1. Let \( V \) be a vector space over \( F \) and \( S, T \in \text{End}(V) \).
   (a) Show that \( I - S \circ T \) is injective iff \( I - T \circ S \) is injective.

   **Solution.** The trick is to observe that
   \[
   S \circ (I - T \circ S) = (I - S \circ T) \circ S.
   \]
   Suppose \( I - S \circ T \) is injective. Let \( v \in \ker(I - T \circ S) \). Then \( (I - T \circ S)(v) = 0_V \). Hence
   \[
   (I - S \circ T)(S(v)) = S((I - T \circ S)(v)) = S(0_V) = 0_V
   \]
   and since \( I - S \circ T \) is injective, we deduce that
   \[
   S(v) = 0_V.
   \]
   So
   \[
   T \circ S(v) = T(0_V) = 0_V;
   \]
   and so
   \[
   v = v - 0_V = v - T \circ S(v) = (I - T \circ S)(v) = 0_V
   \]
   where the last equality follows from our choice of \( v \in \ker(I - T \circ S) \). Hence \( \ker(I - T \circ S) = \{0_V\} \) and we deduce that \( I - T \circ S \) is injective by Theorem 4.12.
   (b) \( T \) is called **nilpotent** if
   \[
   T^n = O
   \]
   for some \( n \in \mathbb{N} \). Show that if \( T \) is nilpotent, then \( I - T \) is bijective. What is \( (I - T)^{-1} \)?

   **Solution.** Let \( S = I + T + T^2 + \cdots + T^{n-1} \) (by the geometric series heuristic in lecture). Then
   \[
   (I - T) \circ S = (I + T + \cdots + T^{n-1}) - (T + T^2 + \cdots + T^n) = I - T^n = I - O = I
   \]
   and likewise for \( S \circ (I - T) \). Hence
   \[
   (I - T)^{-1} = S
   \]
   and \( I - T \) is bijective.

2. Let \( V \) be a vector space over \( \mathbb{R} \) and \( T \in \text{End}(V) \) be an involution, ie. \( T^2 = I \). Define
   \[
   V_+ := \{v \in V \mid T(v) = v\} \quad \text{and} \quad V_- := \{v \in V \mid T(v) = -v\}.
   \]
   (a) Show that \( V_+ \) and \( V_- \) are subspaces of \( V \).

   **Solution.** Let \( \alpha_1, \alpha_2 \in \mathbb{R} \) and \( v_1, v_2 \in V_+ \). Then \( T(v_1) = v_1, T(v_2) = v_2 \), and
   \[
   T(\alpha_1 v_1 + \alpha_2 v_2) = \alpha_1 T(v_1) + \alpha_2 T(v_2)
   = \alpha_1 v_1 + \alpha_2 v_2.
   \]
   Hence \( \alpha_1 v_1 + \alpha_2 v_2 \in V_+ \). Likewise for \( V_- \).
(b) Show that

\[ V_+ \oplus V_- = V. \]

**Solution.** Note that

\[ \mathbf{v} = \frac{1}{2}(I + T)(\mathbf{v}) + \frac{1}{2}(I - T)(\mathbf{v}). \]  \hspace{1cm} (2.1)

Since \( T^2 = I, \)

\[ T\left(\frac{1}{2}(I + T)(\mathbf{v})\right) = T\left(\frac{1}{2}\mathbf{v} + \frac{1}{2}T(\mathbf{v})\right) = \frac{1}{2}T(\mathbf{v}) + \frac{1}{2}T^2(\mathbf{v}) = \frac{1}{2}T(\mathbf{v}) + \frac{1}{2}I(\mathbf{v}) = \frac{1}{2}(I + T)(\mathbf{v}) \]

and likewise,

\[ T\left(\frac{1}{2}(I - T)(\mathbf{v})\right) = T\left(\frac{1}{2}\mathbf{v} - \frac{1}{2}T(\mathbf{v})\right) = \frac{1}{2}T(\mathbf{v}) - \frac{1}{2}T^2(\mathbf{v}) = \frac{1}{2}T(\mathbf{v}) - \frac{1}{2}I(\mathbf{v}) = -\frac{1}{2}(I - T)(\mathbf{v}). \]

