Linear Algebra Done Right

Jiuru Lyu

June 18, 2025

Contents

1 Vector Spaces

1.1 $\mathbb{R}^n$ and $\mathbb{C}^n$

Definition 1.1.1 (Complex Number). A *complex number* is an ordered pair (a, b), where $a, b \in \mathbb{R}$, but we write it as $a + b\mathrm{i}$.
Notation 1.1.2. $\mathbb{C} \coloneqq \{a + b\mathrm{i} \mid a, b \in \mathbb{R}\}$
Definition 1.1.3 (Addition & Multiplication).

$$(a + b\mathrm{i}) + (c + d\mathrm{i}) = (a + c) + (b + d)]\mathrm{i}$$

$$(a + b\mathrm{i})(c + d\mathrm{i}) = (ac - bd) + (ad + bc)\mathrm{i}$$

Theorem 1.1.4 Properties of Complex Arithmetic

1. commutativity: $\alpha + \beta = \beta + \alpha; \quad \alpha\beta = \beta\alpha, \quad \forall \alpha, \beta \in \mathbb{C}$.
2. associativity: $(\alpha + \beta) + \lambda = \alpha + (\beta + \lambda); \quad (\alpha\beta)\lambda = \alpha(\beta\lambda), \quad \forall \alpha, \beta, \lambda \in \mathbb{C}$.
3. identities: $\lambda + 0 = \lambda; \quad \lambda \cdot 1 = \lambda, \forall \lambda \in \mathbb{C}$.
4. additive inverse: $\forall \alpha \in \mathbb{C}, \exists$ unique $\beta \in \mathbb{C}$ *s.t.* $\alpha + \beta = 0$.
5. multiplicative inverse: $\forall \alpha \in \mathbb{C}, \alpha \neq 0, \exists$ unique $\beta \in \mathbb{C}$ *s.t.* $\alpha\beta = 1$.
6. distributivity: $\lambda(\alpha + \beta) = \lambda\alpha + \lambda\beta, \quad \forall \lambda, \alpha, \beta \in \mathbb{C}$.

Definition 1.1.5 (Subtraction). If $-\alpha$ is the additive inverse of α, *subtraction* on $\mathbb{C}$ is defined by

$$\beta - \alpha = \beta + (-\alpha).$$

Definition 1.1.6 (Division). For $\alpha \neq 0$, let $\dfrac{1}{\alpha}$ denote the multiplicative inverse of α. Then, *division* on $\mathbb{C}$ is defined by

$$\frac{\beta}{\alpha} = \beta \cdot \left(\frac{1}{\alpha}\right)$$

Notation 1.1.7. $\mathbb{F}$ is either $\mathbb{R}$ or $\mathbb{C}$.
Definition 1.1.8 (List/Tuple). Suppose n is a non-negative integer. A list of length n is an ordered collection of n elements separated by commas and surrounded by parentheses: $(x_1, x_2, x_3, \cdots, x_n)$. Two lists are equal if and only if they have the same length and the same elements in the same order.

Remark. *Lists must have a FINITE length.*

Definition 1.1.9 ($\mathbb{F}^n$ and Coordinate). $\mathbb{F}^n$ is the set of all lists of length n of elements of $\mathbb{F}$:

$$\mathbb{F}^n \coloneqq \{(x_1, \cdots, x_n) \mid x_i \in \mathbb{R} \forall i = 1, \cdots, n\},$$

where x_i is the i^{th} *coordinate* of $(x_1, \cdots, x_n)$.

Example 1.1.10 $\mathbb{R}^2 = \{(x, y) \mid x, y \in \mathbb{R}\}$ and $\mathbb{R}^3 = \{(x, y, z) \mid x, y, z \in \mathbb{R}\}$.

Definition 1.1.11 (Addition on $\mathbb{F}^n$). *Addition* on $\mathbb{F}^n$ is defined by adding corresponding coordinates:

$$(x_1, \cdots, x_n) + (y_1, \cdots, y_n) = (x_1 + y_1, \cdots, x_n + y_n).$$

Theorem 1.1.12 Commutativity of Addition on $\mathbb{F}^n$
If $x, y \in \mathbb{F}^n$, then $x + y = y + x$.

Proof 1. Suppose $x = (x_1, \cdots, x_n)$ and $y = (y_1, \cdots, y_n)$. Then

$$\begin{aligned} x + y &= (x_1 + y_1, \cdots, x_n + y_n) \\ &= (y_1 + x_1, \cdots, y_n + x_n) = y + x. \end{aligned}$$

■

Definition 1.1.13 (Zero). Let 0 denote the list of length n whose coordinates are all 0: $0 \coloneqq (0, \cdots, 0)$.
Definition 1.1.14 (Additive Inverse on $\mathbb{F}^n$). For $x \in \mathbb{F}^n$, the additive inverse of x, denoted $-x$, is the vector $-x \in \mathbb{F}^n$ *s.t.* $x + (-x) = 0$.
Definition 1.1.15 (Scalar Multiplication in $\mathbb{F}^n$). The product of a number $\lambda \in \mathbb{F}$ and a vector $x \in \mathbb{F}^n$ is computed by multiplying each coordinate of the vector by λ :

$$\lambda x = \lambda(x_1, \cdots, x_n) = (\lambda x_1, \cdots, \lambda x_n),$$

where $x = (x_1, \cdots, x_n) \in \mathbb{F}^n$.

Theorem 1.1.16 Properties of Arithmetic Operations on $\mathbb{F}^n$

1. $(x + y) + z = x + (y + z) \quad \forall x, y, z \in \mathbb{F}^n$
2. $(ab)x = a(bx) \quad \forall x \in \mathbb{F}^n$ and $\forall a, b \in \mathbb{F}$.
3. $1 \cdot x = x \quad \forall x \in \mathbb{F}^n$ and $1 \in \mathbb{F}$.
4. $\lambda(x + y) = \lambda x + \lambda y \quad \forall \lambda \in \mathbb{R}$ and $\forall x, y \in \mathbb{F}^n$.
5. $(a + b)x = ax + bx \quad \forall a, b \in \mathbb{F}$ and $\forall x \in \mathbb{F}^n$.

1.2 Definition of Vector Space

Definition 1.2.1 (Addition on V). An *addition* on V is a function $(u, v) \mapsto u + v$ for all $u, v \in V$.
Definition 1.2.2 (Scalar Multiplication on V). A *scalar multiplication* on V is a function $(\lambda, v) \mapsto \lambda v$ for all $\lambda \in \mathbb{F}$ and $v \in V$.
Definition 1.2.3 (Vector Space). A *vector space* is a set V along with an addition on V and a scalar multiplication *s.t.* the following properties hold:

1. commutativity: $u + v = v + u \quad \forall u, v \in V$
2. associativity: $(u + v) + w = u + (v + w)$ and $(ab)v = a(bv) \quad \forall u, v, w \in V$ and $\forall a, b \in \mathbb{F}$.
3. additive identity: $\exists 0 \in V$ *s.t.* $v + 0 = v \quad \forall v \in V$.
4. additive inverse: $\exists w \in V$ *s.t.* $v + w = 0 \quad \forall v \in V$.
5. multiplicative identity: $\exists 1 \in V$ *s.t.* $1 \cdot v = v \quad \forall v \in V$.
6. distributive properties: $a(u + v) = au + av$ and $(a + b)v = av + bv \quad \forall u, v \in V$ and $a, b \in \mathbb{F}$.

Definition 1.2.4 (Vector). Elements of a vector space are called *vectors* or points.
Notation 1.2.5. V is a vector space over $\mathbb{F}$.
Definition 1.2.6 (Real and Complex Vector Space). A vector space over $\mathbb{R}$ is called a *real vector space*, and a vector space over $\mathbb{C}$ is called a *complex vector space*.

Theorem 1.2.7 Unique Additive Identity of Vector Spaces
A vector space has a unique additive identity.

Proof 1. Suppose 0 and $0'$ are both additive identities for some vector space V. So,

$$\begin{aligned} 0' &= 0' + 0 && \textit{Since } 0 \textit{ is an additive identity} \\ &= 0 + 0' && \textit{commutativity} \\ &= 0. && \textit{Since } 0' \textit{ is an additive identity} \end{aligned}$$

Then, $0' = 0$. ∎

Theorem 1.2.8 Unique Additive Inverse of Vector Spaces
A vector in a vector space has a unique additive inverse.

Proof 2. Let V be a vector space. Suppose w and w' are additive inverses of v for some $v \in V$. Note that

$$\begin{aligned} w &= w + 0 \\ &= w + (v + w') \\ &= (w + v) + w \\ &= 0 + w' = w'. \end{aligned}$$

∎

Notation 1.2.9. Let $v, w \in V$. Then, $-v$ denotes the additive inverse of v.
Definition 1.2.10 (Subtraction). $w - v$ is defined to be $w + (-v)$.

Theorem 1.2.11
$0 \cdot v = 0 \quad \forall v \in V.$

Proof 3. Since $v \in V$, we know

$$\begin{aligned} 0 \cdot v = (0+0)v = 0 \cdot v + 0 \cdot v \\ 0 \cdot v + (-0 \cdot v) = 0 \cdot +0 \cdot +(-0 \cdot v) \\ 0 = 0 \cdot v \end{aligned}$$

■

Theorem 1.2.12
$a \cdot 0 = 0 \quad \forall a \in \mathbb{F}.$

Proof 4. For $a \in \mathbb{F}$, we have

$$\begin{aligned} a \cdot 0 = a \cdot (0+0) = a \cdot 0 + a \cdot 0 \\ a \cdot 0 + (-a \cdot 0) = a \cdot 0 + a \cdot 0 + (-a \cdot 0) \\ 0 = a \cdot 0. \end{aligned}$$

■

Theorem 1.2.13
$(-1)v = -v \quad \forall v \in V.$

Proof 5. For $v \in V$, we have

$$v + (-1)v = 1 \cdot v + (-1) \cdot v = (1 + (-1)) \cdot v = 0 \cdot v = 0.$$

Therefore, by definition, $(-1)v = -v$. ■

Notation 1.2.14. $\mathbb{F}^S$

1. If S is a set, then $\mathbb{F}^S$ denotes the set of functions from S to $\mathbb{F}$.
2. For $f, g \in \mathbb{F}^S$, the sum $f + g \in \mathbb{F}^S$ is the function defined by $(f+g)(x) = f(x) + g(x) \quad \forall x \in S$.
3. For $\lambda \in \mathbb{F}$ and $f \in \mathbb{F}^S$, the product $\lambda f \in \mathbb{F}^S$ is the function defined by $(\lambda f)(x) = \lambda f(x) \quad \forall x \in S$.

Theorem 1.2.15
$\mathbb{F}^S$ is a vector space.

1.3 Subspace

Definition 1.3.1 (Subspace). A subset U of V is called a *subspace* of V if U is also a vector space using the same addition and scalar multiplication as on V.

Theorem 1.3.2 Conditions for a Subspace
A subset U of V is a subspace of V if and only if U satisfies the following conditions:

1. additive identity: $0 \in U$;
2. closed under addition: $u, w \in U \implies u + w \in U$;
3. closed under scalar multiplication: $a \in \mathbb{F}$ and $u \in U \implies au \in U$.

Proof 1.
($\Rightarrow$) Suppose U is a subspace of V. By definition, U is then a vector space, and so those conditions are automatically satisfied. □
($\Leftarrow$) Suppose U satisfies the three conditions. Since U is a subset of V, U automatically has *associativity*, *commutativity*, *multiplicative identity*, and *distributivity*. So, we want to check U has additive inverse and additive identities.
For additive identity, we know $0 \in U$, by assumption.
For additive inverse, by condition #3, we know $-u = (-1)u \in U$.
Then, U is a vector space. ■

Example 1.3.3 If $b \in \mathbb{F}$, then $\{(x_1, x_2, x_3, x_4) \in \mathbb{F}^4 \mid x_3 = 5x_4 + b\}$ is a subspace of $\mathbb{F}^4$ if and only if $b = 0$.
Proof 2.
($\Rightarrow$) Suppose $U = \{(x_1, x_2, x_3, x_4) \in \mathbb{F}^4 \mid x_3 = 5x_4 + b\}$ is a subspace of $\mathbb{F}^4$. Then, $0 = (0, 0, 0, 0) \in U$. So, $0 = 5 \cdot 0 + b$, or $b = 0$. □
($\Leftarrow$) Suppose $b = 0$. Then, $x_3 = 5x_4$. So, $U = \{(x_1, x_2, 5x_4, x_4) \in \mathbb{F}^4\}$

1. $0 = (0, 0, 0, 0) \in U$
2. Note that
$$(x_1, x_2, 5x_4, x_4) + (y_1, y_2, 5y_4, y_4) = (x_1 + y_1, x_2 + y_2, 5(x_4 + y_4), x_4 + y_4) \in U$$
So, addition is closed under U.
3. $\forall a \in \mathbb{F}$, we have
$$a(x_1, x_2, 5x_4, x_4) = (ax_1, ax_2, 5(ax_4), ax_4) \in U$$
Then, U is a subspace of $\mathbb{F}^4$.

■

Example 1.3.4 The set of continuous real-valued functions on interval $[0, 1]$ is a subspace of $\mathbb{R}^{[0,1]}$.
Proof 3.

1. 0 (zero mapping)$\in U$
2. Set f and $g \in \mathcal{C}[0,1]$, the set of continuous functions on interval $[0,1]$. Then, $f+g \in \mathcal{C}[0,1]$.
3. From Calculus, we know that $\forall a \in \mathbb{F}, \quad af \in \mathcal{C}[0,1]$.

■

Definition 1.3.5 (Sum of Subspaces). Suppose $U_1, \cdots, U_m$ are subspaces of V. The *sum* of $U_1, \cdots, U_m$, denoted as $U_1 + \cdots + U_m$, is the set of all possible sums of elements of $U_1, \cdots, U_m$:

$$U_1 + \cdots + U_m = \{u_1 + \cdots + u_m \mid u_i \in U_i \quad \forall i = 1, \cdots, m\}.$$

Example 1.3.6 Suppose $U = \{(x,0,0) \in \mathbb{F}^3 \mid x \in \mathbb{F}\}$ and $W = \{(0,y,0) \in \mathbb{F}^3 \mid y \in \mathbb{F}\}$, then

$$U + W = \{(x,y,0) \in \mathbb{F}^3 \mid x, y \in \mathbb{F}\}.$$

Theorem 1.3.7
Suppose $U_1, \cdots, U_m$ are subspaces of V. Then, $U_1 + \cdots + U_m$ is the *smallest subspace* of V containing $U_1, \cdots, U_m$.

Proof 4. Suppose $U_1, \cdots, U_m$ are subspaces of U. Let $U_1 + \cdots + U_m = \{u_1 + \cdots + u_m \mid u_j \in U_j, j = 1, \cdots m\}$. Suppose $w_j \in U_j$, then $w_1 + \cdots + w_m \in U_1 + \cdots + U_m$.

1. $U_1 + \cdots + U_m$ is a subspace of V.
   (a) Note that
   $$(u_1 + \cdots + u_m) + (w_1 + \cdots + w_m) = (u_1 + w_1) + \cdots + (u_m + w_m) \in U_1 + \cdots + U_m,$$
   so $U_1 + \cdots + U_m$ is closed under addition.
   (b) Similarly, $U_1 + \cdots + U_m$ is closed under scalar multiplication.
   (c) Note that U_j is a subspace, so $0 \in U_j$. Hence, $(0, \cdots, 0) = 0 \in U_1 + \cdots + U_m$. □
2. Now, we want to show this subspace is the smallest subspace containing $U_1, \cdots, U_m$. That is, we want to show $\forall\, W \supseteq U_1 \cup \cdots \cup U_m$, we have $W \supseteq U_1 + \cdots + U_m$.

   Note that $U_j \subseteq U_1 + \cdots + U_m$, so we have $(U_1 \cup U_2 \cup \cdots \cup U_m) \subseteq U_1 + \cdots + U_m$. This means $U_1 + \cdots + U_m$ must contain $U_1, \cdots, U_m$. Let W be some subspace containing $U_1, \cdots, U_m$. Then, for $j = 1, \cdots, m$, we have $u_j \in U_j$, which indicates $u_j \in W$. Therefore, $u_1 + \cdots + u_m \in V$ and thus $U_1 + \cdots + U_m \subseteq W$.

   Since W was arbitrary, we've shown $\forall\, W$ that contains $U_1, \cdots, U_m$, $U_1 + \cdots + U_m \subseteq W$. Therefore, $U_1 + \cdots + U_m$ is the smallest.

■

Definition 1.3.8 (Direct Sum). Suppose $U_1, \cdots, U_m$ are subspaces of V. $U_1 + \cdots + U_m$ is called a *direct sum* if each element of $U_1 + \cdots + U_m$ can be written in only one way as a sum $u_1 + \cdots + u_m$, where $u_j \in U_j$.

Notation 1.3.9. If $U_1 + \cdots + U_m$ is a direct sum, then we use $U_1 \oplus \cdots \oplus U_m$ to denote it.

Example 1.3.10 Let $U = \{(x, y, 0) \in \mathbb{F}^3 \mid x, y \in \mathbb{F}\}$ and $W = \{(0, 0, z) \in \mathbb{F}^3 \mid z \in \mathbb{F}\}$. Then, $\mathbb{F}^3 = U \oplus W$.

Proof 5. Note that $U + W = \{(x, y, z) \mid x, y, z \in \mathbb{F}\} = \mathbb{F}^3$. Suppose

$$(x, y, z) = (x, y, 0) + (0, 0, z), \tag{1}$$

for some $x, y, z \in \mathbb{F}$ and

$$(x, y, z) = (x', y', 0) + (0, 0, z') \tag{2}$$

for some $x', y', z' \in \mathbb{F}$. Then, (1)$-$(2):

$$(0, 0, 0) = (x - x', y - y', 0) + (0, 0, z - z') = (x - x', y - y', z - z').$$

Then, $x - x' = y - y' = z - z' = 0$, which indicates $x = x',\ y = y',\ z = z'$. So, by definition $U + W$ is a direct sum, or $\mathbb{F}^3 = U \oplus W$. ∎

Example 1.3.11 Suppose U_j is the subspace of $\mathbb{F}^n$ *s.t.*

$$\begin{aligned}
U_1 &= \{x, 0, 0, \cdots, 0 \mid x \in \mathbb{F}\} \\
U_2 &= \{0, x, 0, \cdots, 0 \mid x \in \mathbb{F}\} \\
&\vdots \\
U_n &= \{0, 0, 0, \cdots, x \mid x \in \mathbb{F}\}
\end{aligned}$$

Then, $\mathbb{F}^n = U_1 \oplus U_2 \oplus \cdots \oplus U_n$.

Proof 6. Note that $\mathbb{F}^n = U_1 + U_2 + \cdots + U_n$ is evident. Now, we'll prove that $U_1 + U_2 + \cdots + U_n$ is a direct sum. Consider $x = (x_1, x_2, \cdots, x_n) \in \mathbb{F}^n$. Assume that

$$x = (x_1, 0, \cdots, 0) + \cdots + (0, \cdots, 0, x_n) \tag{3}$$

and

$$x = (x_1', 0, \cdots, 0) + \cdots + (0, \cdots, 0, x_n') \tag{4}$$

Then, from (3)-(4), we know that

$$0 = (x_1 - x_1', \cdots, x_n - x_n') = (0, 0, \cdots, 0).$$

Then, $\forall i = 1, \cdots, n$ we have $x_i - x_i' = 0$, or $x_i = x_i'$. Therefore, by definition, we know $U_1 + \cdots + U_n$ is a direct sum. ∎

Example 1.3.12 Let

$$U_1 = \{(x, y, 0) \mid x, y \in \mathbb{F}\}$$
$$U_2 = \{(0, 0, z) \mid z \in \mathbb{F}\}$$
$$U_3 = \{(0, y, y) \mid y \in \mathbb{F}\}$$

Show that $U_1 + U_2 + U_3$ is not a direct sum.

Proof 7. Consider $(0, 0, 0) \in \mathbb{F}^3$. Note that

$$(0, 0, 0) = (0, 0, 0) + (0, 0, 0) + (0, 0, 0)$$

and

$$(0, 0, 0) = (0, 1, 0) + (0, 0, 1) + (0, -1, -1).$$

Then, $U_1 + U_2 + U_3$ is not a direct sum by definition. ■

Theorem 1.3.13
Suppose $U_1, \cdots, U_m$ are subspaces of V. Then,$U_1 + \cdots + U_m$ is a direct sum if and only if the only way to write 0 as a sum $u_1 + \cdots + u_m$ is by taking each $u_j = 0$.

Proof 8.
($\Rightarrow$) Since $U_1 + \cdots + U_m$ is a direct sum, by definition, the only way to write $0 \in \mathbb{F}^n$ is to write it as

$$0 = 0 + \cdots + 0 \qquad \text{where } 0 \in U_i \forall i = 1, \cdots, m. \qquad \square$$

($\Leftarrow$) Suppose the only way to write 0 as a sum $u_1 + \cdots + u_m$ is by taking each $u_j = 0$. Assume that for some $v \in V$, we have

$$v = u_1 + \cdots + u_m, \quad u_j \in U_j \tag{5}$$

and

$$v = u_1' + \cdots + u_m', \quad u_j' \in U_j. \tag{6}$$

Then, by (5)-(6), and according to the conclusion from Example 1.3.11, we have

$$0 = (u_1 - u_1') + \cdots + (u_m - u_m') = 0 + \cdots + 0.$$

So, $\forall i \in 1, \cdots, m$, we have $u_i - u_i' = 0$. that is, $u_i = u_i'$. So, $\forall v \in V$, there is only one way to write v as a sum of $u_1 + \cdots_u + m$. Therefore, by definition, $U_1 + \cdots + U_m$ is a direct sum. ■

Theorem 1.3.14
Suppose U amd W are subspaces of V. Then, $U + W$ is a direct sum if and only if $U \cap W = \{0\}$.

Proof 9.
($\Rightarrow$) Suppose $U + W$ is a direct sum. Assume $v \in U \cap W$. Then, $v \in U$ and $v \in W$. By definition of subspace, we know $-v \in W$ as well. Note that

$$0 = v + (-v) \in U \cap W.$$

Then, by Theorem 1.3.13, we know that the only representation of $0 \in U \cap W$ is $0 = 0 + 0$ since $U \cap W$ is a direct sum. Hence, it must be that $v = -v = 0$, and thus $U \cap W = \{0\}$. □

($\Leftarrow$) Suppose $U \cap W = \{0\}$. Let $u \in U$ and $w \in W$ *s.t.* $u + w = 0$. Then, we have $u = -w$. Since $-w \in W$, we know $u = -w \in W$. By $u \in U$ and $u \in W$, we know that $u \in U \cap W = \{0\}$. Therefore, $0 = 0 + 0$ is the only to represent $0 \in U + W$. By Theorem 1.3.13, we know $U + W$ is a direct sum. ■

Remark. *When extending Theorem 1.3.14 to 3 subspaces* U_1, U_2, U_3, *we cannot conclude* $U_1 \oplus U_2 \oplus U_3$ *if we have* $U_1 \cap U_2 = U_1 \cap U_3 = U_2 \cap U_3 = \{0\}$. *See Example 1.3.12 as a counterexample.*

2 Finite-Dimensional Vector Spaces

2.1 Span and Linear Independence

Notation 2.1.1. We usually write list of vectors without using parentheses.

Example 2.1.2 $(4, 1, 6), (9, 5, 7)$ is a list of vectors of length 2 in $\mathbb{R}^3$.

Definition 2.1.3 (Linear Combination). A *linear combination* of a list $v_1, \cdots, v_m$ of vectors in V is a vector of the form

$$a_1 v_1 + \cdots + a_m v_m,$$

where $a_1, \cdots, a_m \in \mathbb{F}$.

Example 2.1.4 Since $(17, -4, 2) = 6(2, 1, -3) + 5(1, -2, 4)$, we say $(17, -4, 2)$ is a linear combination of $(2, 1, -3), (1, -2, 4)$.

Definition 2.1.5 (Span).

$$\operatorname{span}(v_1, \cdots, v_m) = \{a_1 v_1 + \cdots + a_m v_m \mid a_1 \cdots a_m \in \mathbb{F}\}.$$

Example 2.1.6 Consider $\operatorname{span}(e_1, e_2, e_3)$:

$$\begin{aligned}\operatorname{span}(e_1, e_2, e_3) &= \{a_1 e_1 + a_2 e_2 + a_3 e_3 \mid a_1, a_2, a_3 \in \mathbb{F}\} \\ &= \{(a_1, a_2, a_3) \mid a_1, a_2, a_3 \in \mathbb{F}\} = \mathbb{R}^3.\end{aligned}$$

Theorem 2.1.7
The span of a list of vectors in V is the smallest subspace of V containing all the vectors in the list.

Proof 1. To prove this theorem, we will prove two parts: span is a subspace and span is the smallest subspace.

1. Span is a subspace of V.
   (a) By definition of span, we know $\operatorname{span}(v_1, \cdots, v_m) = \{a_1 v_1 + \cdots + a_m v_m \mid a_1, \cdots, a_m \in \mathbb{F}\}$. If we set $a_1, \cdots, a_m = 0$, then we have $0 = 0v_1 + \cdots + 0v_m$. So, $0 \in \operatorname{span} v_1, \cdots, v_m$.
   (b) Let $a_1 v_1 + \cdots + a_m v_m \in \operatorname{span}(v_1, \cdots, v_m)$ and $b_1 v_1 + \cdots + b_m v_m \in \operatorname{span}(v_1, \cdots, v_m)$. Then,

   $$(a_1 v_1 + \cdots + a_m v_m) + (b_1 v_1 + \cdots + b_m v_m) = (a_1 + b_1)v_1 + \cdots + (a_m + b_m)v_m.$$

   Since $(a_1 + b_1), \cdots, (a_m + b_m) \in \mathbb{F}$, we know $(a_1 + b_1)v_1 + \cdots + (a_m + b_m)v_m \in \operatorname{span}(v_1, \cdots, v_m)$.
   (c) Let $\lambda \in \mathbb{F}$ and $a_1 v_1 + \cdots + a_m v_m \in \operatorname{span}(v_1, \cdots, v_m)$. Then,

   $$\lambda(a_1 v_1 + \cdots + a_m v_m) = \lambda a_1 v_1 + \cdots + \lambda a_m v_m.$$

Since $\lambda a_1, \cdots, \lambda a_m \in \mathbb{F}$, we know that $\lambda(a_1 v_1 + \cdots + a_m v_m) \in \operatorname{span}(v_1, \cdots, v_m)$.

Therefore, we have proven that span is a subspace of V. $\square$

2. Now, we want to show that span is the smallest subspace.

   Let U be a subspace of V containing $v_1, \cdots, v_m$. If we can show that $\operatorname{span}(v_1, \cdots, v_m) \subseteq U$, we then know span is the smallest subspace containing $v_1, \cdots, v_m$. Since U is a subspace containing $v_1, \cdots, v_m$, it is closed under addition and scalar multiplication. So, $a_1 v_1 + \cdots + a_m v_m \in \operatorname{span}(v_1, \cdots, v_m)$. Therefore, $\operatorname{span}(v_1, \cdots, v_m) \subseteq U$.

$\blacksquare$

Definition 2.1.8 (Span as a Verb). If $\operatorname{span}(v_1, \cdots, v_m) = V$, we say $v_1, \cdots, v_m$ *spans* V.
Definition 2.1.9 (Finite-Dimensional Vector Space). A vector space V is called *finite-dimensional* if $\exists$ a list of vectors, say $v_1, \cdots, v_m$ *s.t.* $\operatorname{span}(v_1, \cdots, v_m) = V$. In the following of this notes, we will use *f-d* as a shortcut for saying "finite-dimensional."
Definition 2.1.10 (Infinte-Dimensional Vector Space). A vector space V is infinite-dimensional if it is not *f-d*. This is equivalent to say that $\forall$ lists of vectors in V, they do not span V.
Definition 2.1.11 (Polynomial Functions). A function $p : \mathbb{F} \to \mathbb{F}$ is called a *polynomial* with coefficients in $\mathbb{F}$ if $\exists\, a_0, \cdots, a_m \in \mathbb{F}$ *s.t.* $p(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_m z^m \quad \forall z \in \mathbb{F}$.
Notation 2.1.12. We use $\mathcal{P}(\mathbb{F})$ to denote the set of all polynomial with coefficients in $\mathbb{F}$.

Theorem 2.1.13
$\mathcal{P}(\mathbb{F})$ is a vector space over $\mathbb{F}$.

Proof 2. Recall the definition of $\mathbb{F}^{\mathbb{F}}$. We will show $\mathcal{P}(\mathbb{F})$ is a subspace of $\mathbb{F}^{\mathbb{F}}$.

1. $0 = 0 + 0z + \cdots + 0z^m \in \mathcal{P}(\mathbb{F})$.
2. Suppose $p(z) = a_m z^m + \cdots + a_1 z + a_0$ and $q(z) = b_n z^n + \cdots + b_1 z + b_0 \in \mathcal{P}(\mathbb{F})$. WLOG, suppose $m > n$, then we have $p(z) + q(z) = a_m z^m + \cdots + (a_n + b_n) z^n + \cdots + (a_0 + b_0) \in \mathcal{P}(\mathbb{F})$.
3. Suppose $\lambda \in \mathbb{F}$. Then, $\lambda p(z) = \lambda(a_m z^m + \cdots + a_1 z + a_0) = \lambda a_m z^m + \cdots + \lambda a_0 \in \mathcal{P}(\mathbb{F})$.

Hence, we've shown $\mathcal{P}(\mathbb{F})$ is a subspace over $\mathbb{F}$. $\blacksquare$
Definition 2.1.14 (Degree of a Polynomial). A polynomial $p \in \mathcal{P}(\mathbb{F})$ is said to have *degree* m if $\exists$ scalars $a_0, \cdots, a_m \in \mathbb{F}$ with $a_m \neq 0$ *s.t.* $p(z) = a_m z^m + \cdots + a_1 z + a_0 \quad \forall z \in \mathbb{F}$. We write $\deg p = m$. Specially, $\deg 0 := -\infty$ and $\deg a_0 := 0$ when $a_0 \neq 0$.
Definition 2.1.15 ($\mathcal{P}_m(\mathbb{F})$). For $m \in \mathbb{N}^+$, $\mathcal{P}_m(\mathbb{F})$ denotes the set of all polynomial with coefficients in $\mathbb{F}$ and degree$\leq m$. i.e.,

$$\mathcal{P}_m(\mathbb{F}) := \{p \in \mathcal{P}(\mathbb{F}) \mid \deg p \leq m\}.$$

Example 2.1.16 For each $m \in \mathbb{N}$, $\mathcal{P}_m(\mathbb{F})$ is a *f-d* vector space.
Proof 3. Note that $\mathcal{P}_m(\mathbb{F})$ is a vector space because it is a subspace of $\mathcal{P}(\mathbb{F})$. Suppose $p(z) \in \mathcal{P}_m(\mathbb{F})$, then $p(z) = a_0 + a_1 z + \cdots + a_m z^m \in \operatorname{span}(1, z, \cdots, z^m)$. Then, by definition, $\mathcal{P}_m(\mathbb{F})$ is *f-d*. $\blacksquare$

Remark. *In this proof, we are abusing notation by letting z^k to denote a function.*

Example 2.1.17 $\mathcal{P}(\mathbb{F})$ is infinite-dimensional.

Proof 4. For any list of vectors in $\mathcal{P}(\mathbb{F})$, by definition of list, the length of it is finite. Suppose the highest degree in this list is m. Consider a polynomial with degree of $m+1$: z^{m+1}. Since z^{m+1} cannot be written as linear combinations of the list of polynomials, we know the list does not span $\mathcal{P}(\mathbb{F})$. So, $\mathcal{P}(\mathbb{F})$ is infinite-dimensional. ■

Definition 2.1.18 (Linear Independence). A list $v_1, \cdots, v_m$ of vectors in V is called *linearly independent* (L.I.) if the only choice of $a_1, \cdots, a_m \in \mathbb{F}$ that makes $a_1v_1+\cdots+a_mv_m = 0$ is $a_1 = \cdots = a_m = 0$. Specially, the empty list () is declared to be L.I..

Definition 2.1.19 (Linear Dependence). $v_1, \cdots, v_m$ is called *linearly dependent* if it is not L.I.. Or, equivalently, $v_1, \cdots, v_m$ is *linearly dependent* if $\exists\, a_1, \cdots, a_m \in \mathbb{F}$ not all 0 *s.t.* $\sum_{i=0}^{m} a_iv_i = 0$.

Example 2.1.20 Let $v_1, \cdots, v_m \in V$. If v_j is a linear combination of other v's, then $v_1, \cdots, v_m$ is linearly dependent.

Proof 5. By assumption, $v_j = a_1v_1 + \cdots + a_{j-1}v_{j-1} + a_{j+1}v_{j+a} + \cdots + a_mv_m$ for some a_i not all 0. So, $0 = a_1v_1 + \cdots + a_{j-1}v_{j-1} + a_{j+1}v_{j+1} + \cdots + a_mv_m - v_j$, a linear combination of $v_1, \cdots, v_m$. Since $-v_j$ has a coefficient of $-1 \neq 0$, by definition, $v_1, \cdots, v_m$ is not L.I.. ■

Lemma 2.1.21 Linear Dependence Lemma Suppose $v_1, \cdots, v_m$ is a linearly dependent list in V. Then, $\exists\, j \in \{1, \cdots, m\}$ *s.t.* the following hold:

1. $v_j \in \operatorname{span}(v_1, \cdots, v_{j-1})$
2. if the j^{th} term is removed from $v_1, \cdots, v_m$, the span of the remaining list equals $\operatorname{span}(v_1, \cdots, v_m)$.

Proof 6.

1. Since $v_1, \cdots, v_m$ is linearly dependent, $a_1v_1 + \cdots + a_mv_m = 0$, for some $a_i \neq 0$. Let j be the maximized index *s.t.* $a_j \neq 0$. Then, $a_{j+1} = \cdots = a_m = 0$, by this assumption. Hence,

$$\begin{aligned} a_jv_j &= -a_1v_1 - \cdots - a_{j-1}v_{j-1} - a_{j+1}v_{j+1} - \cdots - a_mv_m \\ &= -a_1v_1 - \cdots - a_{j-1}v_{j-1} \\ v_j &= -\frac{a_1}{a_j}v_1 - \cdots - \frac{a_{j-1}}{a_j}v_{j-1}. \end{aligned}$$

Since $-\frac{a_1}{a_j}, \cdots, -\frac{a_{j-1}}{a_j} \in \mathbb{F}$, we know $v_j \in \operatorname{span}(v_1, \cdots, v_{j-1})$. □

2. Consider

$$\begin{aligned} \operatorname{span}(v_1, \cdots, v_j, \cdots, v_m) &= \operatorname{span}(v_1, \cdots, -\frac{a_1}{a_j}v_1 - \cdots - \frac{a_{j-1}}{a_j}v_{j-1}, \cdots, v_m) \\ &= \operatorname{span}(v_1, \cdots, v_{j-1}, v_{j+1}, \cdots, v_m). \end{aligned}$$

■

Remark. *By using this Lemma 2.1.21, we can do lots of proofs using the "step" strategy. Namely, we start to remove vectors from a list that are linearly dependent to obtain a* L.I. *list. However, this "step" strategy can only be used when dealing with FINITE-dimensional vector spaces.*

Theorem 2.1.22
Let V be a *f-d* vector space. Let $\operatorname{span}(w_1, \cdots, w_n) = V$. Let $u_1, \cdots, u_m$ be L.I.. Then, $m \leq n$.

Proof 7.
$\boxed{\text{Step } 1}$ Note that $u_1, w_1, \cdots, w_n$ is linearly dependent because $u_1 \in V = \operatorname{span}(w_1, \cdots, w_n)$. Then, by Lemma 2.1.21, we can remove one of the w's, say w_{j1}. Then, the list becomes

$$\{u_1, w_1, \cdots, w_n\} \setminus \{w_{j1}\}.$$

$\boxed{\text{Step } 2}$ Adjoin u_2. Apply the same reasoning, since $\operatorname{span}(\{u_1, w_1, \cdots, w_n\} \setminus \{w_{j1}\}) = V$, we know $\{u_1, u_2, w_1, \cdots, w_n\} \setminus \{w_{j1}\}$ is linearly dependent. Since $u_2 \notin \operatorname{span}(u_1)$, Lemma 2.1.21 is not applicable to u_2. Now, we can remove another w from the list, say w_{j2}. The list becomes

$$\{u_1, u_2, w_1, \cdots, w_n\} \setminus \{w_{j1}, w_{j2}\}.$$

$\vdots$

$\boxed{\text{Step } m}$ After m steps, we list will become

$$\{u_1, \cdots, u_m, w_1, \cdots, w_n\} \setminus \{w_{j1}, \cdots, w_{jm}\}.$$

Since $\operatorname{span}(\{u_1, \cdots, u_m, w_1, \cdots, w_n\} \setminus \{w_{j1}, \cdots, w_{jm}\}) = V$, this list is still linearly dependent, so by Lemma 2.1.21, we know $\exists\, w$ to be removed. Therefore, $n \geq m$. ■

Theorem 2.1.23
Every subspace of a *f-d* vector space is *f-d.*

Proof 8. Suppose V to be a *f-d* vector space and U to be a subspace of V.
$\boxed{\text{Step } 1}$ If $U = \{0\}$, then U is *f-d.* If $U \neq \{0\}$, then choose $v_i \in U$ *s.t.* $v_1 \neq 0$.

$\vdots$

$\boxed{\text{Step } j}$ If $U = \operatorname{span}(v_1, \cdots, v_{j-1})$, then U is *f-d.* If $U \neq \operatorname{span}(v_1, \cdots, v_{j-1})$, then choose $v_j \in U$ *s.t.* $v_j \notin \operatorname{span}(v_1, \cdots, v_{j-1})$.

By Lemma 2.1.21 and Theorem 2.1.22, we know this process will eventually terminate because the vector list that spans U cannot be longer than any spanning list of V. Therefore, U is *f-d.* ■

2.2 Bases

Definition 2.2.1 (Basis). A *basis* of V is a list of vectors in V that is L.I. and spans V.

Example 2.2.2

1. The standard basis of $\mathbb{F}^n$:

$$(1,0,\cdots,0),(0,1,0,\cdots,0),\cdots,(0,\cdots,0,1).$$

2. $(1,1,0),(0,0,1)$ is a basis of V, where $V=\{(x,x,y)\in\mathbb{F}^3 \mid x,y\in\mathbb{F}\}$.

   Proof 1.

   (a) Suppose $a_1(1,1,0)+a_2(0,0,1)=0$, we have $(a_1,a_1,a_2)=0$. So, it must be $a_1=a_2=0$. Therefore, $(1,1,0),(0,0,1)$ is L.I.. □

   (b) Suppose $(x,x,y)\in V$. Note that $(x,x,y)=x(1,1,0)+y(0,0,1)$, then, $V=\operatorname{span}((1,1,0),(0,0,1))$.

   Therefore, we've proven $(1,1,0),(0,0,1)$ is a basis of V according to the definition of basis. ■

Theorem 2.2.3 Criterion for Basis
A list $v_1,\cdots,v_n\in V$ is a basis list of V if and only if every $v\in V$ can be written uniquely in the form $v=a_1v_1+\cdots+a_nv_n$, where $a_i\in\mathbb{F}$.

Proof 2.
($\Rightarrow$) Let $v_1,\cdots,v_n$ be a basis of V. Let $v\in V$. By definition of basis, $V=\operatorname{span}(v_1,\cdots,v_n)$. So, $v\in\operatorname{span}(v_1,\cdots,v_n)$, and thus $v=a_1v_1+\cdots+a_nv_n$ for some $a_i\in\mathbb{F}$. Assume for the sake of contradiction that $v=b_1v_1+\cdots+b_nv_n$ for some $b_i\neq a_i\in\mathbb{F}$. Then,

$$\begin{aligned} v-v&=(a_1-b_1)v_1+\cdots+(a_n-b_n)v_n\\ 0&=(a_1-b_1)v_1+\cdots+(a_n-b_n)v_n. \end{aligned}$$

Since $v_1,\cdots,v_n$ is a basis, it is L.I.. So, $0=0v_1+\cdots+0v_n$. Therefore, we know $a_1-b_1=\cdots=a_n-b_n=0$. That is, $a_1=b_1,\cdots,a_n=b_n$. ⋇ This is a contradiction with the assumption that $\exists\, a_i\neq b_i$. Hence, it must be that $v=a_1v_1+\cdots+a_nv_n$ is unique. □

($\Leftarrow$) Suppose $v=a_1v_1+\cdots+a_nv_n$ is the unique representation $\forall\, v\in V$. Then, $v\in\operatorname{span}(v_1,\cdots,v_n)$. Since $v\in V$, then $V\subseteq\operatorname{span}(v_1,\cdots,v_n)$. However, $v_1,\cdots,v_n\in V$, so $\operatorname{span}(v_1,\cdots,v_n)\subseteq V$. Therefore, $\operatorname{span}(v_1,\cdots,v_n)=V$. To show $v_1,\cdots,v_n$ is L.I., further consider $0=a_1v_1+\cdots+a_nv_n$. Since $0\in V$, by assumption, $\exists$ a unique way to write 0 as $a_1v_1+\cdots+a_nv_n$, and that unique way is to take every $a_i=0$. Hence, by definition, we know $v_1,\cdots,v_n$ is L.I.. Since $v_1,\cdots,v_n$ is L.I. and $\operatorname{span}(v_1,\cdots,v_n)=V$, we know $v_1,\cdots,v_n$ is a basis list of V. ■

Theorem 2.2.4
Every spanning list can be reduced to a basis of the vector space.

