1}QA) && \text{(By Problem 2.3.13)} \\ &= tr(I_nA) \\ &= tr(A). \end{aligned}$$

□

(b) How would you define the trace of a linear operator on a finite-dimensional vector space? Justify that your definition is well-defined.

For finite-dimensional vector space V and $T \in \mathcal{L}(V)$ a linear operator, the $tr(T)$ is the trace of the matrix representation of the linear operator in any basis of V. The basis used to represent the transform is allowed to be arbitrary because as shown in part (a), similar representations of the linear transform T does not change the value of $tr(T)$.

Problem 3.1.6. Let A be an $m \times n$ matrix. Prove that if B can be obtained from A by an elementary row [column] operation, then B^t can be obtained from A^t by the corresponding elementary column [row] operation.

Proof. Let A be an $m \times n$ matrix. Suppose B can be obtained from A by an elementary row [column] operation. That is, there exists $E \in M_{m\times m}(\mathbb{F})$ [$E \in M_{n\times n}(\mathbb{F})$] such that $B = EA$ [$B = AE$]. Then, by taking the transpose of both sides,

$$\begin{aligned} B^t &= (EA)^t \quad [(AE)^t] \\ &= A^tE^t. \quad [E^tA^t] \end{aligned}$$

Therefore, B^t can now be obtained from A^t by the corresponding elementary column [row] operation. □

Problem 3.2.14. Let $T, U : V \to W$ be linear transformations.

(a) Prove that $R(T+U) \subseteq R(T) + R(U)$. (See the definition of the sum of subsets of a vector space.)

Proof. Let $w \in R(T+U)$. Then there exists some $v \in V$ such that $w = (T+U)(v)$. By the definition of direct sum, we can say $w = T(v) + U(v)$. Since $T(v) \in R(T)$ and $U(v) \in R(U)$, then we have $w \in R(T) + R(U)$. □

(b) Prove that if W is finite-dimensional, then $rank(T+U) \leq rank(T) + rank(U)$.

Proof. Let W be finite dimensional. Then,

$$\begin{aligned} rank(T+U) &= dim(R(T+U)) \\ &\leq dim(R(T) + R(U)) && \text{(From the part above)} \\ &= dim(R(T)) + dim(R(U)) - dim(R(T) \cap R(U)) \\ &\leq dim(R(T)) + dim(R(U)) \\ &= rank(T) + rank(U). \end{aligned}$$

□

Problem 3.2.22. Let B be an $n \times m$ matrix with rank m. Prove that there exists an $m \times n$ matrix A such that $AB = I_m$.

Proof. Let B be an $n \times m$ matrix with rank m. Then $m \leq n$. By Corollary 1, there exists $X \in M_{n\times n}(\mathbb{F})$ and $Y \in M_{m\times m}(\mathbb{F})$ such that $D = XBY$, where

$$D = \begin{pmatrix} I_m & 0_1 \\ 0_2 & 0_3 \end{pmatrix} \in M_{n\times m}(\mathbb{F}).$$

So, $D = \begin{pmatrix} I_m \\ 0 \end{pmatrix}$, where $0 \in M_{(n-m)\times m}(0)$. Since X and Y are invertible, $B = X^{-1}DY^{-1}$. Let $A = YD^tX$. Then,

$$\begin{aligned} AB &= YD^tXX^{-1}DY^{-1} \\ &= YD^tI_nDY^{-1} \\ &= YD^tDY^{-1} \\ &= YI_mY^{-1} \\ &= YY^{-1} \\ &= I_m. \end{aligned}$$

□

Problem 3.3.10. Prove or give a counterexample to the following statement: If the coefficient matrix of a system of m linear equations in n unknowns has rank m, then the system has a solution.