Hence

\[ \frac{1}{2}(I + T)(\mathbf{v}) \in V_+, \quad \frac{1}{2}(I - T)(\mathbf{v}) \in V_- , \]

and (2.1) implies that \( V = V_+ + V_- . \) If \( \mathbf{v} \in V_+ \cap V_- , \) then

\[ \mathbf{v} = T(\mathbf{v}) = -\mathbf{v} \]

implies that \( \mathbf{v} = 0_V. \) Hence \( V_+ \cap V_- = \{ 0_V \} \) and so \( V = V_+ \oplus V_- . \)

**3.** Let \( A \in \mathbb{R}^{m \times n} \) and \( b \in \mathbb{R}^m. \)

(a) Show that

\[ \text{nullsp}(A^T A) = \text{nullsp}(A). \]

**Solution.** Let \( \mathbf{x} \in \text{nullsp}(A). \) Then \( A\mathbf{x} = 0. \) Therefore, \( A^T A \mathbf{x} = A^T 0 = 0. \) Therefore \( \mathbf{x} \in \text{nullsp}(A^T A). \) So \( \text{nullsp}(A) \subseteq \text{nullsp}(A^T A). \) Conversely, let \( \mathbf{x} \in \text{nullsp}(A^T A). \) Then \( A^T A \mathbf{x} = 0. \) Multiplying on the left by \( \mathbf{x}^T, \) we get

\[ \mathbf{x}^T A^T A \mathbf{x} = \mathbf{x}^T 0 = 0. \]

Observe that the LHS may be written as

\[ \mathbf{x}^T A^T A \mathbf{x} = (A \mathbf{x})^T A \mathbf{x} = \langle A \mathbf{x}, A \mathbf{x} \rangle = \| A \mathbf{x} \|^2 \]

and the last term is 0 iff \( A \mathbf{x} = 0 \) by a property of norms. Therefore \( \mathbf{x} \in \text{nullsp}(A). \) So \( \text{nullsp}(A^T A) \subseteq \text{nullsp}(A). \)

(b) Show that

\[ \text{colsp}(A^T A) = \text{colsp}(A^T). \]

**Solution.** By a theorem in the lecture (relating consistency of linear systems to column space), \( A^T A \mathbf{x} = A^T b \) has a solution iff \( A^T b \in \text{colsp}(A^T A). \) Applying the rank-nullity theorem to \( A^T A, \) we get

\[ \text{rank}(A^T A) = n - \text{nullity}(A^T A). \]  \hspace{1cm} (3.2)

By (a),

\[ \text{nullity}(A^T A) = \text{nullity}(A). \]  \hspace{1cm} (3.3)

Applying the rank-nullity theorem to \( A \) yields

\[ \text{nullity}(A) = n - \text{rank}(A). \]  \hspace{1cm} (3.4)

By the equality of row rank and column rank, we have

\[ \text{rank}(A) = \text{rank}(A^T). \]  \hspace{1cm} (3.5)

Combining (3.2)–(3.5), we get

\[ \text{dim}(\text{colsp}(A^T A)) = \text{dim}(\text{colsp}(A^T)). \]  \hspace{1cm} (3.6)

Now observe that if \( \mathbf{y} \in \text{colsp}(A^T A), \) then \( \mathbf{y} = A^T \mathbf{Ax} \) for some \( \mathbf{x} \in \mathbb{R}^n \) and so \( \mathbf{y} = A^T \mathbf{w} \) for \( \mathbf{w} = \mathbf{Ax} \in \mathbb{R}^m. \) In other words,

\[ \text{colsp}(A^T A) \subseteq \text{colsp}(A^T). \]  \hspace{1cm} (3.7)
(3.6) and (3.7) together implies that
\[ \text{colsp}(A^T A) = \text{colsp}(A^T). \]

(c) Deduce that
\[ A^T A\mathbf{x} = A^T \mathbf{b} \]
always has a solution (even if \( A\mathbf{x} = \mathbf{b} \) has no solution).