Proof 3. Suppose $V = \text{span}(v_1, \cdots, v_n)$. If $v_i = 0$, we just remove v_i. So, let's suppose $v_i \neq 0$.
$\boxed{\text{Step } 1}$ If $v_2 \in \text{span}(v_1)$, delete it. If $v_2 \notin \text{span}(v_2)$, keep it.

$\vdots$

$\boxed{\text{Step } j}$ If $v_j \in \text{span}(v_1, \cdots, v_{j-1})$, delete it. If $v_j \notin \text{span}(v_1, \cdots, v_{j-1})$, keep it.

$\vdots$

$\boxed{\text{Step } n}$ After n steps, we will have a "sub-list" from the original list *s.t.* it spans V and is L.I.. Therefore, the basis list is contained in the spanning list. ■

Corollary 2.2.5 Every *f-d* vector space has a basis.

Proof 4. By definition, *f-d* vector space always has a spanning list. By Theorem 2.2.4, a spanning list contain a basis. ■

> **Theorem 2.2.6**
> Every linearly independent list of vectors in a *f-d* vector space can be extended to a basis of the vector space.

Proof 5. Suppose $u_1, \cdots, u_m$ is L.I. in a *f-d* vector space of V. Let $w_1, \cdots, w_n$ be a basis of V. Then, $u_1, \cdots, u_m, w_1, \cdots, w_n$ spans V. According to Lemma 2.1.21 and Theorem 2.1.22, we can reduce $u_1, \cdots, u_m, w_1, \cdots, w_m$ to some list of $u_1, \cdots, u_m$ and some w's. ■

> **Theorem 2.2.7**
> Suppose V is *f-d* and U is a subspace of V. Then, there is a subspace W of V *s.t.* $V = U \oplus W$.

Proof 6. Since V is *f-d*, U, as V's subspace, is also *f-d*. So, $\exists$ a basis of U, say $u_1, \cdots, u_m$. Then, $u_1, \cdots, u_m$ is L.I. and $\in V$. By Theorem 2.2.6, this list can be extended to a basis

$$u_1, \cdots, u_m, w_1, \cdots, w_n \text{ of } V.$$

Let $W = \text{span}(w_1, \cdots, w_n)$. We'll show $V = U \oplus W$.

1. WTS: $V = U + W$. Suppose $v \in V$. Then,

$$v = \underbrace{a_1u_1 + \cdots + a_mu_m}_{\in U} + \underbrace{b_1w_1 + \cdots + b_nw_n}_{\in W}.$$

   So, $v \in U + W$, or $V = U + W$. □

2. WTS: $U \cap W = \{0\}$. Suppose $v \in U \cap W$. Then, $v \in U$ and $v \in W$. So,

$$v = a_1u_1 + \cdots + a_mv_m = b_1w_1 + \cdots + b_nw_n.$$

   Hence,

$$a_1u_1 + \cdots + a_mu_m - b_1w_1 - \cdots - b_nw_n = 0. \tag{7}$$

   Since by assumption, $u_1, \cdots, u_m, w_1, \cdots, w_n$ is a basis of V, so $u_1, \cdots, u_m, w_1, \cdots, w_n$ is L.I.. Therefore, the only way for Equation (7) to hold is when $a_1 = \cdots = a_m = b_1 = \cdots = b_n = 0$. Hence, $v = 0u_1 + \cdots + u_m = 0$. That is, $U \cap W = \{0\}$.

Therefore, we've shown that $V = U \oplus W$. ■

2.3 Dimension

Theorem 2.3.1
Let B_1 and B_2 be two bases of V, then B_1 and B_2 have the same length.

Proof 1. Since B_1 is L.I. in V and B_2 spans V, by Theorem 2.1.22, we know $\operatorname{len}(B_1) \leq \operatorname{len}(B_2)$. Interchanging the roles of B_1 and B_2, we have $\operatorname{len}(B_2) \leq \operatorname{len}(B_1)$. So, we have $\operatorname{len}(B_1) = \operatorname{len}(B_2)$. ■

Definition 2.3.2 (Dimension). The *dimension* of a *f-d* vector space V is the length of any basis of V.

Notation 2.3.3. We use $\dim V$ to denote the dimension of a *f-d* vector space V.

Example 2.3.4 $\dim \mathbb{F}^n = n$ and $\dim \mathcal{P}_m(\mathbb{F}) = m + 1$ $(1, z, z^2, \cdots, z^m)$.

Theorem 2.3.5
If V is *f-d* and U is a subspace of V, then $\dim U \leq \dim V$.

Proof 2. Let B_1 be a basis of U and B_2 be a basis of V. Then, B_1 is a L.I. list of V and B_2 spans V. Then, By Theorem 2.1.22, we know that $\operatorname{len}(B_1) \leq \operatorname{len}(B_2)$. So, by definition of dimension, we know $\dim U \leq \dim V$. ■

Extension. *If V is* f-d *and U is a subspace of V, given $U \subsetneq V$, then $\dim U < \dim V$.*

Proof 3. Let $u_1, \cdots, u_m$ be a basis of U. Since $U \subsetneq V$, we know $V - U \neq \varnothing$. So, choose $v \in V - U$. Then, $v \notin \operatorname{span}(u_1, \cdots, u_m)$. Therefore, $u_1, \cdots, u_m, v$ is L.I. in V. That is

$$\begin{aligned}\dim V &\geq \dim(\operatorname{span}(u_1, \cdots, u_m, v)) \\ &> \dim(\operatorname{span}(u_1, \cdots, u_m)) \\ &= \dim U.\end{aligned}$$

■

Theorem 2.3.6
Let V be *f-d*, then every L.I. list of vectors in V with length $\dim V$ is a basis of V.

Proof 4. Let $v_1, \cdots, v_n \in V$ be L.I.. Let $n = \dim V$. When extending the list to basis, we get

$$\{v_1, m \cdots, v_n\} \cup \varnothing$$

as a basis of V. That is, $v_1, \cdots, v_n$ has already been a basis of V. ■

Remark. *The proof given above is not that straight-forward, so we are giving an easier-understanding proof as follows.*

Proof 5. Suppose for the sake of contradiction that $\exists v_1, \cdots, v_n \in V$ not a basis of V for $n = \dim V$. Then, $\operatorname{span}(v_1, \cdots, v_n) \neq V$. That is, $\exists\ v_{n+1}$ *s.t.* $v_{n+1} \notin \operatorname{span}(v_1, \cdots, v_n)$. Adding v_{n+1} to the vector list, we have $v_1, \cdots, v_n, v_{n+1}$ is L.I.. By Theorem 2.3.5, we know $\operatorname{len}(v_1, \cdots, v_{n+1}) = n + 1 \leq \dim V$. ⁎ This contradicts with the fact that $\dim V = n < n + 1$. So, our assumption is incorrect, and it must be that $v_1, \cdots, v_n$ is a basis of V. ■

Theorem 2.3.7
Suppose V is *f-d*. Then, every spanning list of vectors in V with length $\dim V$ is a basis of V.

Example 2.3.8 Show that $1, (x-5)^2, (x-5)^3$ is a basis of the subspace U of $\mathcal{P}_3(\mathbb{R})$ defined by

$$U = \{p \in \mathcal{P}_3(\mathbb{R}) \mid p'(5) = 0\}.$$

Proof 6. Consider $a_1 + a_2(x-5)^2 + a_3(x-5)^3 = 0$, we will get $a_1 = a_2 = a_3 = 0$ easily from the equation. Then, $1, (x-5)^2, (x-5)^3$ is L.I.. So, by Theorem 2.3.5, we know $\dim U \geq 3$. Since $U \subsetneq \mathcal{P}_3(\mathbb{R})$, we have $\dim U < \dim \mathcal{P}_3(\mathbb{R}) = 4$. Therefore, $\dim U = 3 = \operatorname{len}(1, (x-5)^2, (x-5)^3)$. By Theorem 2.3.6, we know $1, (x-5)^2, (x-5)^3$ is a basis of U. ■

Theorem 2.3.9
If U_1 and U_2 are subspaces of a *f-d* vector space, then

$$\dim(U_1 + U_2) = \dim(U_1) + \dim(U_2) - \dim(U_1 \cap U_2).$$

Proof 7. Let $u_1, \cdots, u_m$ be a basis of $U_1 \cap U_2$, then $\dim(U_1 \cap U_2) = m$. Also, $u_1, \cdots, u_m$ is L.I. in U_1, so we can extend it to a basis of U_1 as $u_1, \cdots, u_m, v_1, \cdots, v_j$. Then, $\dim(U_1) = m + j$. Similarly, extending $u_1, \cdots, u_m$ to a basis of U_2, we will get $u_1, \cdots, u_m, w_1, \cdots, w_k$. So, $\dim(U_2) = m + k$. Now, we want to show $u_1, \cdots, u_m, v_1, \cdots, v_j, w_1, \cdots, w_k$ is a basis of $U_1 + U_2$.

1. Since $U_1, U_2 \subseteq \operatorname{span}(u_1, \cdots, u_m, v_1, \cdots, v_j, w_1, \cdots, w_k)$, we know that

$$\operatorname{span}(u_1, \cdots, u_m, v_1, \cdots, v_j, w_1, \cdots, w_k) = U_1 + U_2. \qquad \square$$

2. Suppose $a_1u_1 + \cdots + a_mu_m + b_1v_1 + \cdots + b_jv_j + c_1w_1 + \cdots + c_kw_k = 0$. Then we know that

$$c_1w_1 + \cdots + c_kw_k = -a_1u_1 - \cdots - a_mu_m - b_1v_1 - \cdots - b_jv_j.$$

Since $c_1w_1 + \cdots + c_kw_k \in U_2$, and $-a_1u_1 - \cdots - a_mu_m - b_1v_1 - \cdots - b_jv_j \in U_1$, we know that $c_1w_1 + \cdots + c_kw_k \in U_1 \cap U_2$. Therefore, $c_1w_1 + \cdots + c_kw_k = d_1u_1 + \cdots + d_mu_m$. Since $u_1, \cdots, u_m, w_1, \cdots, w_k$ is L.I., we know $c_1 = \cdots = c_k = 0$. So, $-a_1u_1 - \cdots - a_mu_m - b_1v_1 - \cdots - b_jv_j = 0$. Since $u_1, \cdots, u_m, v_1, \cdots, v_j$ is L.I., we have $a_1 = \cdots = a_m = b_1 = \cdots = b_j = 0$. Therefore, we've proven $u_1, \cdots, u_m, v_1, \cdots, v_j, w_1, \cdots, w_k$ is L.I. and thus is a basis of $U_1 + U_2$. $\square$

Since $u_1, \cdots, u_m, v_1, \cdots, v_j, w_1, \cdots, w_k$ is a basis of $U_1 + U_2$, we know $\dim(U_1 + U_2) = m + j + k$. Further note that

$$\begin{aligned}
\dim(U_1) + \dim(U_2) - \dim(U_1 \cap U_2) &= (m + j) + (m + k) - m \\
&= m + j + k \\
&= \dim(U_1 + U_2).
\end{aligned}$$

■

3 Linear Maps

Notation 3.0.1. In this section, we use V and W to denote vector spaces over $\mathbb{F}$.

3.1 The Vector Space of Linear Maps

Definition 3.1.1 (Linear Map). A *linear map* from V to W is a function $T : V \to W$ with the following properties:

- additivity: $T(u+v) = Tu + Tv \qquad \forall u, v \in V$.
- homogeneity: $T(\lambda v) = \lambda(Tv) \qquad \forall \lambda \in \mathbb{F}$ and $\forall v \in V$.

Notation 3.1.2. The set of all linear maps from V to W is denoted by $\mathcal{L}(V, W)$.

Example 3.1.3

1. Zero-mapping: $0 \in \mathcal{L}(V, W)$ is defined by $0v = 0$.
2. Identity-mapping: $I \in \mathcal{L}(V, V)$ is defined by $Iv = v$.
3. Differentiation: $D \in \mathcal{L}(\mathcal{P}(\mathbb{R}), \mathcal{P}(\mathbb{R}))$ is defined by $Dp = p'$.

   Proof 1. Note that $(f+g)' = f' + g'$ and $(\lambda f)' = \lambda f'$. ■
4. Integration: $T \in \mathcal{L}(\mathcal{P}(\mathbb{R}), \mathbb{R})$ is defined by $Tp = \int_0^1 p(x)\,\mathrm{d}x$

   Proof 2. Note that $\int_0^1 (f+g) = \int_0^1 f + \int_0^1 g$ and $\int_0^1 \lambda f = \lambda \int_0^1 f$. ■
5. Backward shift: $T \in \mathcal{L}(\mathbb{F}^\infty, \mathbb{F}^\infty)$ as $T(x_1, x_2, x_3, \cdots) = (x_2, x_3, \cdots)$.

   Proof 3. Note that

$$\begin{aligned} T(x_1, x_2, x_3, \cdots) + T(y_1, y_2, y_3, \cdots) &= (x_2, x_3, \cdots) + (y_2, y_3, \cdots) \\ &= (x_2 + y_2, x_3 + y_3, \cdots) \\ &= T(x_1 + y_1, x_2 + y_2, x_3 + y_3, \cdots). \end{aligned}$$

   Therefore, T is additive. Homogeneity of T is travial and thus omitted here. ■
6. From $\mathbb{F}^n$ to $\mathbb{F}^m$, we define $T \in \mathcal{L}(\mathbb{F}^n, \mathbb{F}^m)$ as

$$T(x_1, \cdots, x_n) = (A_{1,1}x_1 + \cdots + A_{1,n}x_n, \cdots, A_{m,1}x_1 + \cdots + A_{m,n}x_n),$$

   where $A_{j,k} \in \mathbb{F} \quad \forall j = 1, \cdots, m$ and $k = 1, \cdots, n$.

Theorem 3.1.4
Suppose $v_1, \cdots, v_n$ is a basis of V and $w_1, \cdots, w_n \in W$. Then, $\exists$ a unique linear map $T : V \to W$ *s.t.* $Tv_j = w_j \quad \forall j = 1, \cdots, n$.

Remark. *If T in Theorem 3.1.1 is a linear mapping, we should have*

1. $T(v_1 + \cdots + v_n) = Tv_1 + \cdots + Tv_n = w_1 + \cdots + w_n$, *by additivity of T, and*
2. $T(\lambda_j v_j) = \lambda_j T v_j$, *by homogeneity of T.*

Combine the two properties, we should have

$$T(\lambda_1 v_1 + \cdots + \lambda_n v_n) = \lambda_1 T v_1 + \cdots = \lambda_n T v_n = \lambda_1 w_1 + \cdots + \lambda_n w_n.$$

This remark will be very helpful in our following proof of the theorem.

Proof 4. Let's define $T : V \to W$ by $T(c_1 v_1 + \cdots + c_n v_n) = c_1 w_1 + \cdots + c_n w_n$, where $c_1, \cdots, c_n$ are arbitrary elements of $\mathbb{F}$. Now, we want to show that T is a linear mapping.

Suppose $u, v \in V$, $u = a_1 v_1 + \cdots + a_n v_n$, and $v = c_1 v_1 + \cdots + c_n v_n$. Then, we have

$$\begin{aligned} T(u+v) &= T((a_1 + c_1)v_1 + \cdots + (a_n + c_n)v_n) \\ &= (a_1 + c_1)w_1 + \cdots + (a_n + c_n)w_n \\ &= (a_1 w_1 + \cdots + a_n w_n) + (c_1 w_1 + \cdots + c_n w_n) \\ &= Tu + Tv. \qquad \square \end{aligned}$$

Now, we want to show T has homogeneity. Suppose $\lambda \in \mathbb{F}$. Then, we know

$$\begin{aligned} T(\lambda v) &= T(\lambda c_1 v_1 + \cdots + \lambda c_n v_n) \\ &= \lambda c_1 w_1 + \cdots + \lambda c_n w_n \\ &= \lambda(c_1 w_1 + \cdots + c_n w_n) \\ &= \lambda T v. \qquad \square \end{aligned}$$

Also, we want to show that this T satisfy the condition the theorem is asking (i.e., $Tv_j = w_j$). Note that when $c_j = 0$ and other c's equal 0, we will get $Tv_j = w_j$. $\square$

Finally, we will prove the uniqueness of this T. Suppose that $T' \in \mathcal{L}(V, W)$ and $T' v_j = w_j$. Let $c_1, \cdots, c_n \in \mathbb{F}$. Then, $T'(c_j v_j) = c_j w_j$. So, we know that $T'(c_1 v_1 + \cdots + c_n v_n) = c_1 w_1 + \cdots + c_n w_n$. However, by definition, we know $c_1 w_1 + \cdots + c_n w_n = T(c_1 w_1 + \cdots + c_n v_n)$. So, we can conclude that $T'(c_1 v_1 + \cdots + c_n v_n) = T(c_1 w_1 + \cdots + c_n v_n)$. Thus, $T' = T$, and thus the T we defined above is unique in $\mathcal{L}(V, W)$. ■

Definition 3.1.5 (Addition and Scalar Multiplication on $\mathcal{L}(V, W)$). Suppose $S, T \in \mathcal{L}(V, W)$ and $\lambda \in \mathbb{F}$. Then, the *addition* is defined as $(S + T)(v) := Sv + Tv$, and the *scalar multiplication* is defined as $(\lambda T)(v) := \lambda(Tv) \quad \forall v \in V$.

> **Theorem 3.1.6**
> $\mathcal{L}(V, W)$ is a vector space.

Proof 5.

1. additive identity: Note that the zero-mapping $0 \in \mathcal{L}(V, W)$ satisfies the following equation:

$$(0 + T)(v) = 0v + Tv = 0 + Tv = Tv. \qquad \square$$

2. commutativity: Note that

$$(S+T)(v) = Sv + Tv = Tv + Sv = (T+S)(v). \qquad \square$$

3. associativity: Let $S, T, R \in \mathcal{L}(V, W)$. Then,

$$\begin{aligned}((S+T)+R)(v) = (S+T)(v) + Rv &= Sv + Tv + Rv \\ &= Sv + (Tv + Rv) \\ &= Sv + (T+R)(v) \\ &= (S + (T+R))(v).\end{aligned}$$

Let $a, b \in \mathbb{F}$. Then,

$$((ab)T)(v) = T(abv) = T(a(bv)) = aT(bv) = (a(bT))(v). \qquad \square$$

4. multiplicative identity: Note we have $1 \in \mathbb{F}$ *s.t.*

$$(1 \cdot T)(v) = T(1 \cdot v) = Tv. \qquad \square$$

5. additive inverse: Note that

$$(T + (-T))(v) = Tv + (-T)(v) = Tv + T(-v) = T(v - v) = T0 = 0. \qquad \square$$

6. distributivity: Note that

$$a(T+S)(v) = a(Tv + Sv) = aTv + aSv,$$

and

$$(a+b)Tv = T((a+b)v) = T(av + bv) = T(av) + T(bv) = aTv + bTv.$$

■

Definition 3.1.7 (Product of Linear Maps). If $T \in \mathcal{L}(U, V)$ and $S \in \mathcal{L}(V, W)$, then the *product* $ST \in \mathcal{L}(U, W)$ is defined by $(ST)(u) = S(Tu) \quad \forall u \in U$.

Remark. *Compare this definition with composite functions. ST is only defined when T maps into the domain of S.*

Theorem 3.1.8 Algebraic Properties of Products of Linear Maps

1. associativity: $(T_1T_2)T_3 = T_1(T_2T_3)$.
2. identity: $TI = IT = T$, where I is the identity mapping
3. distributive properties: $(S_1 + S_2)T = S_1T + S_2T$ and $S(T_1 + T_2) = ST_1 + ST_2$.

Proof 6. First, we want to show the associativity. Note that

$$[(T_1T_2)T_3](v) = (T_1T_2)(T_3v) = (T_1)(T_2(T_3v)) = (T_1)[(T_2T_3)(v)]. \qquad \square$$

Then, we want to show the identity. This proof can be done using the following diagram:

$$\begin{tikzcd} V \arrow[r, "T"] & W \arrow[d, "I_W"] \\ V \arrow[u, "I_V"] & W \end{tikzcd} \qquad \square$$

Finally, we will show the distributive properties. Note that

$$\begin{aligned}[(S_1 + S_2)T](v) = (S_1 + S_2)(Tv) &= S_1(Tv) + S_2(Tv) \\ &= (S_1T)(v) + (S_2T)(v) \\ &= (S_1T + S_2T)(v).\end{aligned}$$

Similarly, we can show

$$\begin{aligned}[S(T_1 + T_2)](v) = S[(T_1 + T_2)(v)] &= S(T_1v + T_2v) \\ &= S(T_1v) + S(T_2v) \\ &= (ST_1)(v) + (ST_2)(v) \\ &= (ST_1 + ST_2)(v).\end{aligned}$$

■

Example 3.1.9 Suppose $D \in \mathcal{L}(\mathcal{P}(\mathbb{R}), \mathcal{P}(\mathbb{R}))$ is the differentiation map, and $T \in \mathcal{L}(\mathcal{P}(\mathbb{R}), \mathcal{P}(\mathbb{R}))$ be defined by $(Tp)(x) = x^2p(x)$. Show that $DT \neq TD$.

Proof 7. Note that $(DT)p = D(Tp) = D(x^2p(x)) = 2xp(x) + x^2p'(x)$. Similarly, we can compute a general formula for TD: $(TD)p = T(Dp) = T(p') = x^2p'(x)$. Since $2xp(x) + x^2p'(x) \neq x^2p'(x)$, we know $DT \neq TD$. ■

Theorem 3.1.10
Let $T \in \mathcal{L}(V, W)$, then $T(0) = 0$.

Proof 8. Since $T(0) = T(0 + 0) = T(0) + T(0)$, we know $0 = T(0)$, or $T(0) = 0$. ■

Corollary 3.1.11 If $T(0) \neq 0$, then $T \notin \mathcal{L}(V, W)$.

3.2 Null Spaces and Ranges

Definition 3.2.1 (Null Space/Kernel). For $T \in \mathcal{L}(V,W)$, the *null space* of T, denoted $\operatorname{null} T$, is the subset of V consisting of those vectors that T maps to 0: $\operatorname{null} T = \{v \in V \mid Tv = 0\}$.

Remark. *Sometimes, null space of T is also called the kernal of T, denoted as $\ker T$.*

Example 3.2.2

1. Null space of zero-mapping: Let T be the zero mapping from V to W. Since $Tv = 0 \quad \forall v \in V$, we know $\operatorname{null} T = V$.
2. $D \in \mathcal{L}(\mathcal{P}(\mathbb{R}), \mathcal{P}(\mathbb{R}))$ as $Dp = p'$: $\operatorname{null} D = \{a \mid a \in \mathbb{R}\}$.
3. $T \in \mathcal{L}(\mathbb{F}^\infty, \mathbb{F}^\infty)$ as $T(x_1, x_2, x_3, \cdots) = (x_2, x_3, \cdots)$: $\operatorname{null} T = \{(a, 0, 0, \cdots) \mid a \in \mathbb{F}\}$.

Theorem 3.2.3
Suppose $T \in \mathcal{L}(V,W)$. Then, $\operatorname{null} T$ is a subspace of V.

Proof 1.

1. Note that $T(0) = 0$, so $0 \in \operatorname{null} T$. □
2. Suppose $u, v \in \operatorname{null} T$. Then, $Tu = Tv = 0$. So, $T(u+v) = Tu + Tv = 0 + 0 = 0$. Hence, $u + v \in \operatorname{null} T$. □
3. Suppose $u \in \operatorname{null} T$ and $\lambda \in \mathbb{F}$. Then, $Tu = 0$. So, $T(\lambda u) = \lambda Tu = \lambda \cdot 0 = 0$. Therefore, $\lambda u \in \operatorname{null} T$.

■

Definition 3.2.4 (Injective/Injection). A function $T : V \to W$ is called *injective* of $Tu = Tv$ implies $u = v$.

Remark. *Sometimes, the contrapositive will be much more helpful: T is injective if $u \neq v$, then $Tu \neq v$.*

Theorem 3.2.5
Let $T \in \mathcal{L}(V,W)$. Then, T is injective if and only if $\operatorname{null} T = \{0\}$.

Proof 2.
($\Rightarrow$) Suppose T is an injective. We've already known that $\{0\} \subseteq \operatorname{null} T$. Then, we need to show $\operatorname{null} T \subseteq \{0\}$. Suppose $v \in \operatorname{null} T$, then $Tv = 0$. However, since T is an injection, and $Tv = T0 = 0$, then we have $v = 0$. So, $\operatorname{null} T \subseteq \{0\}$. Therefore, it's sufficient to say $\operatorname{null} T = \{0\}$. □

($\Leftarrow$) Suppose $\operatorname{null} T = \{0\}$. Suppose $u, v \in V$ and $Tu = Tv$. Then, $Tu - Tv = T(u - v) = 0$. Hence, $u - v \in \operatorname{null} T$. By $\operatorname{null} T = \{0\}$, we know $u - v = 0$, so $u = v$. Then, T is an injection. ■

Definition 3.2.6 (Range/Image). For $T \in \mathcal{L}(V,W)$, the range of T is the subset of W consisting of those vectors that are of the form Tv for some $v \in V$: $\operatorname{range} T = \{Tv \mid v \in V\}$.

Theorem 3.2.7
If $T \in \mathcal{L}(V,W)$, then $\operatorname{range} T$ is a subspace of W.

Proof 3.

1. Since $T(0) = 0$, we know $0 \in \operatorname{range} T$. $\square$

2. Suppose $w_1, w_2 \in \operatorname{range} T$. Then, $\exists v_1, v_2 \in V$ *s.t.* $Tv_1 = w_1$ and $Tv_2 = w_2$. Then, $w_1 + w_2 = Tv_1 + Tv_2 = T(v_1 + v_2)$. Since $v_1 + v_2 \in V$, we have $w_1 + w_2 = T(v_1 + v_2) \in \operatorname{range} T$. $\square$

3. Suppose $w \in \operatorname{range} T$ and $\lambda \in \mathbb{F}$. Then, $\exists v \in V$ *s.t.* $w = Tv$. So, $\lambda w = \lambda(Tv) = T(\lambda v)$. Since $\lambda v \in V$, $\lambda w = T(\lambda v) \in \operatorname{range} T$.

$\blacksquare$

Definition 3.2.8 (Surjective/Surjection). A function $T : V \to W$ is called *surjective* if $\operatorname{range} T = W$.

Remark. *A function $T : V \to W$ is called a* bijection*, or is bijective, if it is both injective and surjective.*

Theorem 3.2.9 Fundamental Theorem of Linear Maps
Suppose V is *f-d* and $T \in \mathcal{L}(V, W)$. Then, $\operatorname{range} T$ is *f-d* and

$$\dim V = \dim \operatorname{null} T + \dim \operatorname{range} T.$$

Proof 4. Let $u_1, \cdots, u_m$ be a basis of $\operatorname{null} T$. Then, $\dim \operatorname{null} T = m$. By Theorem 3.2.3, we know $\operatorname{null} T$ is a basis of V, so we can extend the basis to a basis of V: $u_1, \cdots, u_m, v_1, \cdots, v_n$. Thus, $\dim V = m + n$. *WTS:* $\dim \operatorname{range} T = n$. *Further WTS:* $Tv_1, \cdots, Tv_n$ *is a basis of* $\operatorname{range} T$.

Suppose $v \in V$. Then

$$v = a_1u_1 + \cdots + a_mu_m + b_1v_1 + \cdots + b_nv_n.$$

Since $u_1, \cdots, u_m \in \operatorname{null} T$, we know $Tu_1, \cdots, Tu_m = 0$. Therefore,

$$Tv = a_1Tu_1 + \cdots + a_mTu_m + b_1Tv_1 + \cdots + b_nTv_n = b_1Tv_1 + \cdots + b_nTv_n.$$

Hence, $\operatorname{span}(Tv_1, \cdots, Tv_n) = \operatorname{range} T$, and thus $\operatorname{range} T$ is *f-d. Now, WTS:* $Tv_1, \cdots, Tv_n$ *is* L.I..

Consider $c_1Tv_1 + \cdots + c_nTv_n = 0$. Then, $T(c_1v_1 + \cdots + c_nv_n) = 0$. Hence, $c_1v_1 + \cdots + c_nv_n \in \operatorname{null} T$. Since $u_1, \cdots, u_m$ is a basis of $\operatorname{null} T$, we know

$$c_1v_1 + \cdots + c_nv_n = d_1u_1 + \cdots + d_mu_m \quad \text{\textit{f.s.} } d_i \in \mathbb{F}.$$

So,

$$c_1v_1 + \cdots + c_nv_n - d_1u_1 - \cdots - d_mu_m = 0. \tag{8}$$

However, by assumption, we know $v_1, \cdots, v_n, u_1, \cdots, u_m$ is a basis of V, and thus it is L.I.. So, the only way to make Equation (8) hold is by taking $c_1 = \cdots = c_n = -d_1 = \cdots = -d_m = 0$. Therefore, we've shown $Tv_1, \cdots, Tv_n$ is L.I., and thus is a basis of $\operatorname{range} T$. Then, $\dim \operatorname{range} T = n$.

So, we've shown that $\dim \operatorname{null} T + \dim \operatorname{range} T = m + n = \dim V$. $\blacksquare$

Theorem 3.2.10
Suppose V and W are *f-d* vector spaces *s.t.* $\dim V > \dim W$. Then, no linear map from V to W is injective.

Proof 5. Let $T \in \mathcal{L}(V, W)$. By the Fundamental Theorem of Linear Maps, we have $\dim V = \dim \operatorname{null} T + \dim \operatorname{range} T$. Then, we know

$$\begin{aligned}\dim \operatorname{null} T &= \dim V - \dim \operatorname{range} T \\ &\geq \dim V - \dim W > 0 \qquad [\dim \operatorname{range} T \leq \dim W]\end{aligned}$$

This implies that $\operatorname{null} T \neq \{0\}$. So, T is not injective by Theorem 3.2.5. ■

Theorem 3.2.11
Suppose V and W are *f-d* vector space *s.t.* $\dim V < \dim W$. Then, no linear map from V to W is surjective.

Proof 6. We know

$$\begin{aligned}\dim \operatorname{range} T &= \dim V - \dim \operatorname{null} T \\ &\leq \dim V < \dim W\end{aligned}$$

Then, T cannot be surjective by definition. ■

Example 3.2.12 Solving Linear Systems Using Linear Maps I
For a homogenous system of linear equations,

$$\begin{cases} A_{1,1}x_1 + \cdots + A_{1,n}x_n = 0 \\ \qquad\qquad \vdots \\ A_{m,1}x_1 + \cdots + A_{m,n}x_n = 0 \end{cases},$$

where $A_{j,k} \in \mathbb{F}$ and $(x_1, \cdots, x_n) \in \mathbb{F}^n$, we can defined a linear map $T : \mathbb{F}^n \to \mathbb{F}^m$ as

$$T(x_1, \cdots, x_n) = \left(\sum_{k=1}^{n} A_{1,k}x_k, \cdots, \sum_{k=1}^{n} A_{m,k}x_k \right).$$

Apparently, $(x_1, \cdots, x_n) = 0$ is a solution to the system, but the question is "If there are any non-zero solutions for this linear system?"

Theorem 3.2.13
A homogeneous system of linear equations with more variables than equations has non-zero solutions.

Proof 7. Suppose $T \in \mathcal{L}(V, W)$. Then, $\dim V = n$ and $\dim W = m$. Suppose $n > m$. So, $\dim V > \dim W$. By the Theorem 3.2.5, we know T is not injective. ■

Example 3.2.14 Solving Linear Systems Using Linear Maps II

For an inhomogeneous system of linear equations

$$\begin{cases} \sum_{k=1}^{n} A_{1,k}x_k = c_1 \\ \vdots \\ \sum_{k=1}^{n} A_{m,k}x_k = c_m \end{cases},$$

where $A_{j,k} \in \mathbb{F}$ and $(c_1, \cdots, c_m) \in \mathbb{F}^m$ and $(x_1, \cdots, x_n) \in \mathbb{F}^n$, we can define $T : \mathbb{F}^n \to \mathbb{F}^m$ by

$$T(x_1, \cdots, x_m) = \left(\sum_{k=1}^{n} A_{1,k}x_k, \cdots, \sum_{k=1}^{n} A_{m,k}x_k = c_1 \right).$$

However, in this case, $(x_1, \cdots, x_n) = 0$ may not be a solution to the system.

Theorem 3.2.15
An inhomogeneous system of linear equations with more equations than variables has no solution for some choice of the constant terms.

Proof 8. Suppose $T \in \mathcal{L}(V, W)$. So, $\dim V = n$ and $\dim W = m$. Suppose $n < m$. Then, $\dim V < \dim W$. By Theorem 3.2.11, we know T is not surjective. ■

3.3 Matrices

Definition 3.3.1 (Matrix). Let $m, n \in \mathbb{Z}^+$. An *m-by-n matrix* A is a rectangular array of elements of $\mathbb{F}$ with m rows and n columns:

$$A = \begin{pmatrix} A_{1,1} & \cdots & A_{1,n} \\ \vdots & & \vdots \\ A_{m,1} & \cdots & A_{m,n} \end{pmatrix}.$$

The notation $A_{j,k}$ denotes the entry in row j, column k of A.

Definition 3.3.2 (Matrix of a Linear Map). Suppose $T \in \mathcal{L}(V, W)$ and $v_1, \cdots, v_n$ is a basis of V and $w_1, \cdots, w_m$ is a basis of W. The *matrix of T* with respect to these bases is the $m \times n$ matrix $\mathcal{M}(T)$ whose $A_{j,k}$ are defined by

$$Tv_k = A_{1,k}w_1 + \cdots + A_{m,k}w_m.$$

If the bases are not clear from the context, then the notation $\mathcal{M}(T, (v_1, \cdots, v_n), (w_1, \cdots, w_m))$ is used.

Example 3.3.3 Suppose $T \in \mathcal{L}(\mathbb{F}^2, \mathbb{F}^3)$ is defined by $T(x, y) = (x + 3y, 2x + 5y, 7x + 9y)$. Find the matrix of T with respect to the standard bases of $\mathbb{F}^2$ and $\mathbb{F}^3$.

Solution 1.

Note that $T(1, 0) = (1, 2, 7)$ and $T(0, 1) = (3, 5, 9)$. Then,

$$\mathcal{M}(T) = \begin{pmatrix} 1 & 3 \\ 2 & 5 \\ 7 & 9 \end{pmatrix}.$$

□

Example 3.3.4 Suppose $D \in \mathcal{L}(\mathcal{P}_3(\mathbb{R}), \mathcal{P}_2(\mathbb{R}))$ is the differentiation map defined by $Dp = p'$. Find the matrix of D with respect to the standard bases of $\mathcal{P}_3(\mathbb{R})$ and $\mathcal{P}_2(\mathbb{R})$.

Solution 2.

Standard bases of $\mathcal{P}_3(\mathbb{R}) : 1, x, x^2, x^3$. Standard bases of $\mathcal{P}_2(\mathbb{R}) : 1, x, x^2$. Since $(x^n)' = nx^{n-1}$, so we have

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

So, we have

$$\mathcal{M}(D) = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}.$$

□

Definition 3.3.5 (Matrix Addition). The *sum of two matrices of the same size* is the matrix obtained by adding corresponding entries in the matrices:

$$\begin{pmatrix} A_{1,1} & \cdots & A_{1,n} \\ \vdots & & \vdots \\ A_{m,1} & \cdots & A_{m,n} \end{pmatrix} + \begin{pmatrix} C_{1,1} & \cdots & C_{1,n} \\ \vdots & & \vdots \\ C_{m,1} & \cdots & C_{m,n} \end{pmatrix} = \begin{pmatrix} A_{1,1}+C_{1,1} & \cdots & A_{1,n}+C_{1,n} \\ \vdots & & \vdots \\ A_{m,1}+C_{m,1} & \cdots & A_{m,n}+C_{m,n} \end{pmatrix}.$$

Theorem 3.3.6
Suppose $S, T \in \mathcal{L}(V, W)$. Then, $\mathcal{M}(S+T) = \mathcal{M}(S) + \mathcal{M}(T)$.

Proof 3. Let $v_1, \cdots, v_n$ be a basis of V and $w_1, \cdots, w_n$ be a basis of W. Suppose $\mathcal{M}(S) = A$ and $\mathcal{M}(T) = C$. Then, if $1 \leq k \leq n$, we have

$$\begin{aligned} (S+T)v_k &= Sv_k + Tv_k \\ &= (A_{1,k}w_1 + \cdots + A_{m,k}w_m) + (C_{1,k}w_1 + \cdots + C_{m,k}w_m) \\ &= (A_{1,k} + C_{1,k})w_1 + \cdots + (A_{m,k} + C_{m,k})w_m. \end{aligned}$$

Hence, we have $\mathcal{M}(S+T) = \mathcal{M}(S) + \mathcal{M}(T)$. ■

Definition 3.3.7 (Scalar Multiplication of a Matrix). The *product of a scalar and a matrix* is the matrix obtained by multiplying each entry in the matrix by the scalar:

$$\lambda \begin{pmatrix} A_{1,1} & \cdots & A_{1,n} \\ \vdots & & \vdots \\ A_{m,1} & \cdots & A_{m,n} \end{pmatrix} = \begin{pmatrix} \lambda A_{1,1} & \cdots & \lambda A_{1,n} \\ \vdots & & \vdots \\ \lambda A_{m,1} & \cdots & \lambda A_{m,n} \end{pmatrix}.$$

In other words, $(\lambda A)_{j,k} = \lambda A_{j,k}$.

Theorem 3.3.8
Suppose $\lambda \in \mathbb{F}$ and $T \in \mathcal{L}(V, W)$. Then, $\mathcal{M}(\lambda T) = \lambda \mathcal{M}(T)$.

Proof 4. Let $v_1, \cdots, v_n$ be a basis of V and $\mathcal{M}(T) = A$. When $1 \leq k \leq v$, note that

$$\begin{aligned} (\lambda T)v_k &= \lambda(Tv_k) \\ &= \lambda(A_{1,k}w_1 + \cdots + A_{m,k}w_m) \\ &= (\lambda A_{1,k})w_1 + \cdots + (\lambda A_{m,k})w_m. \end{aligned}$$

So, $\mathcal{M}(\lambda T) = \lambda \mathcal{M}(T)$. ■

Notation 3.3.9. $\mathbb{F}^{m,n} \coloneqq$ the set of all $m \times n$ matrices with entries in $\mathbb{F}$.

Theorem 3.3.10
Suppose $m, n \in \mathbb{Z}^+$. With addition and scalar multiplication defined above, $\mathbb{F}^{m,n}$ is a vector space and $\dim \mathbb{F}^{m,n} = mn$.

Proof 5. It is trivial to prove $\mathbb{F}^{m,n}$ is a vector space. □

Define $A_{j,k}$ as the matrix with 1 on its j^{th} row, k^{th} column and 0 elsewhere. Then, we can see that $A_{j,k}$ for $j = 1, \cdots, m$ and $k = 1, \cdots, n$ is a basis for $\mathbb{F}^{m,n}$. So, $\dim \mathbb{F}^{m,n} = m \cdot n$. ■

Definition 3.3.11 (Matrix Multiplication). Suppose A is an $m \times n$ matrix and C is an $n \times p$ matrix. Then, AC is defined to be the $m \times p$ matrix whose entry in row j. column k is given by

$$(AC)_{j,k} = \sum_{r=1}^{n} A_{j,r} C_{r,k}.$$

Remark. *Matrix multiplication is not commutative. i.e.,* $AC \neq CA$*. However, it is distributive and associative.*

Theorem 3.3.12
If $T \in \mathcal{L}(U, V)$ and $S \in \mathcal{L}(V, W)$, then $\mathcal{M}(ST) = \mathcal{M}(S)\mathcal{M}(T)$.