Proof. Let $A \in M_{m\times n}(\mathbb{F})$ represent the coefficient matrix of a system of m linear equations in n unknowns. Let $rank(A) = m$. Then, the equation $Ax = b$ has a solution if and only if $b \in R(L_A)$. Since $dim(R(L_A)) = m$, then $R(L_A) \subseteq \mathbb{F}^m$. Thus, $b \in \mathbb{F}^m$. Let $A_1, \ldots, A_n$ denote the columns of A. Then, $R(L_A) = span(\{A_1, \ldots, A_n\})$. This gives us

$$\begin{aligned}
Ax = b \text{ has a solution } &\iff b \in span(\{A_1, \ldots, A_n\}) \\
&\iff span(\{A_1, \ldots, A_n\}) = span(\{A_1, \ldots, A_n, b\}) \\
&\iff dim(span(\{A_1, \ldots, A_n\})) = dim(span(\{A_1, \ldots, A_n, b\})).
\end{aligned}$$

Since we know $m = dim(span(\{A_1, \ldots, A_n\}))$, then $m = dim(span(\{A_1, \ldots, A_n, b\}))$. Since we know $b \in \mathbb{F}^m$, then $Ax = b$ must have a solution. □

Problem 4.1.10. The classical adjoint of a 2×2 matrix $A \in M_{2\times 2}(\mathbb{F})$ is the matrix

$$C = \begin{pmatrix} A_{22} & -A_{12} \\ -A_{21} & A_{11} \end{pmatrix}.$$

Prove that

(a) $CA = AC = (det(A))I$.

Proof.

$$\begin{aligned}
CA &= \begin{pmatrix} A_{22} & -A_{12} \\ -A_{21} & A_{11} \end{pmatrix} \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \\
&= \begin{pmatrix} A_{22}A_{11} + -A_{12}A_{21} & A_{22}A_{12} + -A_{12}A_{22} \\ -A_{21}A_{11} + A_{11}A_{21} & -A_{21}A_{12} + A_{11}A_{22} \end{pmatrix} \\
&= (A_{22}A_{11} - A_{12}A_{21}) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\
&= (det(A))I.
\end{aligned}$$

□

(b) $det(C) = det(A)$.

Proof.

$$\begin{aligned}
det(C) &= A_{22}A_{11} - A_{12}A_{21} \\
&= A_{11}A_{22} - A_{12}A_{21} \\
&= det(A).
\end{aligned}$$

□

(c) The classical adjoint of A^t is C^t.

Proof. Since $A^t = \begin{pmatrix} A_{11} & A_{21} \\ A_{12} & A_{22} \end{pmatrix}$. Then, the classical adjoint of A^t would be $\begin{pmatrix} A_{22} & -A_{21} \\ -A_{12} & A_{11} \end{pmatrix}$. Then, taking the transpose, you get $\begin{pmatrix} A_{22} & -A_{12} \\ -A_{21} & A_{11} \end{pmatrix}$, which is C. Thus, the classical adjoint of A^t is C^t. □

(d) If A is invertible, then $A^{-1} = (det(A))^{-1}C$.

Proof.

$$\begin{aligned} (det(A))^{-1}C &= \frac{1}{(A_{11}A_{22} - A_{12}A_{21})} \begin{pmatrix} A_{22} & -A_{12} \\ -A_{21} & A_{11} \end{pmatrix} \\ &= A^{-1}. && \text{(By Theorem 4.2)} \end{aligned}$$

□

Problem 4.2.25. Prove that $det(kA) = k^n det(A)$ for any $A \in M_{n\times n}(\mathbb{F})$.

Proof. Let $A \in M_{n\times n}(\mathbb{F})$. We know that for a matrix B obtained by multiplying a row from A by k, we have $det(B) = kdet(A)$. Since kA is the result of multiplying each row in A by k, we have $det(kA) = k^n det(A)$. □

Problem 4.3.12. A matrix $Q \in M_{n\times n}(\mathbb{R})$ is called orthogonal if $QQ^t = I$. Prove that if Q is orthogonal, then $det(Q) = \pm 1$.

Proof. Let $Q \in M_{n\times n}(\mathbb{R})$ be orthogonal. Note, $Q^t = Q^{-1}$. Then

$$\begin{aligned} 1 &= det(I) \\ &= det(QQ^t) \\ &= det(Q)det(Q^t) \\ &= (det(Q))^2. \end{aligned}$$

Therefore, $det(Q) = \pm 1$. □

Problem 4.3.13. For $M \in M_{n\times n}(\mathbb{C})$, let $\overline{M}$ be the matrix such that $(\overline{M})_{ij} = \overline{M_{ij}}$ for all i, j, where $\overline{M_{ij}}$ is the complex conjugate of M_{ij}.