SOLUTION. This follows from (b): Since \( A^T \mathbf{b} \in \text{colsp}(A^T) \), so \( A^T \mathbf{b} \in \text{colsp}(A^T A) \). Hence \( A^T A\mathbf{x} = A^T \mathbf{b} \) always have a solution.

(d) Show that (a), (b), and (c) are false in general over arbitrary fields, i.e. for \( A \in \mathbb{F}^{m \times n} \) and \( \mathbf{b} \in \mathbb{F}^m \).

SOLUTION. Let \( \mathbb{F} = \mathbb{Z}/2\mathbb{Z} \). Let
\[ A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \in (\mathbb{Z}/2\mathbb{Z})^{2 \times 2} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \in (\mathbb{Z}/2\mathbb{Z})^2. \]

Note that
\[ A^T A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}. \]

Hence
\[ \text{nullsp}(A^T A) = (\mathbb{Z}/2\mathbb{Z})^2 \neq \left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\} = \text{nullsp}(A) \]
and
\[ \text{colsp}(A^T A) = \left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right\} \neq \left\{ \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\} = \text{colsp}(A^T). \]

Also
\[ A^T A\mathbf{x} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \neq \begin{bmatrix} 1 \\ 1 \end{bmatrix} = A^T \mathbf{b} \]
for all \( \mathbf{x} \in (\mathbb{Z}/2\mathbb{Z})^2 \).

4. Let \( V \) be a vector space over \( \mathbb{R} \) and let \( \langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{R} \) be an inner product. Let \( S \) be a subset (not necessarily a subspace) of \( V \). We define the orthogonal annihilator of \( S \), denoted \( S^\perp \), to be the set
\[ S^\perp = \{ \mathbf{v} \in V \mid \langle \mathbf{v}, \mathbf{w} \rangle = 0 \text{ for all } \mathbf{w} \in S \}. \]

(a) Show that \( S^\perp \) is always a subspace of \( V \).

SOLUTION. Let \( \mathbf{v}_1, \mathbf{v}_2 \in S^\perp \). Then \( \langle \mathbf{v}_1, \mathbf{w} \rangle = 0 \) and \( \langle \mathbf{v}_2, \mathbf{w} \rangle = 0 \) for all \( \mathbf{w} \in S \). So for any \( \alpha, \beta \in \mathbb{R} \),
\[ \langle \alpha \mathbf{v}_1 + \beta \mathbf{v}_2, \mathbf{w} \rangle = \alpha \langle \mathbf{v}_1, \mathbf{w} \rangle + \beta \langle \mathbf{v}_2, \mathbf{w} \rangle = 0 \]
for all \( \mathbf{w} \in S \). Hence \( \alpha \mathbf{v}_1 + \beta \mathbf{v}_2 \in S^\perp \).

(b) Show that \( \text{span}(S) \subseteq (S^\perp)^\perp \).

SOLUTION. Let \( \mathbf{w} \in S \). For any \( \mathbf{v} \in S^\perp \), we have \( \langle \mathbf{v}, \mathbf{w} \rangle = 0 \) by definition of \( S^\perp \). Since this is true for all \( \mathbf{v} \in S^\perp \) and \( \langle \mathbf{w}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{w} \rangle \), we see that \( \langle \mathbf{w}, \mathbf{v} \rangle = 0 \) for all \( \mathbf{v} \in S^\perp \), i.e. \( \mathbf{w} \in S^\perp \). Hence \( S \subseteq (S^\perp)^\perp \). Since \( (S^\perp)^\perp \) is a subspace by (a) (it is the orthogonal annihilator of \( S^\perp \)), we have that \( \text{span}(S) \subseteq (S^\perp)^\perp \).

(c) Show that if \( S_1 \) and \( S_2 \) are subsets of \( V \) and \( S_1 \subseteq S_2 \), then \( S_2^\perp \subseteq S_1^\perp \).

SOLUTION. Let \( \mathbf{v} \in S_2^\perp \). Then \( \langle \mathbf{v}, \mathbf{w} \rangle = 0 \) for all \( \mathbf{w} \in S_2 \) and so for all \( \mathbf{w} \in S_1 \) (since \( S_1 \subseteq S_2 \)). Hence \( \mathbf{v} \in S_1^\perp \).