Notation 3.3.13. Suppose A is an $m \times n$ matrix.

1. If $1 \leq j \leq m$, then $A_{j,\cdot}$ denotes the $1 \times n$ matrix consisting of row j of A.
2. If $1 \leq k \leq n$, then $A_{\cdot,k}$ denotes the $m \times 1$ matrix consisting of column k of A.

In other words,

$$A = \begin{pmatrix} A_{1,1} & \cdots & A_{1,n} \\ \vdots & & \vdots \\ A_{m,1} & \cdots & A_{m,n} \end{pmatrix}; \qquad A_{j,\cdot} = \begin{pmatrix} A_{j,1} & \cdots & A_{j,n} \end{pmatrix} \in \mathbb{F}^{1,n}; \qquad A_{\cdot,k} = \begin{pmatrix} A_{1,k} \\ \vdots \\ A_{m,k} \end{pmatrix} \in \mathbb{F}^{m,1}.$$

Theorem 3.3.14 Practical Interpretations of Matrix Multiplication

1. Suppose A is an $m \times n$ matrix and C is an $n \times p$ matrix. Then, $(AC)_{j,k} = A_{j,\cdot} C_{\cdot,k}$ for $1 \leq j \leq m$ and $1 \leq k \leq p$.
2. Suppose A is an $m \times n$ matrix and C is an $n \times p$ matrix. Then, $(AC)_{\cdot,k} = AC_{\cdot,k}$ for $1 \leq k \leq p$.
3. Suppose A is an $m \times n$ matrix and $C = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}$ is an $n \times 1$ matrix. Then,

$$AC = c_1 A_{\cdot,1} + \cdots + c_n A_{\cdot,n}.$$

In other words, AC is a linear combination of the columns of A, with the scalars that multiply the columns coming from C.

Example 3.3.15

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} \begin{pmatrix} 5 \\ 1 \end{pmatrix} = 5 \begin{pmatrix} 1 \\ 3 \\ 5 \end{pmatrix} + 1 \begin{pmatrix} 2 \\ 4 \\ 6 \end{pmatrix} = \begin{pmatrix} 7 \\ 19 \\ 31 \end{pmatrix}.$$

3.4 Invertibility and Isomorphic Vector Spaces

Definition 3.4.1 (Invertible). A linear map $T \in \mathcal{L}(V, W)$ is called *invertible* if $\exists$ a linear map $S \in \mathcal{L}(W, V)$ *s.t.* ST equals the identity map on I and TS equals the identity map on W.
Definition 3.4.2 (Inverse). A linear map $S \in \mathcal{L}(W, V)$ satisfying $ST = I$ and $TS = I$ is called an *inverse* of T.

> **Theorem 3.4.3**
> An invertible linear map has a unique inverse.

Proof 1. Suppose $T \in \mathcal{L}(V, W)$ is invertible. Let S_1 and S_2 be inverses of T.Then,

$$S_1 = S_1 I = S_1(TS_2) = (S_1 T)S_2 = IS_2 = S_2.$$

Thus, $S_1 = S_2$, and so inverse is unique. ∎
Notation 3.4.4. If T is invertible, then its inverse is denoted by T^{-1}.

> **Theorem 3.4.5**
> A linear map is invertible if and only if it is injective and surjective.

Proof 2.
($\Rightarrow$) Let $T \in \mathcal{L}(V, W)$ be invertible. Then, $TT^{-1} = I_W$ and $T^{-1}T = T_V$. Let $Tv = 0$. Note that $(T^{-1}T)v = 0$, so $Iv = 0$ and thus $v = 0$. Therefore, $\operatorname{null} T = \{0\}$, and so T is an injection.
To show T is surjective, suppose $w \in W$. Note that since $T^{-1} \in \mathcal{L}(W, V)$, $T^{-1}w \in V$. So,

$$T(T^{-1}w) = (TT^{-1})w = T_W w = w \in W.$$

Therefore, $T^{-1}w$ is the $v \in V$ we intend to find. Hence, T is also a surjection. □
($\Leftarrow$) Let T be surjective and injective. For $w \in W$, define $Sw \in V$ *s.t.* $T(Sw) = w$. So, we know Sw is unique. Since $(T \circ S)w = w$, we know $(T \circ S) = I_W$. Consider $(S \circ T)v = S(Tv)$, we have $T(S(Tv)) = Tv$, by definition of S. Since T is injective, we know $S(Tv) = V$. So, $(S \circ T)v = v$, and thus $ST = T_V$. Therefore T is invertible.
Now, we want to show S is a linear map. Let $w_1, w_2 \in W$, then

$$T(S(w_1 + w_2)) = (TS)(w_1 + w_2) = I_W(w_1 + w_2) = w_1 + w_2.$$

By definition, $w_1 + w_2 = T(Sw_1) + T(Sw_3) = T(Sw_1 + Sw_2)$. So, $T(S(w_1 + w_2)) = T(Sw_1 + Sw_2)$. By T is an injection, we have $S(w_1 + w_2) = Sw_1 + Sw_2$. So, S is additive. Further consider

$$T(S(\lambda w)) = \lambda w = \lambda(T(Sw)) = T(\lambda Sw)$$

for some $w \in W$. Again, since T is injective, $S(\lambda w) = \lambda Sw$. So, S has homogeneity. Then, S is a linear map. ∎
Definition 3.4.6 (Isomorphism). An *isomorphism* is an invertible linear map.
Definition 3.4.7 (Isomorphic). Two vector spaces are called *isomorphic* if there is an isomorphism from one vector space onto the other one.
Notation 3.4.8. If two vector spaces V and W are isomorphic, we denote them as $V \cong W$.

Theorem 3.4.9
Suppose V and W are *f-d* vector spaces, then $V \cong W$ if and only if $\dim V = \dim W$.

Proof 3.
($\Rightarrow$) Suppose $V \cong W$. By Fundamental Theorem of Linear Maps, we know

$$\dim V = \dim \operatorname{null} T + \dim \operatorname{range} T.$$

Since $V \cong W$, T is invertible and thus is injective and surjective. So, $\dim \operatorname{null} T = 0$ and $\dim \operatorname{range} T = \dim W$. Therefore, $\dim V = 0 + \dim W = \dim W$. $\square$

($\Leftarrow$) Suppose $\dim V = \dim W$. Suppose $v_1, \cdots, v_n$ and $w_1, \cdots, w_n$ are bases of V and W, respectively. Then, $\dim V = \dim W = n$. Here, we want to define a bijection between V and W. Let T be defined as $Tv_i = wi \quad (i = 1, \cdots, n)$.

Let $Tv = 0$. Then, $T(a_1v_1 + \cdots + a_nv_n) = 0$. So, by definition, $a_1w_1 + \cdots + a_nw_n = 0$. Since $w_1, \cdots, w_n$ is a basis, we have $a_1 = \cdots = a_n = 0$. So, $\operatorname{null} T = \{0\}$, and thus T is an injection.

Let $w \in W$ be any vector. Then, we know $w = c_1w_1 + \cdots + c_nw_n$. Note that, by definition of T, we have $T(c_1v_1 + \cdots + c_nv_n) = c_1w_1 + \cdots + c_nw_n$. Hence, $\forall w \in W, \exists v = c_1v_1 + \cdots + c_nv_n \in V$ *s.t.* $Tv = w$. Therefore, T is a surjection.

Finally, it is trivial to show that T is indeed a linear map, and so the proof is complete. ■

Theorem 3.4.10
Suppose $v_1, \cdots, v_n$ is a basis of V and $w_1, \cdots, w_m$ is a basis of W. then, $\mathcal{M}$ is an isomorphism between $\mathcal{L}(V, W)$ and $\mathbb{F}^{m,n}$.

Proof 4. We already know $\mathcal{M}$ is linear, so we just need to show $\mathcal{M}$ is a bijection.

To prove $\mathcal{M}$ is injective, consider $\mathcal{M}(T) = 0$ for some $T \in \mathcal{L}(V, W)$. So, we get $Tv_k = 0$. Since $v_1, \cdots, v_n$ is a basis of V, we know $Tv = 0 \quad \forall v \in V$. Then, T is the zero-mapping, or $T = 0$. Therefore, $\operatorname{null} \mathcal{M} = \{0\}$.

To show $\mathcal{M}$ is surjective, suppose $A \in \mathbb{F}^{m,n}$. Let T be a linear map from V to W *s.t.*

$$Tv_k = \sum_{j=1}^{m} A_{j,k}w_j, \quad k = 1, \cdots, n.$$

Obviously, $\mathcal{M}(T) = A$, and thus $\operatorname{range} \mathcal{M} = \mathbb{F}^{m,n}$. So, $\mathcal{M}$ is also a surjection. ■

Theorem 3.4.11
Suppose V and W are *f-d*. Then, $\mathcal{L}(V, W)$ is *f-d* and $\dim \mathcal{L}(V, W) = (\dim V)(\dim W)$.

Proof 5. By Theorem 3.4.10 and Theorem 3.4.9, we know $\dim \mathcal{L}(V, W) = \dim \mathbb{F}^{m,n}$. Further by Theorem 3.3.10, we know $\dim \mathbb{F}^{m,n} = (m)(n)$. As $\dim V = n$ and $\dim W = m$, so we have

$$\dim \mathcal{L}(V, W) = (\dim V)(\dim W).$$

■

Definition 3.4.12 (Matrix of a Vector, $\mathcal{M}(v)$). Suppose $v \in V$ and $v_1, \cdots, v_n$ is a basis of V. The *matrix* *of* v with respect to this basis is the $n \times 1$ matrix

$$\mathcal{M}(v) = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix},$$

where $c_1, \cdots, c_n$ are scalars *s.t.* $v = c_1 v_1 + \cdots + c_n v_n$.

Theorem 3.4.13 $\mathcal{M}(T)_{\cdot,k} = \mathcal{M}(v_k)$
Suppose $T \in \mathcal{L}(V, W)$ and $v_1, \cdots, v_n$ is a basis of V and $w_1, \cdots, w_m$ is a basis of W. Let $1 \leq k \leq n$. Then, the k^{th} column of $\mathcal{M}(T)$, which is denoted by $\mathcal{M}(T)_{\cdot,k}$, equals $\mathcal{M}(v_k)$.

Proof 6. This theorem is an immediate result by definitions of matrix of a linear mapping and a vector. ■

Theorem 3.4.14
Suppose $T \in \mathcal{L}(V, W)$ and $v \in V$. Suppose $v_1, \cdots, v_n$ is a basis of V and $w_1, \cdots, w_m$ is a basis of W. Then, $\mathcal{M}(Tv) = \mathcal{M}(T)\mathcal{M}(v)$.

Proof 7. Note that $v = c_1 v_1 + \cdots + c_n v_n$, so we have $Tv = c_1 T v_1 + \cdots + c_n T v_n$. So, by Theorem 3.4.13, we know

$$\begin{aligned} \mathcal{M}(Tv) &= c_1 \mathcal{M}(Tv_1) + \cdots + c_n \mathcal{M}(Tv_n) \\ &= c_1 \mathcal{M}(T)_{\cdot,1} + \cdots + c_n \mathcal{M}(T)_{\cdot,n} \\ &= \mathcal{M}(T)\mathcal{M}(v). \end{aligned}$$

The final equality holds due to our interpretation of matrix multiplication as column linear combinations (Theorem 3.3.14(3)) ■

Remark. *$\mathcal{M} : \mathbb{F}^n \to \mathbb{F}^{n,1}$ is an isomorphism:*

$$v = c_1 v_1 + \cdots + c_n v_n \longmapsto \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}.$$

Proof 8. Suppose $\mathcal{M}(v) = 0 : \ \mathcal{M}(c_1 v_1 + \cdots + c_n v_n) = 0$. So, we have $c_1 w_1 + \cdots + c_n w_n = 0$. Since $w_1, \cdots, w_n$ is a basis, $c_1 = \cdots = c_n = 0$. So, $v = 0$. Therefore, $\operatorname{null} \mathcal{M} = \{0\}$, and so $\mathcal{M}$ is injective. □

Now, prove $\mathcal{M}$ is surjective. Note that $\forall \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}$, we have $\mathcal{M}(c_1 v_1 + \cdots + c_n v_n) = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}$. So, $\mathcal{M}$ is a surjection. □

Finally, its' trivial to prove $\mathcal{M}$ is a linear map. □

Since $\mathcal{M}$ is both surjective and injective, $\mathcal{M}$ is an isomorphism. ■

Definition 3.4.15 (Operator). A linear map from a vector space to itself is called an *operator*.

Notation 3.4.16. The notation $\mathcal{L}(V)$ denotes the set of all operators on V. So, $\mathcal{L}(v) = \mathcal{L}(V, V)$.

Theorem 3.4.17
Suppose V is *f-d* and $T \in \mathcal{L}(V)$. Then, the following are equivalent: (a) T is invertible; (b) T is injective; and (c) T is surjective.

Proof 9.

1. Clearly (a) implies (b). $\square$

2. Suppose (b): T is injective. So, $\operatorname{null} T = \{0\}$. Then, by Fundamental Theorem of Linear Maps, we know
$$\dim V = \dim \operatorname{null} T + \dim \operatorname{range} T = 0 + \dim \operatorname{range} T.$$
Since $\dim \operatorname{range} T = \dim V$, we know T is surjective. $\square$

3. Suppose (c): T is surjective. So, $\operatorname{range} T = V$. Then, by Fundamental Theorem of Linear maps, we have
$$\dim \operatorname{null} T = \dim V - \dim \operatorname{range} T = 0.$$
So, $\operatorname{null} T = \{0\}$, and thus T is injective. Since T is surjective and injective, T is invertible.

■

Example 3.4.18 Show that for each polynomial $q \in \mathcal{P}(\mathbb{R})$, there exists a polynomial $p \in \mathcal{P}(\mathbb{F})$ such that $((x^2 + 5x + 7)p)'' = q$.

Proof 10. We know that every non-zero polynomial must have a degree of m. So, we can think of this problem under $\mathcal{P}_m(\mathbb{R})$. Note that
$$((x^2 + 5x + 7)p)'' = 2p + (4x + 10)p' + (x^2 + 5x + 7)p'' = q.$$
Therefore, the degree of p and q should be the same. Define $T : \mathcal{P}_m(\mathbb{R}) \to \mathcal{P}_m(\mathbb{R})$ as
$$Tp = ((x^2 + 5x + 7)p)''.$$
Then, T is an operator on $\mathcal{P}_m(\mathbb{R})$. Consider $Tp = 0$. We have $ax + b = (x^2 + 5x + 7)p$. Note that only when $p = 0$, the equation above holds. So, it must be that $p = 0$ when $Tp = 0$. That is, $\operatorname{null} T = \{0\}$, and so T is injective. By Theorem 3.4.18, we know T is also surjective, and so our proof is complete. ■

3.5 Duality

Definition 3.5.1 (Linear Functional). A *linear functional* on V is a linear map from V to $\mathbb{F}$. That is, a linear functional is an element of $\mathcal{L}(V, \mathbb{F})$.

Example 3.5.2

1. Fix $(c_1, \cdots, c_n) \in \mathbb{F}^n$. Define $\varphi : \mathbb{F}^n \to \mathbb{F}$ by $\varphi(x_1, \cdots, x_n) = c_1 x_1 + \cdots + c_n x_n$. Then, φ is a linear functional on $\mathbb{F}^n$.
2. Define $\varphi : \mathcal{P}(\mathbb{R}) \to \mathbb{R}$ as $\varphi(p) = 3p''(5) + 7p(4)$.
3. Define $\varphi : \mathcal{P}(\mathbb{R}) \to \mathbb{R}$ as $\varphi(p) = \displaystyle\int_0^1 p(x)\mathrm{d}x$.

Definition 3.5.3 (Dual Space/V'/V^*). The *dual space* of V, denoted as V', is the vector space of all linear functionals on V. In other words, $V' = \mathcal{L}(V, \mathbb{F})$.

Theorem 3.5.4
Suppose V is *f-d*. Then, V' is also *f-d* and $\dim V' = \dim V$.

Proof 1. Note that for a general linear map, $\mathcal{L}(V, W) \cong \mathbb{F}^{m,n}$. So, $\mathcal{L}(V, \mathbb{F}) = V' \cong \mathbb{F}^{1,n}$. Hence,

$$\dim V' = \dim \mathbb{F}^{1,n} = 1 \cdot n = n = \dim V.$$

■

Definition 3.5.5 (Dual Basis). If $v_1, \cdots, v_n$ is a basis of V, then the *dual basis* of $v_1, \cdots, v_n$ is the list $\varphi_1, \cdots, \varphi_n$ of elements of V', where each φ_j is the linear functional on V *s.t.*

$$\varphi_j(v_k) = \begin{cases} 1 & \text{if } k = j \\ 0 & \text{if } k \neq j \end{cases}.$$

Example 3.5.6 Find the dual basis of $e_1, \cdots, e_n \in \mathbb{F}^n$

Solution 2.

$$\begin{array}{cccc} \varphi_1(e_1) = 1 & \varphi_2(e_1) = 0 & \cdots & \varphi_n(e_1) = 0 \\ \varphi_1(e_2) = 0 & \varphi_2(e_2) = 1 & \cdots & \varphi_n(e_2) = 0 \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_1(e_n) = 1 & \varphi_2(e_n) = 0 & \cdots & \varphi_n(e_n) = 1 \end{array}$$

Define φ_j as

$$\varphi_j(x) = \varphi_j(x_1, \cdots, x_n) = x_1\varphi_j(e_1) + \cdots + x_j\varphi_j(e_j) + \cdots + x_n\varphi_j(e_n) = x_j.$$

□

Theorem 3.5.7
Suppose V is *f-d*. Then, the dual basis of a basis of V is a basis of V'.

Proof 3. Suppose $v_1, \cdots, v_n$ is a basis of V and $\varphi_1, \cdots, \varphi_n$ denotes the dual basis. Since we've shown $\dim V = \dim V'$ in Theorem 3.5.4, we only need to show $\varphi_1, \cdots, \varphi_n$ is L.I.. Select $c_1\varphi_1 + \cdots + c_n\varphi_n = 0$. Then,

$$(c_1\varphi_1 + \cdots + c_n\varphi_n)(v) = 0 \quad \forall\, v \in V.$$

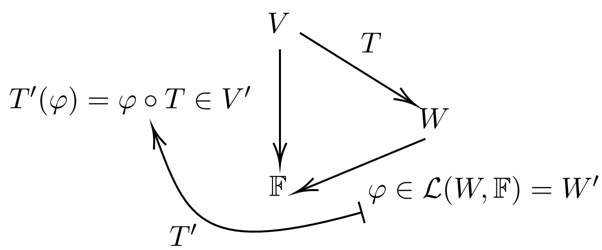
Suppose $v = v_1 + \cdots + v_n$, then

$$(c_1\varphi_1 + \cdots + c_n\varphi_n)(v_j) = c_j \quad \text{for } j = 1, \cdots, n.$$

So, $(c_1\varphi_1 + \cdots + c_n\varphi_n)(v) = c_1 + \cdots + c_n = 0$. So, it must be that $c_1 = \cdots = c_n = 0$. Therefore, $\varphi_1, \cdots, \varphi_n$ is L.I. and our proof is complete. ■

Definition 3.5.8 (Dual Map). If $T \in \mathcal{L}(V, W)$, then the *dual map* of T is the linear map $T' \in \mathcal{L}(W', V')$ defined by $T'(\varphi) = \varphi \circ T$ for $\varphi \in W'$.

Remark. *The following diagram represents dual map (but not an exact representation).*



Also, dual map is a linear map, so it is additive and homogeneous.

1. $T'(\varphi + \psi) = (\varphi + \psi) \circ T = \varphi \circ T + \psi \circ T = T'(\varphi) + T'(\psi)$.
2. $T'(\lambda\varphi) = (\lambda\varphi) \circ T = \lambda(\varphi \circ T) = \lambda T'(\varphi)$.

Example 3.5.9 Suppose $D : \mathcal{P}(\mathbb{R}) \to \mathcal{P}(\mathbb{R})$ as $Dp = p'$.

1. Define a linear functional $\varphi : \mathcal{P}(\mathbb{R}) \to \mathbb{R}$ as $\varphi(p) = p(3)$. Find $D'(\varphi)$.

   Solution 4.

   $$(D'(\varphi))(p) = (\varphi \circ D)(p) = \varphi(Dp) = \varphi(p') = p'(3).$$

   □

2. Define $\varphi : \mathcal{P}(\mathbb{R}) \to \mathbb{R}$, a linear functional, as $\varphi(p) = \int_0^1 p(x)\,\mathrm{d}x$. Find $D'(\varphi)$.

   Solution 5.

   $$(D'(\varphi))(p) = (\varphi \circ D)(p) = \varphi(Dp) = \varphi(p') = \int_0^1 p'(x)\,\mathrm{d}x = p(1) - p(0).$$

   □

Theorem 3.5.10 Algebraic Properties of Dual Maps

1. $(S+T)' = S' + T' \quad \forall\, S, T \in \mathcal{L}(V, W)$
2. $(\lambda T)' = \lambda T' \quad \forall\, T \in \mathcal{L}(V, W)$
3. $(ST)' = T'S' \quad \forall\, T \in \mathcal{L}(U, V)$ and $S \in \mathcal{L}(V, W)$

Proof 6.

1. $(S+T)' \in \mathcal{L}(W', V')$. Let $\varphi \in W'$. Then,

$$(S+T)'(\varphi) = \varphi \circ (S+T) = \varphi \circ S + \varphi \circ T = S'(\varphi) + T'(\varphi) = (S' + T')(\varphi). \qquad \square$$

2. $(\lambda T)' \in \mathcal{L}(W', V')$. Let $\varphi \in W'$. Then,

$$(\lambda T)'(\varphi) = \varphi \circ (\lambda T) = \lambda(\varphi \circ T) = \lambda T'(\varphi) = (\lambda T')(\varphi). \qquad \square$$

3. $(ST)' \in \mathcal{L}(W', U')$. Let $\varphi \in W'$. Then,

$$(ST)'(\varphi) = \varphi \circ (ST) = \varphi \circ (S \circ T) = (\varphi \circ S) \circ T = (S'(\varphi)) \circ T = T'(S'(\varphi)) = (T'S')(\varphi).$$

■

Definition 3.5.11 (Transpose/A^t). The transpose of a matrix A, denoted A^t, is the matrix obtained from A by interchanging the rows and columns. i.e., $(A^t)_{k,j} = A_{j,k}$.

Remark. *Transpose is additive and homogeneous. That is,* $(A + C)^t = A^t + C^t$ *and* $(\lambda A)^t = \lambda A^t$.

Theorem 3.5.12
If A is an $m \times n$ matrix and C is an $n \times p$ matrix, then $(AC)^t = C^t A^t$.

Proof 7. Note that

$$(AC)^t_{k,j} = (AC)_{j,k} = \sum_{r=1}^{n} A_{j,r} C_{r,k} = \sum_{r=1}^{n} (C^t)_{k,r} (A^t)_{r,j} = (C^t A^t)_{k,j}$$

■

Theorem 3.5.13
Suppose $T \in \mathcal{L}(V, W)$. Then, $\mathcal{M}(T') = (\mathcal{M}(T))^t$.

Proof 8. Suppose $v_1, \cdots, v_n$ is a basis of V, $w_1, \cdots, w_m$ is a basis of W, $\varphi_1, \cdots, \varphi_n$ is a basis of V', and $\psi_1, \cdots, \psi_m$ is a basis of W'. Let $A = \mathcal{M}(T)$ and $C = \mathcal{M}(T')$. Since $T'(\psi_j) = C_{1,j}\varphi_1 + \cdots + C_{n,j}\varphi_n$ and $T'(\psi_j) = \psi_j \circ T$, we have $\psi_j \circ T = C_{1,j}\varphi_1 + \cdots + C_{n,j}\varphi_n$. Consider

$$(\psi_j \circ T)(v_k) = (C_{1,j}\varphi_1 + \cdots + C_{n,j}\varphi_n)(v_k) = C_{k,j}\varphi_k(v_k) = C_{k,j}.$$

Also, we have

$$(\psi_j \circ T)(v_k) = \psi_j(Tv_k) = \psi_j(A_{1,k}w_1 + \cdots + A_{m,k}w_m) = \psi_j(A_{j,k}w_j) = A_{j,k}(\varphi_j(w_j)) = A_{j,k}.$$

Therefore, we have $A_{j,k} = C_{k,j}$, and thus $A = C^t$. So, $\mathcal{M}(T) = (\mathcal{M}(T'))^t$. ■

Definition 3.5.14 (Annihilator/U^0). For $U \subseteq V$, the *annihilator* of U, denoted as U^0, is defined by

$$U^0 = \{\varphi \in V' \mid \varphi(u) = 0 \quad \forall u \in U\}.$$

Theorem 3.5.15
Suppose $U \subseteq V$. Then U^0 is a subspace of V'.

Proof 9.

1. $0 \in U^0$: Since $0(u) = 0 \quad \forall u \in U$, then $0 \in U^0$. □
2. Let $\varphi, \psi \in U^0$. Then,
$$(\varphi + \psi)(u) = \varphi(u) + \psi(u) = 0.$$
So, $\varphi + \psi \in U^0$. □
3. Let $\lambda \in \mathbb{F}$ and $\varphi \in U^0$. Then
$$(\lambda\varphi)(u) = \lambda\varphi(u) = \lambda \cdot 0 = 0.$$
So, $\lambda\varphi \in U^0$.

■

Lemma 3.5.16 Suppose V is *f-d* vector space. If U is a subspace of V and $S \in \mathcal{L}(U, W)$, then there exists $T \in \mathcal{L}(V, W)$ *s.t.* $Tu = Su \quad \forall u \in U$.

Proof 10. Suppose $u_1, \cdots, u_m$ is a basis of U. Then, we can extend it to a basis of V as $u_1, \cdots, u_m, v_{m+1}, \cdots, v_n$. Define $T \in \mathcal{L}(V, W)$ as $Tu_i = Su_i$, $Tv_j = 0$, where $i = 1, \cdots, m$ and $j = m+1, \cdots, n$. Note that

$$\begin{aligned} Tu &= T(a_1u_1 + \cdots + a_mu_m) \\ &= a_1Tu_1 + \cdots + a_mTu_m \\ &= a_1Su_1 + \cdots + a_mSu_m \\ &= S(a_1u_1 + \cdots + a_mu_m) = Su. \end{aligned}$$

Therefore, we've found such a T. ■

Theorem 3.5.17
Let V be *f-d* and U be a subspace of V, then $\dim U + \dim U^0 = \dim V$.

Proof 11. Let $i \in \mathcal{L}(U, V)$ as $i(u) = u \quad \forall u \in U$. Then, $i' \in \mathcal{L}(V', U')$. So, by Fundamental Theorem of Linear Map, we know

$$\dim V' = \dim \operatorname{null} i' + \dim \operatorname{range} i'. \tag{9}$$

By Theorem 3.5.4, we know $\dim V = \dim V'$ Note that $U^0 = \{\varphi \in V' \mid \varphi(u) = 0 \quad \forall u \in U\}$ and

$$\begin{aligned} \operatorname{null} i' &= \{\varphi \in V' \mid i'(\varphi) = 0\} \\ &= \{\varphi \in V' \mid \varphi \circ i = 0\} \\ &= \{\varphi \in V' \mid (\varphi \circ i)(u) = 0 \quad \forall u \in U\} \\ &= \{\varphi \in V' \mid \varphi(u) = 0 \quad \forall u \in U\} \end{aligned}$$

So, $U^0 = \operatorname{null} i'$, and thus $\dim \operatorname{null} i' = \dim U^0$.

Further, if $\varphi \in U'$, then $\varphi : U \to \mathbb{F}$. By Lemma 3.5.16, φ can be extended to $\psi \in V'$ with $\psi(u) = \varphi(u) \quad \forall u \in U$. Note that $i'(\psi) = \psi \circ i$, so $(\psi \circ i)(u) = \psi(u) = \varphi(u) \quad \forall u \in U$. Then, $\exists \psi \in V'$ *s.t.* $i'(\psi) = \varphi$. So, $\varphi \in \operatorname{range} U'$. So, $\dim \operatorname{range} i' = \dim U' = \dim U$.

Substitute $\dim V' = \dim V$, $\dim \operatorname{null} i' = \dim U^0$, and $\dim \operatorname{range} i' = \dim U$ to Equation (9), we get

$$\dim V = \dim U^0 + \dim U.$$

■

Theorem 3.5.18 The Null Space of T'
Suppose V and W are *f-d* and $T \in \mathcal{L}(V, W)$. Then,

1. $\operatorname{null} T' = (\operatorname{range} T)^0$
2. $\dim \operatorname{null} T' = \dim \operatorname{null} T + \dim W - \dim V$

Proof 12.

1. ($\subseteq$) Suppose $\varphi \in \operatorname{null} T' \subseteq W'$. Then, $T'(\varphi) = \varphi \circ T = 0 \in V'$. So, we know

   $$(\varphi \circ T)(v) = 0 \quad \forall v \in V. \quad \text{i.e., } \varphi(Tv) = 0.$$

   Note that $Tv \in \operatorname{range} T$. By definition, we have $\varphi \in (\operatorname{range} T)^0$ □

   ($\supseteq$) Suppose $\varphi \in (\operatorname{range} T)^0$. Then, $\varphi(w) = 0 \quad \forall w \in \operatorname{range} T$. That is, $\varphi(Tv) = 0 \quad \forall v \in V$. So, $(\varphi \circ T)(v) = 0 \quad \forall v \in V$. Hence, we know $\varphi \circ T = T'(\varphi) = 0 \in V'$. Thus, $\varphi \in \operatorname{null} T'$ ■

2. $$\begin{aligned} \dim \operatorname{null} T' &= \dim(\operatorname{range} T)^0 \\ &= \dim W - \dim \operatorname{range} T \\ &= \dim W - (\dim V - \dim \operatorname{null} T) \\ &= \dim W - \dim V + \dim \operatorname{null} T. \end{aligned}$$

■

Theorem 3.5.19
Suppose V and W are *f-d* and $T \in \mathcal{L}(V, W)$. Then, T is surjective if and only if T' is injective.

Proof 13.

($\Rightarrow$) Suppose T is surjective. Then, $\dim \operatorname{range} T = W$. So, $(\operatorname{range} T)^0 = \{0\}$. Hence,

$$\dim \operatorname{null} T' = \dim(\operatorname{range} T)^0 = 0.$$

Thus, T' is injective. $\square$

($\Leftarrow$) Suppose T' is injective. Then,

$$\dim \operatorname{null} T' = 0.$$

So, $\dim(\operatorname{range} T)^0 = \dim \operatorname{null} T' = 0$. Then, $(\operatorname{range} T)^0 = \{0\}$. So, $\dim \operatorname{range} T = W$, and thus T is surjective. $\blacksquare$

Theorem 3.5.20 The Range of T'
Suppose V and W are *f-d* and $T \in \mathcal{L}(V, W)$. Then,

1. $\dim \operatorname{range} T' = \dim \operatorname{range} T$
2. $\operatorname{range} T' = (\operatorname{null} T)^0$

Proof 14.

1. By Fundamental Theorem of Linear Map, we have

$$\begin{aligned}\dim \operatorname{range} T' &= \dim W' - \dim \operatorname{null} T' \\ &= \dim W' - \dim(\operatorname{range} T)^0 \\ &= \dim W' - \dim W' + \dim \operatorname{range} T \\ &= \dim \operatorname{range} T.\end{aligned}$$

$\blacksquare$

2. Suppose $\varphi \in \operatorname{range} T' \subseteq V'$. Then, $\exists \psi \in W'$ *s.t.* $T'(\psi) = \psi \circ T = \varphi$. Let $v \in \operatorname{null} T$. Then,

$$\varphi(v) = (\psi \circ T)(v) = \psi(Tv) = \psi(0) = 0.$$

Then, $\varphi \in (\operatorname{null} T)^0$. So, $\operatorname{range} T' \subseteq (\operatorname{null} T)^0$. $\square$

Note that

$$\dim \operatorname{range} T' = \dim \operatorname{range} T = \dim V - \dim \operatorname{null} T = \dim(\operatorname{null} T)^0.$$

Then, $\operatorname{range} T' \subseteq (\operatorname{null} T)^0$ and $\dim \operatorname{range} T' = \dim(\operatorname{null} T)^0$, so it must be that $\operatorname{range} T' = (\operatorname{null} T)^0$.

$\blacksquare$

Theorem 3.5.21
Suppose V and W are *f-d* and $T \in \mathcal{L}(V, W)$. Then, T is injective if and only if T' is surjective.

Proof 15.
($\Rightarrow$) If T is injective, $\operatorname{null} T = \{0\}$. So,

$$\dim \operatorname{null} T = \dim V - \dim(\operatorname{null} T)^0 = \dim V - \dim \operatorname{range} T' = 0.$$

So, $\dim \operatorname{range} T' = \dim V = \dim V'$. Then, T' is surjective. $\square$

($\Leftarrow$) If T' is surjective, $\dim \operatorname{range} T' = \dim V' = \dim V$. So,

$$\dim \operatorname{null} T = \dim V - \dim(\operatorname{null} T)^0 = \dim V - \dim \operatorname{range} T' = 0.$$

Then, $\operatorname{null} T = \{0\}$, and so T is injective. ■

Definition 3.5.22 (Row Rank & Column Rank). Suppose A is an $m \times n$ matrix with entries in $\mathbb{F}$.

1. The *row rank* of A is the dimension of the span of the rows of A in $\mathbb{F}^{1,n}$.
2. The *column rank* of A is the dimension of the span of the columns of A in $\mathbb{F}^{m,1}$.

Theorem 3.5.23
Suppose V and W are *f-d* and $T \in \mathcal{L}(V, W)$. Then, $\dim \operatorname{range} T$ equals the column rank of $\mathcal{M}(T)$.

Proof 16. Suppose $v_1, \cdots, v_n$ is a basis of V and $w_1, \cdots, w_m$ is a basis of W. Then,

$$Tv_k = A_{1,k}w_1 + \cdots + A_{m,k}w_m$$

and thus

$$\mathcal{M}(Tv_k) = \begin{pmatrix} A_{1,k} \\ \vdots \\ A_{m,k} \end{pmatrix} \in \mathbb{F}^{m,1}$$

Therefore, $\mathcal{M}(T) = \begin{pmatrix} \mathcal{M}(Tv_1) & \cdots & \mathcal{M}(Tv_n) \end{pmatrix}$. Note that $\operatorname{range} T = \operatorname{span}(Tv_1, \cdots, Tv_n)$.

Define $\mathcal{M} : \operatorname{span}(Tv_1, \cdots, Tv_n) \to \operatorname{span}(\mathcal{M}(Tv_1), \cdots, \mathcal{M}(Tv_n))$ as $w \mapsto \mathcal{M}(w)$.

1. $\mathcal{M}$ is surjective: Note that

   $$c_1\mathcal{M}(Tv_1) + \cdots + c_n\mathcal{M}(Tv_n) = \mathcal{M}(c_1Tv_1 + \cdots + c_nTv_n).$$

   Since $c_1Tv_1 + \cdots + c_nTv_n \in \operatorname{range} T$, we know $\mathcal{M}$ is surjective. □

2. $\mathcal{M}$ is injective: Let

   $$\mathcal{M}(c_1Tv_1 + \cdots + c_nTv_n) = 0. \tag{10}$$

   We can reduce $c_1Tv_1 + \cdots + c_nTv_n$ to a basis $Tv_{j_1}, \cdots, Tv_{j_m}$. Then, Equation (10) becomes $\mathcal{M}(a_1Tv_{j_1} + \cdots + a_mTv_{j_m}) = 0$. By definition of matrix, we know $\begin{pmatrix} a_1 \\ \vdots \\ a_m \end{pmatrix} = 0$. So, $a_1 = \cdots = a_m = 0$ and $a_1Tv_{j_1} + \cdots + a_mTv_{j_m} = 0$. So, $\mathcal{M}$ is injective. □

Since $\mathcal{M}$ is both surjective and injective, $\mathcal{M}$ is a bijection. Thus, $\mathcal{M}$ is an isomorphism between $\operatorname{span}(Tv_1, \cdots, Tv_n)$ and $\operatorname{span}(\mathcal{M}(Tv_1), \cdots, \mathcal{M}(Tv_n))$. In other words,

$$\operatorname{span}(Tv_1, \cdots, Tv_n) \cong \operatorname{span}(\mathcal{M}(Tv_1), \cdots, \mathcal{M}(Tv_n)).$$

Then, $\dim \operatorname{span}(Tv_1, \cdots, Tv_n) = \dim \operatorname{span}(\mathcal{M}(Tv_1), \cdots, \mathcal{M}(Tv_n))$. That is,

$$\dim \operatorname{range} T = \text{column rank of } T.$$

■

Theorem 3.5.24 Row Rank Equals Column Rank
Suppose $A \in \mathbb{F}^{m,n}$. Then, the row rank of A equals the column rank of A.

Proof 17. Define $T : \mathbb{F}^{n,1} \to \mathbb{F}^{m,1}$ by $Tx = Ax$. Then, $\mathcal{M}(T) = A$, where $\mathcal{M}(T)$ is computed with respect to the standard basis of $\mathbb{F}^{n,1}$ and $\mathbb{F}^{m,1}$. Note that

$$\begin{aligned}
\text{column rank of } A &= \text{column rank of } \mathcal{M}(T) \\
&= \dim \operatorname{range} T && \text{Theorem 3.5.23} \\
&= \dim \operatorname{range} T' && \text{Theorem 3.5.20(1)} \\
&= \text{column rank of } \mathcal{M}(T') \\
&= \text{column rank of } A^t && \text{Theorem 3.5.13} \\
&= \text{row rank of } A
\end{aligned}$$

■

Definition 3.5.25 (Rank). The *rank* of a matrix $A \in \mathbb{F}^{m,n}$ is the column rank of A, denoted as $\operatorname{rank} A$.

3.6 Quotients of Vector Spaces

Definition 3.6.1 ($v + U$/Affine Subset). Suppose $v \in V$ and U is a subspace of V. Then

$$v + U \coloneqq \{v + u \mid u \in U\}.$$

An *affine subset* of V is a subset of V of the form $v + U$ for some $v \in V$ and some subspace U of V. The affine subset is said to be *parallel* to U.

Definition 3.6.2 (Quotient Space, V/U). Suppose U is a subspace of V. Then the quotient space V/U is the set of all affine subsets of V parallel to U. In other words,

$$V/U \coloneqq \{v + U \mid v \in V\}.$$

Example 3.6.3 If $U = \left\{(x, 2x) \in \mathbb{R}^2 \mid x \in \mathbb{R}\right\}$, then $\mathbb{R}^2/U$ is the set of all lines in $\mathbb{R}^2$ with slope of 2.

Theorem 3.6.4
Suppose U is a subspace of V and $v, w \in V$. Then, the following are equivalent:

1. $v - w \in U$
2. $v + U = w + U$
3. $(v + U) \cap (w + U) \neq \varnothing$

Proof 1.

1. We want to show (1) $\implies$ (2). Suppose $v - w \in U$. Note that $v + u = w + ((v - w) + u)$. Since $v - u$ and $u \in U$, we have $(v - w) + u \in U$. So, $v + u \in w + U$. Similarly, we can show that $w + u \in v + U$. Then, we have $v + U = w + U$. $\square$
2. Now, we want to show (2) $\implies$ (3): Suppose $v + U = w + U$. Then, we have $(v + U) \cap (w + U) \neq \varnothing$, which is evident from the assumption. $\square$
3. Finally, we will show (3) $\implies$ (1). Suppose $(v + U) \cap (w + U) \neq \varnothing$. Then, $\exists u_1, u_2 \in U$ *s.t.* $v + u_1 = w + u_2$. So we have $v - w = u_2 - u_1 \in U$.

$\blacksquare$

Definition 3.6.5 (Addition & Scalar Multiplication on V/U). Suppose U is a subspace of V. Then, *addition* and *scalar multiplication* is defined on V/U by

$$(v + U) + (w + U) = (v + w) + U$$

and

$$\lambda(v + U) = (\lambda v) + U$$

for $v, w \in U$ and $\lambda \in \mathbb{F}$.

Theorem 3.6.6
Suppose U is a subspace of V. Then, V/U, with the operations of addition and scalar multiplication defined above, is a vector space.

Proof 2.