(a) Prove that $det(\overline{M}) = \overline{det(M)}$.

Proof. Let $M \in M_{n\times n}(\mathbb{C})$. Using induction, we will show $det(\overline{M}) = \overline{det(M)}$.

(Base Case): Let $n = 1$. Then $det(\overline{M}) = \overline{M_{11}} = \overline{det(M)}$.

(Induction Step): Let $n \geq 2$. Then

$$\begin{aligned} det(\overline{M}) &= \sum_{j=1}^{n}(-1)^{1+j}\overline{M_{1j}}(det(\tilde{\overline{M}}_{1j})) \\ &= \sum_{j=1}^{n}(-1)^{1+j}\overline{M_{1j}}(\overline{det(\tilde{M}_{1j})}) && \text{(By the inductive hypothesis)} \\ &= \overline{det(M)}. \end{aligned}$$

$\square$

(b) A matrix $Q \in M_{n\times n}(\mathbb{C})$ is called unitary if $QQ^* = I$, where $Q^* = \overline{Q^t}$. Prove that if Q is a unitary matrix, then $|det(Q)| = 1$.

Proof. Let $Q \in M_{n\times n}(\mathbb{C})$ be unitary. Then there exists $Q^* \in M_{n\times n}(\mathbb{C})$ such that $Q^* = \overline{Q^t}$ and $QQ^* = I_n$. Then

$$\begin{aligned} 1 &= det(I_n) \\ &= det(QQ^*) \\ &= det(Q)det(Q^*) \\ &= det(Q)\overline{det(Q^t)} && \text{(By (a))} \\ &= det(Q)\overline{det(Q)} \\ &= |det(Q)|. && \text{(Since } z\overline{z} = |z| = a^2+b^2 \text{ for } z = a+bi \in \mathbb{C}) \end{aligned}$$

$\square$

Problem 4.3.16. Use determinants to prove that if $A, B \in M_{n\times n}(\mathbb{F})$ are such that $AB = I$, then A is invertible (and hence $B = A^{-1}$).

Proof. Suppose $A, B \in M_{n\times n}(\mathbb{F})$ such that $AB = I_n$. Then

$$\begin{aligned} 1 &= det(I) \\ &= det(AB) \\ &= det(A)det(B). \end{aligned}$$

Thus, $det(A) \neq 0$ and as discussed in section 4.3, A is invertible, and thus $A^{-1} = B$. $\square$

Problem 5.1.20. Let A be an $n \times n$ matrix with characteristic polynomial

$$f(t) = (-1)^n t^n + a_{n-1}t^{n-1} + \cdots + a_1 t + a_0.$$

Prove that $f(0) = a_0 = det(A)$. Deduce that A is invertible if and only if $a_0 \neq 0$.

Proof. Let A be an $n \times n$ matrix with characteristic polynomial $f(t) = (-1)^n t^n + a_{n-1}t^{n-1} + \cdots + a_1 t + a_0$. Then, from the definition of the characteristic polynomial, $f(t) = det(A - tI_n)$ implies $f(0) = det(A - 0 \cdot I_n) = det(A)$ and $f(t) = (-1)^n t^n + a_{n-1}t^{n-1} + \cdots + a_1 t + a_0$ implies $f(0) = a_0$. Thus, $f(0) = a_0 = det(A)$. Finally, from Properties of the Determinant, Section 4.4, we know A is invertible if and only if $det(A) \neq 0$. Therefore, A is invertible if and only if $a_0 \neq 0$. □

Problem 5.1.21. Let A and $f(t)$ be as in Exercise 20.

(a) Prove that $f(t) = (A_{11} - t)(A_{22} - t) \dots (A_{nn} - t) + q(t)$, where $q(t)$ is a polynomial of degree at most $n - 2$. Hint: Apply mathematical induction to n.

Proof.