(d) Show that \( ((S^\perp)^\perp)^\perp = S^\perp \).

SOLUTION. Applying (c) to the inclusion \( S \subseteq (S^\perp)^\perp \) in the proof of (b), we get
\[ ((S^\perp)^\perp)^\perp \subseteq S^\perp. \]

Apply (b) to \( S^\perp \), we get
\[ \text{span}(S^\perp) \subseteq ((S^\perp)^\perp)^\perp. \]
But by (a), \( \text{span}(S^\perp) = S^\perp \). Hence we get equality.

(e) Show that either \( S \cap S^\perp \) must be either the empty set \( \emptyset \) or the zero subspace \( \{0_V\} \).

**Solution.** If \( S \cap S^\perp \neq \emptyset \), then let \( v \in S \cap S^\perp \). Since \( v \in S^\perp \), we have \( \langle v, w \rangle = 0 \) for all \( w \in S \). Since \( v \in S \), in particular, \( \|v\|^2 = \langle v, v \rangle = 0 \), implying that \( v = 0_V \). In other words, the only vector in \( S \cap S^\perp \) is \( 0_V \). So \( S \cap S^\perp = \{0_V\} \). On the other hand, if \( 0_V \notin S \), then \( 0_V \notin S \cap S^\perp \) and so \( S \cap S^\perp = \emptyset \) (if not, then the previous argument gives a contradiction).

(f) Show that if \( W \) is a subspace, then \( V = W \oplus W^\perp \). This is called the orthogonal complement of \( W \).

**Solution.** Since \( W \) and \( W^\perp \) are both subspaces (the latter follows from (a)), \( 0_V \in W \) and \( 0_V \in W^\perp \). So by (e), we have that \( W \cap W^\perp = \{0_V\} \). It remains to show that \( W + W^\perp = V \). Clearly \( W + W^\perp \subseteq V \). For the converse, let \( v \in V \) and let \( \{w_1, \ldots, w_r\} \) be an orthonormal basis of \( W \). Consider
\[
\mathbf{x} := \langle v, w_1 \rangle w_1 + \langle v, w_2 \rangle w_2 + \cdots + \langle v, w_r \rangle w_r.
\]
and
\[
\mathbf{y} := v - \mathbf{x}.
\]
Clearly \( \mathbf{x} \in W \). We claim that \( \mathbf{y} \in W^\perp \). For any \( w \in W \), we could write
\[
w = \langle w, w_1 \rangle w_1 + \langle w, w_2 \rangle w_2 + \cdots + \langle w, w_r \rangle w_r
\]
since \( \{w_1, \ldots, w_r\} \) is an orthonormal basis of \( W \). So
\[
\langle \mathbf{y}, w \rangle = \langle v - \mathbf{x}, w \rangle = \langle v, w \rangle - \langle \mathbf{x}, w \rangle = \langle v, \sum_{i=1}^{r} \langle w, w_i \rangle w_i \rangle - \left( \sum_{i=1}^{r} \langle \mathbf{x}, w_i \rangle w_i, w \right) = \sum_{i=1}^{r} \langle v, w_i \rangle \langle w, w_i \rangle - \sum_{i=1}^{r} \langle \mathbf{x}, w_i \rangle \langle w_i, w \rangle = 0
\]
since \( \langle w, w_i \rangle = \langle w_i, w \rangle \). Hence \( v = x + y \in W + W^\perp \).

(g) Show that if \( W \) is a subspace, then \( W = (W^\perp)^\perp \).

**Solution.** As in the proof of (b), we have \( W \subseteq (W^\perp)^\perp \). But by (f), we have
\[
W \oplus W^\perp = V = W^\perp \oplus (W^\perp)^\perp
\]
and so
\[
\dim(W) + \dim(W^\perp) = \dim(W^\perp) + \dim((W^\perp)^\perp)
\]
and so
\[
\dim(W) = \dim((W^\perp)^\perp).
\]
Hence \( W = (W^\perp)^\perp \).