1. Addition on V/U makes sense.

   Note the addition can be written in the language of mapping as $+ : V/U \times V/U \to V/U$. So, we have $(v+U, w+U) \mapsto (v+w)+U$. Suppose $\exists\, \hat{v}.\hat{w} \in V$ *s.t.* $v+U = \hat{v}+U$ and $w+U = \hat{w}+U$. Note that $v-\hat{v} \in U$ and $w-\hat{w} \in U$ by Theorem 3.6.4. Then, $(v-\hat{v})+(w-\hat{w}) \in U$. So, we have $(v+w)-(\hat{v}+\hat{w}) in U$. Further, by Theorem 3.6.4, we have

   $$(v+w)+U = (\hat{v}+\hat{w})+U. \qquad \square$$

2. Scalar multiplication on V/U makes sense.

   We can write the scalar multiplication on V/U as a mapping: $\cdot : \mathbb{F} \times V/U \to V/U$ defined as $(\lambda, v+U) \mapsto \lambda v + U$. Suppose $\exists\, \hat{v} \in V$ *s.t.* $v+U = \hat{v}+U$. So we know $v - \hat{v} \in U$, and thus $\lambda(v-\hat{v}) = \lambda v - \lambda\hat{v} \in U$. By Theorem 3.6.4, we then have $(\lambda v)+U = (\lambda \hat{v})+U$. Thus, the scalar multiplication makes sense. $\square$

3. additive identity: $0+U = U$. $\square$

4. additive inverse: $(-v)+U$. $\square$

5. commutativity:

   $$\begin{aligned}(v+U)+(w+U) &= (v+w)+U = (w+v)+U \\ &= (w+U)+(v+U). \qquad \square\end{aligned}$$

6. associativity:

   $$\begin{aligned}[(v+U)+(w+U)]+(x+U) &= [(v+w)+U]+(x+U) \\ &= [(v+w)+x]+U \\ &= [v+(w+x)]+U \\ &= (v+U)+[(w+x)+U] \\ &= (v+U)+[(x+U)+(x+U)]. \qquad \square\end{aligned}$$

7. multiplicative identity: $1 \cdot (v+U) = (1 \cdot v)+U = v+U$. $\square$

8. distributivity:

   $$\begin{aligned}a[(v+U)+(w+U)] &= a[(v+w)+U] \\ &= a(v+w)+U \\ &= (av+aw)+U \\ &= (av+U)+(aw+U) \\ &= a(v+U)+a(w+U).\end{aligned}$$

$$\begin{aligned}(a+b)(v+U) &= (a+b)v+U \\ &= (av+bv)+U \\ &= (av+U)+(bv+U) \\ &= a(v+U)+b(v+U)\end{aligned}$$

■

Definition 3.6.7 (Quotient Map). Suppose U is a subspace of V. The *quotient map* π is the linear map $\pi : V \to V/U$ defined by $\pi(v) := v+U \quad \forall v \in V$.

Remark. *Here are some properties of the quotient map:*

1. $\pi(v)$ *is defined* $\forall v \in V$. *Thus,* π *is surjective.*
2. $\operatorname{null}\pi = \{v \in V \mid \pi(v) = 0\}$. *If* $\pi(v)=0$, *then* $v+U = U = 0+U$. *So,* $v-0 \in U$ *by Theorem 3.6.4. Then,* $v \in U$. *So,* $\operatorname{null}\pi \subseteq U$. *Further,* $\forall v \in U$, *if* $\pi(v) = 0$, *then* $v \in \operatorname{null}\pi$, *then* $U \subseteq \operatorname{null}\pi$. *So,* $U = \operatorname{null}\pi$.
3. $\pi(v+w) = (v+w)+U = (v+U)+(w+U) = \pi(v)+\pi(w)$.
4. $\pi(\lambda v) = (\lambda v)+U = \lambda(v+U) = \lambda\pi(v)$.

Theorem 3.6.8
Suppose V is *f-d* and U is a subspace of V. Then

$$\dim V/U = \dim V - \dim U.$$

Proof 3. By Fundamental Theorem of Linear Map, we have

$$\dim V = \dim \operatorname{null}\pi + \dim \operatorname{range}\pi. \tag{11}$$

Since $\operatorname{null}\pi = U$ from the Remark, we have $\dim\operatorname{null}\pi = \dim U$. Further, since π is surjective as mentioned in the Remark, $\operatorname{range}\pi = V/U$. Hence, $\dim\operatorname{range}\pi = \dim V/U$. Therefore, Equation (11) becomes

$$\dim V = \dim U + \dim V/U,$$

or we have

$$\dim V/U = \dim V - \dim U$$

■

Definition 3.6.9 ($\tilde{T}$). Suppose $T \in \mathcal{L}(V,W)$. Define $\tilde{T} : V/(\operatorname{null} T) \to W$ by $\tilde{T}(v+\operatorname{null}T) - Tv$.

Proof 4.

1. This definition makes sense

   Suppose $u,v \in V$ *s.t.* $u+\operatorname{null}T = v+\operatorname{null}T$. By Theorem 3.6.4, we know $u-v \in \operatorname{null}T$. Then, $T(u-v) = 0$, or $Tu = Tv$. □

2. $\tilde{T}$ is a linear map.

$$\begin{aligned}\tilde{T}[(u+\operatorname{null}T)+(v+\operatorname{null}T)] &= \tilde{T}[(u+v)+\operatorname{null}T] \\ &= T(u+v) \\ &= Tu+Tv = \tilde{T}(u+\operatorname{null}T)+\tilde{T}(v+\operatorname{null}T). \qquad \square\end{aligned}$$

$$\begin{aligned}\tilde{T}[\lambda(u + \operatorname{null} T)] &= \tilde{T}(\lambda u + \operatorname{null} T)\\ &= T(\lambda u)\\ &= \lambda T u\\ &= \lambda T(u + \operatorname{null} T).\end{aligned}$$

■

Theorem 3.6.10
Suppose $T \in \mathcal{L}(V, W)$. Then,

1. $\tilde{T}$ is injective.
2. $\operatorname{range} \tilde{T} = \operatorname{range} T$.
3. $V/(\operatorname{null} T) \cong \operatorname{range} T$.

Proof 5.

1. Suppose $v \in V$ and $\tilde{T}(v + \operatorname{null} T) = 0$. Then, $Tv = 0$. So, $v \in \operatorname{null} T$, or $v - 0 \in \operatorname{null} T$. By Theorem 3.6.4, we then have $v + \operatorname{null} T = 0 + \operatorname{null} T$. Then, it implies $\operatorname{null} \tilde{T} = 0$. So, $\tilde{T}$ is injective. □
2. By definition of $\tilde{T}$, it must be $\operatorname{range} \tilde{T} = \operatorname{range} T$. □
3. Note that $\dim V/(\operatorname{null} T) = \dim \operatorname{null} \tilde{T} + \dim \operatorname{range} \tilde{T} = 0 + \dim \operatorname{range} T$. Then, by Theorem 3.4.9, we know two vector spaces are isomorphic if and only if their dimensions are equal. Then,

$$V(\operatorname{null} T) \cong \operatorname{range} T.$$

■

4 Eigenvectors and Invariant Subspaces

4.1 Invariant Subspaces

Theorem 4.1.1
Suppose V is *f-d* with $\dim V = n \geq 1$. Then, $\exists$ $1-$dimensional subspaces $U_1, \cdots, U_n$ of V *s.t.*

$$V = U_1 \oplus \cdots \oplus U_n.$$

Proof 1. Choose a basis $v_1, \cdots, v_n$ of V. Then, we know $V = \operatorname{span}(v_1) + \cdots + \operatorname{span}(v_n)$. Also, $\forall v \in V$, we have $v = a_1 v_1 + \cdots + a_n v_n$ with $a_j v_j \in \operatorname{span}(v_j)$. Set $a_1 v_1 + \cdots + a_n v_n = 0$. Since $v_1, \cdots, v_n$ is a basis, it must be $a_1 = \cdots = a_n = 0$. Then,

$$V = \operatorname{span}(v_1) \oplus \cdots \oplus \operatorname{span}(v_n).$$

■

Theorem 4.1.2
Suppose $U_1, \cdots, U_m$ are *f-d* subspaces of V *s.t.* $U_1 + \cdots + U_m$ is a direct sum. Then, $U_1 \oplus \cdots \oplus U_m$ is *f-d* and

$$\dim U_1 \oplus \cdots \oplus U_m = \dim U_1 + \cdots + \dim U_m.$$

Proof 2. Suppose $u_{k,1}, \cdots, u_{k.j_k}$ is a basis of the subspace U_k. Then, any vector in $\bigoplus_{i=1}^{m} U_i$ is in the form of $u_1 + \cdots + u_m, \quad u_j \in U_j$. Also,

$$u_i = \sum_{k=1}^{j_i} a_{i,k} u_{i,k}.$$

So,

$$u_1 + \cdots + u_m = \sum_{k=1}^{j_1} a_{1,k} u_{1,k} + \cdots + \sum_{k=1}^{j_m} a_{m,k} u_{m,k}.$$

Then, $u_1 + \cdots + u_m$ is a linear combination of $u_{1,1}, \cdots, u_{j,m}$. So, the direct sum is *f-d.* □

Further, suppose

$$\sum_{k=1}^{j_1} a_{1,k} u_{1,k} + \cdots + \sum_{k=1}^{j_m} a_{m,k} u_{m,k} = 0.$$

Since $U_1 + \cdots + U_m$ is a direct sum, it must be

$$\sum_{k=1}^{j_1} a_{1,k} u_{1,k} = \cdots = \sum_{k=1}^{j_m} a_{a,k} u_{m,k} = 0.$$

Since we selected bases, $a_{1,k} = \cdots = a_{m,k} = 0$. So, $u_{1,1}, \cdots, u_{m,j_m}$ is a basis of $U_1 \oplus \cdots \oplus U_m$. Then,

$$\dim U_1 \oplus \cdots \oplus U_m = \dim U_1 + \cdots + \dim U_m.$$

■

Definition 4.1.3 (Invariant Subspace). Suppose $T \in \mathcal{L}(V)$. A subspace U of V is called *invariant* under T if $u \in U$ implies $Tu \in U$.

Example 4.1.4 Suppose $T \in \mathcal{L}(V)$. Show that each of the following subspaces of V is invariant under T:

1. $\{0\}$

   Proof 3. $T0 = 0 \in \{0\}$ ■

2. V

   Proof 4. $u \in V \implies Tu \in V$ ■

3. $\text{null}\, T$

   Proof 5. $u \in \text{null}\, T \implies Tu = 0 \in \text{range}\, T$ ■

4. $\text{range}\, T$

   Proof 6. $u \in \text{range}\, T \implies Tu \in \text{range}\, T$ ■

Example 4.1.5 Suppose $T \in \mathcal{L}(\mathcal{P}(\mathbb{R}))$ is defined by $Tp = p'$. Then, $\mathcal{P}_4(\mathbb{R})$ is invariant under T.

Proof 7. Note that $Tp_4) \in \mathcal{P}_4(\mathbb{R})$. Then, $\mathcal{P}_4(\mathbb{R})$ is invariant under T. ■

Definition 4.1.6 (Eigenvalue). Suppose $T \in \mathcal{L}(V)$. A number $\lambda \in \mathbb{F}$ is called an *eigenvalue* of T if $\exists\, v \in V$ *s.t.* $v \neq 0$ and $Tv = \lambda v$.

Corollary 4.1.7 T has a $1-$dimensional invariant subspace if and only if T has an eigenvalue.

Proof 8.

($\Rightarrow$) Suppose $\text{span}(v)$ is invariant under T. Let U be defined as $U = \{\lambda v \mid \lambda \in \mathbb{F}\} = \text{span}(v)$. Then. U is the invariant subspace under T and $\dim U = 1$. Then, $\forall v \in V$, we have $Tv \in U$. Hence, $\exists \lambda \in \mathbb{F}$ *s.t.* $Tv = \lambda v$. Then, λ is an eigenvalue. □

($\Leftarrow$) Suppose $\lambda \in \mathbb{F}$ is an eigenvalue. Then, $Tv = \lambda v$. Hence, $\text{span}(v)$ is a 1 =dimensional invariant subspace under T. ■

Theorem 4.1.8 Equivalent Conditions to be an Eigenvalue
Suppose V is *f-d*, $T \in \mathcal{L}(V)$, and $\lambda \in \mathbb{F}$. Then, the following are equivalent:

1. λ is an eigenvalue of T.
2. $T - \lambda I$ is not injective.
3. $T - \lambda I$ is not surjective.
4. $T - \lambda I$ is not invertible.

Proof 9.

1. (1) $\implies$ (2): Suppose λ is an eigenvalue of T. Then, $\exists v \in V$ *s.t.* $v \neq 0$ and $Tv - \lambda v$. So, $Tv - \lambda v = (T - \lambda I)v = 0$. Since $v \neq 0$, $\text{null}\,(T - \lambda I) \neq \{0\}$, and thus T is not injective. □

2. Note that $T - \lambda I$ is an operator by itself. By Theorem 3.4.17, we know (2), (3), and (4) are equivalent.

3. (4) $\implies$ (1): Suppose $T - \lambda I$ is not invertible. Then, it is not injective. So, $\exists v \neq 0$ *s.t.* $(T - \lambda I)v = 0$. That is, $Tv - \lambda Iv = Tv - \lambda v = 0$. So, $Tv = \lambda v$. Then, λ is an eigenvalue of T.

■

Definition 4.1.9 (Eigenvector). Suppose $T \in \mathcal{L}(V)$ and $\lambda \in \mathbb{F}$ is an eigenvalue of T. A vector $v \in V$ is called an *eigenvector* of T corresponding to λ if $v \neq 0$ and $Tv = \lambda v$.

Corollary 4.1.10 A vector $v \in V$ with $v \neq 0$ is an eigenvector of T with respect to λ if and only if $v \in \text{null}\,(T - \lambda I)$.

Proof 10. Note that $Tv = \lambda v$ if and only if $(T - \lambda I)v = 0$. ■

Example 4.1.11 Suppose $T \in \mathcal{L}(\mathbb{F}^2)$ is defined by $T(w, z) = (-z, w)$.

1. Find the eigenvalues and eigenvectors of T if $\mathbb{F} = \mathbb{R}$.

   Solution 11.

   Let $T(2, z) = \lambda(w, z)$. So, $(-z, w) = (\lambda w, \lambda z)$. Then, solve $\begin{cases} -z = \lambda w \\ w = \lambda z \end{cases}$.

   Then, we have $\lambda^2 z + z = 0$. If $z \neq 0$, $\lambda^2 + 1 = 0$. This equation has no solutions on $\mathbb{R}$. So T has no eigenvalues. If $w = 0, z = 0$, then $T(w, z) = T(0.0) = T0$. By definition, T has no eigenvalues. □

2. Find the eigenvalues and eigenvectors of T if $\mathbb{F} = \mathbb{C}$.

   Solution 12.

   Applying similar rational, $z \neq 0$ and solve $\lambda^2 + 1 = 0$. Then, we have $\lambda = \pm \mathrm{i}$. If $\lambda = \mathrm{i}$, then $-z = \mathrm{i}w$. So, $v = (w, z) = (w, -\mathrm{i}w)$. If $\lambda - \mathrm{i}$, then $-z = -\mathrm{i}w$, or $z = \mathrm{i}w$. So, $v = (w, \mathrm{i}w)$. □

Theorem 4.1.12
Let $T \in \mathcal{L}(V)$. Suppose $\lambda_1, \cdots, \lambda_m$ are distinct eigenvalues of T and $v_1, \cdots, v_m$ are corresponding eigenvectors. Then, $v_1, \cdots, v_m$ is L.I..

Proof 13. Suppose for the sake of contradiction that $v_1, \cdots, v_m$ is linearly dependent. Let k be the smallest positive integer *s.t.* $v_k \in \text{span}(v_1, \cdots, v_{k-1})$. Then, $v_k = a_1 v_1 + \cdots + a_{k-1} v_{k-1}$. Applying T, we have

$$\lambda_k v_k = a_1 \lambda_1 v_1 + \cdots + a_{k-1} \lambda_{k-1} v_{k-1}. \tag{12}$$

Since $v_k = a_1 v_1 + \cdots + a_{k-1} v_{k-1}$, we also have

$$\lambda_k v_k = a_1 \lambda_k v_1 + \cdots + a_{k-1} \lambda_k v_{k-1}. \tag{13}$$

So, by Equation (13)-(12), we have

$$0 = a_1(\lambda_k - \lambda_1) v_1 + \cdots + a_{k-1}(\lambda_k - \lambda_{k-1}) v_{k-1}.$$

By assumption, $v_1, \cdots, v_{k-1}$ is L.I.. Then, it must be that $a_1 = \cdots = a_{k-1} = 0$ since $\lambda_1, \cdots, \lambda_k$ are distinct eigenvalues. Therefore, $v_k = a_1 v_1 + \cdots + a_{k-1} v_{k-1} = 0$. ⋇ This contradicts with the fact that v_k is an eigenvector, which cannot be 0. So,it must be that $v_1, \cdots, v_m$ are L.I. ■

Theorem 4.1.13
Suppose V is *f-d*. Then, each operator on V has at most $\dim V$ distinct eigenvalues.

Proof 14. Let $T \in \mathcal{L}(V)$. Suppose $\lambda_1, \cdots, \lambda_m$ are distinct eigenvalues of T. Let $v_1, \cdots, v_m$ be corresponding eigenvectors. By Theorem 4.1.12, we know $v_1, \cdots, v_m$ is L.I.. Further by Theorem 2.3.5, we know $\dim \operatorname{span}(v_1, \cdots, v_m) \leq \dim V$. That is, $m \leq \dim V$ as desired. ■

4.2 Eigenvectors and Upper-Triangular Matrices

Definition 4.2.1 (T^m). Suppose $T \in \mathcal{L}(V)$ and m is a positive integer. Then, T^m is defined by

$$T^m := \underbrace{T \cdots T}_{m \text{ times}}.$$

Specially, T^0 is defined to be the identity operator I on V. Further, if T is invertible with inverse T^{-1}, then T^{-m} is defined by $T^{-m} := (T^{-1})^m$.

Theorem 4.2.2

$$T^m T^n = T^{m+n}; \qquad (T^m)^n = T^{mn}.$$

Definition 4.2.3 ($p(T)$). Suppose $T \in \mathcal{L}(V)$ and $p \in \mathcal{P}(\mathbb{F})$ is a polynomial given by

$$p(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_m z^m, \quad z \in \mathbb{F}.$$

Then, $p(T)$ is the operator defined by

$$p(T) := a_o I + a_1 T + a_2 T^2 + \cdots + a_m T^m.$$

Example 4.2.4 Suppose $D \in \mathcal{L}(\mathcal{P}(\mathbb{R}))$ is the differentiation operator defined by $Dq = q'$ and p is the polynomuial defined by $p(x) = 7 - 3x + 5x^2$. Find $p(D)$ and $(p(D))q$.

Solution 1.

$$\begin{aligned} p(D) &= 7I - 3D + 5D^2 \\ (p(D))q &= (7I - 3D + 5D^2)q \\ &= 7Iq - 3Dq + 5D^2 q \\ &= 7q - 3q' + 5q''. \end{aligned}$$

□

Theorem 4.2.5
If we fix an operator $T \in \mathcal{L}(V)$, then the function from $\mathcal{P}(\mathbb{F})$ to $\mathcal{L}(V)$ given by $p \mapsto p(T)$ is linear.

Proof 2. Suppose $f : \mathcal{P}(\mathbb{F}) \to \mathcal{L}(V)$ is defined by $p \mapsto p(T)$. Suppose

$$p = a_0 + a_1 z + \cdots + a_m z^m \mapsto a_0 I + a_1 T + \cdots + a_m T^m$$

and

$$q = b_0 + b_1 z + \cdots + b_m z^m \mapsto b_0 I + b_1 T + \cdots + b_m T^m.$$

Then,

$$\begin{aligned} f(p+q) &= (a_0+b_0)I + (a_1+b_1)T + \cdots + (a_m+b_m)T^m \\ &= (a_0I + a_1T + \cdots + a_mT^m) + (b_0I + b_1T + \cdots + b_mT^m) \\ &= f(p) + f(q). \end{aligned}$$

Further, suppose $\lambda \in \mathbb{F}$, then

$$\begin{aligned} f(\lambda p) &= \lambda a_0 I + \lambda a_1 T + \cdots + \lambda a_m T^m \\ &= \lambda(a_0 I + a_1 T + \cdots + a_m T^m) \\ &= \lambda f(p). \end{aligned}$$

■

Definition 4.2.6 (Product of Polynomials). If $p, q \in \mathcal{P}(\mathbb{F})$, then $pq \in \mathcal{P}(\mathbb{F})$ is the polynomial defined by $(pq)(z) := p(z)q(z)$ for $z \in \mathbb{F}$.

Remark. $(pq)(z) = p(z)q(z) = q(z)p(z) = (qp)(z)$ *for* $z \in \mathbb{F}$.

Theorem 4.2.7 Multiplicative Properties
Suppose $p, q \in \mathcal{P}(\mathbb{F})$ and $T \in \mathcal{L}(V)$. Then

1. $(pq)(T) = p(T)q(T)$
2. $p(T)q(T) = q(T)p(T)$

Proof 3.

1. Suppose $p(z) = \sum_{j=0}^{m} a_j z^j$ and $q(z) = \sum_{k=0}^{n} b_k z^k$. Then

$$(pq)(z) = p(z)q(z) = \sum_{j=0}^{m} a_j z^j \sum_{k=0}^{n} b_k z^k = \sum_{j=0}^{m}\sum_{k=0}^{n} a_j b_k z^{j+k}$$

So, by definition, we have

$$p(T)q(T) = \sum_{j=0}^{m}\sum_{k=0}^{n} a_j b_k T^{j+k} = \left(\sum_{j=0}^{m} a_j T^j\right) \cdot \left(\sum_{k=0}^{n} b_k T^k\right) = p(T)q(T). \qquad \square$$

2. Similar to the Remark,

$$p(T)q(T) = (pq)(T) = (qp)(T) = q(T)p(T).$$

■

Theorem 4.2.8 Fundamental Theorem of Algebra
Every non-constant polynomial with complex coefficients has a zero.

Theorem 4.2.9 Existence of Eigenvalues
Every operator on a *f-d*, non-zero, complex vector space has an eigenvalue.

Proof 4. Let V be a complex vector space with dimension $n > 0$. Suppose $T \in \mathcal{L}(V)$. Choose $v \in V$ *s.t.* $v \neq 0$. Then, $v, Tv, T^2v, \cdots, T^nv$ is linearly dependent because $\dim V = n$ but the length of the list is $n+1 > n$. Hence, $\exists\, a_0, a_1, \cdots, a_n$ not all $0 \in \mathbb{C}$ *s.t.*

$$0 = a_0v + a_1Tv + \cdots + a_nT^nv \tag{14}$$

By Fundamental Theorem of Algebra (Theorem 4.2.8), we have

$$a_0 + a_1z + \cdots + a_nz^n = c(z-\lambda_1)\cdots(z-\lambda_m)$$

with $c \in \mathbb{C},\ c \neq 0$, and $\lambda_j \in \mathbb{C}$. Then, Equation (14) becomes

$$\begin{aligned} 0 &= a_0v + a_1Tv + \cdots + a_nT^nv \\ &= (a_0I + a_1T + \cdots + a_nT^n)v \\ &= c(T-\lambda_1I)\cdots(T-\lambda_mI)v \end{aligned}$$

Since $v \neq 0$ and $c \neq 0$, it must be some $T - \lambda_iI = 0$. Thus, $T = \lambda_iI$, and λ_i is an eigenvalue of T. ■

Definition 4.2.10 (Diagonal of a Matrix). The *diagonal of a square matrix* consists of the entires along the line from the upper left corner to the bottom right corner.

Definition 4.2.11 (Upper-Triangular Matrix). A matrix is called *upper-triangular* if all the entires below the diagonal equal 0. Typically, we present an upper triangular matrix in the form

$$\begin{pmatrix} \lambda_1 & & * \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}.$$

Theorem 4.2.12 Conditions for Upper-Triangular Matrix
Suppose $T \in \mathcal{L}(V)$ and $v_1, \cdots, v_n$ is a basis of V. Then, the following are equivalent:

1. the matrix of T with respect to $v_1, \cdots, v_n$ is upper triangular.
2. $Tv_j \in \operatorname{span}(v_1, \cdots, v_j)$ for each $j = 1, \cdots, n$
3. $\operatorname{span}(1, \cdots, v_j)$ is invariant under T for each $j = 1, \cdots, n$.

Proof 5.

1. First, we will show (1) $\iff$ (2).

   Suppose $\mathcal{M}(T) = \begin{pmatrix} A_{1,1} & \cdots & A_{1,n} \\ & \ddots & \vdots \\ 0 & & A_{n,n} \end{pmatrix}$. Then,

$$\begin{aligned} Tv_1 &= A_{1,1}v_1 \\ Tv_2 &= A_{1,2}v_1 + a_{2,2}v_2 \\ &\vdots \\ Tv_j &= A_{1,j}v_1 + \cdots + A_{j,j}v_j. \end{aligned}$$

So, $Tv_j \in \operatorname{span}(v_1, \cdots, v_j)$. The reverse implication is trivial to prove. □

2. (3) $\implies$ (2) is obvious and trivial to prove.

3. Lastly, we want to show (2) $\implies$ (3).

   Note that for each fixed $j = 1, \cdots, n$, we have

$$\begin{aligned} &Tv_1 \in \operatorname{span}(v_1) \subseteq \operatorname{span}(v_1, \cdots, v_j) \\ &Tv_2 \in \operatorname{span}(v_1, v_2) \subseteq \operatorname{span}(v_1, \cdots, v_j) \\ &\qquad \vdots \\ &Tv_j \in \operatorname{span}(v_1, \cdots, v_j) \end{aligned}$$

   Let $v \in \operatorname{span}(v_1, \cdots, v_j)$. Then, v is a linear combination of $v_1, \cdots, v_j$, then

$$Tv \in \operatorname{span}(v_1, \cdots, v_j).$$

   That is, $\operatorname{span}(v_1, \cdots, v_j)$ is invariant under T.

■

Definition 4.2.13 (Quotient Operator). Suppose $T \in \mathcal{L}(V)$ and U is a subspace of V invariant under T. The *quotient operator* $T/U \in \mathcal{L}(V/U)$ is defined by $(T/U)(v+U) \coloneqq Tv + U$.

Proof 6. The definition makes sense, and here is the proof. If $v + U = w + U$, then $v - w \in U$. So, $T(v-w) \in U$ since U is invariant. That is, $Tv - Tw \in U$. Then, $Tv + U = Tw + U$. ■

Theorem 4.2.14
Suppose U is a subspace of V. Let $v_1 + U, \cdots, v_m + U$ be a basis of V/U and $u_1, \cdots, u_n$ be a basis of U. Then, $v_1, \cdots, v_m, u_1, \cdots, u_n$ is a basis of V.

Proof 7. Let $v \in V$. Then $v + U \in V/U$. So, $v + U = a_1v_1 + \cdots + a_mv_m + U$, uniquely. Then, by Theorem 3.6.4, we have $v - (a_1v_1 + \cdots + a_mv_m) \in U$. Therefore, $v - (a_1v_1 + \cdots + a_mv_m) = b_1u_1 + \cdots + b_nu_n$, uniquely. So, $v = a_1v_1 + \cdots + a_mv_m + b_1u_1 + \cdots + b_nu_n$. uniquely. By definition, $v_1, \cdots, v_m, u_1, \cdots, u_n$ is a basis of V. ■

Theorem 4.2.15
Suppose V is a *f-d* complex vector space and $T \in \mathcal{L}(V)$. Then, T has an upper-triangular matrix with respect to some basis of V.

Proof 8.
$\boxed{\text{Base Case}}$ When $\dim V = 1$, the implication holds.

$\boxed{\text{Inductive Steps}}$ Suppose the implication is true for some complex vector space with dimension of $n-1$. Let $\dim V = n$ and v_1 be any eigenvector of T. Suppose $U = \operatorname{span}(v_1)$. Then, U is invariant under T. Note that $\dim V/U = \dim V - \dim U = n-1$, so we can use the inductive hypothesis on the quotient operator $T/U \in \mathcal{L}(V/U)$. Then, $\exists$ a basis $v_2 + U, \cdots, v_n + U \in V/U$ *s.t.* T/U has an upper-triangular matrix. By Theorem 4.2.12, we have

$$(T/U)(v_j + U) \in \operatorname{span}(v_2 + U, \cdots, c = v_j + U) \quad \text{for } j \in \{2, \cdots, n\}.$$

So, $Tv_j + U = (c_2v_2 + \cdots + c_jv_j) + U$. Then,

$$Tv_j - (c_2v_2 + \cdots + c_jv_j) \in U = \operatorname{span}(v_1).$$

So, $Tv_j - (c_2v_2 + \cdots + c_jv_j) = c_1v_1$ for some $c_1 \in \mathbb{F}$. Then, $Tv_j = c_1v_1 + c_2v_2 + \cdots + c_jv_j$. So, $Tv_j \in \operatorname{span}(v_1, \cdots, v_j)$ for $j \in \{1, \cdots, n\}$. Since by Theorem 4.2.14, $v_1, \cdots, v_n$ is a basis of V, further by Theorem 4.2.12, T has an upper-triangular matrix with respect to $v_1, \cdots, v_n$. So, the implication is true for $\dim V = n$.

Since the implication is true for $\dim V = 1$ and is true for $\dim V = n$ whenever it is hold for $\dim V = n - 1$, by the Principle of Mathematical Induction, the implication is true for all positive integers n. Hence, the proof is complete. ■

4.3 Eigenspaces and Diagonal Matrices

Definition 4.3.1 (Diagonal Matrix). A *diagonal matrix* is a square matrix that is 0 everywhere except possibly along the diagonal.

Definition 4.3.2 (Eigenspace, $E(\lambda, T)$). Suppose $T \in \mathcal{L}(V)$ and $\lambda \in \mathbb{F}$. The *eigenspace* of T corresponding to λ, denoted $E(\lambda, T)$, is defined by

$$E(\lambda, T) = \text{null}\,(T - \lambda I).$$

In other words, $E(\lambda, T)$ is the set of all eigenvectors of T corresponding to λ, along with the 0 vector.

Theorem 4.3.3 Sum of Eigenspaces is a Direct Sum
Suppose V is *f-d* and $T \in \mathcal{L}(V)$. Suppose also that $\lambda_1, \cdots, \lambda_m$ are distinct eigenvalues of T. Then

$$E(\lambda_1, T) + \cdots + E(\lambda_m, T)$$

is a direct sum. Further

$$\dim E(\lambda_1, T) + \cdots + \dim E(\lambda_m, T) \leq \dim V.$$

Proof 1. Suppose $u_1 + \cdots + u_m = 0$, where $u_j \in E(\lambda_j, T)$. If some $u_i \neq 0$, then $u_1 + \cdots + u_m$ can never be 0 because $u_1, \cdots, u_m$, as eigenvectors corresponding to distinct eigenvalues, is L.I.. Hence, the only way for $u_1 + \cdots + u_m$ to be 0 is by taking $u_1 = \cdots = u_m = 0$. Hence, we know $E(\lambda_1, T) + \cdots + E(\lambda_m, T)$ is a direct sum. $\square$

By Theorem 4.1.2, we know

$$\begin{aligned}\dim E(\lambda_1, T) + \cdots + \dim E(\lambda_m, T) &= \dim E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T) \\ &\leq \dim V.\end{aligned}$$

■

Definition 4.3.4 (Diagonalizable). An operator $T \in \mathcal{L}(V)$ is called *diagonalizable* if the operator has a diagonal matrix with respect to some basis of V.

Theorem 4.3.5 Conditions Equivalent to Diagonalizability
Suppose V is *f-d* and $T \in \mathcal{L}(V)$. Let $\lambda_1, \cdots, \lambda_m$ denote the distinct eigenvalues of T. Then, the following are equivalent:

1. T is diagonalizable.
2. V has a basis consisting of eigenvectors of T.
3. $\exists\, 1-$dimensional subspaces $U_1, \cdots, U_n$ of V, each invariant under T, *s.t.* $V = U_1 \oplus \cdots \oplus U_n$.
4. $V = E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T)$.
5. $\dim V = \dim E(\lambda_1, T) + \cdots + \dim E(\lambda_m, T)$.

Remark. *To prove this theorem, we will prove (1) $\iff$ (2), (2) $\iff$ (3), (2) $\implies$ (4), (4) $\implies$ (5), and (5) $\implies$ (2).*

Proof 2.

1. (1) $\iff$ (2): By definition, we know T is diagonalizable if and only if $\exists$ a basis $v_1, \cdots, v_n$ of T *s.t.*

$$\mathcal{M}(T) = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix},$$

which holds if and only if we have $Tv_1 = \lambda_1 v_1, \cdots, Tv_n = \lambda_n v_n$. i.e., $v_1, \cdots, v_n$ are eigenvectors of T. $\square$

2. (2) $\implies$ (3): Suppose $v_1, \cdots, v_n$ is a basis of V. Then, for some $v \in V$, we have $v = a_1 v_1 + \cdots + a_n v_n$. So, we know $V = \text{span}(v_1) + \cdots + \text{span}(v_n)$. Further, let $a_1 v_1 + \cdots + a_m v_m = 0$. Since $v_1, \cdots, v_n$ is a basis, it must be $a_1 = \cdots = a_m = 0$. So, there is only one way to represent 0. So,

$$V = \text{span}(v_1) \oplus \cdots \oplus \text{span}(v_n).$$

Now, we want to show each $\text{span}(v_j)$ is invariant. Consider $T(c_j v_j) = c_j T v_j = c_j \lambda_j v_j \in \text{span}(v_j)$. So, $\text{span}(v_j)$ is invariant. $\square$

3. (3) $\implies$ (2): Suppose $\exists$ $1-$dimensional subspaces $U_1, \cdots, U_n$ of V, each invariant under T, *s.t.* $V = U_1 \oplus \cdots \oplus U_n$. Then, $\forall v \in V$, we have $v = a_1 u_1 + \cdots + a_n u_n$ uniquely. Then, $u_1, \cdots, u_n$ is a basis of V. Since $U_1, \cdots, U_n$ are $1-$dimensional invariant subspaces, $u_1, \cdots, u_n$ are the eigenvalues. $\square$

4. (2) $\implies$ (4): Suppose V has a basis consisting of eigenvectors of T. Then, $v = a_1 v_1 + \cdots + a_n v_n$ is a linear combination of eigenvectors of T. By definition, $E(\lambda_j, T)$ contains the eigenvectors corresponding to λ_j. Further since $\lambda_1, \cdots, \lambda_m$ is distinct, corresponding eigenvectors are L.I.. Then, $E(\lambda_j, T) \cap E(\lambda_i, T) = \{0\}$ if $i \neq j$. Then, we have

$$v = a_1 v_1 + \cdots + a_n v_n \in E(\lambda_1, T) + \cdots + E(\lambda_m, T).$$

Hence, $V = \mathbb{E}(\lambda_1, T) + \cdots + E(\lambda_m, T)$. Further by Theorem 4.3.3, we have

$$V = \mathbb{E}(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T). \quad \square$$

5. (4) $\implies$ (5): This conclusion can be deduced from Theorem 4.3.3 and its proof.

6. (5) $\implies$ (2): Suppose $\dim V = \dim E(\lambda_1, T) + \cdots + \dim E(\lambda_m, T)$. Select B_j, the basis of $E(\lambda_j, T)$ for $j = 1, \cdots, m$. Denote $\dim V = n$. Then, if we put these bases together as $B_1, \cdots, B_m$, we can write the collection as $v_1, \cdots, v_n$. Suppose $a_1 v_1 + \cdots + a_n v_n = 0$. Let u_j denote the sum of all the terms $a_k v_k$ *s.t.* $v_k \in E(\lambda_j, T)$. Then, the equation becomes $u_1 + \cdots + u_m = 0$ and each $u_j \in E(\lambda_j, T)$. Since eigenvectors corresponding to distinct eigenvalues are L.I., it must be that $u_1 = \cdots = u_m = 0$. Further, by definition of u_j, and since $u_k's$ are bases of $E(\lambda_j, T)$, it must be $a_1 = \cdots = a_n = 0$. So, we know $v_1, \cdots, v_n$ is L.I.. Further, since $\text{len}(v_1, \cdots, v_n) = n = \dim V$, we know that $v_1, \cdots, v_n$ is a basis of V.

$\blacksquare$

Theorem 4.3.6
If $T \in \mathcal{L}(V)$ has $\dim V$ distinct eigenvalues, then T is diagonalizable.

Proof 3. Suppose $T \in \mathcal{L}(V)$ has $\dim V$ distinct eigenvalues: $\lambda_1, \cdots, \lambda_{\dim V}$. Then, it has $v_1, \cdots, v_{\dim V}$ as corresponding eigenvectors and is L.I.. Note that $\operatorname{len}(v_1, \cdots, v_{\dim V}) = \dim V$. So, $v_1, \cdots, v_{\dim V}$ is a basis of V. By Theorem 4.3.5, with respect to this basis consisting of eigenvectors, T has a diagonal matrix. ■

Example 4.3.7 The *Fibonacci Sequence* $F_1, F_2, \cdots$ is defined by

$$F_1 = F_2 = 3 \qquad \text{and} \qquad F_n = F_{n-2} + F_{n-1} \quad \text{for } n \geq 3.$$

Define $T \in \mathcal{L}(\mathbb{R}^2)$ by $T(x, y) = (y, x + y)$.

1. Show that $T^n(0, 1) = (F_n, F_{n+1})$ for each $n \in \mathbb{Z}^+$.

   Proof 4.

   - Base Case: Note that $T(0, 1) = (1, 1) = (F_1, F_2)$.
   - Inductive Process: Suppose $T^{n-1}(0, 1) = (F_{n-1}, F_n)$. Then,

   $$\begin{aligned} T^n = [T(T^{n-1})](0,1) &= T[T^{n-1}(0,1)] \\ &= T(F_{n-1}, F_n) \\ &= (F_n, F_{n-1} + F_n) \\ &= (F_n, F_{n+1}). \end{aligned}$$

   So, $T^n(0, 1) = (F_n, F_{n+1}) \quad \forall\, n \in \mathbb{Z}^+$ by Principle of Mathematical Induction.

   ■

2. Find the eigenvalues of T.

   Solution 5.

   Suppose $T(x, y) = \lambda(x, y)$. So, $(y, x+y) = (\lambda x, \lambda y)$. Solve $\begin{cases} y = \lambda x \\ x + y = \lambda y \end{cases}$. That is, $x + \lambda x = \lambda^2 x$, or $\lambda^2 x - \lambda x - x = 0$. It follows $x \neq 0$, so solving $\lambda^2 - \lambda - 1 = 0$, we get

   $$\lambda_1 = \frac{1 + \sqrt{5}}{2} \qquad \text{and} \qquad \lambda_2 = \frac{1 - \sqrt{5}}{2}.$$

   □

3. Since T has two eigenvalues, T should have a basis of $\mathbb{R}^2$ consisting of eigenvectors. Find the basis.

   Solution 6.

When $\lambda_1 = \dfrac{1+\sqrt{5}}{2}$, we have $y = \lambda x = \dfrac{1+\sqrt{5}}{2}x$. So, $v_1 = \left(x, \dfrac{1+\sqrt{5}}{2}x\right) = x\left(1, \dfrac{1+\sqrt{5}}{2}\right)$. That is,

$$v_1 = \left(1, \frac{1+\sqrt{5}}{2}\right).$$

Similarly, we have

$$v_2 = \left(1, \frac{1-\sqrt{5}}{2}\right).$$

Further, it follows that

$$\mathcal{M}(T, v_1, v_2) = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}.$$

□

4. Find F_n using an expression of n only.

Solution 7.

Note that $(0,1) = \dfrac{1}{\sqrt{5}}(v_1 - v_2)$. So, we have

$$\begin{aligned}
T^n(0,1) &= T^n\left(\frac{1}{\sqrt{5}}(v_1 - v_2)\right) \\
&= \frac{1}{\sqrt{5}}T^n(v_1 - v_2) \\
&= \frac{1}{\sqrt{5}}(T^n v_1 - T^n v_2) \\
&= \frac{1}{\sqrt{5}}(\lambda_1^n v_1 - \lambda_2^n v_2) \\
&= \frac{1}{\sqrt{5}}\left(\left(\frac{1+\sqrt{5}}{2}\right)^n \left(1, \frac{1+\sqrt{5}}{2}\right) - \left(\frac{1-\sqrt{5}}{2}\right)^n \left(1, \frac{1-\sqrt{5}}{2}\right)\right) \\
&= \frac{1}{\sqrt{5}}\left(\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n, \left(\frac{1+\sqrt{5}}{2}\right)^{n+1} - \left(\frac{1-\sqrt{5}}{2}\right)^{n+1}\right) \\
&= (F_n, F_{n+1}).
\end{aligned}$$

So, we have

$$F_n = \frac{1}{\sqrt{5}}\left(\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n\right).$$

□

5 Inner Product Spaces

5.1 Inner Products and Norms

Definition 5.1.1 (Dot Product). For $x, y \in \mathbb{R}^n$, the *dot product* of x and y, denoted $x \cdot y$, is defined by

$$x \cdot y = x_1 y_1 + \cdots + x_n y_n,$$

where $x = (x_1, \cdots, x_n)$ and $y = (y_1, \cdots, y_n)$.