– Base Case: Let $n = 2$. Then

$$f(t) = \begin{vmatrix} A_{11} - t & A_{12} \\ A_{21} & A_{22} - t \end{vmatrix} = (A_{11} - t)(A_{22} - t) - A_{12}A_{21}.$$

– Induction Step: Let the statement be true for some $n-1 \times n-1$ matrix A. Then

$$f(t) = (A_{11} - t) \begin{vmatrix} A_{22} - t & A_{23} & \dots & A_{2n} \\ A_{21} & A_{22} - t & \dots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n2} & A_{n3} & \dots & A_{nn} - t \end{vmatrix} + \sum_{j=2}^{n} (-1)^{j+1} A_{1j} det((A \tilde{-} tI)_{1j}).$$

By inductive hypothesis, we get a polynomial $q''(t)$ with degree at most $n-2$ and polynomial $q'(t)$ with degree at most $n - 3$. Thus, we have

$$f(t) = (A_{11} - t)(A_{22} - t) \dots (A_{nn} - t) + (A_{11} - t)q'(t) + \sum_{j=1}^{n} (-1)^{j+1} A_{1j} q''(t).$$

By defining $q(t) = (A_{11} - t)q'(t) + \sum_{j=1}^{n} (-1)^{j+1} A_{1j} q''(t)$, then $q(t)$ is a polynomial of degree at most $n - 2$. Therefore, $f(t) = f(t) = (A_{11} - t)(A_{22} - t) \dots (A_{nn} - t) + q(t)$.

□

(b) Show that $tr(A) = (-1)^{n-1} a_{n-1}$.

Proof. Consider the coefficient of t^{n-1} in the expansion of the characteristic polynomial $f(t) = (A_{11} - t)(A_{22} - t) \dots (A_{nn} - t) + q(t)$. Since it is only obtained from multiplying $n-1$ diagonal elements of $det(A - tI_n)$ by the scalar of the remaining diagonal element, then the coefficient of t^{n-1} is $(-1)^{n-1}tr(A)$. So, $a_{n-1} = (-1)^{n-1}tr(A)$, or $tr(A) = (-1)^{n-1}a_{n-1}$. □

Problem 5.2.11. Let A be an $n \times n$ matrix that is similar to an upper triangular matrix and has the distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k$ with corresponding multiplicities $m_1, m_2, \ldots, m_k$. Prove the following statements.

(a) $tr(A) = \sum_{i=1}^{k} m_i \lambda_i$

Proof. Factoring the characteristic polynomial $f(t)$ into linear factors, we have $f(t) = (-1)^n (t-\lambda_1)^{m_1}(t-\lambda_2)^{m_2} \ldots (t-\lambda_k)^{m_k}$. Since the trace is the coefficient of the t^{n-1} term, then by multiplying the right hand side to its full extent, we can see that the t^{n-1} coefficient is $m_1\lambda_1 + m_2\lambda_2 + \cdots + m_k\lambda_k$. Therefore, $tr(A) = \sum_{i=1}^{k} m_i \lambda_i$. □

(b) $det(A) = (\lambda_1)^{m_1}(\lambda_2)^{m_2} \ldots (\lambda_k)^{m_k}$.

Proof. Since $f(0) = det(A)$ and $f(t) = (-1)^n (t-\lambda_1)^{m_1}(t-\lambda_2)^{m_2} \ldots (t-\lambda_k)^{m_k}$, then $det(A) = \prod_{i=1}^{k} \lambda_i^{m_i}$. □

Problem 5.2.13. Let T be an invertible linear operator on a finite-dimensional vector space V.

(a) Recall that for any eigenvalue λ of T, λ^{-1} is an eigenvalue of T^{-1} (Exercise 9 of Section 5.1). Prove that the eigenspace of T corresponding to λ is the same as the eigenspace of T^{-1} corresponding to λ^{-1}.