5. Let \( (V, \langle \cdot, \cdot \rangle) \) be an inner product space over \( \mathbb{R} \). Let \( W \) be a non-trivial subspace of \( V \). By Problem 4(f), we have that
\[
V = W \oplus W^\perp.
\]
Note that by Homework 2, Problem 3, this means that every \( v \in V \) can be written uniquely as
\[
v = w + w',
\]
where \( w \in W \) and \( w' \in W^\perp \). We will define a function \( P : V \rightarrow V \) by \( P(v) = w \) for every \( v \in V \) according to the decomposition in (5.8). \( P \) is called the **orthogonal projection** onto \( W \).
(a) Show that \( P \in \text{End}(V) \).

**Solution.** Let \( v_1, v_2 \in V \) and let \( v_i = w_i + w'_i \) where \( w_i \in W \) and \( w'_i \in W^\perp \), \( i = 1, 2 \). Let \( \alpha_1, \alpha_2 \in \mathbb{R} \). Then

\[
\alpha_1 v_1 + \alpha_2 v_2 = (\alpha_1 w_1 + \alpha_2 w_2) + (\alpha_1 w'_1 + \alpha_2 w'_2)
\]

where \( \alpha_1 w_1 + \alpha_2 w_2 \in W \) and \( \alpha_1 w'_1 + \alpha_2 w'_2 \in W^\perp \). So

\[
P(\alpha_1 v_1 + \alpha_2 v_2) = P((\alpha_1 w_1 + \alpha_2 w_2) + (\alpha_1 w'_1 + \alpha_2 w'_2))
\]

\[
= \alpha_1 w_1 + \alpha_2 w_2
\]

\[
= \alpha_1 P(v_1) + \alpha_2 P(v_2).
\]

Hence \( P \) is linear.

(b) Show that \( P \) is idempotent, i.e. \( P^2 = P \).

**Solution.** Let \( v \in V \) and let \( v = w + w' \) where \( w \in W \) and \( w' \in W^\perp \). Then

\[
P^2(v) = P(P(v)) = P(w) = w = P(v).
\]

Since this holds for every \( v \in V \), \( P^2 = P \).

(c) Show that \( \langle P(v), v \rangle \geq 0 \) for every \( v \in V \).

**Solution.** Let \( v \in V \) and let \( v = w + w' \) where \( w \in W \) and \( w' \in W^\perp \). By the definition of \( W^\perp \), \( \langle w, w' \rangle = 0 \). Hence,

\[
\langle P(v), v \rangle = \langle w, w + w' \rangle = \langle w, w \rangle + \langle w, w' \rangle = \|w\|^2 \geq 0.
\]

(d) Show that \( \|P(v)\| \leq \|v\| \) for every \( v \in V \).

**Solution.** This follows from (f).

\[
\|v\|^2 = \|P(v)\|^2 + \|(I - P)(v)\|^2 \geq \|P(v)\|^2
\]

since \( \|(I - P)(v)\|^2 \geq 0 \). Taking positive square root yields \( \|v\| \geq \|P(v)\| \).

(e) Show that \( I - P \) is the orthogonal projection onto \( W^\perp \).

**Solution.** By Problem 4(g), \( (W^\perp)^\perp = W \). So

\[
V = W^\perp \oplus (W^\perp)^\perp = W^\perp \oplus W.
\]

If \( v = w + w' \) where \( w \in W = (W^\perp)^\perp \) and \( w' \in W^\perp \), then

\[
w' = v - w = I(v) - P(v) = (I - P)(v).
\]

So \( I - P \) is the projection onto \( W^\perp \).

(f) Show that

\[
\|v\|^2 = \|P(v)\|^2 + \|(I - P)(v)\|^2.
\]

**Solution.** Let \( v \in V \) and let \( v = w + w' \) where \( w \in W \) and \( w' \in W^\perp \). Note that \( \langle w, w' \rangle = 0 \). So by Pythagoras theorem,

\[
\|v\|^2 = \|w + w'\|^2
\]

\[
= \|w\|^2 + \|w'\|^2
\]

\[
= \|P(v)\|^2 + \|(I - P)(v)\|^2.
\]