Theorem 5.1.2 Properties of dot Product

1. $x \cdot x = x_1^2 + \cdots + x_n^2 \geq 0 \quad \forall x \in \mathbb{R}^n$.
2. $x \cdot x = 0$ if and only if $x = 0$.
3. For $y \in \mathbb{R}^n$, define $f : \mathbb{R}^n \to \mathbb{R}$ as $x \mapsto x \cdot y$. Then, f is linear.
4. $\forall x, y \in \mathbb{R}^n,\ x \cdot y = y \cdot x$.

Proof 1. Properties #1, #2, and #4 are trivial to prove, so the proof is omitted. Here we complete a proof for property #3, linearity of dot product. Let $f : \mathbb{R}^n \to \mathbb{R}$ be defined as $x \mapsto x \cdot y$ for a fixed $y \in \mathbb{R}^n$. Note that

$$\begin{aligned} f(a+b) = (a+b) \cdot y = (a_1 + b_1)y_1 + \cdots + (a_n + b_n)y_n) \\ = (a_1 y_1 + \cdots + a_n y_n) + (b_1 y_1 + \cdots + b_n y_n) \\ = f(a) + f(b). \end{aligned}$$

Further, notice that

$$\begin{aligned} f(\lambda x) = (\lambda x) \cdot y = (\lambda x_1) y_1 + \cdots + (\lambda x_n) y_n \\ = \lambda(x_1 y_x + \cdots + x_n y_n = \lambda f(x). \end{aligned}$$

■

Remark. *For $w, z \in \mathbb{C}^n$, we define the dot product of w and z, denoted as $\langle w, z \rangle$, as*

$$\langle w, z \rangle = w_1 \overline{z_1} + \cdots + w_n \overline{z_n}.$$

Definition 5.1.3 (Inner Product). An *inner product* on V is a function that takes each ordered pair (u, v) of elements of V to a number $\langle u, v \rangle \in \mathbb{F}$ and has the following properties:

1. positivity: $\langle v, v \rangle \geq 0 \quad \forall v \in V$.
2. definiteness: $\langle v, v \rangle = 0$ if and only if $v = 0$.
3. additivity in first slot: $\langle u + v, w \rangle = \langle u, w \rangle + \langle v, w \rangle \quad \forall u, v, w \in V$.
4. homogeneity in first slot: $\langle \lambda u, v \rangle = \lambda \langle u, v \rangle \quad \forall \lambda \in \mathbb{F}$ and $\forall u, v \in V$.
5. conjugate symmetry: $\langle u, v \rangle = \overline{\langle v, u \rangle} \quad \forall u, v \in V$.

Example 5.1.4 Here, we offer some examples of inner product. Note that there might be multiple inner products over a vector space, as long as the following the definition and properties given in Definition 5.1.3.

1. Consider $\mathbb{C}[-1,1]$, the set of continuous real-valued functions on the interval $[-1,1]$. An inner product can be defined as $\langle f, g\rangle = \int_{-1}^{1} f(x)g(x)\,\mathrm{d}x$.

   Proof 2.

   (a) $\langle f, f\rangle = \int_{-1}^{1} f^2(x)\,\mathrm{d}x \geq 0$.

   (b) $\langle f, f\rangle = 0$ if and only if $f(x) = 0$.

   (c) Note that
   $$\begin{aligned}\langle f+g, h\rangle &= \int_{-1}^{1}[f(x)+g(x)]h(x)\,\mathrm{d}x \\ &= \int_{-1}^{1} f(x)h(x)+g(x)h(x)\,\mathrm{d}x \\ &= \int_{-1}^{1} f(x)h(x)\,\mathrm{d}x + \int_{-1}^{1} g(x)h(x)\,\mathrm{d}x \\ &= \langle f, h\rangle + \langle g, h\rangle.\end{aligned}$$

   (d) $\langle \lambda f, g\rangle = \int_{-1}^{1} \lambda f(x)g(x)\,\mathrm{d}x = \lambda\int_{-1}^{1} f(x)g(x)\,\mathrm{d}x = \lambda\langle f, g\rangle$.

   (e) $\langle g, f\rangle = \int_{-1}^{1} g(x)f(x)\,\mathrm{d}x = \int_{-1}^{1} f(x)g(x)\,\mathrm{d}x = \langle f, g\rangle = \overline{\langle f, g\rangle}$.

   ■

2. An inner product on $\mathcal{P}(\mathbb{R})$ can be defined as $\langle p, q\rangle = \int_{0}^{\infty} p(x)q(x)e^{-x}\,\mathrm{d}x$

   Proof 3. The definition makes sense. Consider the inner product as $\langle\ \rangle : \mathcal{P}(\mathbb{R})\times\mathcal{P}(\mathbb{R})\to\mathbb{R}$. Note that $\infty\notin\mathbb{R}$. So we need to show the improper integral converges to a finite number under any circumstances. Consider
   $$\frac{x^2 p(x)q(x)}{e^x} = \frac{p(x)q(x)e^{-x}}{\frac{1}{x^2}}.$$
   Note that
   $$\lim_{x\to\infty}\frac{p(x)q(x)e^{-x}}{\frac{1}{x^2}} = 0$$
   Further since $\int_0^\infty \frac{1}{x^2}\,\mathrm{d}x$ converges as it is a p-series with $p = 2 > 1$, we know it must be $\int_0^\infty p(x)q(x)e^{-x}\,\mathrm{d}x$ converges as well, by the comparison test. The remaining job is to show this definition of $\langle\ \rangle$ indeed retain the five properties as required in Definition 5.1.3, which is trivial and so is omitted. ■

Definition 5.1.5 (Inner Product Space). An *inner product space* is a vector space V along with an inner product on V.

Example 5.1.6 Euclidean Inner Product on $\mathbb{F}^n$ is defined as

$$\langle (w_1, \cdots, w_n), (z_1, \cdots, z_n) \rangle = w_1 \overline{z_1} + \cdots + w_n \overline{z_n},$$

where $(w_1, \cdots, w_n), (z_1, \cdots, z_n) \in \mathbb{F}^n$.

Notation 5.1.7. For the rest of this Chapter, without otherwise specification, V denotes an inner product space over $\mathbb{F}$.

Remark. *If not explicitly defined, the inner product is the Euclidean inner product as defined in Example 5.1.6.*

Theorem 5.1.8 Basic Properties of an Inner Product

1. For each fixed $u \in V$, the function that takes v to $\langle v, u \rangle$ is a linear map from V to $\mathbb{F}$.
2. $\langle 0, u \rangle = 0$ for every $u \in V$.
3. $\langle u, 0 \rangle = 0$ for ever $u \in V$.
4. $\langle u, v + w \rangle = \langle u, v \rangle + \langle u, w \rangle$ for all $u, v, w \in V$.
5. $\langle u, \lambda v \rangle = \overline{\lambda} \langle u, v \rangle \quad \forall \lambda \in \mathbb{F}$ and $u, v \in V$.

Proof 4.

1. Define $f : V \to \mathbb{F}$ as $v \mapsto \langle v, u \rangle$ for some fixed $u \in V$. Then,

$$f(v + w) = \langle v + w, u \rangle = \langle v, u \rangle + \langle w, u \rangle = f(v) + f(w).$$

$$f(\lambda v) = \langle \lambda v, u \rangle = \lambda \langle v, u \rangle = \lambda f(v). \qquad \square$$

2. Since f is a linear map, then $f(0) = \langle 0, u \rangle = 0. \qquad \square$
3. Note that $\langle u, 0 \rangle = \overline{\langle 0, u \rangle} = \overline{0} = 0. \qquad \square$
4. Notice

$$\begin{aligned} \langle u, v + w \rangle = \overline{\langle v + w, u \rangle} &= \overline{\langle v, u \rangle + \langle w, u \rangle} \\ &= \overline{\langle v, u \rangle} + \overline{\langle w, u \rangle} \\ &= \langle u, v \rangle + \langle u, w \rangle. \qquad \square \end{aligned}$$

5. Observe that

$$\begin{aligned} \langle u, \lambda v \rangle = \overline{\langle \lambda v, u \rangle} &= \overline{\lambda \langle v, u \rangle} \\ &= \overline{\lambda} \cdot \overline{\langle v, u \rangle} = \overline{\lambda} \langle u, v \rangle. \end{aligned}$$

■

Definition 5.1.9 (Norm). Suppose V is a vector space. Then, the *norm* of v is a real-valued function $\| \, \| : V \to \mathbb{R}$ satisfying the following properties:

1. $\|v\| \geq 0$ and $\|v\| = 0$ if and only if $v = 0$.

2. $\|\alpha v\| = |\alpha|\|v\| \quad \forall \alpha \in \mathbb{F}$ and $v \in V$.

3. triangle inequality: $\|u + v\| \leq \|u\| + \|v\| \quad \forall u, v \in \mathbb{F}$.

Definition 5.1.10 (Norm Induced by An Inner Product). For $v \in V$, $\|v\| = \sqrt{\langle v, v\rangle}$ is a *norm* on V.

Remark. *We will prove Definition 5.1.10 is indeed a definition of norm that satisfies the conditions indicated by Definition 5.1.9 throughout the rest of this section.*

Theorem 5.1.11 Basic Properties of Norms
Let $v \in V$. Then,

1. $\|v\| = 0$ if and only if $v = 0$.

2. $\|\lambda v\| = |\lambda|\|v\| \quad \forall \lambda \in \mathbb{F}$.

Proof 5.

1. $\|v\| = 0$ if and only if $\sqrt{\langle v, v\rangle} = 0$, which is equivalent to $\langle v, v\rangle = 0$. By properties of an inner product, $\langle v, v\rangle = 0$ if and only if $v = 0$. So, the proof is complete. □

2. Consider
$$\|\lambda v\|^2 = \langle \lambda v, \lambda v\rangle = \lambda \cdot \overline{\lambda}\langle v, v\rangle = |\lambda|^2\langle v, v\rangle.$$
So, $\|\lambda v\| = \sqrt{|\lambda|^2\langle v, v\rangle} = |\lambda|\|v\|$.

■

Definition 5.1.12 (Orthogonal). Two vectors $u, v \in V$ are called *orthogonal* if $\langle u, v\rangle = 0$.

Theorem 5.1.13 Orthogonality and 0

1. 0 is orthogonal to every vector in V.

2. 0 is the only vector in V that is orthogonal to itself.

Proof 6.

1. As $\langle 0, u\rangle = 0 \quad \forall u \in V$, the proof is complete. □

2. Note that $\langle v, v\rangle = 0$ if and only if $v = 0$, so we complete the proof. □

■

Theorem 5.1.14 Pythagorean Theorem
Suppose u and v are orthogonal vectors in V, then
$$\|u + v\|^2 = \|u\|^2 + \|v\|^2.$$

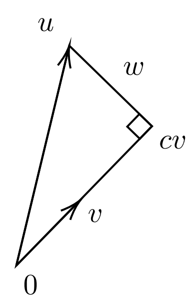
Proof 7. Note that

$$\begin{aligned}\|u+v\|^2 &= \langle u+v, u+v\rangle \\ &= \langle u, u+v\rangle + \langle v, u+v\rangle \\ &= \langle u,u\rangle + \langle u,v\rangle + \langle v,u\rangle + \langle v,v\rangle.\end{aligned}$$

Since u and v are orthogonal, $\langle u,v\rangle = \langle v,u\rangle = 0$. So, $\|u+v\|^2 = \langle u,u\rangle + \langle v,v\rangle = \|u\|^2 + \|v\|^2$. ■

Theorem 5.1.15 An Orthogonal Decomposition
Suppose $u, v \in V$, with $v \neq 0$. Set $c = \dfrac{\langle u,v\rangle}{\|v\|^2}$ and $w = u - \dfrac{\langle u,v\rangle}{\|v\|^2}v$. Then, $\rangle w, v\rangle = 0$ and $u = cv + w$.

Proof 8.



The idea is the find c, w *s.t.* $\langle v, w\rangle = 0$ and $w = u - cv$. That is, $u = w + cv$. Since $\langle v, w\rangle = 0$, then we have

$$\langle v, u - cv\rangle = 0 = \langle u - cv, v\rangle = \langle u, v\rangle - c\|v\|^2.$$

So,

$$c = \frac{\langle u,v\rangle}{\|v\|^2}$$

and

$$w = u - cv = u - \frac{\langle u,v\rangle}{\|v\|^2}v.$$

■

Theorem 5.1.16 Cauchy-Schwarz Inequality
Suppose $u, v \in V$. Then,

$$|\langle u,v\rangle| \leq \|u\|\|v\|.$$

This inequality is an equality if and only if one of u, v is a scalar multiples of the other.

Proof 9. If $v = 0$, then $|\langle u,v\rangle| = 0 = \|u\|\|v\|$. So, we can assume $v \neq 0$. Consider the orthogonal decomposition,

$$u = \frac{\langle u,v\rangle}{\|v\|^2} \cdot v + w.$$

Then, by the Pythagorean Theorem, we have

$$\begin{aligned}\|u\|^2 = \left\|\frac{\langle u,v\rangle}{\|v\|^2}\cdot v\right\|^2 + \|w\|^2 &= \frac{|\langle u,v\rangle|^2}{\|v\|^4}\|v\|^2 + \|w\|^2 \\ &= \frac{|\langle u,v\rangle|^2}{\|v\|^2} + \|w\|^2 \geq \frac{|\langle u,v\rangle|^2}{\|v\|^2}\end{aligned}$$

As $\|v\|^2 > 0$, we have $\|u\|^2\|v\|^2 \geq |\langle u,v\rangle|^2$. Further since $\|u\| \geq 0,\ \|v\| \geq 0$, and $|\langle u,v\rangle| \geq 0$, then

$$|\langle u,v\rangle| \leq \|u\|\|v\|.$$

The equality holds if and only if $\|w\|^2 = 0$. That is, $w = 0$ from the orthogonal decomposition. In other words, u and v are linearly dependent. ■

Theorem 5.1.17 Triangle Inequality
Suppose $u, v \in V$. Then

$$\|u+v\| \leq \|u\| + \|v\|.$$

This inequality is an equality if and only if one of u, v is a non-negative multiple of the other.

Proof 10. Note that

$$\begin{aligned}\|u+v\|^2 &= \langle u+v, u+v\rangle \\ &= \langle u,u\rangle + \langle u,v\rangle + \langle v,u\rangle + \langle v,v\rangle \\ &= \langle u,u\rangle + \langle v,v\rangle + \langle u,v\rangle + \overline{\langle u,v\rangle} \\ &= \|u\|^2 + \|v\|^2 + 2\,\mathrm{Re}\left(\langle u,v\rangle\right) \\ &\leq \|u\|^2 + \|v\|^2 + 2|\langle u,v\rangle| \\ &\leq \|u\|^2 + \|v\|^2 + 2\|u\|\|v\| && \text{Cauchy-Schwarz Inequality} \\ &= (\|u\| + \|v\|)^2.\end{aligned}$$

Since $\|u+v\| \geq 0,\ \|u\| \geq 0$, and $\|v\| \geq 0$, we have

$$\|u+v\| \leq \|u\| + \|v\|.$$

The equality holds if and only if $\langle u,v\rangle = \|u\|\|v\|$. That is, when u and v are linearly dependent to each other. ■

Remark. *After proving this triangle inequality, we finally, and officially, complete our proof to show the norm induced by an inner product as stated in Definition 5.1.10 is indeed a norm satisfying the formal definition of norms as stated in Definition 5.1.9.*

Theorem 5.1.18 Parallelogram Equality
Suppose $u, v \in V$. Then

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

Proof 11. Note that

$$\begin{aligned}\|u+v\|^2 + \|u-v\|^2 &= \langle u+v, u+v\rangle + \langle u-v, u-v\rangle \\ &= \langle u,u\rangle + \langle v,v\rangle + \langle u,v\rangle + \langle v,u\rangle + \langle u,u\rangle + \langle v,v\rangle - \langle u,v\rangle - \langle v,u\rangle \\ &= \|u\|^2 + \|u\|^2 + \|v\|^2 + \|v\|^2 \\ &= 2\big(\|u\|^2 + \|v\|^2\big).\end{aligned}$$

■

Theorem 5.1.19
Suppose V is a real inner product space. Then,

$$\langle u,v\rangle = \frac{\|u+v\|^2 - \|u-v\|^2}{4}.$$

Proof 12. Note that

$$\begin{aligned}\|u+v\|^2 - \|u-v\|^2 &= \langle u+v, u+v\rangle - \langle u-v, u-v\rangle \\ &= \|u\|^2 + \|v\|^2 + 2\langle u,v\rangle - \big(\|u\|^2 + \|v\|^2 - 2\langle u,v\rangle\big) \\ &= 4\langle u,v\rangle.\end{aligned}$$

So, we have

$$\langle u,v\rangle = \frac{\|u+v\|^2 - \|u-v\|^2}{4}.$$

■

Theorem 5.1.20
Suppose V is a complex inner product space. Then,

$$\langle u,v\rangle = \frac{\|u+v\|^2 - \|u-v\|^2 + \|u+\mathrm{i}v\|^2\mathrm{i} - \|u-\mathrm{i}v\|^2\mathrm{i}}{4}.$$

Proof 13. Note that

$$\begin{aligned}&\langle u+v, u+v\rangle - \langle u-v, u-v\rangle + \langle u+\mathrm{i}v, u+\mathrm{i}v\rangle\mathrm{i} - \langle u-\mathrm{i}v, u-\mathrm{i}v\rangle\mathrm{i} \\ &= 2\langle u,v\rangle + 2\langle v,u\rangle + (2\langle u,\mathrm{i}v\rangle + 2\langle \mathrm{i}v,u\rangle)\mathrm{i} \\ &= 2\langle u,v\rangle + 2\langle v,u\rangle + (-2\mathrm{i}\langle u,v\rangle + 2\mathrm{i}\langle v,u\rangle)\mathrm{i} \\ &= 2\langle u,v\rangle + 2\langle v,u\rangle + 2\langle u,v\rangle - 2\langle v,u\rangle \\ &= 4\langle u,v\rangle.\end{aligned}$$

so, we have

$$\langle u,v\rangle = \frac{\|u+v\|^2 - \|u-v\|^2 + \|u+\mathrm{i}v\|^2\mathrm{i} - \|u-\mathrm{i}v\|^2\mathrm{i}}{4}.$$

■

Theorem 5.1.21
Let U be a vector space. If $\| \ \|$ is a norm on U satisfying the parallelogram equality, then there is an inner product $\langle \ \rangle$ on U *s.t.* $\|u\| = \sqrt{\langle u, u \rangle} \quad \forall u \in U$.

5.2 Orthonormal Bases

Definition 5.2.1 (Orthonormal). A list of vectors is called *orthonormal* if each vector in the list has norm 1 and is orthogonal to all the other vectors in the list. In other words, a list $e_1, \cdots, e_m$ of vectors in V is orthonormal if

$$\langle e_j, e_k \rangle = \begin{cases} 1 & \text{if } j = k \\ 0 & \text{if } j \neq k \end{cases}.$$

Theorem 5.2.2
If $e_1, \cdots, e_m$ is an orthonormal list of vectors in V, then

$$\|a_1 e_1 + \cdots + a_m e_m\|^2 = |a_1|^2 + \cdots + |a_m|^2 \quad \forall a_1, \cdots, a_m \in \mathbb{F}.$$

Proof 1. Note that

$$\langle a_1 e_1, a_2 e_2 + \cdots + a_m e_m \rangle = \langle a_1 e_1, a_2 e_2 \rangle + \cdots + \langle a_1 e_1, a_m e_m \rangle = 0.$$

So, by the Pythagorean Theorem, we have

$$\begin{aligned} \|a_1 e_1 + \cdots + a_m e_m\|^2 &= \|a_1 e_1\|^2 + \|a_2 e_2 + \cdots + a_m e_m\|^2 \\ &= \|a_1 e_1\|^2 + \|a_2 e_2\|^2 + \cdots + \|a_m e_m\|^2 \\ &= |a_1|^2 + |a_2|^2 + \cdots + |a_m|^2. \end{aligned}$$

■

Theorem 5.2.3
Every orthonormal list of vectors is L.I..

Proof 2. Suppose $e_1, \cdots, e_m$ is an orthonormal list of vectors in V. Then, $\|a_1 e_1 + \cdots + a_m e_m\|^2 = 0$. By Theorem 5.2.2, it is equivalent to $|a_1|^2 + \cdots + |a_m|^2 = 0$. Since each $|a_j| \geq 0$, it must be $a_j = 0$ for all $j = 1, \cdots, m$. Therefore, the orthonormal list is L.I.. ■

Definition 5.2.4 (Orthonormal Basis). An *orthonormal basis* of Vc is an orthonormal list of vectors in V that is also a basis of V.

Theorem 5.2.5
Every orthonormal list of vectors in V with length $\dim V$c is an orthonormal basis of V.

Proof 3. By Theorem 5.2.3, any orthonormal list of vectors must be L.I.. Further since it has length $\dim V$, it is a basis of V. So, by definition, it is an orthonormal basis of V. ■

Theorem 5.2.6
Suppose $e_1, \cdots, e_n$ is an orthonormal basis of V and $v \in V$. Then, $v = \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_n \rangle e_n$, and $\|v\|^2 = |\langle v, e_1 \rangle|^2 + \cdots + |\langle v, e_n \rangle|^2$.

Proof 4. Suppose $v \in V$ and $v = a_1e_1 + \cdots + a_ne_n$. Then,

$$\langle v, e_j\rangle = \langle a_1e_1 + \cdots + a_ne_n, e_j\rangle = \langle a_je_j, e_j\rangle = a_j.$$

So, we have

$$v = \langle v, e_1\rangle e_1 + \cdots + \langle v, e_n\rangle e_n.$$

Further, by Theorem 5.2.2, we have

$$\|v\|^2 = |\langle v, e_1\rangle|^2 + \cdots + |\langle v, e_n\rangle|^2.$$

■

Theorem 5.2.7 Gram-Schmidt Procedure
Suppose $v_1, \cdots, v_m$ is L.I. list of vectors in V. Let $e_1 = \dfrac{v_1}{\|v_1\|}$. For $j = 2, \cdots, m$, define e_j inductively by

$$e_j = \frac{v_j - \langle v_j, e_1\rangle e_1 - \cdots - \langle v_j, e_{j-1}\rangle e_{j-1}}{\|v_j - \langle v_j, e_1\rangle e_1 - \cdots - \langle v_j, e_{j-1}\rangle e_{j-1}\|}. \tag{15}$$

Then, $e_1, \cdots, e_m$ is an orthonormal list of vectors in V *s.t.* $\operatorname{span}(v_1, \cdots, v_j) = \operatorname{span}(e_1, \cdots, e_j)$ for $j = 1, \cdots, m$.

Proof 5. To prove that Gram-Schmidt Procedure indeed produces an orthonormal list of vectors in V, we will use prove by mathematical induction.

$\boxed{\text{Base Case}}$ Suppose $j = 1$, then $\operatorname{span}(v_1) = \operatorname{span}(e_1)$ since v_1 is a positive multiple of e_1. So, the conclusion holds when $j = 1$.

$\boxed{\text{Inductive Steps}}$ Suppose for some $1 < j < m$, we have $\operatorname{span}(v_1, \cdots, v_{j-1}) = \operatorname{span}(e_1, \cdots, e_{j-1})$. Since$v_1, \cdots, v_m$ is L.I., we know $v_j \notin \operatorname{span}(v_1, \cdots, v_{j-1})$. That is, $v_j \notin \operatorname{span}(e_1, \cdots, e_{j-1})$. (*If* $v_j \in \operatorname{span}(e_1, \cdots, e_{j-1})$, *then* $v_j = \langle v_j, e_1\rangle e_1 + \cdots + \langle v_j, e_{j-1}\rangle e_{j-1}$.) *Then, we are dividing by* 0 *in Equation (15).* So, we are not dividing by 0 in Equation (15). Dividing a vector by its norm produces a new vector with norm 1, so $\|e_j\| = 1$. Now, we want to verify e_j is orthogonal to $e_1, \cdots, e_{j-1}$. Pick some k *s.t.* $1 \leq k < j$. Then

$$\begin{aligned}
\langle e_j, e_k\rangle &= \left\langle \frac{v_j - \langle v_j, e_1\rangle e_1 - \cdots - \langle v_j, e_{j-1}\rangle e_{j-1}}{\|v_j - \langle v_j, e_1\rangle e_1 - \cdots - \langle v_j, e_{j-1}\rangle e_{j-1}\|}, e_k \right\rangle \\
&= \frac{\langle v_j - \langle v_j, e_1\rangle e_1 - \cdots - \langle v_j, e_{j-1}\rangle e_{j-1}, e_k\rangle}{\|\langle v_j - \langle v_j, e_1\rangle e_1 - \cdots - \langle v_j, e_{j-1}\rangle e_{j-1}\|} \\
&= \frac{\langle v_j, e_k\rangle - \langle v_j, e_k\rangle}{\|\langle v_j - \langle v_j, e_1\rangle e_1 - \cdots - \langle v_j, e_{j-1}\rangle e_{j-1}\|} \\
&= 0
\end{aligned}$$

Then, $e_1, \cdots, e_j$ is an orthonormal basis, and $v_j \in \operatorname{span}(e_1, \cdots, e_j)$ since v_j is a linear combination of $e_1, \cdots, e_j$ by Equation (15). Further, we have

$$\dim \operatorname{span}(v_1, \cdots, v_j) = \dim \operatorname{span}(e_1, \cdots, e_j)$$

and

$$\operatorname{span}(v_1, \cdots, v_j) \subseteq \operatorname{span}(e_1, \cdots, e_j).$$

That is, exactly, $\mathrm{span}(v_1, \cdots, v_j) = \mathrm{span}(e_1, \cdots, e_j)$. ■

Theorem 5.2.8
Every *f-d* inner product space has an orthonormal basis.

Proof 6. Suppose V is *f-d*, and select a basis of V. Apply Gram-Schmidt Procedure (Theorem 5.2.7) to this basis, we then have an orthonormal basis of V. ■

Theorem 5.2.9
Suppose V is *f-d*. Then, every orthonormal list of vectors in V can be extended to an orthonormal basis of V.

Proof 7. Suppose $e_1, \cdots, e_m$ is an orthonormal list of vectors in V. Then, $e_1, \cdots, e_m$ is L.I. and can be extended to a basis $e_1, \cdots, e_m, v_1, \cdots, v_n$ of V. Apply Gram-Schmidt Procedure to this basis, we get an orthonormal list $e_1, \cdots, e_m, f_1, \cdots, f_n$. Here, $e_1, \cdots, e_m$ is unchanged since they are already orthonormal. Then, $e_1, \cdots, e_m, f_1, \cdots, f_n$ is an orthonormal basis of V. ■

Theorem 5.2.10
Suppose $T \in \mathcal{L}(V)$. If T has an upper-triangular matrix with respect to some basis of V, then T has an upper-triangular matrix with respect to some orthonormal basis of V.

Proof 8. Suppose $\mathcal{M}(T)$ is upper-triangular with respect to a basis $v_1, \cdots, v_n$ of V. Then, we know $\mathrm{span}(v_1, \cdots, v_j)$ is invariant under T for $j = 1, \cdots, n$. Apply Gram-Schmidt Procedure to $v_1, \cdots, v_n$, we will get an orthonormal basis $e_1, \cdots, e_n$ of V. Further, since $\mathrm{span}(e_1, \cdots, e_j) = \mathrm{span}(v_1, \cdots, v_j)$ for $j = 1, \cdots, n$, we know $\mathrm{span}(e_1, \cdots, e_j)$ is invariant under T. Therefore, T has an upper-triangular matrix with respect to the orthonormal basis $e_1, \cdots, e_n$. ■

Theorem 5.2.11 Schur's Theorem
Suppose V is a *f-d* complex vector space and $T \in \mathcal{L}(V)$. Then, T has an upper-triangular matrix with respect to some orthonormal basis of V.

Proof 9. Since V is a *f-d* complex vector space, T must have an upper-triangular matrix with respect to some basis of V. Further, by Theorem 5.2.10, T must have an upper-triangular matrix with respect to an orthonormal basis of V. ■

Example 5.2.12 The function $\varphi : \mathbb{F}^3 \to \mathbb{F}$ defined by

$$\varphi(z_1, z_2, z_3) = 2z_1 - 5z_2 + z_3$$

is a linear functional on $\mathbb{F}^3$. We could write this linear functional in the form $\varphi(z) = \langle z, u\rangle$ for every $z \in \mathbb{F}^3$, where $u = \langle 2, -5, 1\rangle$.

Theorem 5.2.13 Riesz Representation Theorem
Suppose V is *f-d* and φ is a linear functional on V. Then, there is a unique vector $u \in V$ *s.t.* $\varphi(v) = \langle v, u \rangle$ for every $v \in V$.

Proof 10. Let $e_1, \cdots, e_n$ be an orthonormal basis of V. Then, for all $v \in V$, we have

$$v = \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_n \rangle e_n.$$

So,

$$\begin{aligned} \varphi(v) &= \varphi(\langle v, e_1 \rangle e_1 + \cdots + \langle v, e_n \rangle e_n) \\ &= \langle v, e_1 \rangle \varphi(e_1) + \cdots + \langle v, e_n \rangle \varphi(e_n) \\ &= \langle v, \overline{\varphi(e_1)} e_1 \rangle + \cdots + \langle v, \overline{\varphi(e_n)} e_n \rangle \\ &= \langle v, \overline{\varphi(e_1)} e_1 + \cdots + \overline{\varphi(e_n)} e_n \rangle. \end{aligned}$$

Suppose $\exists u_1, u_2 \in V$ *s.t.* $\varphi(v) = \langle v, u_1 \rangle = \langle v, u_2 \rangle$. Then, $\langle v, u_1 \rangle - \langle v, u_2 \rangle = \langle v, u_1 - u_2 \rangle = 0$. Let $v = u_1 - u_2$, then we have $\langle u_1 - u_2, u_1 - u_2 \rangle = 0$. So, it must be $u_1 = u_2$. Therefore, $\exists$ a unique $u \in V$ and

$$u = \overline{\varphi(e_1)} e_1 + \cdots + \overline{\varphi(e_n)} e_n \textit{ s.t. } \varphi(v) = \langle v, u \rangle \quad \forall v \in V.$$

■

Example 5.2.14 Find $u \in \mathcal{P}_2(\mathbb{R})$ *s.t.* $\displaystyle\int_{-1}^{1} p(t)(\cos(\pi t))\, \mathrm{d}t = \int_{-1}^{1} p(t)u(t)\, \mathrm{d}t$ for every $p \in \mathcal{P}_2(\mathbb{R})$.

Remark. *Define an inner product on* $\mathcal{P}_2(\mathbb{R})$ *as* $\langle p, q \rangle = \displaystyle\int_{-1}^{1} p(x)q(x)\, \mathrm{d}x$ *to solve this problem.*

Solution 11.

Let $\varphi \in \mathcal{L}(\mathcal{P}_2(\mathbb{R}), \mathbb{R})$ be defined as $\varphi(t) = \displaystyle\int_{-1}^{1} p(t)(\cos(\pi t))\, \mathrm{d}t$. Note that $1, x, x^2$ is a basis of $\mathcal{P}_2(\mathbb{R})$. To find an orthonormal basis of $\mathcal{P}_2(\mathbb{R})$, apply Gram-Schmidt Procedure, we have

$$e_1 = \frac{1}{\|1\|} = \frac{1}{\sqrt{\displaystyle\int_{-1}^{1} 1 \cdot 1\, \mathrm{d}t}} = \sqrt{\frac{1}{2}}.$$

Since $x - \langle x, e_1 \rangle e_1 = x - \displaystyle\int_{-1}^{1} x\sqrt{\frac{1}{2}}\, \mathrm{d}x \cdot \sqrt{\frac{1}{2}} = x$, and $\|x\| = \sqrt{\displaystyle\int_{-1}^{1} x^2\, \mathrm{d}x} = \sqrt{\frac{2}{3}}$, we have

$$e_2 = \frac{x}{\sqrt{\frac{2}{3}}} = \sqrt{\frac{3}{2}} x.$$

Further, consider

$$\begin{aligned} x^2 - \langle x^2, e_1 \rangle e_1 - \langle x^2, e_2 \rangle e_2 &= x^2 - \int_{-1}^{1} x^2 \sqrt{\frac{1}{2}}\, \mathrm{d}x \cdot \sqrt{\frac{1}{2}} - \int_{-1}^{1} x^2 \sqrt{\frac{3}{2}} x\, \mathrm{d}x \cdot \sqrt{\frac{3}{2}} x \\ &= x^2 - \frac{1}{3}, \end{aligned}$$

and note that

$$\left\| x^2 - \frac{1}{3} \right\| = \sqrt{\int_{-1}^{1} \left(x^2 - \frac{1}{3} \right)^2 \mathrm{d}x} = \sqrt{\int_{-1}^{1} x^4 - \frac{2}{3}x^2 + \frac{1}{9} \, \mathrm{d}x} = \sqrt{\frac{8}{45}}.$$

So, we have

$$e_3 = \frac{x^2 - \frac{1}{3}}{\sqrt{\frac{8}{45}}} = \sqrt{\frac{45}{8}} \left(x^2 - \frac{1}{3} \right).$$

That is, $e_1 = \sqrt{\frac{1}{2}}, \ e_2 = \sqrt{\frac{3}{2}}x, \ e_3 = \sqrt{\frac{45}{8}} \left(x^2 - \frac{1}{3} \right)$ is an orthonormal basis of $\mathcal{P}_2(\mathbb{R})$. Then, we have

$$\varphi(e_1) = \int_{-1}^{1} \sqrt{\frac{1}{2}} \cos(\pi t) \, \mathrm{d}t = \sqrt{\frac{1}{2}} \int_{-1}^{1} \cos(\pi t) \, \mathrm{d}t = 0$$

$$\varphi(e_2) = \int_{-1}^{1} \sqrt{\frac{3}{2}} t \cos(\pi t) \, \mathrm{d}t = \sqrt{\frac{3}{2}} \int_{-1}^{1} t \cos(\pi t) \, \mathrm{d}t = 0$$

$$\begin{aligned}
\varphi(e_3) &= \int_{-1}^{1} \sqrt{\frac{45}{8}} \left(t^2 - \frac{1}{3} \right) \cos(\pi t) \, \mathrm{d}t \\
&= \sqrt{\frac{45}{8}} \int_{-1}^{1} t^2 \cos(\pi t) \, \mathrm{d}t - \sqrt{\frac{45}{8}} \cdot \frac{1}{3} \underbrace{\int_{-1}^{1} \cos(\pi t) \, \mathrm{d}t}_{0} \\
&= \sqrt{\frac{45}{8}} \int_{-1}^{1} t^2 \cos(\pi t) \, \mathrm{d}t \\
&= \sqrt{\frac{45}{8}} \left(-\frac{4}{\pi^2} \right).
\end{aligned}$$

So, by Theorem 5.2.15 and its proof, we know

$$\begin{aligned}
u = \varphi(e_1)e_1 + \varphi(e_2)e_2 + \varphi(e_3)e_3 &= 0 + 0 + \sqrt{\frac{45}{8}} \left(-\frac{4}{\pi^2} \right) \sqrt{\frac{45}{8}} \left(x^2 - \frac{1}{3} \right) \\
&= \frac{45}{8} \left(-\frac{4}{\pi^2} \right) \left(x^2 - \frac{1}{3} \right) \\
&= -\frac{45}{2\pi^2} \left(x^2 - \frac{1}{3} \right).
\end{aligned}$$

$\square$

5.3 Orthogonal Complements and Minimization Problems

Definition 5.3.1 (Orthogonal Complement, $U^\perp$). If U is a subset of V, then the *orthogonal complement* of U, denoted $U^\perp$, is the set of all vectors in V that are orthogonal to every vector in U:

$$U^\perp = \{v \in v \mid \langle v, u \rangle = 0 \quad \forall u \in U\}.$$

Theorem 5.3.2 Basic Properties of Orthogonal Complements

1. If U is a subset of V, then $U^\perp$ is a subspace of V.
2. $\{0\}^\perp = V$.
3. $V^\perp = \{0\}$.
4. If U is a subset of V, then $U \cap U^\perp \subseteq \{0\}$.
5. If U and W are subsets of V and $U \subseteq W$, then $W^\perp \subseteq U^\perp$.

Proof 1.

1. Let $v, w \in U^\perp$. Then $\langle v + w, u \rangle = \langle v, w \rangle + \langle w, u \rangle = 0 + 0 = 0$. So, $v + w \in U^\perp$. Further, suppose $\lambda \in \mathbb{F}$. Then $\langle \lambda v, u \rangle = \lambda \langle v, u \rangle = \lambda \cdot 0 = 0$. So, $\lambda v \in U^\perp$. Finally since $\langle 0, u \rangle = 0$, we know $0 \in U^\perp$. Then, $U^\perp$ is a subspace of V. □
2. Since $\langle v, 0 \rangle = 0 \quad \forall v \in V$, we know $\{0\}^\perp = V$. □
3. Suppose $v \in V^\perp$. Then, $\langle v, v \rangle = 0$. By property of an inner product, it must be that $v = 0$. So, $V^\perp = \{0\}$. □
4. Suppose U is a subset of V. Let $v \in U \cap U^\perp$. Then, $v \in U$ and $v \in U^\perp$. So, $\langle v, v \rangle = 0$. Then, it must be that $v = 0$. So, $U \cap U^\perp \subseteq \{0\}$. □
5. Suppose U and W are subsets of V with $U \subseteq W$. Suppose $v \in W^\perp$. Then, $\langle v, u \rangle = 0 \quad \forall u \in W$. Since $U \subseteq W$, we have $\langle v, w \rangle = 0 \quad \forall u \in U$. That is, $v \in U^\perp$. Then, we have $W^\perp \subseteq U^\perp$.

■

Theorem 5.3.3
Suppose U is a *f-d* subspace of V. Then, $V = U \oplus U^\perp$.

Proof 2. Suppose $u \in U$ and $w \in U^\perp$. Then, $\forall v \in V$, we have $v = cu + w$ for some $c \in \mathbb{F}$ and $\langle u, w \rangle = 0$. Then, we have $V = U + U^\perp$. Further, by Theorem 5.3.2(4), $U \cap U^\perp = \{0\}$ since U and $U^\perp$ are all subspaces of V. Hence, $V = U \oplus U^\perp$. ■

Corollary 5.3.4 Suppose V is *f-d* and U is a subspace of V. Then, $\dim U^\perp = \dim V - \dim U.$

Theorem 5.3.5
Suppose U is a *f-d* subspace of V. Then, $U = (U^\perp)^\perp$.

Proof 3.

($\subseteq$). Suppose $u \in U$. Then, $\langle u, v\rangle = 0 \quad \forall v \in U^\perp$. Then, $u \in (U^\perp)^\perp$. That is, $U \subseteq (U^\perp)^\perp$. □

($\supseteq$). Suppose $v \in (U^\perp)^\perp$. Then, $v = u + w$ for some $u \in U$ and $w \in U^\perp$. Then, $w = v - u \in (U^\perp)^\perp$. Since $U \subseteq (U^\perp)^\perp$, we know $u \in U^\perp$. Then, $v - u \in (U^\perp)^\perp$. Hence, $v - u \in U^\perp \cap (U^\perp)^\perp$. That is, $v - u$ is orthogonal to itself. So, it must be that $v - u = 0$ or $v = u$. Since $u \in U$ and $v \in U$, we have shown that $(U^\perp)^\perp \subseteq U$. ■

Definition 5.3.6 (Orthogonal Projection, P_U). Suppose U is a *f-d* subspace of V. Then orthogonal projection of V onto U is the operator $P_U \in \mathcal{L}(V)$ defined as follows: For $v \in V$, write $v = u + w$, where $u \in U$ and $w \in U^\perp$. Then, $P_U v = u$.