Proof. Let T be an invertible linear operator on a finite-dimensional vector space V. Let λ be an eigenvalue of T. Let $x \in E_\lambda$, where E_λ is the eigenspace of T corresponding to λ. Also, let $E'_{\lambda^{-1}}$ be the eigenspace of T^{-1} corresponding to λ^{-1}. Then

$$T(T^{-1}(x)) = x = \lambda^{-1}\lambda x = \lambda^{-1}T(x) = T(\lambda^{-1}x),$$

which implies $T^{-1}(x) = \lambda^{-1}x$. Thus, $E_\lambda \subseteq E'_{\lambda^{-1}}$.

Now, let $x \in E'_{\lambda^{-1}}$. Then, similarly,

$$T^{-1}(T(x)) = x = \lambda\lambda^{-1}x = \lambda T^{-1}(x) = T^{-1}(\lambda x),$$

which implies $T(x) = \lambda x$. Thus, $E'_{\lambda^{-1}} \subseteq E_\lambda$. Therefore, $E_\lambda = E'_{\lambda^{-1}}$. □

(b) Prove that if T is diagonalizable, then T^{-1} is diagonalizable.

Proof. Suppose T is diagonalizable and $dim(V) = n$. Then there exists an ordered basis β such that $[T]_\beta$ is diagonal. So,

$$[T]_\beta = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}.$$

Since each eigenvalue is nonzero and for any eigenvalue λ_i of T, λ_i^{-1} is an eigenvalue of T^{-1}, we get

$$\left[T^{-1}\right]_\beta = \left(\left[T\right]_\beta\right)^{-1} = \begin{pmatrix} \lambda_1^{-1} & & 0 \\ & \ddots & \\ 0 & & \lambda_n^{-1} \end{pmatrix}.$$

Therefore, T^{-1} must be diagonalizable. □

Problem 5.4.11. Let T be a linear operator on a vector space V, let v be a nonzero vector in V, and let W be the T-cyclic subspace of V generated by v. Prove that

(a) W is T-invariant.

Proof. Let T be a linear operator on a vector space V, let v be a nonzero vector in V, and let W be the T-cyclic subspace of V generated by v. Suppose $dim(W) = k$. Then $\{v, Tv, T^2v, \dots, T^{k-1}v\}$ is a basis for W. Then, let $x = a_0v + a_1Tv + a_2T^2v + \cdots + a_{k-1}T^{k-1}v$ for some $a_0, \dots, a_{k-1} \in \mathbb{F}$. Thus,

$$T(x) = a_0Tv + a_1T^2v + \cdots + a_{k-1}T^kv.$$

Since $T^nv \in W$ for all $n \in \mathbb{Z}^+$, then $x \in span(W)$. Thus, W is T-invariant. □

(b) Any T-invariant subspace of V containing v also contains W.

Proof. Let T be a linear operator on a vector space V, let v be a nonzero vector in V, and let W be the T-cyclic subspace of V generated by v. Let W' be a T-invariant subspace of V that contains v. Let $k = dim(W)$ with basis $\{v, Tv, T^2v, \dots, T^{k-1}v\}$. Since W' is T-invariant, then $T^n(W') \subseteq W'$ for all $n \in \mathbb{Z}^+$. Thus, we can find $v, Tv, T^2v, \dots, T^{k-1}v$ in W'. Thus, $W \subseteq W'$ and $span(W) \subseteq span(W')$. □

Problem 5.4.17. Let A be an $n \times n$ matrix. Prove that

$$dim(span(\{I_n, A, A^2, \dots\})) \leq n.$$

Proof. Let T be a linear operator on a vector space V, let v be a nonzero vector in V, and let W be the T-cyclic subspace of V generated by v. Let W' be a T-invariant subspace of V that contains v. We'll show $A^m \in span(\{I_n, A, A^2, \dots, A^{n-1}\})$ for all $m \in \mathbb{N}$. Clearly, this is true for $m < n$. We will use induction on m to show $A^m \in span(\{I_n, A, A^2, \dots, A^{n-1}\})$ for all $m \geq n$.