Remark. *By Theorem 5.3.3, $V = U \oplus U^\perp$, which ensures each $v \in V$ can be uniquely represented in the form of $u + w$ with $u \in U$ and $w \in U^\perp$, and thus P_U is well-defined.*

Example 5.3.7 Suppose $x \in V$ with $x \neq 0$ and $U = \operatorname{span}(x)$. Show that

$$P_U v = \frac{\langle v, x\rangle}{\|x^2\|} x \qquad \forall v \in V.$$

Proof 4. Suppose $v \in V$. Then,

$$v = \frac{\langle v, x\rangle}{\|x^2\|} x + \left(v - \frac{\langle v, x\rangle}{\|x^2\|} x\right),$$

where $\frac{\langle v, x\rangle}{\|x^2\|} x \in \operatorname{span}(x)$ and $v - \frac{\langle v, x\rangle}{\|x^2\|} x \in U^\perp$. Thus, $P_U v = \frac{\langle v, x\rangle}{\|x^2\|} x$. ■

Theorem 5.3.8 Properties of Orthogonal Projections
Suppose U is a *f-d* subspace of V and $v \in V$. Then,

1. $P_U \in \mathcal{L}(V)$.
2. $P_U u = u \quad \forall u \in U$.
3. $P_U w = 0 \quad \forall w \in U^\perp$.
4. range $P_U = U$.
5. null $P_U = U^\perp$.
6. $v - P_U v \in U^\perp$.
7. $P_U^2 = P_U$.
8. $\|P_U v\| \leq \|v\|$.
9. for every orthonormal basis $e_1, \cdots, e_m$ of U,

$$P_U v = \langle v, e_1\rangle e_1 + \cdots + \langle v, e_m\rangle e_m.$$

Proof 5.

1. Suppose $v_1 = u_1 + w_1$ and $v_2 = u_2 + w_2$, where $v_1, v_2 \in V$, $u_1, u_2 \in U$, and $w_1, w_2 \in U^\perp$. Then, $v_1 + v_2 = (u_1 + u_2) + (w_1 + w_2)$, where $u_1 + u_2 \in U$ and $w_1 + w_2 \in U^\perp$. So,

$$P_U(v_1 + v_2) = u_1 + u_2 = P_U v_1 + P_U v_2.$$

   Additionally, suppose $\lambda \in \mathbb{F}$. Then, $\lambda v_1 = \lambda u_1 + \lambda w_1$, where $\lambda u_1 \in U$ and $\lambda w_1 \in U^\perp$. Then,

$$P_U(\lambda v_1) = \lambda u_1 = \lambda P_U(v_1). \qquad \square$$

2. Suppose $u \in U$. Then, $u = u + 0$, where $u \in U$ and $0 \in U^\perp$. So, $P_U u = u$. $\square$

3. Suppose $w \in U^\perp$. Then, $w = 0 + w$, where $0 \in U$ and $w \in U^\perp$. So, $P_U w = 0$. $\square$

4. By definition of P_U, we have $\operatorname{range} P_U \subseteq U$. By Theorem 5.3.8(2), we know $U \subseteq \operatorname{range} P_U$. So, $\operatorname{range} P_U = U$. $\square$

5. By Theorem 5.3.8(3), we have $U^\perp \subseteq \operatorname{null} P_U$. Further note if $v \in \operatorname{null} P_U$, then $v = 0 + v$ with $0 + u$ and $v \in U^\perp$. So, $\operatorname{null} P_U \subseteq U^\perp$. That is, $\operatorname{null} P_U = U$. $\square$

6. If $v = u + w$ with $u \in U$ and $w \in U^\perp$, then

$$v - P_U v = v - u = w \in U^\perp. \qquad \square$$

7. If $v = u + w$ with $u \in U$ and $w \in U^\perp$, then

$$(P_U^2)v = P_U(P_U v) = P_U u = u = P_U v.$$

   So, $P_U^2 = P_U$. $\square$

8. If $v = u + w$ with $u \in U$ and $w \in U^\perp$, then we have

$$\|P_U v\|^2 = \|u\|^2 \leq \|u\|^2 + \|w\|^2 = \|v\|^2$$

   by the Pythagorean Theorem. $\square$

9. If $v = u + w$ with $u \in U$ and $w \in U^\perp$, then

$$v = u + w = \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_m \rangle e_m + (v - \langle v, e_1 \rangle e_1 - \cdots - \langle v, e_m \rangle e_m).$$

   Since $e_1, \cdots, e_m$ is an orthonormal basis of U, we have $\langle v e_1 \rangle e_1 + \cdots + \langle v, e_m \rangle e_m \in U$. Now, consider

$$\begin{aligned}
\big\langle \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_m \rangle e_m, v - \langle v, e_1 \rangle e_1 - \cdots - \langle v, e_m \rangle e_m \big\rangle &= \big\langle \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_m \rangle e_m, v \big\rangle - \|u\|^2 \\
&= \langle v, e_1 \rangle \langle e_1, v \rangle + \cdots + \langle v, e_m \rangle \langle e_m, v \rangle - \|u\|^2 \\
&= \langle v, e_1 \rangle \overline{\langle v, e_1 \rangle} + \cdots + \langle v, e_m \rangle \overline{\langle v, e_m \rangle} - \|u\|^2 \\
&= |\langle v, e_1 \rangle|^2 + \cdots + |\langle v, e_m \rangle|^2 - \|u\|^2 \\
&= \|u\|^2 - \|u\|^2 = 0 \quad (By\, Theorem\, 5.2.2)
\end{aligned}$$

   Then, $v - \langle v, e_1 \rangle e_1 - \cdots - \langle v, e_m \rangle e_m \in U^\perp$. So, we have $P_U v = \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_m \rangle e_m$.

$\blacksquare$

Theorem 5.3.9 Minimizing the Distance to a Subspace
Suppose U is a *f-d* subspace of V, $v \in V$, and $u \in U$. Then, $\|v - P_U v\| \leq \|v - u\|$. The inequality is an equality if and only if $u = P_U v$.

Proof 6. Note that $\|v - P_U v\|^2 \leq \|v - P_U v\|^2 + \|P_U v - u\|^2$ since $\|P_U v - u\|^2 \geq 0$. Further, since $v - P_U v \in U^\perp$ by Theorem 5.3.8(6) and $P_U v - u \in U$ by the Pythagorean Theorem, we have

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

Then, $\|u - P_U v\|^2 \leq \|v = P_U v\|^2 + \|P_U v - u\|^2 = \|v - u\|^2$. Since $\|v - P_U v\|^2 \geq 0$ and $\|v - u\|^2 \geq 0$, we have $\|v - P_U v\| \leq \|v - u\|$. The equality holds if and only if $\|P_U v - u\|^2 = 0$. That is, $\|P_U v - u\| = 0$, $P_U v - u = 0$, or $P_U v = u$. ■

Example 5.3.10 In $\mathbb{R}^4$, set $U = \mathrm{span}((1,1,0,0),(1,1,1,2))$. Find $u \in U$ *s.t.* $\|u - (1,2,3,4)\|$ is as small as possible.

Solution 7.

By Theorem 5.3.9, we need to find $P_U v = \langle v, e_1 \rangle e_1 + \langle v, e_2 \rangle e_2$. Thus, we need to use Gram-Schmidt Procedure to find e_1 and e_2:

$$e_1 = \frac{1}{\sqrt{2}}(1,1,0,0) \qquad \text{and} \qquad e_2 = \frac{1}{\sqrt{5}}(0,0,1,2).$$

Set $v = (1,2,3,4)$, we have

$$\begin{aligned} P_U v &= \langle (1,2,3,4), \frac{1}{\sqrt{2}}(1,1,0,0) \rangle \frac{1}{\sqrt{2}}(1,1,0,0) + \langle (1,2,3,4), \frac{1}{\sqrt{5}}(0,0,1,2) \rangle \frac{1}{\sqrt{5}}(0,0,1,2) \\ &= \left(\frac{3}{2}, \frac{3}{2}, \frac{11}{5}, \frac{22}{5} \right). \end{aligned}$$

□

6 Operators on Inner Product Spaces

6.1 Self-Adjoint and Normal Operators

Definition 6.1.1 (Adjoint, T^*). Suppose $T \in \mathcal{L}(V, W)$. The *adjoint* of T is the function $T^* : W \to V$ *s.t.*

$$\langle Tv, w \rangle = \langle v, T^* w \rangle$$

for every $v \in V$ and every $w \in W$.

Theorem 6.1.2
If $T \in \mathcal{L}(V, W)$, then $T^* \in \mathcal{L}(W, V)$.

Proof 1.

1. The definition of adjoint makes sense.

   Suppose $T \in \mathcal{L}(V, W)$. Fix $w \in W$. Let $f : V \to \mathbb{F}$ be defined as $v \mapsto \langle Tv, w \rangle$. Then, f is a linear functional on V. Note that

   $$\begin{aligned} f(au + bv) = \langle T(au + bv), w \rangle &= \langle aTu + bTv, w \rangle \\ &= a\langle Tu, w \rangle + b\langle Tv, w \rangle \\ &= af(u) + b(fv). \end{aligned}$$

   By Riesz Representation Theorem, we know $f(v) = \langle v, \Delta \rangle$ for some $\Delta \in V$. We call this unique Δ as $T^* w$. That is, for each $w \in W$, $\exists$ unique $T^* w \in V$. So, T^* is well-defined as a function from W to V. $\square$

2. Adjoint is a linear map.

   Suppose $w_1, w_2 \in W$. If $v \in V$, then

   $$\begin{aligned} \langle v, T^*(w_1 + w_2) \rangle = \langle Tv, w_1 + w_2 \rangle &= \langle Tv, w_1 \rangle + \langle Tv, w_2 \rangle \\ &= \langle v, T^* w_1 \rangle + \langle v, T^* w_2 \rangle \\ &= \langle v, T^* w_1 + T^* w_2 \rangle. \end{aligned}$$

   So, $T^*(w_1 + w_2) = T^* w_1 + T^* w_2$. $\square$

   Now fix $w \in W$ and $\lambda \in \mathbb{F}$. If $v \in V$, then

   $$\begin{aligned} \langle v, T^*(\lambda w) \rangle = \langle Tv, \lambda w \rangle &= \overline{\lambda}\langle Tv, w \rangle \\ &= \overline{\lambda}\langle v, T^* w \rangle \\ &= \langle v, \lambda T^* w \rangle. \end{aligned}$$

   So, we know $T^*(\lambda w) = \lambda T^* w$. $\square$

   Thus, we've shown T^* is a linear map as desired.

$\blacksquare$

Example 6.1.3 Define $T : \mathbb{R}^3 \to \mathbb{R}^2$ by $T(x_1, x_2, x_3) = (x_2 + 3x_3, 2x_1)$. Find a formula for T^*.

Solution 2.

Define $T^* : \mathbb{R}^2 \to \mathbb{R}^3$. Let $y = (y_1, y_2) \in \mathbb{R}^2$. Then,

$$\begin{aligned}\langle x, T^* y\rangle = \langle Tx, y\rangle &= y_1x_2 + 3y_1x_3 + 2x_1y_2 \\ &= \langle (x_1, x_2, x_3), (2y_2, y_1, 3y_1)\rangle.\end{aligned}$$

Thus, $T^* : \mathbb{R}^2 \to \mathbb{R}^3$ is defined as $T^*(y_1, y_2) = (2y_2, y_1, 3y_1)$. □

Example 6.1.4 Fix $u \in V$ and $x \in W$. Define $T \in \mathcal{L}(V, W)$ by $Tv = \langle v, u\rangle x$ for every $v \in V$. Find a formula for T^*.

Solution 3.

Define $T^* \in \mathcal{L}(W, V)$. Consider

$$\begin{aligned}\langle v, T^* w\rangle = \langle Tv, w\rangle &= \big\langle \langle v, u\rangle x, w\big\rangle \\ &= \langle v, u\rangle\langle x, w\rangle \\ &= \big\langle v, \langle w, x\rangle u\big\rangle.\end{aligned}$$

So, we have $T^* w = \langle w, x\rangle u$. □

Theorem 6.1.5 Properties of the Adjoint

1. $(S + T)^* = S^* + T^* \quad \forall S, T \in \mathcal{L}(V, W)$.
2. $(\lambda T)^* = \overline{\lambda} T^* \quad \forall \lambda \in \mathbb{F}$ and $T \in \mathcal{L}(V, W)$.
3. $(T^*)^* = T \quad \forall T \in \mathcal{L}(V, W)$.
4. $I^* = I$, where I is the identity operator on V.
5. $(ST)^* = T^* S^* \quad \forall T \in \mathcal{L}(V, W)$ and $S \in \mathcal{L}(W, U)$.

Proof 4.

1. Consider

$$\begin{aligned}\langle v, (S + T)^* w\rangle = \langle (S + T)v, w\rangle &= \langle Sv, w\rangle + \langle Tv, w\rangle \\ &= \langle v, S^* w\rangle + \langle v, T^* w\rangle \\ &= \langle v, S^* w + T^* w\rangle \\ &= \langle v, (S^* + T^*)w\rangle.\end{aligned}$$

   So, we have $(S + T)^* w = (S^* + T^*)w \quad \forall w \in W$. □

2. Note that

$$\begin{aligned}\langle v, (\lambda T)^* w\rangle = \langle (\lambda T)v, w\rangle &= \lambda\langle Tv, w\rangle \\ &= \lambda\langle v, T^* w\rangle \\ &= \langle v, \overline{\lambda} T^* w\rangle.\end{aligned}$$

So, we get $(\lambda T)^* w = \overline{\lambda} T^* w \quad \forall w \in W.$ $\square$

3. Consider
$$\begin{aligned} \langle v, (T^*)^* w\rangle = \langle T^* v, w\rangle &= \overline{\langle w, T^* v\rangle} \\ &= \overline{\langle Tw, v\rangle} \\ &= \langle v, Tw\rangle. \end{aligned}$$
So, it is $(T^*)^* w = Tw \quad \forall w \in W.$ $\square$

4. Note we have
$$\langle v, I^* w\rangle = \langle Iv, w\rangle = \langle v, w\rangle.$$
So, $I^* w = w \quad \forall w \in W$. That is, $I^* = I.$ $\square$

5. We have
$$\begin{aligned} \langle v, (ST)^* w\rangle = \langle (ST)v, w\rangle &= \langle S(Tv), w\rangle \\ &= \langle Tv, S^* w\rangle \\ &= \langle v, T^*(S^* w)\rangle. \end{aligned}$$
So, $(ST)^* w = T^*(S^* w) = (T^* S^*) w \quad \forall w \in W.$

■

Theorem 6.1.6 Null Space and Range of T^*
Suppose $T \in \mathcal{L}(V, W)$. Then,

1. $\text{null}\, T^* = (\text{range}\, T)^{\perp}$.
2. $\text{range}\, T = (T^*)^{\perp}$.
3. $\text{null}\, T = (\text{range}\, T^*)^{\perp}$.
4. $\text{range}\, T^* = (\text{null}\, T)^{\perp}$.

Proof 5.

1. Suppose $w \in \text{null}\, T^*$. Then, $T^* w = 0$. So, $\langle v, T^* w\rangle = 0$. That is, $\langle Tv, w\rangle = 0 \quad \forall v \in 0$. Then, w is orthogonal to any Tv. That is, $w \in (\text{range}\, T)^{\perp}$. Conversely, if $w \in (\text{range}\, T)^{\perp}$, we have $\langle Tv, w\rangle = 0$, and thus $\langle v, T^* w\rangle = 0$, or $T^* w = 0$. That is, $w \in \text{null}\, T^*$. Hence, $\text{null}\, T^* = (\text{range}\, T)^{\perp}$. $\square$

2. Note that $(\text{null}\, T^*)^{\perp} = \left((\text{range}\, T)^{\perp}\right)^{\perp} = \text{range}\, T.$ $\square$

3. Suppose $v \in \text{null}\, T$. Then, $Tv = 0$, and $\langle Tv, w\rangle = 0$. So, $\langle v, T^* w\rangle = 0 \quad \forall w \in W$. Then, v is orthogonal to every vectors in $T^* w$. So, $v \in (\text{range}\, T^*)^{\perp}$. In the other way around, if we assume $v \in (\text{range}\, T^*)^{\perp}$, then $\langle v, T^* w\rangle = \langle Tv, w\rangle = 0$. So, $Tv = 0$, and thus $v \in \text{null}\, T$. Hence, we have $\text{null}\, T = (\text{range}\, T^*)^{\perp}$. $\square$

4. Consider $(\text{null}\, T)^{\perp} = \left((\text{range}\, T^*)^{\perp}\right)^{\perp} = \text{range}\, T^*$.

■

Definition 6.1.7 (Conjugate Transpose). The *conjugate transpose* of an $m \times n$ matrix is the $n \times m$ matrix obtained by interchanging the rows and columns and then taking the conjugate of each entry.

Theorem 6.1.8
Let $T \in \mathcal{L}(V, W)$. Suppose $e_1, \cdots, e_n$ is an orthonormal basis of V and $f_1, \cdots, f_m$ is an orthonormal basis of W. Then, $\mathcal{M}(T^*, (f_1, \cdots, f_m), (e_1, \cdots, e_m))$ is the conjugate transpose of $\mathcal{M}(T, (e_1, \cdots, e_n), (f_1, \cdots, f_m))$.

Proof 6. Suppose $\mathcal{M}(T)$ denote the matrix $\mathcal{M}(T, (e_1, \cdots, e_n), (f_1, \cdots, f_m))$ and let $\mathcal{M}(T^*)$ denote the matrix $\mathcal{M}(T^*, (f_1, \cdots, f_m), (e_1, \cdots, e_m))$. Then, note that $Te_k = \langle Te_k, f_1 \rangle f_1 + \cdots + \langle Te_k, f_m \rangle f_m$. So,

$$(\mathcal{M}(T))_{j,k} = \langle Te_k, f_j \rangle.$$

Further, consider $T^* f_k = \langle T^* f_k, e_1 \rangle e_1 + \cdots + \langle T^* f_k, e_n \rangle e_n$. That is,

$$\begin{aligned}(\mathcal{M}(T^*))_{j,k} = \langle T^* f_k, e_j \rangle &= \overline{\langle e_j, T^* f_k \rangle} \\ &= \overline{\langle Te_j, f_k \rangle} \\ &= \overline{(\mathcal{M}(T))_{k,j}}\end{aligned}$$

So, we've shown that $\mathcal{M}(T^*)$ is the conjugate transpose of $\mathcal{M}(T)$. ■

Definition 6.1.9 (Self-Adjoint). An operator $T \in \mathcal{L}(V)$ is called *self-adjoint* if $T = T^*$. In other words, $T \in \mathcal{L}(V)$ is self-adjoint if and only if $\langle Tv, w \rangle = \langle v, Tw \rangle \quad \forall v, w \in V$.

Theorem 6.1.10
The sum of two self-adjoint operators is self-adjoint, and the product of a real scalar and a self-adjoint operator is self-adjoint.

Proof 7.

1. Suppose $T, S \in \mathcal{L}(V)$ are self-adjoint. Then,

$$(S + T)^* = S^* + T^* = S + T.$$

   So, $S + T$ is self-adjoint. □

2. Let $\lambda \in \mathbb{R}$. Then,

$$(\lambda T)^* = \lambda T^* = \lambda T.$$

   So, λT is self-adjoint.

■

Theorem 6.1.11
Every eigenvalue of a self-adjoint operator is real.

Proof 8. Suppose T is a self-adjoint operator on V. Let λ be an eigenvalue of T, and let v be a non-zero vector in V *s.t.* $Tv = \lambda v$. Then,

$$\lambda \|v\|^2 = \langle \lambda v, v \rangle = \langle Tv, v \rangle = \langle v, Tv \rangle = \langle v, \lambda v \rangle = \overline{\lambda} \|v\|^2.$$

So, it must be $\lambda = \overline{\lambda}$, which means λ is real. ■

Theorem 6.1.12
Suppose V is a complex inner product space and $T \in \mathcal{L}(V)$. Suppose $\langle Tv, v\rangle = 0 \quad \forall v \in V$. Then, $T = 0$.

Proof 9. Note that

$$\begin{aligned}\langle Tu, w\rangle &= \frac{1}{4}\Big[\big\langle T(u+w), u+w\big\rangle - \big\langle T(u-w), u-w\big\rangle\Big] \\ &\quad + \frac{\mathrm{i}}{4}\Big[\big\langle T(u+\mathrm{i}w), u+\mathrm{i}w\big\rangle - \big\langle T(u-\mathrm{i}w), (u-\mathrm{i}w)\big\rangle\Big] \\ &= 0 \quad \forall u, w \in V.\end{aligned}$$

Let $w = Tu \in V$. Then, $\langle Tu, Tu\rangle = 0$. That is, $Tu = 0 \quad \forall u \in V$. So, $T = 0$. ■

Theorem 6.1.13
Suppose V is a complex inner product space and $T \in \mathcal{L}(V)$. Then, T is self-adjoint if and only if $\langle Tv, v\rangle \in \mathbb{R} \quad \forall v \in V$.

Proof 10.
($\Rightarrow$) Let $v \in V$. Then,

$$\begin{aligned}\langle Tv, v\rangle - \overline{\langle Tv, v\rangle} &= \langle Tv, v\rangle - \langle v, Tv\rangle \\ &= \langle Tv, v\rangle - \langle T^*v, v\rangle \\ &= \langle (T - T^*)v, v\rangle\end{aligned} \tag{16}$$

If $\langle Tv, v\rangle \in \mathbb{R} \quad \forall v \in V$, then Equation (16)$= 0$. That is, $\langle (T - T^*)v, v\rangle = 0 \quad \forall v \in V$. So, $T - T^* = 0$, or $T = T^*$. That is, T is self-adjoint. □

($\Leftarrow$) Conversely, if T is self-adjoint, then Equation (16)$= 0$. That is, $\langle Tv, v\rangle = \overline{\langle Tv, v\rangle} = 0$, or we have $\langle Tv, v\rangle = \overline{\langle Tv, v\rangle}$. This is equivalent to the conclusion $\langle Tv, v\rangle \in \mathbb{R}$. ■

Theorem 6.1.14
Suppose T is a self-adjoint operator on V *s.t.* $\langle Tv, v\rangle = 0 \quad \forall v = V$. Then, $T = 0$.

Proof 11. We've already shown this to be true under a complex inner product space. Thus, we can assume V is a real inner product space. If $u, w \in V$, then

$$\begin{aligned}\langle Tu, w\rangle &= \frac{1}{4}\langle T(u+w), u+w\rangle - \langle T(u-w), u-w\rangle \\ &= 0 \quad \forall u, w \in V.\end{aligned}$$

Let $w = Tu$. Then, $\langle Tu, Tu\rangle = 0$, or $Tu = 0 \quad \forall u \in V$. So, $T = 0$. ■

Definition 6.1.15 (Normal Operator). An operator on an inner product space is called *normal* if it commutes with its adjoint. In other words, $T \in \mathcal{L}(V)$ is normal if $TT^* = T^*T$.

Example 6.1.16 Let T be the operator on $\mathbb{F}^2$ whose matrix with respect to the standard basis is

$$\begin{pmatrix} 2 & -3 \\ 3 & 2 \end{pmatrix}.$$

Show that T is not self-adjoint but is still normal.

Proof 12. Since $\mathcal{M}(T) = \begin{pmatrix} 2 & -3 \\ 3 & 2 \end{pmatrix}$ and $\mathcal{M}(T^*) = \begin{pmatrix} 2 & 3 \\ -3 & 2 \end{pmatrix}$, then $\mathcal{M}(T) \neq \mathcal{M}(T^*)$, and thus it is not self-adjoint. However, note that

$$\begin{pmatrix} 2 & -3 \\ 3 & 2 \end{pmatrix}\begin{pmatrix} 2 & 3 \\ -3 & 2 \end{pmatrix} = \begin{pmatrix} 13 & 0 \\ 0 & 13 \end{pmatrix}$$

and

$$\begin{pmatrix} 2 & 3 \\ -3 & 2 \end{pmatrix}\begin{pmatrix} 2 & -3 \\ 3 & 2 \end{pmatrix} = \begin{pmatrix} 13 & 0 \\ 0 & 13 \end{pmatrix}.$$

So, by definition, T is normal. ■

Theorem 6.1.17
An operator $T \in \mathcal{L}(V)$ is normal if and only if $\|Tv\| = \|T^*v\| \quad \forall v \in V$.

Proof 13. Note that

$$\begin{aligned} T \text{ is normal} &\iff T^*T - TT^* = 0 \\ &\iff \big\langle (T^*T - TT^*)v, v \big\rangle = 0 \quad \forall v \in V \\ &\iff \langle T^*Tv, v\rangle = \langle TT^*v, v\rangle \quad \forall v \in V \\ &\iff \langle Tv, Tv\rangle = \langle T^*v, T^*v\rangle \quad \forall v \in V \\ &\iff \|Tv\|^2 = \|T^*v\|^2 \quad \forall v \in V. \end{aligned}$$

Since $\|Tv\| \geq 0$ and $\|T^*v\| \geq 0$, it is equivalent to

$$\|Tv\| = \|T^*v\| \quad \forall v \in V.$$

■

Theorem 6.1.18
Suppose $T \in \mathcal{L}(V)$ is normal and $v \in V$ is an eigenvector of T with eigenvalue λ. Then, v is also an eigenvector of T^* with eigenvalue $\overline{\lambda}$.

Proof 14. Note that $(T - \lambda I)^* = T^* - \overline{\lambda} I$. Consider $(T - \lambda I)(T - \lambda I)^* = TT^* - \overline{\lambda} T - \lambda T^* + \lambda\overline{\lambda}$ and $(T - \lambda I)^*(T - \lambda I) = T^*T - \overline{\lambda} T - \lambda T^* + \lambda\overline{\lambda}$. Since, T is normal, $TT^* = T^*T$. So.

$$(T - \lambda I)(T - \lambda I)^* = (T - \lambda I)^*(T - \lambda I).$$

That is, $T - \lambda I$ is also normal. So, by Theorem 6.1.17, we have

$$\|(T - \lambda I)v\| = \|(T^* - \overline{\lambda} I)v\| = 0.$$

That is, $T^*v = \overline{\lambda} v$, or v is an eigenvector of T^* with eigenvalue $\overline{\lambda}$. ■

Theorem 6.1.19
Suppose $T \in \mathcal{L}(V)$ is normal. Then, eigenvectors of T corresponding to distinct eigenvalues are orthogonal.

Proof 15. Suppose α, β are distinct eigenvalues of T, with corresponding eigenvectors u, v. Then, $Tu = \alpha u$ and $Tv = \beta v$. By Theorem 6.1.18, we have $T^* v = \overline{\beta} v$. So, we have

$$\begin{aligned}(\alpha - \beta)\langle u, v\rangle &= \langle \alpha u, v\rangle - \langle u, \overline{\beta} v\rangle \\ &= \langle Tu, v\rangle - \langle U, T^* v\rangle \\ &= \langle Tu, v\rangle - \langle Tu, v\rangle \\ &= 0.\end{aligned}$$

Since $\alpha \neq \beta$, it must be $\langle u, v\rangle = 0$. So, u and v are orthogonal. ■

6.2 The Spectral Theorem

Theorem 6.2.1 Complex Spectral Theorem
Suppose $\mathbb{F} = \mathbb{C}$ and $T \in \mathcal{L}(V)$. Then, the following are equivalent:

1. T is normal.
2. V has an orthonormal basis consisting of eigenvectors of T.
3. T has a diagonal matrix with respect to some orthonormal basis of V.

Proof 1. Note that (2) $\iff$ (3) is obvious by Theorem 4.3.5. No we need to show (3) $\iff$ (1) to complete the proof. □

Suppose (3). Then, $\mathcal{M}(T)$ is diagonal. That is, $\mathcal{M}(T^*)$ is also diagonal. Then, $\mathcal{M}(T)\mathcal{M}(T^*) = \mathcal{M}(T^*)\mathcal{M}(T)$. That is $\mathcal{M}(TT^*) = \mathcal{M}(T^*T)$, or $TT^* = T^*T$. So, T is normal. □

Suppose (1). That is, T is normal. Then, by Schur's Theorem, $\exists$ an orthonormal basis $e_1, \cdots, e_n$ of V *s.t.* $\mathcal{M}(T, (e_1, e_n))$ is an upper triangular matrix. Suppose

$$\mathcal{M}(T, (e_1, \cdots, e_n)) = \begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ & \ddots & \vdots \\ 0 & & a_{n,n} \end{pmatrix}.$$

Then,

$$\mathcal{M}(T^*, (e_1, \cdots, e_n)) = \begin{pmatrix} \overline{a_{1,1}} & & 0 \\ \vdots & \ddots & \\ \overline{a_{1,n}} & \cdots & \overline{a_{n,n}} \end{pmatrix}.$$

Then, $Te_1 = a_{1,1}e_1$ and $T^*e_1 = \overline{a_{1,1}}e_1 + \cdots + \overline{a_{1,n}}e_n$. Further, note that $\|Te_1\|^2 = |a_{1,1}|^2$ and $\|T^*e_1\|^2 = |a_{1,1}|^2 + \cdots + |a_{1,n}|^2$. Since $\|Te_1\|^2 = \|T^*e_1\|^2$, we have $|a_{1,1}|^2 = |a_{1,1}|^2 + \cdots + |a_{1,n}|^2$. Then, it must be that $|a_{1,2}|^2 + \cdots + |a_{1,n}|^2 = 0$. Applying this procedure to $\|Te_2\|^2 = \|T^*e_2\|^2, \cdots, \|Te_n\|^2 = \|T^*e_n\|^2$, we have $|a_{j,k}| = 0$ when $j \neq k$. So, $\mathcal{M}(T)$ is a diagonal matrix. ■

Lemma 6.2.2 Invertible Quadratic Expressions Suppose $T \in \mathcal{L}(V)$ is self-adjoint and $b, c \in \mathbb{R}$ are *s.t.* $b^2 < 4c$. Then, $T^2 + bT + cI$ is invertible.

Proof 2. Let $v \in V$ *s.t.* $v \neq 0$. Note that

$$\begin{aligned} \langle (T^2 + bT + cI)v, v\rangle &= \langle T^2 v, v\rangle + b\langle Tv, v\rangle + c\langle v, v\rangle \\ &= \langle Tv, Tv\rangle + b\langle Tv, v\rangle + c\|v\|^2 && T\ is\ self-adjoint \\ &\geq \|Tv\|^2 - |b|\|Tv\|\|v\| + c\|v\|^2 && Cauchy-Schwarz \\ &= \left(\|Tv\| - \frac{|b|\|v\|}{2}\right)^2 + \left(c - \frac{b^2}{4}\right)\|v\|^2 \\ &> 0 && b^2 < 4c \end{aligned}$$

Then, $\forall v \neq 0,\ \langle (T^2 + bT + cI)v, v\rangle > 0$. So, it must be that $(T^2 + bT + cI)v = 0$ if and only if $v = 0$. Then, $\text{null}\,(T^2 + bT + cI) = \{0\}$. Thus, $T^2 + bT + cI$ is injective, and thus it is invertible. ■

Lemma 6.2.3 Suppose $V \neq \{0\}$ and $T \in \mathcal{L}(V)$ is a self-adjoint operator. Then, T has an eigenvalue.

Proof 3. Let $m = \dim V$ and choose $v \in V$. Then, $v, Tv, \cdots, T^n v$ cannot be L.I. because we have $n + 1 > \dim V$ vectors in the list. So, $\exists a_0, \cdots, a_n \in \mathbb{R}$ *s.t.* $a_0 v + a_1 Tv + \cdots + a_n T^n v = 0$. Make the a's the coefficient of a polynomial then

$$a_0 + a_1 x + \cdots + a_n x^n = c(x^2 + b_1 x + c_1) \cdots (x^2 + b_M x + c_M)(x - \lambda_1) \cdots (x - \lambda_m),$$

where c is a non-zero real number, each $b_j, c_j, \lambda_j \in \mathbb{R}$, each $b_j < 4c_j$, and $m + M \geq 1$. Then, we have

$$\begin{aligned} 0 &= a_0 v + a_1 Tv + \cdots + a_n T^n v \\ &= (a_0 I + a_1 T + \cdots + a_n T^n) v \\ &= c(T^2 + b_1 T + c_1 I) \cdots (T^2 + b_M T + c_M I)(T - \lambda_1 I) \cdots (T - \lambda_m I). \end{aligned}$$

By Lemma 6.2.2, $T^2 + b_j T + c_j I$ is invertible. Since $c \neq 0$, it must be that $0 = (T - \lambda_1 I) \cdots (T - \lambda_m I)$. Hence, $T - \lambda_j I$ is not injective for at least one j. So, T has at least one eigenvalue. ■

Definition 6.2.4 (Restriction Operator, $T|_U$). Suppose $T \in \mathcal{L}(V)$ and U is an invariant subspace of V under T. Then, the *restriction operator*, $T|_U \in \mathcal{L}(V)$, is defined as $T|_U(u) = Tu$ for $u \in U$.

Theorem 6.2.5
Suppose $T \in \mathcal{L}(V)$ is self-adjoint and U is a subspace of V that is invariant under T. Then,

1. $U^\perp$ is invariant under T;
2. $T|_U \in \mathcal{L}(U)$ is self-adjoint;
3. $T|_{U^\perp} \in \mathcal{L}(U^\perp)$ is self-adjoint.

Proof 4.

1. Suppose $v \in U^\perp$ and $u \in U$. Then, $\langle v, Tu \rangle = \langle Tv, u \rangle = 0$ since U is invariant under T (and hence $Tu \in U$) and $v \in U^\perp$. Then, we have $Tv \in U^\perp \quad \forall v \in U^\perp$, proving $U^\perp$ is an invariant subspace under T. □
2. Note that if $u, v \in U$, then

$$\langle (T|_U)u, v \rangle = \langle Tu, v \rangle = \langle u, Tv \rangle = \langle u, (T|_U)v \rangle.$$

   Therefore, $T|_U$ is self-adjoint. □
3. Replace U with $U^\perp$ in (2) and apply the conclusion from (1), and we complete the proof.

■

Theorem 6.2.6 Real Spectral Theorem
Suppose $\mathbb{F} = \mathbb{R}$ and $T \in \mathcal{L}(V)$. Then, the following are equivalent:

1. T is self-adjoint;
2. V has an orthonormal basis consisting of eigenvectors of T.
3. T has a diagonal matrix with respect to some orthonormal basis of V.

Proof 5. Similar to the complex case, (2) $\iff$ (3) is obvious. So, we will show (3) $\implies$ (1) and (1) $\implies$ (2) to complete the proof. □

Suppose (3) holds. Then, $\mathcal{M}(T)$ is diagonal. So, we have $\mathcal{M}(T)^t = \mathcal{M}(T)$. That is, $T = T^*$, and thus T is self-adjoint. $\square$

Suppose (1) holds. We will use mathematical induction on $\dim V$. $\boxed{\text{Base Case}}$ When $\dim V = 1$. Clearly, (1) $\implies$ (2). $\boxed{\text{Inductive Steps}}$ Assume $\dim V > 1$ and (1) $\implies$ (2) holds for all cases with dimension $\dim V - 1$. Let u be an eigenvector of T with $\|u\| = 1$. Let $U = \operatorname{span}(u)$. Then, $\dim U = 1$. Since $V = U \oplus U^{\perp}$, we know $\dim U^{\perp} = \dim V - \dim U = \dim V - 1$. So, (1) $\implies$ (2) holds on $U^{\perp}$. That is, $\exists$ an orthonormal basis of $U^{\perp}$ consisting of eigenvectors of $T|_{U^{\perp}}$. Now, add u to this orthonormal basis, we get a basis of V. Further since $u \in U$, this basis is an orthonormal basis of V consisting of eigenvectors of T. $\blacksquare$

6.3 Positive Operators and Isometries

Definition 6.3.1 (Positive Operator). An operator $T \in \mathcal{L}(V)$ is called *positive* if T is self-adjoint and $\langle Tv, v\rangle \geq 0 \quad \forall v \in V$.

Definition 6.3.2 (Square Root). An operator R is called a *square root* of an operator T if $R^2 = T$.

Example 6.3.3 Suppose $T \in \mathcal{L}(\mathbb{R}^3)$ and $R \in \mathcal{L}(\mathbb{R}^3)$ be defined as $T(z_1, z_2, z_3) = (z_3, 0, 0)$ and $R(z_1, z_2, z_3) = (z_2, z_3, 0)$. Then, R is a square root of T.

Proof 1. Since $R^2(z_1, z_2, z_3) = R(z_2, z_3, 0) = (z_3, 0, 0) = T(z_1, z_2, z_3)$, R is a square root of T. ■

Theorem 6.3.4 Characterization of Positive Operators
Let $T \in \mathcal{L}(V)$. Then, the following are equivalent:

1. T is positive;
2. T is self-adjoint and all the eigenvalues of T are non-negative;
3. T has a positive square root;
4. T has a self-adjoint square root;
5. $\exists$ an operator $R \in \mathcal{L}(V)$ *s.t.* $T = R^*R$.

Proof 2.

(1) $\Longrightarrow$ (2): Since T is positive, then T is self-adjoint. Let λ be an eigenvalue of T. Then, $Tv = \lambda v$ for some $v \in V$. Then, $\langle Tv, v\rangle = \langle \lambda v, v\rangle = \lambda\langle v, v\rangle = \lambda\|v\|^2$. Since T is positive, $\langle Tv, v\rangle \geq 0$. Further since $\|v\|^2 \geq 0$, it must also be $\lambda \geq 0$. So, we complete the proof. □

(2) $\Longrightarrow$ (3): Suppose T is self-adjoint and all the eigenvalues of T are non-negative. By the Spectrum Theorem, $\exists$ an orthonormal basis $e_1, \cdots, e_n$, where $e_1, \cdots, e_n$ are eigenvectors of T. Let $\lambda_1, \cdots, \lambda_n$ be the corresponding eigenvalues, where $\lambda_j \geq 0$. Let $R \in \mathcal{L}(V)$ *s.t.* $Re_j = \sqrt{\lambda_j}e_j$. Then

$$\begin{aligned} \langle Rv, v\rangle &= \left\langle a_1\sqrt{\lambda_1}e_1 + \cdots + a_n\sqrt{\lambda_n}e_n, a_1e_1 + \cdots + a_ne_n \right\rangle \\ &= |a_1|^2\sqrt{\lambda_1} + \cdots + |a_n|^2\sqrt{\lambda_n} \geq 0. \end{aligned}$$

Further, we can verity R is self-adjoint (proof omitted). So, R is positive by definition. Note that

$$R^2e_j = R\Big(\sqrt{\lambda_j}e_j\Big) = \sqrt{\lambda_j}\sqrt{\lambda_j}e_j = \lambda_je_j = Te_j.$$

So, R is a square root of T. □

(3) $\Longrightarrow$ (4): Suppose T has a positive square root. By definition, positive operators are self-adjoint. □

(4) $\Longrightarrow$ (5): Suppose T has a self-adjoint square root. Then, we have $R \in \mathcal{L}(V)$ *s.t.* $R^* = R$ and $R^2 = T$. That is, $T = R^2 = RR = R^*R$. □

(5) $\Longrightarrow$ (1): Suppose $\exists$ an operator $R \in \mathcal{L}(V)$ *s.t.* $T = T^*T$. Then,

$$T^* = (R^*R)^* = R^*(R^*)^* = R^*R = T.$$

So, T is self-adjoint. Now, since

$$\langle Tv, v\rangle = \langle R^*Rv, v\rangle = \langle Rv, Rv\rangle = \|Rv\|^2 \geq 0,$$

we have T is a positive operator. ■

Theorem 6.3.5
Each positive operator on V has a unique positive square root.

Proof 3. Let T be a positive operator on V. Select v to be an eigenvector of T with corresponding eigenvalue of λ. Then, we have $Tv = \lambda v$. Let R be a positive square root of T. Apply Spectrum Theorem to R, then $\exists$ an orthonormal basis $e_1, \cdots, e_n$, where $e_1, \cdots, e_n$ are eigenvectors of R. Then, $\exists \lambda_1, \cdots, \lambda_n \geq 0$ *s.t.* $Re_j = \sqrt{\lambda_j} e_j$. Suppose $v \in V$ and $v = a_1e_1 + \cdots + a_ne_n$. Then,

$$Rv = a_1\sqrt{\lambda_1}e_1 + \cdots + a_n\sqrt{\lambda_n}e_n \quad \text{and} \quad R^2v = a_1\lambda_1e_1 + \cdots + a_n\lambda_ne_n.$$

Further, $Tv = \lambda v = \lambda a_1e_1 + \cdots + \lambda a_ne_n$. Since $R^2v = Tv$, we know

$$a_1(\lambda_1 - \lambda)e_1 + \cdots + a_n(\lambda_n - \lambda)e_n = 0.$$

Since $e_1, \cdots, e_n$ is an orthonormal basis, for each $j = 1, \cdots, n$, we have $a_j(\lambda_j - \lambda) = 0$. So, it must be $a_j = 0$ or $\lambda_j = \lambda$. If $a_j = 0$, then we can delet it from the representation of v. So,

$$v = \sum_{\{j|\lambda_j = \lambda\}} a_je_j$$

Hence,

$$Rv = \sum_{\{j|\lambda_j=\lambda\}} a_j\sqrt{\lambda}e_j = \sqrt{\lambda}v.$$

■

Definition 6.3.6 (Isometry). An operator $S \in \mathcal{L}(V)$ is called an *isometry* if $\|Sv\| = \|v\| \quad \forall v \in V$. In other words, an operator is an isometry if ti preserves norms.

Example 6.3.7 Let $\lambda_1, \cdots, \lambda_n \in \mathbb{F}$ with $|\lambda_j|$ and $S \in \mathcal{L}(V)$ *s.t.* $Se_j = \lambda_je_j$ for some orthonormal bases $e_1, \cdots, e_n$ of V. Then, S is an isometry.

Proof 4. Let $v \in V$. Then, $v = \langle v, e_1\rangle e_1 + \cdots + \langle v, e_n\rangle e_n$. So, $\|v\|^2 = |\langle v, e_1\rangle|^2 + \cdots + |\langle v, e_n\rangle|^2$. Further, $Sv = \lambda_1\langle v, e_1\rangle e_1 + \cdots + \lambda_n\langle v, e_n\rangle e_n$, and thus $\|Sv\|^2 = |\lambda_1|^2|\langle v, e_1\rangle|^2 + \cdots + |\lambda_n|^2|\langle v, e_n\rangle|^2$. Since $|\lambda_j| = 1$, we know

$$\|Sv\|^2 = |\langle v, e_1\rangle|^2 + \cdots + |\langle v, e_n\rangle|^2 = \|v\|^2.$$

So, $\|Sv\| = \|v\|$ since $\|Sv\| \geq 0$ and $\|v\| \geq 0$. That is, by definition, S is an isometry. ■

Theorem 6.3.8 Characterization of Isometries
Suppose $S \in \mathcal{L}(V)$. Then, the following are equivalent:

1. S is an isometry.
2. $\langle Su, Sv \rangle = \langle u, v \rangle \quad \forall u, v \in V$;
3. $Se_1, \cdots, Se_n$ is orthonormal for every orthonormal list of vectors $e_1, \cdots, e_n$ in V;
4. $\exists$ an orthonormal basis $e_1, \cdots, e_n$ of V *s.t.* $Se_1, \cdots, Se_n$ is orthonormal;
5. $S^*S = I$;
6. $SS^* = I$;
7. S^* is an isometry;
8. S is invertible and $S^{-1} = S^*$.

Proof 5.
(1) $\Longrightarrow$ (2): Note that

$$\begin{aligned}\langle Su, Sv \rangle = \frac{\|Su + Sv\|^2 - \|Su - Sv\|^2}{4} &= \frac{\|S(u+v)\|^2 - \|S(u-v)\|^2}{4} \\ &= \frac{\|u+v\|^2 - \|u-v\|^2}{4} \\ &= \langle u, v \rangle \qquad \square\end{aligned}$$

(2) $\Longrightarrow$ (3): We have

$$\langle Se_i, Se_j \rangle = \langle e_i, e_j \rangle = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

So, $Se_1, \cdots, Se_n$ are orthonormal. $\square$

(3) $\Longrightarrow$ (4): Suppose $e_1, \cdots, e_m$ is orthonormal. We can extend it to a basis of V: $e_1, \cdots, e_m, v_{m+1}, \cdots, v_n$. Then, apply the Gram-Schmidt Procedure, we get an orthonormal basis, $e_1, \cdots, e_m, e_{m+1}, \cdots, e_n$ of V. $\square$

(4) $\Longrightarrow$ (5): Suppose $e_1, \cdots, e_n$ is an orthonormal basis of V. Then,

$$\langle S^*Se_j, e_k \rangle = \langle Se_j, Se_k \rangle = \langle e_j, e_j \rangle.$$

Suppose $u, v \in V$ *s.t.* $u = a_1e_1 + \cdots + a_ne_n$ and $v = b_1e_1 + \cdots + b_ne_n$. Then,

$$\begin{aligned}\langle S^*Su, v \rangle = \langle Su, Sv \rangle &= \langle S(a_1e_1 + \cdots + a_ne_n), S(b_1e_1 + \cdots + b_ne_n) \rangle \\ &= \langle a_1Se_1 + \cdots + a_nSe_n, b_1Se_1 + \cdots + b_nSe_n \rangle \\ &= \langle a_1Se_1, b_1Se_1 \rangle + \cdots + \langle a_nSe_n, b_nSe_n \rangle \\ &= a_1\overline{b_1}\|Se_1\|^2 + \cdots + a_n\overline{b_n}\|Se_n\|^2 \\ &= a_1\overline{b_1} + \cdots + a_n\overline{b_n} \\ &= \langle u, v \rangle.\end{aligned}$$

So, $S^*Su = u$, or $S^*S = I$. $\square$

(5) $\implies$ (6): Suppose $S^*S = I$. Then, $S = S^*$. So, $SS^* = I$. $\square$

(6) $\implies$ (7): Suppose $S^*S = I$. Then,

$$\|S^*v\|^2 = \langle S^*v, S^*v\rangle = \langle SS^*v, v\rangle = \langle v, v\rangle = \|v\|^2. \qquad \square$$

(7) $\implies$ (8): Suppose S^* is an isometry. Then, we know $S^*S = I$ and $SS^* = I$ by the proofs done above. So, S is invertible, and $S^{-1} = S^*$. $\square$

(8) $\implies$ (1): Finally, suppose S is invertible and $S^{-1} = S^*$. Then, $S^*S = I$. Note that

$$\|Sv\|^2 = \langle Sv, Sv\rangle = \langle S^*Sv, v\rangle = \langle v, v\rangle = \|v\|^2.$$

$\blacksquare$

Theorem 6.3.9
Suppose V is a complex inner product space and $S \in \mathcal{L}(V)$. Then, S is an isometry if and only if $\exists$ an orthonormal basis of V consisting of eigenvectors of S whose corresponding eigenvalues all have absolute value of 1.

Proof 6.

($\Rightarrow$): By the Spectrum Theorem, $\exists$ an orthonormal basis $e_1, \cdots, e_n$, where $e_1, \cdots, e_n$ are eigenvectors of S. Suppose $\lambda_1, \cdots, \lambda_n$ are the corresponding eigenvalues. Then, we have

$$\|Se_j\| = \|\lambda_j e_j\| = |\lambda_j|.$$

Since S is an isometry, $\|Se_j\| = \|e_j\| = 1$. So, $|\lambda_j| = \|Se_j\| = 1$. $\square$

($\Leftarrow$): This direction is proven in Example 6.3.7. $\blacksquare$

6.4 Polar Decomposition and SVD

Notation 6.4.1. If T is a positive operator, then $\sqrt{T}$ denotes the unique positive square root of T.

Remark. *We want to verify that the definition of* $\sqrt{T^*T}$ *is reasonable:* $\langle T^*Tv, v\rangle = \langle Tv, Tv\rangle \geq 0$. *Also,* $(T^*T)^* = T^*T$. *So,* T^*T *is a positive operator, and thus* $\sqrt{T^*T}$ *is well-defined.*

> **Theorem 6.4.2 Polar Decomposition**
> Suppose $T \in \mathcal{L}(V)$. Then, $\exists$ an isometry $S \in \mathcal{L}(V)$ *s.t.* $T = S\sqrt{T^*T}$.

Proof 1.
$\boxed{\text{Step 1}}$ Characteristics of $\operatorname{range}\sqrt{T^*T}$: Note that

$$\begin{aligned}\|Tv\|^2 = \langle Tv, Tv\rangle &= \langle T^*Tv, v\rangle \\ &= \langle \sqrt{T^*T}\sqrt{T^*T}v, v\rangle \\ &= \langle \sqrt{T^*T}v, \sqrt{T^*T}v\rangle \\ &= \|\sqrt{T^*T}v\|^2.\end{aligned}$$

So, $\forall v \in V$, we have $\|Tv\| = \|\sqrt{T^*T}v\|$. Define $S_1 : \operatorname{range}\sqrt{T^*T} \to \operatorname{range} T$ as $S_1\left(\sqrt{T^*T}v\right) = Tv$. Then, we have $\|S_1\sqrt{T^*T}v\| = \|Tv\|$.

1. Now, we want to verify that S_1 is well-defined. Suppose $v_1, v_2 \in V$ *s.t.* $\sqrt{T^*T}v_1 = \sqrt{T^*T}v_2$. Then,

$$\begin{aligned}\|Tv_1 - Tv_2\| = \|T(v_1 - v_2)\| &= \|\sqrt{T^*T}(v_1 - v_2)\| \\ &= \|\sqrt{T^*T}v_1 - \sqrt{T^*T}v_2\| \\ &= 0.\end{aligned}$$

   So, S_1 is well-defined.

2. Further, we want to show S_1 is linear. By using the linearity of T, we can easily prove that S_1 is also linear.

3. Additionally, S_1 is surjective by definition of S_1.

4. Also, S_1 is isometry. Note that $\forall u \in \operatorname{range}\sqrt{T^*T}$, we have $\|S_1u\| = \|u\|$ since $\|\sqrt{T^*T}v\| = \|Tv\|$.

5. Hence, S_1 is injective: Note that $\|S_1v\| = 0$ if and only if $\|v\| = 0$, which is equivalent to $v = 0$. So, $\operatorname{null} S_1 = \{0\}$. $\square$

$\boxed{\text{Step 2}}$ Extend S_1 to an isometry on V. Note that we have $\dim\operatorname{range}\sqrt{T^*T} = \dim\operatorname{range} T$. So, we know $\dim\left(\operatorname{range}\sqrt{T^*T}\right)^\perp = \dim\left(\operatorname{range} T\right)^\perp$. Select an orthonormal basis $e_1, \cdots, e_m$ of $\left(\operatorname{range}\sqrt{T^*T}\right)^\perp$ and an orthonormal basis $f_1, \cdots, f_m$ of $(\operatorname{range} T)^\perp$. Now, let's define $S_2 : \left(\operatorname{range}\sqrt{T^*T}\right)^\perp \to (\operatorname{range} T)^\perp$ as $S_1(a_1e_1 + \cdots + a_me_m) = a_1f_1 + \cdots + a_mf_m$. We can then not only show S_2 is well-defined but also S_2

is linear. Moreover, $\forall w \in \left(\text{range}\sqrt{T^*T}\right)^{\perp}$, if $w = a_1e_1 + \cdots + a_me_m$, we have

$$\begin{aligned}\|S_2w\|^2 = \|S_2(a_1e_1 + \cdots + a_me_m)\|^2 &= \|a_1f_1 + \cdots + a_mf_m\|^2 \\ &= |a_1|^2 + \cdots + |a_m|^2 \\ &= \|a_1e_1 + \cdots + a_me_m\|^2 \\ &= \|w\|^2.\end{aligned}$$

So, $\|S_2w\| = \|w\|$. Now, we define

$$Sv = \begin{cases} S_1v, & v \in \text{range}\sqrt{T^*T} \\ S_2v, & v \in \left(\text{range}\sqrt{T^*T}\right)^{\perp}\end{cases}$$

Note that since $V = \text{range}\sqrt{T^*T} \oplus \left(\text{range}\sqrt{T^*T}\right)^{\perp}$, we can uniquely represent $v \in V$ as $v = u + w$ for some $u \in \text{range}\sqrt{T^*T}$ and $w \in \left(\text{range}\sqrt{T^*T}\right)^{\perp}$. Hence, we can also write the definition of S as $Sv = S_1u + S_2w$. If we select $\sqrt{T^*T}v \in \text{range}\sqrt{T^*T}$, then we have $S\left(\sqrt{T^*T}v\right) = S_1\left(\sqrt{T^*T}v\right) = Tv$. Therefore, $T = S\sqrt{T^*T} \quad \forall v \in V$. □

Finally, we will show S is an isometry. Note that $v = u + w$. So, by Pythagorean Theorem,

$$\begin{aligned}\|Sv\|^2 = \|S_1u + S_2w\|^2\|S_1u\|^2 + \|S_2w\|^2 \\ &= \|u\|^2 + \|w\|^2 \\ &= \|v\|^2.\end{aligned}$$

■

Definition 6.4.3 (Singular Values). Suppose $T \in \mathcal{L}(V)$. The *singular values* of T are the eigenvalues of $\sqrt{T^*T}$, with each eigenvalue λ repeated $\dim E\left(\lambda, \sqrt{T^*T}\right)$ times.

Example 6.4.4 Define $T \in \mathcal{L}(\mathbb{F}^4)$ by

$$T(z_1, z_2, z_3, z_4) = (0, 3z_1, 2z_2, -3z_4).$$

Find the singular values of T.

Solution 2.

Suppose $v = (z_1, z_2, z_3, z_4) \in \mathbb{F}^4$ and $w = (y_1, y_2, y_3, y_4) \in \mathbb{F}^4$. Consider

$$\begin{aligned}\langle v, T^*w\rangle &= \langle Tv, w\rangle \\ &= \langle (0, 3z_1, 2z_2, -3z_4), (y_1, y_2, y_3, y_4)\rangle \\ &= 3z_1\overline{y_2} + 2z_2\overline{y_3} - 3z_4\overline{y_4} \\ &= \langle (z_1, z_2, z_3, z_3), (3y_2, 2y_3, 0, -3y_4)\rangle.\end{aligned}$$

So, $T^*w = T^*(y_1, y_2, y_3, y_4) = (3y_2, 2y_3, 0, -3y_4)$. Then, $T^*T(z_1, z_2, z_3, z_4) = (9z_1, 4z_2, 0, 9z_4)$. Then, $\sqrt{T^*T}(z_2, z_2, z_3, z_4) = (3_z1, 2z_2, 0, 3z_4)$. So, the eigenvalues of $\sqrt{T^*T}$ are 3, 2, and 0. Also,

$$\dim E\left(3, \sqrt{T^*T}\right) = 2, \quad \dim E\left(2, \sqrt{T^*T}\right) = \dim E\left(0, \sqrt{T^*T}\right) = 1.$$

So, the singular values are $3, 3, 2, 0$. □

Theorem 6.4.5 Singular Value Decomposition (SVD)
Suppose $T \in \mathcal{L}(V)$ has singular values $s_1, \cdots, s_n$. Then, $\exists$ orthonormal bases $e_1, \cdots, e_n$ and $f_1, \cdots, f_n$ of V *s.t.*

$$Tv = s_1 \langle v, e_1 \rangle f_1 + \cdots + s_n \langle v, e_n \rangle f_n$$

for every $v \in V$.

Remark. *Relevant Theorem used in proving SVD: Spectrum Theorem, Characterization and Properties of Isometry, and Polar Decomposition.*

Proof 3. Apply the Spectrum Theorem to $\sqrt{T^*T}$, we know $\exists$ an orthonormal basis $e_1, \cdots, e_n$ of V *s.t.*

$$\sqrt{T^*T} e_j = s_j e_j \quad \forall j = 1, \cdots, n.$$

Note that $\forall v \in V$, we have

$$v = \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_n \rangle e_n \tag{17}$$

Apply $\sqrt{T^*T}$ to Equation (17) we have

$$\sqrt{T^*T} v = s_1 \langle v, e_1 \rangle e_1 + \cdots + s_n \langle v, e_n \rangle e_n. \tag{18}$$

By Polar Decomposition, $\exists$ an isometry $S \in \mathcal{L}(V)$ *s.t.* $T = S\sqrt{T^*T}$. Apply S to Equation (18), we get

$$S\left(\sqrt{T^*T} v\right) = s_1 \langle v, e_1 \rangle S e_1 + \cdots + s_n \langle v, e_n \rangle S e_n.$$

By the characteristics of isometry, since $e_1, \cdots, e_n$ is an orthonormal basis, $Se_1, \cdots, Se_n$ is also an orthonormal basis. Let $f_j = Se_j$. Then,

$$Tv = S\sqrt{T^*T} v = s_1 \langle v, e_e \rangle f_1 + \cdots + s_n \langle v, e_n \rangle f_n.$$

■

Theorem 6.4.6
Suppose $T \in \mathcal{L}(V)$. Then, the singular values of T are the non-negative square roots of the eigenvalues of T^*T, with each eigenvalue λ repeated $\dim E(\lambda, T^*T)$ times.

Proof 4. By the Spectrum Theorem, $\exists$ an orthonormal basis $e_1, \cdots, e_n$ and non-negative number $\lambda_1, \cdots, \lambda_n$ *s.t.* $T^*Te_j = \lambda_j e_j \quad \forall j = 1, \cdots, n$. Then, we have $\sqrt{T^*T} e_j = \sqrt{\lambda_j} e_j \quad \forall j = 1, \cdots, n$, which completes the proof. ■

7 Operators on Complex Vector Spaces

7.1 Generalized Eigenvectors, Nilpotent Operators

Theorem 7.1.1
Suppose $T \in \mathcal{L}(V)$. Then,

$$\{0\} \subseteq \operatorname{null} T^0 \subseteq \operatorname{null} T^1 \subseteq \cdots \subseteq \operatorname{null} T^k \subseteq \operatorname{null} T^{k+1} \subseteq \cdots$$

Proof 1. Let $k \in \mathbb{N}^+$. Let $v \in \operatorname{null} T^k$. Then, $T^k v = 0$. Then, we know $T(T^k v) = T^{k+1} v = 0$. So, $v \in \operatorname{null} T^{k+1}$. That is, $\operatorname{null} T^k \subseteq \operatorname{null} T^{k+1}$ as desired. ■

Theorem 7.1.2
Suppose $T \in \mathcal{L}(V)$. Suppose m is a non-negative integer *s.t.* $\operatorname{null} T^m = \operatorname{null} T^{m+1}$. Then,

$$\operatorname{null} T^m = \operatorname{null} T^{m+1} = \operatorname{null} T^{m+2} = \operatorname{null} T^{m+3} = \cdots$$

Proof 2. Let $k \in \mathbb{N}$. We've already shown $\operatorname{null} T^{m+k} \subseteq \operatorname{null} T^{m+k+1}$ in Theorem 7.1.1. Now, let $v \in \operatorname{null} T^{m+k+1}$. So, $T^{m+k+1}(v) = 0$. That is, $T^{m+1}\big(T^k v\big) = 0$. So, $T^k v \in \operatorname{null} T^{m+1} = \operatorname{null} T^m$. In other words, $T^m\big(T^k v\big) = T^{m+k}(v) = 0$. So, $v \in \operatorname{null} T^{m+k}$. Then, $\operatorname{null} T^{m+k+1} \subseteq \operatorname{null} T^{m+k}$. Hence,

$$\operatorname{null} T^{m+k} = \operatorname{null} T^{m+k+1}.$$

■

Theorem 7.1.3
Suppose $T \in \mathcal{L}(V)$. Let $n = \dim V$. Then,

$$\operatorname{null} T^n = \operatorname{null} T^{n+1} = \operatorname{null} T^{n+2} = \cdots$$

Proof 3. Suppose for the sake of contradiction that $\operatorname{null} T^n \neq \operatorname{null} T^{n+1}$. Then,

$$\operatorname{null} T^0 \subsetneq \operatorname{null} T \subsetneq T^2 \subsetneq \cdots \subsetneq \operatorname{null} T^n \subsetneq T^{n+1}.$$

As the symbol $\subsetneq$ means "contained in but not equal to," at each of the strict inclusions in the chain above, the dimension increases by at least 1. That is, $\dim \operatorname{null} T^{n+1} \geq n+1$. ⁎ This is a contradiction because a subspace of V ($\operatorname{null} T^{n+1}$) cannot be a dimension larger than $\dim V = n$. So, it must be that our assumption is wrong, and $\operatorname{null} T^n = \operatorname{null} T^{n+1}$. ■

Theorem 7.1.4
Suppose $T \in \mathcal{L}(V)$. Let $n = \dim V$. Then,

$$V = \operatorname{null} T^n \oplus \operatorname{range} T^n.$$

Proof 4. Note that $\dim V = \dim \operatorname{null} T^n + \dim \operatorname{range} T^n$ by the Fundamental Theorem of Linear Maps. So, we only need to prove $(\operatorname{null} T^n) \cap (\operatorname{range} T^n) = \{0\}$. Let $v \in (\operatorname{null} T^n) \cap (\operatorname{range} T^n)$. Then, $\exists u \in V$ ***s.t.*** $v = T^n u$. Since $v \in \operatorname{null} T^n$, $T^N v = T^n(T^n u) = 0$. That is, $T^{2n} u = T^n v = 0$. Therefore, $u \in$

null T^{2n} = null T^n. So, we now have $T^n u = 0$. Hence, $v = T^n u = 0$. Then, i (null T^n) $\cap$ (range T^n) $= \{0\}$, and thus $V = \text{null } T^n \oplus \text{range } T^n$. ■

Definition 7.1.5 (Generalized Eigenvector). Suppose $T \in \mathcal{L}(V)$ and λ is an eigenvalue of T. A vector $v \in V$ is called a *generalized eigenvector* of T corresponding to λ if $v \neq 0$ and $(T - \lambda I)^j v = 0$ for some positive integer j.

Definition 7.1.6 (Generalized Eigenspace, $G(\lambda, T)$). Suppose $T \in \mathcal{L}(V)$ and $\lambda \in \mathbb{F}$. The *generalized eigenspace* of T corresponding to λ, denoted $G(\lambda, T)$, is defined to be the set of all generalized eigenvectors of T corresponding to λ, along with the 0 vector.

Theorem 7.1.7
Suppose $T \in \mathcal{L}(V)$ and $\lambda \in \mathbb{F}$. Then,

$$G(\lambda, T) = \text{null } (T - \lambda I)^{\dim V}.$$

Proof 5.
($\subseteq$): Let $v \in G(\lambda, T)$. Then, $\exists\, j \in \mathbb{N}^+$ *s.t.*

$$v \in \text{null } (T - \lambda I)^j.$$

Since null $(T - \lambda I)^j \subseteq$ null $(T - \lambda)^{j+1} \subseteq \cdots \subseteq$ null $(T - \lambda I)^{\dim V}$, we have $v \in$ null $(T - \lambda I)^{\dim V}$. So, $G(\lambda, T) \subseteq$ null $(T - \lambda I)^{\dim V}$.

($\supseteq$): Conversely, suppose $v \in$ null $(T - \lambda I)^{\dim V}$. Then,

$$(T - \lambda I)^{\dim V} v = 0.$$

By definition, v is a generalized eigenvector, and so $v \in G(\lambda, T)$. Then, null $(T - \lambda I)^{\dim V} \subseteq G(\lambda, T)$. ■

Theorem 7.1.8
Let $T \in \mathcal{L}(V)$. Suppose $\lambda_1, \ldots, \lambda_m$ are distinct eigenvalues of T and $v_1, \ldots, v_m$ are corresponding generalized eigenvectors. Then, $v_1, \ldots, v_m$ is L.I..

Proof 6. Let $a_1, \ldots, a_m \in \mathbb{C}$ *s.t.*

$$0 = a_1 v_1 + \cdots + a_m v_m. \tag{19}$$

Let k be the largest non-negative integer such that $(T - \lambda_1 I)^k v_1 \neq 0$. Let $w = (T - \lambda_1)^k v_1$, then

$$\begin{aligned}(T - \lambda_1 I)w &= (T = \lambda_1 I)(T - \lambda_1 I)^k v = 0\\ &= (T - \lambda_1 I)^{k+1} v = 0\end{aligned}$$

So, w is an eigenvector, and

$$Tw = \lambda_1 w. \tag{20}$$

Minus λw from both sides of Equation (20), we have

$$(T - \lambda I)w = (\lambda_1 - \lambda)w \qquad \forall v \in \mathbb{F}$$

Then, $(T - \lambda I)^n w = (\lambda_1 - \lambda)^n w,\ \lambda \in \mathbb{F},\ n = \dim V$. Apply the operator $(T - \lambda_1 I)^k (T - \lambda_2 I)^n \cdots (T - \lambda_m I)^m$

to both sides of Equation (19), we have

$$\begin{aligned}
0 &= (T-\lambda_1 I)^k (T-\lambda_2 I)^n \cdots (T-\lambda_m I)^n (a_1 v_1 + \cdots + a_m v_m) \\
&= (T-\lambda_1 I)^k (T-\lambda_2 I)^n \cdots (T-\lambda_m I)^n (a_m v_m) + \cdots + (T-\lambda_1 I)^k (T-\lambda_2 I)^n \cdots (T-\lambda_m I)^n (a_1 v_1) \\
&= (T-\lambda_1 I)^k (T-\lambda_2 I)^n \cdots (T-\lambda_m I)^n (a_1 v_1) \\
&= a_1 (T-\lambda_2 I)^n \cdots (T-\lambda_m I)^n w \\
&= a_1 \underbrace{(T-\lambda_1 I)^k (T-\lambda_2 I)^n \cdots (T-\lambda_m I)^n}_{\neq 0} \underbrace{w}_{\neq 0}
\end{aligned}$$

So, it must be $a_1 = 0$. Apply the same rationale, we can show $a_1 = \cdots = a_m = 0$. Therefore, $v_1, \ldots, v_m$ is L.I. by definition. ■

Definition 7.1.9 (Nilpotent). An operator is called *nilpotent* if some power of it equals 0.

Theorem 7.1.10
Suppose $N \in \mathcal{L}(V)$ is nilpotent. Then, $N^{\dim V} = 0$.

Proof 7. Note that $\operatorname{null}\,(N-0I)^{\dim V} = G(0, N) = V$. So, we have proven $N^{\dim V} = 0$. ■

Lemma 7.1.11 Suppose $N \in \mathcal{L}(V)$ has a basis such that $\mathcal{M}(N)$ is an upper-triangular matrix with its diagonal all 0. Then, N is nilpotent.

Proof 8. Suppose the basis is $v_1, \ldots, v_n$ and

$$A = \mathcal{M}(N) = \begin{pmatrix} 0 & & * \\ & \ddots & \\ & & 0 \end{pmatrix}.$$

Then,

$$\begin{aligned}
&Nv_1 = 0 \\
&Nv_2 = A_{1,2} v_1 + 0, \quad N^2 v_2 = A_{1,2} N v_1 = 0 \\
&\quad\vdots \\
&Nv_n = A_{1,n} v_1 + \cdots + A_{n-1,n} v_{n-1} + 0.
\end{aligned}$$

So, $N^n v_n = A_{1,n} N^{n-1} v_1 + A_{2,n} N^{n-1} v_2 + \cdots + A_{n-1,n} N^{n-1} v_{n-1} = 0$. That is, $N^n = 0$. So, we've shown that N is nilpotent. ■

Theorem 7.1.12 Matrix of a Nilpotent Operator
Suppose N is a nilpotent operator on V. Then, $\exists$ a basis of V with respect to which the matrix of N has the form

$$\begin{pmatrix} 0 & & * \\ & \ddots & \\ 0 & & 0 \end{pmatrix};$$

where all entries on and below the diagonal are 0's.

Proof 9. Let $k \in \mathbb{N} \cup \{0\}$ be the smallest such that $N^k = 0$. So, we have $\operatorname{null} N^k = V$ and $k \leq n$. So,

$N^j \neq 0 \quad \forall j < k$. So, we have

$$\{0\} = \text{null } N^0 \subsetneq \text{null } N^1 \subsetneq \text{null } N^2 \subsetneq \cdots \subsetneq \text{null } N^k.$$

Select $v_1^1, \ldots, v_n^1, v_1^2, \ldots, v_{n_2}^2, \ldots, v_1^k, \ldots, v_{n_k}^k$ as a basis of N. It can be also written as $v_1, \ldots, v_n$.

1. Let j be an index such that $v_j \in \text{null } N$. Then, $Nv_j = 0$.
2. Let j be an index such that $v_j \in \text{null } N^2$. Then, $N^2(v_j) = N(Nv_j) = 0$. So, $Nv_j \in \text{null } N$.

So, $Nv_j = \sum_{\{i | v_i \in \text{null } N\}} A_{i,j} v_j, \quad i < j.$ ■

Theorem 7.1.13
Let $T \in \mathcal{L}(V)$ *s.t.* T is no nilpotent. Suppose $\dim V = n$. Then, $V = \text{null } T^{n-1} \oplus \text{range } T^{n-1}$.

Proof 10. Since T is not nilpotent, $N^n \neq 0$. So, $\text{null } N^n \subsetneq V$. That is,

$$0 \subseteq \text{null } T \subseteq \text{null } T^2 \subseteq \cdots \subseteq \text{null } T^{n-1} \subseteq \text{null } T^n \subsetneq V.$$

So, it must be the case that $\text{null } T^{n-1} = \text{null } T^n$.

Suppose $v \in \left(\text{null } T^{n-1}\right) \cap \left(\text{range } T^{n-1}\right)$. Then, $\exists u \in V$ *s.t.* $T^{n-1}u = v$. Note that

$$T^{n-1}v = T^{n-1}\left(T^{n-1}u\right) = T^{2n-2}u = T^n u = 0.$$

So, $u \in \text{null } T^n = \text{null } T^{n-1}$. That is, $T^{n-1}u = 0$. So, $v = 0$. Then, $\left(\text{null } T^{n-1}\right) \cap \left(\text{range } T^{n-1}\right) = \{0\}$, and thus $V = \text{null } T^{n-1} \oplus \text{range } T^{n-1}$. ■

Theorem 7.1.14
Suppose $T \in \mathcal{L}(V)$, $\alpha, \beta \in \mathbb{F}$ with $\alpha \neq \beta$. Then,

$$G(\alpha, T) \cap G(\beta, T) = \{0\}.$$

Proof 11. Let $v \in G(\alpha, T) \cap G(\beta, T)$ with $v \neq 0$. Then, we know v is a generalized eigenvector of α and β at the same time. However, given $\alpha \neq \beta$, their corresponding generalized eigenvectors should be L.I.. ⋇ This contradicts with the fact that v cannot be L.I. with v. Then, our assumption is wrong, and $G(\alpha, T) \cap G(\beta, T) = \{0\}$. ■

7.2 Decomposition of an Operator

Theorem 7.2.1
Suppose $T \in \mathcal{L}(V)$ and $p = \mathcal{P}(\mathbb{F})$. Then, $\operatorname{null} p(T)$ and $\operatorname{range} p(T)$ are invariant under T.

Proof 1. Let $v \in \operatorname{null} p(T)$. Then, $p(T)(Tv) = T(p(T)v) = T(0) = 0$. So, $\operatorname{null} p(T)$ is invariant under T. Suppose $v \in \operatorname{range} p(T)$, then $\exists u \in V$ ***s.t.*** $p(T)u = v$. Then, $Tv = T(p(T)u) = p(T)(Tu) \in \operatorname{range} p(T)$. So, $\operatorname{range} p(T)$ is also invariant under T. ■

Theorem 7.2.2
Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Let $\lambda_1, \ldots, \lambda_m$ be the distinct eigenvalues of T. Then,

1. $V = G(\lambda_1, T) \oplus \cdots \oplus G(\lambda_m, T)$.
2. each $G(\lambda_j, T)$ is invariant under T.
3. each $(T - \lambda_j I) \mid_{G(\lambda_j,T)}$ is nilpotent.

Proof 2.

1. We will prove it by induction. Obviously, the conclusion follows when $n = 1$. Now, consider $n > 1$. Suppose the conclusion holds for all spaces with dimension $\leq n - 1$.

   WTS: the conclusion is true for $\dim V = n$.

   Consider $V = \operatorname{null}(T - \lambda_1 I)^n \oplus \operatorname{range}(T - \lambda_1 I)^n = G(\lambda_1, T) \oplus U$ if we fix $U = \operatorname{range}(T - \lambda_1 I)^n$. Obviously ,$G(\lambda_1, T) \neq \{0\}$. So, $\dim U < n$, and so our inductive hypothesis is applicable to U. Note that $G(\lambda_i, T) \cap G(\lambda_j, T) = \{0\}$ if $i \neq j$. Then, $\lambda_2, \ldots, \lambda_m$ are eigenvalues of $T \mid_U$. So, $U = G(\lambda_2, T\mid_U) \oplus \cdots \oplus G(\lambda_m, T\mid_U)$. Then, $V = G(\lambda_1, T) \oplus G(\lambda_2, T\mid_U) \oplus \cdots \oplus G(\lambda_m, T\mid_U)$.

   WTS: $G(\lambda_j, T\mid_U) = G(\lambda_j, T)$

   Note that $G(\lambda_j, T\mid_U) \subseteq G(\lambda_j, T)$ is evident. Conversely, suppose $v \in G(\lambda_k, T) \subseteq V$. Then, $v = v_1 + u$ for some $v_1 \in G(\lambda_1, T)$ and $u \in U$. Further, by our inductive hypothesis, we have

   $$u = v_2 + \cdots + v_m \quad \text{for some} \quad v_j \in G(\lambda_j, T\mid_U) \subseteq G(\lambda_j, T).$$

   Then, $v = v_1 + u = v_1 + v_2 + \cdots + v_m \in G(\lambda_k, T)$. That is, $v_1 + \cdots + (v_k - v) + \cdots + v_m = 0$. Then, $v_1 \in G(\lambda_1, T), \ldots, v_k - v \in G(\lambda_k, T), \ldots, v_m \in G(\lambda_m, T)$. Therefore, $v_1, \ldots, v_k - v, \ldots, v_m$ are L.I.. So, it must be that $v_1 = \cdots = v_k - 2 = \cdots = v_m = 0$. So, $v = v_1 + u = 0 + u = u$. Then, $v \in U$. So, $v \in G(\lambda_k, T) \cap U = G(\lambda_k, T\mid_U)$. As k was arbitrary, we've shown $G(\lambda_k, U) \subseteq G(\lambda_k, T\mid_U)$. So, $G(\lambda_j, T\mid_U) = G(\lambda_j, T)$. We complete our proof.

2. Note that $G(\lambda_j, T) = \operatorname{null}(T - \lambda_j I)^n = \operatorname{null} p(T)$ if $p(z) = (z - \lambda_j)^n$. By Theorem 7.2.1, $\operatorname{null} p(T)$ is invariant under T. So, it follows that $G(\lambda_j, T)$ is also invariant under T. □

3. By definition, we have $G(\lambda_j, T) = \operatorname{null}(T - \lambda_j I)^n$. Then, $\left[(T - \lambda_j I)\mid_{G(\lambda_j,T)}\right]^n = 0$. So, by definition, $(T - \lambda_j I)\mid_{G(\lambda_j,T)}$ is nilpotent.

■

Corollary 7.2.3 Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Then, $\exists$ a basis of V consisting of generalized eigenvectors of T.

Definition 7.2.4 (Multiplicity). Suppose $T \in \mathcal{L}(V)$. The *(algebraic) multiplicity* of an eigenvalue λ of T is defined to be the dimension of the corresponding generalized eigenspace $G(\lambda, T)$. In other words, the multiplicity of an eigenvalue λ of T equals $\dim \operatorname{null} (T - \lambda I)^{\dim V}$. The *geometric multiplicity* of an eigenvalue λ of T is $\dim E(\lambda, T)$.

Theorem 7.2.5
Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Then, the sum of the multiplicities of all eigenvalues of T equals $\dim V$.

Proof 3. By Theorem 7.2.2 (1), we know $V = G(\lambda_1, T) \oplus \cdots \oplus G(\lambda_m, T)$. So, we have

$$\dim V = \dim G(\lambda_1, T) + \cdots + \dim G(\lambda_m, T).$$

■

Definition 7.2.6 (Block Diagonal Matrix). A *block diagonal matrix* is a square matrix of the form

$$\begin{pmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_m \end{pmatrix},$$

where $A_1, \ldots, A_m$ are square matrices lying along the diagonal and all the other entries of the matrix equal 0.

Theorem 7.2.7
Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Let $\lambda_1, \ldots, \lambda_m$ be the distinct eigenvalues of T, with multiplicities $d_1, \ldots, d_m$. Then, $\exists$ a basis of V with respect to which T has a black diagonal matrix of the form

$$\begin{pmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_m \end{pmatrix},$$

where each a_j is d_j-by-d_j upper-triangular matrix of the form

$$A_j = \begin{pmatrix} \lambda_j & & * \\ & \ddots & \\ 0 & & \lambda_j \end{pmatrix}.$$

Proof 4. Note that $Tv_k = A_{1,k}v_1 + \cdots + A_{k,k}v_k + \cdots + A_{n,k}v_n$. Also, $(T - \lambda_j I) \mid_{G(\lambda_j, T)}$ is nilpotent. For each $G(\lambda_j, T)$, choose a basis of $G(\lambda_j, T)$ and $\dim G(\lambda_j, T) = d_j$. Then,

$$\mathcal{M}\Big((T - \lambda_j I) \mid_{G(\lambda_j, T)}\Big) = \begin{pmatrix} 0 & & * \\ & \ddots & \\ 0 & & 0 \end{pmatrix}.$$

Since $\mathcal{M}\Big((T-\lambda_j I)\mid_{G(\lambda_j,T)}\Big) = \mathcal{M}\Big(T\mid_{G(\lambda_j,T)}\Big) - \mathcal{M}(\lambda_j I)$, we have

$$\begin{aligned}
\mathcal{M}\Big(T\mid_{G(\lambda_j,T)}\Big) &= \begin{pmatrix} 0 & & * \\ & \ddots & \\ 0 & & 0 \end{pmatrix} + \mathcal{M}(\lambda_j I) \\
&= \begin{pmatrix} 0 & & * \\ & \ddots & \\ 0 & & 0 \end{pmatrix} + \begin{pmatrix} \lambda_j & & * \\ & \ddots & \\ 0 & & \lambda_j \end{pmatrix} \\
&= \begin{pmatrix} \lambda_j & & * \\ & \ddots & \\ 0 & & \lambda_j \end{pmatrix}.
\end{aligned}$$

Put all the bases of $G(\lambda_j, T)$ together, we have completed the proof. ∎

7.3 Characteristic and Minimal Polynomials

Definition 7.3.1 (Characteristic Polynomial). Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Let $\lambda_1, \ldots, \lambda_m$ denote the distinct eigenvalues of T, with multiplicities $d_1, \ldots, d_m$. The polynomial

$$(z-\lambda_1)^{d_1} \cdots (z-\lambda_m)^{d_m}$$

is called the *characteristic polynomial* of T.

Theorem 7.3.2
Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Then,

1. the characteristic polynomial of T has degree $\dim V$;
2. the zeros of the characteristic polynomial of T are eigenvalues of T.

Proof 1.

1. Note that $V = G(\lambda_1, T) \oplus \cdots \oplus G(\lambda_m, T)$. So, $\dim V = d_1 + \cdots + d_m$. That is, the characteristic polynomial of T has degree $\dim V$. □
2. By the definition of characteristic polynomial, it is evidently true.

■

Theorem 7.3.3 Cayley-Hamilton Theorem
Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Let q denote the characteristic polynomial of T. Then, $q(T) = 0$.

Proof 2. Suppose $\lambda_1, \ldots, \lambda_m$ are distinct eigenvalues of T and $d_1, \ldots, d_m$ are their corresponding multiplicities. For each $j = 1, \ldots, m$, we have $(T - \lambda_j I)\mid_{G(\lambda_j, T)}$ is nilpotent. Then, $(T-\lambda_j I)^{d_j}\mid_{G(\lambda_j,T)} = 0$. Since $q(z) = (z-\lambda_1)^{d_1} \cdots (z-\lambda_m)^{d_m}$, we know $q(T) = (T-\lambda_1 I)^{d_1} \cdots (T-\lambda_m I)^{d_m}$. Consider $v \in V$. Since $V = G(\lambda_1, T) \oplus \cdots \oplus G(\lambda_m, T)$, then $v = a_1 v_1 + \cdots + a_m v_m$, where $v_j \in G(\lambda_j, T)$. Then,

$$\begin{aligned} q(T)v &= q(T)(a_1 v_1 + \cdots + a_m v_m) \\ &= a_1 q(T) v_1 + \cdots + a_m q(T) v_m. \end{aligned}$$

For simplicity, consider

$$\begin{aligned} q(T)v_j &= (T-\lambda_1 I)^{d_1} \cdots (T-\lambda_m I)^{d_m} v_j \\ &= (T-\lambda_1 I)^{d_1} \cdots (T-\lambda_m I)^{d_m} (T-\lambda_j I)^{d_j} v_j. \end{aligned}$$

Since $v_j \in G(\lambda_j, T)$, we know $(T-\lambda_j I)^{d_j} v_j = 0$. Then, $q(T)v_j = 0$ for each $j = 1, \ldots, m$. So, $q(T)v = 0$. That is, $q(T) = 0$. ■

Definition 7.3.4 (Monic Polynomial). A *monic polynomial* is a polynomial whose highest-degree coefficient equals 1.