- Base Case: Let $m = n$. By Theorem 5.21 and the Cayley-Hamilton Theorem for matrices, the characteristic polynomial of $L_A\big|_W$, for $a_0, \dots, a_{n-1} \in \mathbb{F}$, is

$$f(t) = det(A - tI) = (-1)^n(a_0 + a_1t + \cdots + a_{n-1}t^{n-1} + t^n)$$

and $f(A) = 0_n$. Thus, $A^n = \sum_{i=0}^{n-1}(-1)^n(-a_i)A^i$. Thus,

$$A^m = A^n \in span(\{I_n, A, A^2, \dots, A^{n-1}\}).$$

- Induction Step: Let $m > n$ and suppose $A^{m'} \in span(\{I_n, A, A^2, \dots, A^{n-1}\})$ for all $m' \leq m$. Then, $A^m, A^n \in span(\{I_n, A, A^2, \dots, A^{n-1}\})$. Thus, there exists scalars $a_0, \dots, a_{n-1} \in \mathbb{F}$ such that $A^m = \sum_{i=0}^{n-1} a_i A^i$. Thus,

$$\begin{aligned} A^{m+1} &= A^m A \\ &= \left(\sum_{i=0}^{n-1} a_i A^i \right) A \\ &= \sum_{i=0}^{n-1} a_i A^{i+1}. \end{aligned}$$

  Since $A^i \in span(\{I_n, A, A^2, \dots, A^{n-1}\})$ for all $i \leq n$, then A^{m+1} is a linear combination of elements in $span(\{I_n, A, A^2, \dots, A^{n-1}\})$. Therefore, $A^{m+1} \in span(\{I_n, A, A^2, \dots, A^{n-1}\})$.

Thus, $span(\{I_n, A, A^2, \dots\}) = span(\{I_n, A, A^2, \dots, A^{n-1}\})$. Therefore,

$$dim(span(\{I_n, A, A^2, \dots\})) = dim(span(\{I_n, A, A^2, \dots, A^{n-1}\})) \leq n.$$

□

Problem 6.1.12. Let $\{v_1, v_2, \dots, v_k\}$ be an orthogonal set in V, and let $a_1, a_2, \dots, a_k$ be scalars. Prove that

$$\left\| \sum_{i=1}^{k} a_i v_i \right\|^2 = \sum_{i=1}^{k} |a_i|^2 \|v_i\|^2.$$

Proof. Let $\{v_1, v_2, \dots, v_k\}$ be an orthogonal set in V, and let $a_1, a_2, \dots, a_k$ be scalars. Then

$$\begin{aligned} \left\| \sum_{i=1}^{k} a_i v_i \right\|^2 &= \left\langle \sum_{i=1}^{k} a_i v_i, \sum_{i=1}^{k} a_i v_i \right\rangle \\ &= \sum_{i=1}^{k} \langle a_i v_i, a_i v_i \rangle + \sum_{i \neq j} \langle a_i v_i, a_j, v_j \rangle \qquad (\text{Note, } \langle v_i, v_j \rangle = c\delta_{ij} \text{ for some } c \in \mathbb{F}) \\ &= \sum_{i=1}^{k} |a_i|^2 \|v_i\|^2. \end{aligned}$$

□

Problem 6.1.17. Let T be a linear operator on an inner product space V, and suppose that $\|T(x)\| = \|x\|$ for all x. Prove that T is injective.

Proof. Let T be a linear operator on an inner product space V, and suppose that $\|T(x)\| = \|x\|$ for all x. Then,

$$\|T(x)\| = \sqrt{\langle T(x), T(x)\rangle} = \sqrt{\langle x, x\rangle} = \|x\|.$$

Thus, $\langle T(x), T(x)\rangle = \langle x, x\rangle$. By Theorem 6.9, we have $\langle x, T^*T(x)\rangle = \langle x, x\rangle$ for all x. Then by Theorem 6.1e, we have $T^*T(x) = x$, which implies $T^*T = I$, and similarly, $TT^* = I$. Therefore, T is invertible, which implies T is bijective. □

Problem 6.2.16a. *Bessel's Inequality.* Let V be an inner product space, and let $S = \{v_1, v_2, \ldots, v_n\}$ be an orthonormal subset of V. Prove that for any $x \in V$ we have

$$\|x\|^2 \geq \sum_{i=1}^{n} |\langle x, v_i\rangle|^2 .$$

Hint: Apply Theorem 6.6 to $x \in V$ and $W = span(S)$. Then use Exercise 10 of Section 6.1.

Proof. Let V be an inner product space, and let $S = \{v_1, v_2, \ldots, v_n\}$ be an orthonormal subset of V. Let $W = span(S)$ and $x \in V$. Then, by Theorem 6.6, there exist unique $u \in W$ and $z \in W^{\perp}$ such that $x = u + z$ and

$$u = \sum_{i=1}^{n} \langle x, v_i\rangle v_i.$$

Thus,

$$\begin{aligned}
\|x\|^2 &= \|u+z\|^2 \\
&= \|u\|^2 + \|z\|^2 && \text{(By previous exercise)} \\
&\geq \|u\|^2 \\
&= \langle u, u\rangle \\
&= \left\langle \sum_{i=1}^{n} \langle x, v_i\rangle v_i, \sum_{i=1}^{n} \langle x, v_i\rangle v_i \right\rangle \\
&= \sum_{i=1}^{n} \langle x, v_i\rangle \overline{\langle x, v_i\rangle} \langle v_i, v_i\rangle \\
&= \sum_{i=1}^{n} \langle x, v_i\rangle \overline{\langle x, v_i\rangle} && \text{(Since } v_i \text{ is orthonormal)} \\
&= \sum_{i=1}^{n} |\langle x, v_i\rangle|^2 .
\end{aligned}$$

□

Lemma 1. Let T be a linear operator on a finite-dimensional inner product space V. Then $R(T^*)^\perp = N(T)$.

Proof. Let T be a linear operator on a finite-dimensional inner product space V. Then $x \in R(T^*)^\perp$ if and only if $\langle x, T^*(y) \rangle = \langle T(x), y \rangle$ for all $y \in V$. Also, $\langle T(x), y \rangle = 0$ if and only if $T(x) = 0$. Thus, $x \in N(T)$. Thus, $x \in R(T^*)^\perp$ if and only if $x \in N(T)$. Therefore, $R(T^*)^\perp = N(T)$. $\square$

Problem 6.3.13. Let T be a linear operator on a finite-dimensional inner product space V. Prove the following results.

(a) $N(T^*T) = N(T)$. Deduce that $rank(T^*T) = rank(T)$.

Proof. Let T be a linear operator on a finite-dimensional inner product space V. Then, there exists a unique linear operator T^*. Let $x \in N(T^*T)$. Then, $T^*T(x) = 0$ implies $0 = \langle T^*T(x), x \rangle = \langle T(x), T(x) \rangle$. That then implies $T(x) = 0$, which implies $x \in N(T)$. So, $N(T^*T) \subseteq N(T)$. Now, suppose $x \in N(T)$. Then, $T^*T(x) = T^*(0) = 0$ since T^* is linear. Thus, $N(T) \subseteq N(T^*T)$. Therefore, $N(T^*T) = N(T)$.

Now, using the dimension theorem,

$$dim(V) = nullity(T^*T) + rank(T^*T) = nullity(T) + rank(T).$$

Since $nullity(T^*T) = nullity(T)$, then $rank(T^*T) = rank(T)$. $\square$

(b) $rank(T) = rank(T^*)$. Deduce from (a) that $rank(TT^*) = rank(T)$.

Proof. Let T be a linear operator on a finite-dimensional inner product space V. Then by Theorem 6.7, $V = R(T) \oplus R(T)^\perp = R(T^*) \oplus R(T^*)^\perp$. Then, by Lemma 1 above, $R(T^*)^\perp = N(T)$. By the dimension theorem, $dim(V) = rank(T) + nullity(T)$. Putting it all together, we have

$$\begin{aligned} dim(V) &= rank(T) + nullity(T) \\ &= dim(R(T^*)) + dim(R(T^*)^\perp) \\ &= dim(R(T^*)) + dim(N(T)) \\ &= rank(T^*) + nullity(T). \end{aligned}$$

Thus, $rank(T) = rank(T^*)$. Now, from (a), we know $rank(T^{**}T^*) = rank(TT^*) = rank(T^*)$. Therefore, putting them together, we get $rank(TT^*) = rank(T)$. $\square$