Theorem 7.3.5
Suppose $T \in \mathcal{L}(V)$. Then, $\exists$ a unique monic polynomial p of smallest degree such that $p(T) = 0$.

Proof 3. Let $\dim V = n$. Then, the list $I, T, T^2, \ldots, T^{n^2}$ is not L.I. in $\mathcal{L}(V)$ because $\mathcal{L}(V)$ has dimension n^2 and we have a list of length $n^2 + 1$. Let m be the smallest positive integer such that the list $I, T, T^2, \ldots, T^m$ is linearly dependent. Then, by the Linear Dependence Lemma, T^m is a linear combination of $I, T, \ldots, T^{m-1}$. So, we have

$$a_0 I + a_1 T + a_2 T^2 + \cdots + a_{m-1} T^{m-1} + T^m = 0 \tag{21}$$

Define a monic $p \in \mathcal{P}(\mathbb{F})$ as $p(z) = a_0 + z_1 z + a_2 z^2 + \cdots + a_{m-1} z^{m-1} + z^m$. Then, Equation (21) implies $p(T) = 0$. Now, we will prove the uniqueness. Suppose $\exists$ a monic $q \in \mathcal{P}(\mathbb{F})$ with $\deg q = m$ *s.t.* $q(T) = 0$. Then, $(p - q)(T) = p(T) - q(T) = 0$ and $\deg(p - q) < m$. Hence, $p = q$. ■

Definition 7.3.6 (Minimal Polynomial). Suppose $T \in \mathcal{L}(V)$. Then, the *minimal polynomial* of T is the unique monic polynomial p of smallest degree such that $p(T) = 0$.

Corollary 7.3.7 By the Cayley-Hamilton Theorem, the minimal polynomial of each $T \in \mathcal{L}(V)$ has degree $\leq \dim V$.

Theorem 7.3.8 Division Algorithm of Polynomials
Suppose $p, s \in \mathcal{P}(\mathbb{F})$ with $s \neq 0$. Then, $\exists$ unique $q, r \in \mathcal{P}(\mathbb{F})$ *s.t.* $p = sq + r$ and $\deg r < \deg s$.

Proof 4. Let $\deg p = n$ and $\deg s = m$. If $n < m$, then $q = 0$ and $r = p$. Now, we assume $n \geq m$. Define $T : \mathcal{P}_{n-m}(\mathbb{F}) \times \mathcal{P}_{m-1}(\mathbb{F}) \to \mathcal{P}_n(\mathbb{F})$ as $T(q, r) = sq + r$. It is easy to verify that T is a linear map. If $(q, r) \in \operatorname{null} T$, then $sq + r = 0$. So, $q = r = 0$. That is, $\dim \operatorname{null} T = 0$ and T is injective. Further, note that $\dim(\mathcal{P}_{n-m}(\mathbb{F}) \times \mathcal{P}_{m-1}(\mathbb{F})) = (n - m + 1) + (m - 1 + 1) = n + 1$ and $\dim \operatorname{range} T = n + 1 = \dim \mathcal{P}_n(\mathbb{F})$. Since $\operatorname{range} T \subseteq \mathcal{P}_n(\mathbb{F})$ and $\dim \operatorname{range} T = \dim \mathcal{P}_n(\mathbb{F})$, we have $\operatorname{range} T = \mathcal{P}_n(\mathbb{F})$. Therefore, T is surjective. ■

Theorem 7.3.9
Suppose $T \in \mathcal{L}(V)$ and $q \in \mathcal{P}(\mathbb{F})$. Then, $q(T) = 0$ if and only if q is a polynomial multiple of the minimal polynomial of T.

Proof 5. Let p be the minimal polynomial of T.
($\Leftarrow$): Suppose $q = sp$. Then, $q(T) = s(T)p(T) = 0$. □
($\Rightarrow$): Suppose $q(T) = 0$. By division algorithm of polynomials, $q = sp + r$ with $\deg r < \deg p$. Then, $q(T) = s(T)p(T) + r(T) = 0$. Note that $p(T) = 0$, so $r(T) = 0$. Then, $r = 0$. It must be $q = sp$. ■

Theorem 7.3.10 Characteristic Polynomial and Minimal Polynomial
Suppose $\mathbb{F} = \mathbb{C}$ and $T \in \mathcal{L}(V)$. Then, the characteristic polynomial of T is a polynomial multiple of the minimal polynomial of T.

Proof 6. Suppose q is a characteristic polynomial of T. Then, by Cayley-Hamilton Theorem, $q(T) = 0$. Further by Theorem 7.3.9, q is a polynomial multiple of the minimal polynomial of T. ■

Theorem 7.3.11
Let $T \in \mathcal{L}(V)$. Then, the zeros of the minimal polynomial of T are precisely the eigenvalues of T.

Remark. *"Precisely" means "is and only is." So, we need to prove the theorem from two directions.*

Proof 7. Suppose $p(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_{m-1} z^{m-1} + z^m$ is the minimal polynomial of T.

($\Rightarrow$): Suppose $p(\lambda) = 0$. *WTS: λ is the eigenvalue.* Since $p(\lambda) = 0$, we have $p(z) = (z - \lambda)q(z)$. Then, $p(T) = (T - \lambda I)q(T) = 0$. Then, $\deg q < \deg p$ and $p(T)v = (T - \lambda I)q(T)v = 0 \quad \forall v \in V$. So, $\exists v \in V$ *s.t.* $q(T)v \neq 0$. So, it must be that $T - \lambda I$ is not injective, and thus λ is an eigenvalue of T. □

($\Leftarrow$): Suppose $\lambda \in \mathbb{F}$ is an eigenvalue of T. Then, $\exists v \in V$ *s.t.* $Tv = \lambda v$ with $v \neq 0$. Consider $T^j v = \lambda^j v$. Then,

$$\begin{aligned} p(T)V &= (a_0 I + a_1 T + \cdots + a_{m-1} T^{m-1} + T^m)v \\ &= (a_0 + a_1 \lambda + \cdots + a_{m-1} \lambda^{m-1} + \lambda^m)v \\ &= p(\lambda)v = 0 \end{aligned}$$

Since $v \neq 0$, it must be $p(\lambda) = 0$. ■

Example 7.3.12 Suppose $T \in \mathcal{L}(\mathbb{C}^3)$ be defined as

$$T(z_1, z_2, z_3) = (6z_1 + 3z_2 + 4z_3, 6z_2 + 2z_3, 7z_3).$$

Then,

$$\mathcal{M}(T) = \begin{pmatrix} 6 & 3 & 4 \\ 0 & 6 & 2 \\ 0 & 0 & 7 \end{pmatrix}.$$

Find the minimal polynomial of T.

Solution 8.

Since $\mathcal{M}(T) = \begin{pmatrix} 6 & 3 & 4 \\ 0 & 6 & 2 \\ 0 & 0 & 7 \end{pmatrix}$, the eigenvalues of T are $6, 6, 7$. The multiplicity of 6 is 2 and that of 7 is 1. So, the characteristic polynomial of T is $q(z) = (z-6)^2(z-7)$. Then, the minimal polynomial is polynomial multiple of $(z-6)(z-7)$. So, the minimal polynomial of T should be $(z-6)(z-7)$ or $(z-6)^2(z-7)$. Note that

$$\begin{aligned} \mathcal{M}\big[(T-6I)^2(T-7I)\big] &= (\mathcal{M}(T-6I))^2 \mathcal{M}(T-7I) \\ &= \begin{pmatrix} 0 & 0 & 10 \\ 0 & 0 & 2 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} -1 & 3 & 4 \\ 0 & -1 & 2 \\ 0 & 0 & 0 \end{pmatrix} \\ &= \end{aligned}$$

and

$$\mathcal{M}[(T-6I)(T-7I)] = \begin{pmatrix} 0 & 3 & 4 \\ 0 & 0 & 2 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} -1 & 3 & 4 \\ 0 & -1 & 2 \\ 0 & 0 & 0 \end{pmatrix} \neq 0.$$

So, $(z-6)^2(z-7)$ is the minimal polynomial of T. □

Example 7.3.13 Find the minimal polynomial of operator $T \in \mathcal{L}(\mathbb{C}^3)$ defined by $T(z_1, z_2, z_3) = (6z_1, 6z_2, 7z_3)$.

Solution 9.

Note that

$$\mathcal{M}(T) = \begin{pmatrix} 6 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 7 \end{pmatrix}.$$

Then, the characteristic polynomial is $q(z) = (z-6)^2(z-7)$. The minimal polynomial could be $(z-6)^2(z-7)$ or $(z-6)(z-7)$. Since

$$\begin{aligned} \mathcal{M}[(T-6I)(T-7I)] &= \mathcal{M}(T-6I)\mathcal{M}(T-7I) \\ &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix} = 0, \end{aligned}$$

the minimal polynomial of T is $(z-6)(z-7)$. □

Theorem 7.3.14
Suppose $T \in \mathcal{L}(V)$. T is invertible if and only if the constant term in the minimal polynomial of T is non-zero.

Proof 10. Let $p(z) = a_0 + a_1 z + \cdots + a_{m-1} z^{m-1} + z^m$ be the minimal polynomial of T.

($\Rightarrow$) We will prove the contrapositive: "If $a_0 = 0$, then T is not invertible." Suppose $a_0 = 0$. Then,

$$p(z) = a_1 z + \cdots + a_{m-1} z^{m-1} + z^m.$$

Then, $p(0) = 0$. So, 0 is an eigenvalue of T. That is, $Tv = 0$ for some $v \neq 0$. Then, T is not injective, and thus is not invertible. □

($\Leftarrow$) We will prove the contrapositive: "If T is not invertible, then $a_0 = 0$." Suppose T is not invertible. Then, T is not injective. So, $\exists v \neq 0$ *s.t.* $Tv = 0$. That is, $Tv = 0 \cdot v$ or 0 is an eigenvalue of T. So, $p(z) = zq(z)$, and thus $a_0 = 0$. ■

Theorem 7.3.15
Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. V has a basis consisting of eigenvectors of T if and only if the minimal polynomial of T has no repeated roots.

7.4 Jordan Form

Example 7.4.1 Let $N \in \mathcal{L}(\mathbb{F}^4)$ be the nilpotent operator $N(z_1, z_2, z_3, z_4) = (0, z_1, z_2, z_3)$. Let $v = (1, 0, 0, 0)$. Then, $Nv = (0, 1, 0, 0)$, $N^2v = (0, 0, 1, 0)$, and $N^3v = (0, 0, 0, 1)$. Note that v, Nv, N^2v, N^3v is a basis of $\mathbb{F}^4$, and the matrix of N with respect to this basis is

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

Example 7.4.2 Let $N \in \mathcal{L}(\mathbb{F}^6)$ be the nilpotent operator defined by

$$N(z_1, z_2, z_3, z_4, z_5, z_6) = (0, z_1, z_2, 0, z_4, 0).$$

Let $v_1 = (1, 0, 0, 0, 0, 0)$, $v_2 = (0, 0, 0, 1, 0, 0)$, and $v_3 = (0, 0, 0, 0, 0, 1)$. Then, we have N^2v_1, Nv_1, Nv_2, v_2, v_3 to be a basis of $\mathbb{F}^6$. The matrix of N with respect to this basis is

$$\left(\begin{array}{ccc} \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} & \begin{matrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{matrix} & \begin{matrix} 0 \\ 0 \\ 0 \end{matrix} \\ \begin{matrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{matrix} & \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} & \begin{matrix} 0 \\ 0 \end{matrix} \\ \begin{matrix} 0 & 0 & 0 \end{matrix} & \begin{matrix} 0 & 0 \end{matrix} & \begin{pmatrix} 0 \end{pmatrix} \end{array}\right)$$

Theorem 7.4.3
Suppose $N \in \mathcal{L}(V)$ is nilpotent. Then, $\exists v_1, \ldots, v_n \in V$ and $m_1, \ldots, m_n \in \mathbb{N}^+$ such that

1. $N^{m_1}v_1, \ldots, Nv_1, v_1, \ldots, N^{m_n}v_n, \ldots, Nv_n, v_n$ is a basis of V;
2. $N^{m_1+1}v_1 = \cdots = N^{m_n+1}v_n = 0$.

Proof 1. We will prove by induction on $\dim V$.

Base Case When $\dim V = 1$, the conclusions obviously hold.

Inductive Steps Assume $\dim V > 1$ and the conclusions hold for all spaces with dimension smaller than $\dim V$. Since N is nilpotent, it is not injective and thus is not surjective. So, $\operatorname{range} N \subsetneq V$. That is, $\dim \operatorname{range} N < \dim V$. Since N is nilpotent, it is not injective and thus is not surjective. So, $\operatorname{range} N \subsetneq V$. that is, $\dim \operatorname{range} N < \dim V$. Apply the inductive hypothesis on $\operatorname{range} N$. Consider $N|_{\operatorname{range} N} \in \mathcal{L}(\operatorname{range} N)$, then $\exists v_1, \ldots, v_n \in \operatorname{range} N$ and $m_1, \ldots, m_n \in \mathbb{N}^+$ such that

$$N^{m_1}v_1, \ldots, Nv_1, v_1, \ldots, N^{m_n}v_n, \ldots, Nv_n, v_n. \tag{22}$$

is a basis of $\operatorname{range} N$, and $N^{m_1+1}v_1 = \cdots = N^{m_n+1}v_n = 0$. For each j, $v_j \in \operatorname{range} N$. Then, $\exists u_j \in$

V *s.t.* $v_j = Nu_j$. So, $N^{k+1}u_j = N^k v_j \quad \forall k \in \mathbb{N}^+$. We now claim the following list of vectors is L.I.:

$$N^{m_1+1}u_1, \ldots, Nu_1, u_1, \ldots, N^{m_n+1}u_n, \ldots, Nu_n, u_n \tag{23}$$

Let $a_1^{m_1+1}N^{m_1+1}u_1 + \cdots + a_1^1 Nu_1 + a_1^0 u_1 + \cdots + a_n^{m_n+1}N^{m_n+1}u_n + \cdots + a_n^1 Nu_n + a_n^0 u_n = 0$. Then,

$$a_1^{m_1+1}N^{m_1}v_1 + \cdots + a_1^1 v_1 + a_1^0 u_1 + \cdots + a_n^{m_n+1}N^{m_n}v_n + \cdots + a_n^1 v_n + a_n^0 u_n = 0. \tag{24}$$

Apply N to both sides of the Equation (24),

$$\underbrace{a_1^{m_1+1}N^{m_1+1}v_1}_{0} + \cdots + a_1^1 Nv_1 + a_1^0 \underbrace{Nu_1}_{v_1} + \cdots + \underbrace{a_n^{m_n+1}N^{m_n+1}v_n}_{0} + \cdots + a_n^1 Nv_n + a_n^0 \underbrace{Nu_n}_{v_n} = 0.$$

So,

$$a_1^{m_1}N^{m_1}v_1 + \cdots + a_1^1 Nv_1 + a_1^0 v_1 + \cdots + a_n^{m_n}N^{m_n}v_n + \cdots + a_n^1 Nv_n + a_n^0 v_n = 0.$$

Since Equation (22) is a basis, it must be all the coefficients equal to 0. Meanwhile, reconsider Equation (24). It becomes

$$a_1^{m_1+1}N^{m_1}v_1 + \cdots + a_n^{m_n+1}N^{m_n}v_n = 0.$$

As $N^{m_1}, \ldots, N^{m_n}$ is included in the list of vector stated in Equation (22), they must also be L.I.. Thus, we have $a_1^{m_1+1} = \cdots = a_n^{m_n+1} = 0$. So, we have proven the claim by showing Equation (23) is indeed a list of L.I. vectors. Now, extend Equation (23) into a bassi of V:

$$N^{m_1+1}u_1, \ldots, Nu_1, u_1, \ldots, N^{m_n+1}u_n, \ldots, Nu_n, u_n, w_1, \ldots, w_p \tag{25}$$

Then, each $Nw_j \in \operatorname{range} N = \operatorname{span}(\text{Equation } (22))$ *s.t.* $Nw_j = Nx_j$. Now, suppose $u_{n+j} = w_j - x_j$, and we have $Nu_{n+j} = 0$. Hence,

$$N^{m_1+1}u_1, \ldots, Nu_1, u_1, \ldots, N^{m_n+1}u_n, \ldots, Nu_n, u_n, u_{n+1}, \ldots, u_{n+p} \tag{26}$$

spans V because it contains each x_j and u_{n+j} and thus w_j. Since Equation (25) and Equation (26) have the same length, Equation (26) is a basis of V satisfying the desired condition. ■

Definition 7.4.4 (Jordan Basis). Suppose $T \in \mathcal{L}(V)$. A basis of V is called a *Jordan basis* of T if $\mathcal{M}(T)$ with respect to this basis has a block diagonal matrix

$$\begin{pmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_p \end{pmatrix},$$

where each A_j is an upper-triangular matrix of the form

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

Theorem 7.4.5 Jordan Form
Suppose V is a complex vector space. If $T \in \mathcal{L}(V)$, then $\exists$ a basis of V that is a Jordan basis for T.

Proof 2. First consider a nilpotent operator $N \in \mathcal{L}(V)$. Suppose $v_1, \ldots, v_n \in \mathcal{L}(V)$ satisfy the condition in Theorem 7.4.3. For each j, note that the list of vectors $N^{m_j} v_j, N^{m_{j-1}} v_j, \ldots, N v_j, v_j$ correspond to a matrix of N as

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

Hence, the conclusion holds for a nilpotent operator. Assume $T \in \mathcal{L}(V)$. Let $\lambda_1, \ldots, \lambda_m$ be distinct eigenvalues of T. Then, we have the generalized eigenspace decomposition:

$$V = G(\lambda_1, T) \oplus \cdots \oplus G(\lambda_m, T),$$

where each $(T - \lambda_j I) \mid_{G(\lambda_j, T)}$ is nilpotent. Thus, some basis of each $G(\lambda_j, T)$ is a Jordan basis of $T - \lambda_j I$. So,

$$\mathcal{M}\Big((T - \lambda_j I) \mid_{G(\lambda_j, T)}\Big) = \begin{pmatrix} 0 & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & 0 \end{pmatrix}$$

and

$$\mathcal{M}\Big(T \mid_{G(\lambda_j, T)}\Big) = \begin{pmatrix} \lambda_j & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda_j \end{pmatrix}.$$

Also, the dimension of the matrix is $\dim G(\lambda_j, T)$. ■

8 Operators on Real Vectors Spaces

8.1 Complexification

Definition 8.1.1 (Complexification of V/$V_{\mathbb{C}}$). Suppose V is a real vector space. The *complexification* of V, denoted $V_{\mathbb{C}}$, equals $V \times V$. An element of $V_{\mathbb{C}}$ is an ordered pair (u, v), where $u, v \in V$, but we will write this as $u + \mathrm{i}v$.

Definition 8.1.2 (Addition & Multiplication on $V_{\mathbb{C}}$).

1. *Addition* on $V_{\mathbb{C}}$ is defined by

$$(u_1 + \mathrm{i}v_1) + (u_2 + \mathrm{i}v_2) = (u_1 + u_2) + \mathrm{i}(v_1 + v_2).$$

   for $u_1, u_2, v_1, v_2 \in V$.

2. *Complex Scalar Multiplication* on $V_{\mathbb{C}}$ is defined by

$$(a + b\mathrm{i})(u + \mathrm{i}v) = (au - bv) + \mathrm{i}(av + bu)$$

   for $a, b \in \mathbb{R}$ and $u, v \in V$.

Theorem 8.1.3
Suppose V is a real vector space. Then, with the definition of addition and scalar multiplication as above, $V_{\mathbb{C}}$ is a complex vector space.

Proof 1.

1. Addition. Let $u_j + \mathrm{i}v_j \in \mathbb{C}$.

   (a) commutativity:

$$\begin{aligned}(u_1 + \mathrm{i}v_1) + (u_2 + \mathrm{i}v_2) &= (u_1 + u_2) + \mathrm{i}(v_1 + v_2)\\ &= (u_2 + u_1) + \mathrm{i}(v_2 + v_1)\\ &= (u_2 + \mathrm{i}v_2) + (u_1 + \mathrm{i}v_1). \qquad \square\end{aligned}$$

   (b) associativity:

$$\begin{aligned}((u_1, v_1) + (u_2, v_2)) + (u_3, v_3) &= (u_1 + u_2, v_1 + v_2) + (u_3, v_3)\\ &= (u_1 + u_2 + u_3, v_1 + v_2 + v_3)\\ &= (u_1 + (u_2 + u_3), v_1 + (v_2 + v_3))\\ &= (u_1, v_1) + ((u_2, v_2) + (u_3, v_3)). \qquad \square\end{aligned}$$

   (c) identity:

$$\begin{aligned}(0, 0) + (u, v) = (0 + u, 0 + v) &= (u + 0, v + 0)\\ &= (u, v) + (0, 0)\\ &= (u, v). \qquad \square\end{aligned}$$

   (d) inverse:

$$(-u, -v) + (u, v) = (-u + u, -v + v) = (0, 0). \qquad \square$$

2. Scalar Multiplication: Let $(u, v) \in V_{\mathbb{C}}$, $a + b\mathrm{i}$ and $c + d\mathrm{i} \in \mathbb{C}$.

   (a) identity:
   $$(1 + 0\mathrm{i})(u + \mathrm{i}v) = u + \mathrm{i}v + 0\mathrm{i}u - 0v = u + \mathrm{i}v. \qquad \square$$

   (b) associativity: can be easily verified. omitted.

   (c) distributivity: can be easily verified. omitted.

■

Theorem 8.1.4
Suppose V is a real vector space.

1. If $v_1 \ldots, v_n$ is a basis of V (as a real vector space), then $v_1, \ldots, v_n$ is a basis of $V_{\mathbb{C}}$ (as a complex vector space).

2. The dimension of $V_{\mathbb{C}}$ (as a complex vector space) equals the dimension of V (as a real vector space).

Proof 2.

1. Suppose $v_1, \ldots, v_n$ is a basis of V. Then, $V = \operatorname{span}(v_1, \ldots, v_n)$. Then, $\operatorname{span}(v_1, \ldots, v_n)$ in $V_{\mathbb{C}}$ contains $v_1, \ldots, v_n, \mathrm{i}v_1, \ldots, \mathrm{i}v_n$. For any $u + \mathrm{i}v \in V_{\mathbb{C}}$, we have
   $$\begin{aligned} u + \mathrm{i}v &= (a_1 v_1 + \cdots + a_n v_n) + \mathrm{i}(b_1 v_1 + \cdots + b_n v_n) \\ &= a_1 v_1 + \cdots + a_n v_n + b_1 \mathrm{i}v_1 + \cdots + b_n \mathrm{i}v_n. \end{aligned}$$
   So, $v_1, \ldots, v_n, \mathrm{i}v_1, \ldots, \mathrm{i}v_n$ spans $V_{\mathbb{C}}$. Note that
   $$\operatorname{span}(v_1, \ldots, v_n, \mathrm{i}v_1, \ldots, \mathrm{i}v_n) = \operatorname{span}(v_1, \ldots, v_n).$$
   Then, we get $V_{\mathbb{C}} = \operatorname{span}(v_1, \ldots, v_n)$. Now, let $\lambda_1 v_1 + \cdots + \lambda_n v_n = 0$ for $\lambda_j \in \mathbb{C}$. Then,
   $$\operatorname{Re}(\lambda_1 v_1) + \cdots + \operatorname{Re}(\lambda_n v_n) = 0 \quad \text{and} \quad \operatorname{Im}(\lambda_1 v_1) + \cdots + \operatorname{Im}(\lambda_n v_n) = 0.$$
   Since $\operatorname{Re}(\lambda_j), \operatorname{Im}(\lambda_j) \in \mathbb{R}$, it must be that
   $$\operatorname{Re}(\lambda_1) = \cdots = \operatorname{Re}(\lambda_n) = 0 \quad \text{and} \quad \operatorname{Im}(\lambda_1) = \cdots = \operatorname{Im}(\lambda_n) = 0.$$
   Then, we have
   $$\lambda_1 = \cdots = \lambda_n = 0.$$
   That is, $v_1, \ldots, v_n$ is L.I.. Hence, $v_1, \ldots, v_n$ is a basis of $V_{\mathbb{C}}$. $\square$

2. We know immediately that (1) implies (2). The proof is complete.

■

Definition 8.1.5 (Complexification of $T/T_{\mathbb{C}}$). Suppose V is a real vector space and $T \in \mathcal{L}(V)$. The *complexification* of T, denoted $T_{\mathbb{C}}$, is the operator $T_{\mathbb{C}} \in \mathcal{L}(V_{\mathbb{C}})$ defined by $T_{\mathbb{C}}(u + \mathrm{i}v) = Tu + \mathrm{i}Tv$ for $u, v \in V$.

Remark. *It can be easily verified that this definition indeed gives an operator* $T_{\mathbb{C}} \in \mathcal{L}(V_{\mathbb{C}})$.

Example 8.1.6 Suppose A is an $n \times n$ matrix of real numbers. Define $T \in \mathcal{L}(\mathbb{R}^n)$ by $Tx = Ax$. Identifying the complexification of $\mathbb{R}^n$ with $\mathbb{C}^n$, we then have $T_{\mathbb{C}} z = Az$ for each $z \in \mathbb{C}^n$.

Theorem 8.1.7
Suppose V is a real vector space with basis $v_1, \dots, v_n$ and $T \in \mathcal{L}(V)$. Then, $\mathcal{M}(T) = \mathcal{M}(T_{\mathbb{C}})$, where both matrices are with respect to the basis $v_1, \dots, v_n$.

Proof 3. Note that

$$T_{\mathbb{C}}(v_k) = T_{\mathbb{C}}(v_k + \mathrm{i} \cdot 0) = Tv_k + \mathrm{i}T0 = Tv_k.$$

So, $\mathcal{M}(T) = \mathcal{M}(T_{\mathbb{C}})$. ■

Theorem 8.1.8
Every operator on a non-zero *f-d* vector space has an invariant subspace of dimension 1 or 2.

Proof 4. We only need to consider the real case. Let $T \in \mathcal{L}(V)$, then $T_{\mathbb{C}} \in \mathcal{L}(V_{\mathbb{C}})$. Then, $T_{\mathbb{C}}$ has an eigenvalue $a + b\mathrm{i}$, and a corresponding eigenvector $u + \mathrm{i}v \in V_{\mathbb{C}}$ *s.t.*

$$T_{\mathbb{C}}(u + \mathrm{i}v) = (a + b\mathrm{i})(u + \mathrm{i}v) \quad \Longrightarrow \quad Tu + \mathrm{i}Tv = (au - bv) + (av + bu)\mathrm{i}$$

So, $Tu = au - bv$ and $Tv = av + bu$. Let $U = \operatorname{span}(u, v)$ in V. Then, $au - bv, av + bu \in U$. Therefore, U is an invariant subspace of V under T. If u, v is L.I., then $\dim U = 2$; if u, v is linearly dependent, then $\dim U = 1$. ■

Theorem 8.1.9
Suppose V is a real vector space and $T \in \mathcal{L}(V)$. Then, the minimal polynomial of $T_{\mathbb{C}}$ equals the minimal polynomial of T.

Proof 5. Suppose V is a real vector space and $T \in \mathcal{L}(V)$. Then,

$$(T_{\mathbb{C}})^n(u + \mathrm{i}v) = T^n u + \mathrm{i}T^n v.$$

Let $p \in \mathcal{P}(\mathbb{R})$ be the minimal polynomial of T. Then, $p(T_{\mathbb{C}}) = (p(T))_{\mathbb{C}}$.

In fact, let $p(x) = a_0 + a_1 x + \cdots + a_n x^n$, then $p(T_{\mathbb{C}}) = a_0 I + a_i T_{\mathbb{C}} + \cdots + a_n T_{\mathbb{C}}^n$. So,

$$\begin{aligned} p(T_{\mathbb{C}})(u + \mathrm{i}v) &= a_0(u + \mathrm{i}v) + a_1 T_{\mathbb{C}}(u + \mathrm{i}v) + \cdots + a_n T_{\mathbb{C}}^n(u + \mathrm{i}v) \\ &= (a_0 u + a_1 Tu + \cdots + a_n T^n u) + \mathrm{i}(a_0 v + a_1 Tv + \cdots + a_n T^n v) \\ &= p(T)(u) + \mathrm{i}p(T)(v) \\ &= (p(T))_{\mathbb{C}}(u + \mathrm{i}v). \end{aligned}$$

So, $p(T_{\mathbb{C}}) = (p(T))_{\mathbb{C}}$.

Since $p(T) = 0$, $(p(T))_{\mathbb{C}} = 0$, and thus $p(T_{\mathbb{C}}) = 0$.

Suppose $q \in \mathcal{P}(\mathbb{C})$ is a monic polynomial and $q(T_{\mathbb{C}})(u) = 0 \quad \forall u \in V$. Let $q(z) = b_0 + b_1 z + \cdots + b_m z^m$, where $b_m = 1$, and $r(z) = \mathrm{Re}(b_0) + \mathrm{Re}(b_1 z) + \cdots + \mathrm{Re}(b_m z^m)$. So, $q(T_{\mathbb{C}}) = b_0 I + b_1 T_{\mathbb{C}} + \cdots + b_m T_{\mathbb{C}}^m = 0$. That is, $(q(T))_{\mathbb{C}} = 0$. So, $(q(T))_{\mathbb{C}}(u+\mathrm{i}v) = q(T)(u) + \mathrm{i}q(T)(v) = 0$. Then, it must be $q(T)(u) = 0 \quad \forall u \in V$. So, $b_0 u + b_1 Tu + \cdots + b_m T^m u = 0$, which is equivalent to $\mathrm{Re}(b_0)u + \mathrm{Re}(b_1)Tu + \cdots \mathrm{Re}(b_m)T^m u = 0$. By definition of $r(T)$, we have $r(T) = 0$.

Also, we have $\deg q = \deg r$. Further given p is the minimal polynomial of T, $\deg r \geq \deg p$. Hence, $\deg q = \deg r \geq \deg p$. Thus, p is also a minimal polynomial of $T_{\mathbb{C}}$. ■

Theorem 8.1.10
Suppose V is a real vector space, $T \in \mathcal{L}(V)$, and $\lambda \in \mathbb{R}$. Then, λ is an eigenvalue of $T_{\mathbb{C}}$ if and only if λ is an eigenvalue of T.

Proof 6. Since the minimals of T and $T_{\mathbb{C}}$ are the same, the zeros of the minimal polynomials will also be the same. Given zeros of the minimal polynomial of T are precisely the eigenvalues of T, the proof is therefore complete. ■

Proof 7.

($\Rightarrow$) Firstly, suppose λ is an eigenvalue of T. Then, $\exists\ v \neq 0$ *s.t.* $Tv = \lambda v$. So, $T_{\mathbb{C}}(v) = \lambda v$, and thus λ is an eigenvalue of T_C. □

($\Leftarrow$) Conversely, suppose λ is an eigenvalue of $T_{\mathbb{C}}$. Then, $\exists u, v \in V$ with $u + \mathrm{i}v \neq 0$ *s.t.*

$$T_{\mathbb{C}}(u + \mathrm{i}v) = \lambda(u + \mathrm{i}v).$$

So, $Tu = \lambda u$ and $Tv = \lambda v$. Then, λ must be an eigenvalue of T. ■

Theorem 8.1.11
Suppose V is a real vector space, $T \in \mathcal{L}(V)$, $\lambda \in \mathbb{C}$, j is an non-negative integer, and $u, v \in V$. Then, $(T_{\mathbb{C}} - \lambda I)^j(u + \mathrm{i}v) = 0$ if and only if $(T_{\mathbb{C}} - \bar{\lambda} I)^j(u - \mathrm{i}v) = 0$.

Proof 8. To prove this theorem, we only have to prove the forward direction. We will prove by induction on j.

Base Case If $j = 0$, then $(T_{\mathbb{C}} - \lambda I)^0 = I$. So, we have $u + \mathrm{i}v = 0$. Then, $u = 0$, and $v = 0$. Therefore, $u - \mathrm{i}v = 0$. □

Inductive Steps Assume $j \geq 1$ and the desired results holds for $j - 1$. That is,

$$(T_{\mathbb{C}} - \lambda I)^{j-1}(u + \mathrm{i}v) \implies (T_{\mathbb{C}} - \bar{\lambda} I)^{j-1}(u - \mathrm{i}v) = 0.$$

Consider

$$(T_{\mathbb{C}} - \lambda I)^{j-1}(T_{\mathbb{C}} - \lambda I)(u + \mathrm{i}v) = 0. \tag{27}$$

Writing $\lambda = a + b\mathrm{i}$, we have

$$\begin{aligned}(T_{\mathbb{C}} - \lambda I)(u + \mathrm{i}v) &= T_{\mathbb{C}}(u + \mathrm{i}v) - (a + b\mathrm{i})(u + \mathrm{i}v) \\ &= (Tu - au + bv) + \mathrm{i}(Tv - bu - av)\end{aligned}$$

and

$$\begin{aligned}(T_{\mathbb{C}} - \bar{\lambda} I)(u + \mathrm{i}v) &= T_{\mathbb{C}}(u + \mathrm{i}v) - (a - b\mathrm{i})(u + \mathrm{i}v) \\ &= (Tu - au + bv) - \mathrm{i}(Tv - bu + av).\end{aligned}$$

So, Eq. (27) becomes

$$(T_{\mathbb{C}} - \lambda I)^{j-1}(Tu - au + bv) + \mathrm{i}(Tv - bu - av) = 0. \tag{28}$$

Apply our inductive hypothesis to Eq. (28), we have

$$(T_{\mathbb{C}} - \bar{\lambda} I)^{j-1}((Tu - au + bv) - \mathrm{i}(Tv - bu + av)) = 0$$

That is, $(T_{\mathbb{C}} - \bar{\lambda} I)^{j-1}((T_{\mathbb{C}} - \bar{\lambda} I)(u + \mathrm{i}v)) = 0$, or $(T_{\mathbb{C}} - \bar{\lambda} I)^{j}(u + \mathrm{i}v) = 0$. ■

Corollary 8.1.12 Suppose V is a real vector space, $T \in \mathcal{L}(V)$, and $\lambda \in \mathbb{C}$. Then, λ is an eigenvalue of $T_{\mathbb{C}}$ if and only if $\bar{\lambda}$ is an eigenvalue of $T_{\mathbb{C}}$.

Proof 9. Take $j = 1$ in Theorem 8.1.11. The proof is completed. ■

Theorem 8.1.13

Suppose V is a real vector space, $T \in \mathcal{L}(V)$, and $\lambda \in \mathbb{C}$ is an eigenvalue of $T_{\mathbb{C}}$. Then, the multiplicity of λ as an eigenvalue of $T_{\mathbb{C}}$ equals the multiplicity of $\bar{\lambda}$ as an eigenvalue of $T_{\mathbb{C}}$.

Proof 10. We only need to show $\dim G(\lambda, T_{\mathbb{C}}) = \dim G(\bar{\lambda}, T_{\mathbb{C}})$. Select $u_1 + \mathrm{i}v_1, \ldots, u_m + \mathrm{i}v_m$ as a basis of $G(\lambda, T_{\mathbb{C}})$. Then,

$$(T_{\mathbb{C}} - \lambda I)^{\dim V}(u_j + \mathrm{i}v_j) = 0 \quad \text{for each } j.$$

Then, $(T_{\mathbb{C}} - \bar{\lambda} I)^{\dim V}(u_j - \mathrm{i}v_j) = 0$ by Theorem 8.1.11. Now, consider $u_1 - \mathrm{i}v_1, \ldots, u_m - \mathrm{i}v_m$. Suppose

$$(a_1 + b_1\mathrm{i})(u_1 - \mathrm{i}v_1) + \cdots + (a_m + b_m\mathrm{i})(u_m - \mathrm{i}v_m) = 0.$$

Then,

$$\sum_{j=1}^{m} a_j u_j + b_j v_j + \mathrm{i}(b_j u_j - a_j v_j) = 0. \tag{29}$$

Note that $(a_j - b_j\mathrm{i})(u_j + \mathrm{i}v_j) = a_j u_j + b_j v_j + \mathrm{i}(b_j u_j - a_j v_j)$. Then, Eq. (29) becomes

$$\sum_{j=1}^{m} \overline{a_j + b_j\mathrm{i}}(u_j + \mathrm{i}v_j) = 0.$$

Since $u_1 + \mathrm{i}v_1, \ldots, u_m + \mathrm{i}v_m$ is a basis, it must be $\overline{a_1 + b_1\mathrm{i}} = \cdots = \overline{a_m + b_m\mathrm{i}} = 0$. So, $a_1 + b_1\mathrm{i} = \cdots = a_m + b_m\mathrm{i} = 0$. Therefore, we have $u_1 - \mathrm{i}v_1, \ldots, u_m - \mathrm{i}v_m$ is L.I.. Now, let $u - \mathrm{i}v \in G(\bar{\lambda}, T_{\mathbb{C}})$. Then,

$$u + \mathrm{i}v = (a_1 - b_1\mathrm{i})(u_1 + \mathrm{i}v_1) + \cdots + (a_m - b_m\mathrm{i})(u_m + \mathrm{i}v_m).$$

So, $u - \mathrm{i}v = (a_1 + b_1\mathrm{i})(u_1 - \mathrm{i}v_1) + \cdots + (a_m + b_m\mathrm{i})(u_m - \mathrm{i}v_m)$. Hence, $G(\bar{\lambda}, T_{\mathbb{C}}) = \operatorname{span}(u_1 - \mathrm{i}v_1, \ldots, u_m - \mathrm{i}v_m)$. Since

$$\dim \operatorname{span}(u_1 + \mathrm{i}v_1, \ldots, u_m + \mathrm{i}v_m) = \dim \operatorname{span}(u_1 - \mathrm{i}v_1, \ldots, u_m - \mathrm{i}v_m),$$

multiplicity of λ equals multiplicity of $\bar{\lambda}$. ■

Theorem 8.1.14

Every operator on an odd-dimensional real vector space has an eigenvalue.

Proof 11. Suppose V is a real vector space with odd dimension. Let $T \in \mathcal{L}(V)$. Then, by Corollary 8.1.12, we know non-real eigenvalues of $T_{\mathbb{C}}$ come in pairs and their multiplicities are the same by Theorem 8.1.13. So,

$$\sum(\text{multiplicity of non-real eigenvalues}) = \text{an even number}.$$

Since $V = G(\lambda_1, T) \oplus \cdots \oplus G(\lambda_m, T)$, we have

$$\sum(\text{multiplicity of all eigenvalues}) = \dim V_{\mathbb{C}} = \dim V = \text{an odd number}.$$

So, there must be at least one real eigenvalues left. ■

Theorem 8.1.15
Suppose V is a real vector space and $T \in \mathcal{L}(V)$. Then, the coefficients of the characteristic polynomial of $T_{\mathbb{C}}$ are all real.

Proof 12. Suppose λ is a non-real eigenvalue of $T_{\mathbb{C}}$ with multiplicity m. Then, $\bar{\lambda}$ is also an eigenvalue of $T_{\mathbb{C}}$ with multiplicity m. Then, characteristic polynomial of $T_{\mathbb{C}}$ must be in the form

$$\begin{aligned}(z-\lambda)^m(z-\bar{\lambda})^m f(z) &= \left(z^2 - (\lambda+\bar{\lambda})z + |\lambda|^2\right)^m f(z)\\ &= \left(z^2 - 2(\mathrm{Re}(\lambda))z + |\lambda|^2\right)^m f(z).\end{aligned}$$

Suppose $f(z) = (z-t_1)^{d_1}\cdots(z-t_r)^{d_r}$ with each $t_j \in \mathbb{R}$. Then, the characteristic polynomial of $T_{\mathbb{C}}$ becomes

$$\left(z^2 - 2(\mathrm{Re}(\lambda))z + |\lambda|^2\right)^m (z-t_1)^{d_1}\cdots(z-t_r)^{d_r},$$

with all real coefficients. ■

Definition 8.1.16 (Characteristic Polynomial). Suppose V is a real vector space and $T \in \mathcal{L}(V)$. Then, the *characteristic polynomial* of T is defined to be the characteristic polynomial of $T_{\mathbb{C}}$.

Corollary 8.1.17 Degree and Zeros of Characteristic Polynomial Suppose V is a real vector space and $T \in \mathcal{L}(V)$. Then,

1. the coefficients of the characteristic polynomial of T are all real;
2. the characteristic polynomial of T has degree $\dim V$;
3. the eigenvalues of T are precisely the real zeros of the characteristic polynomial of T.

Theorem 8.1.18 Cayley-Hamilton Theorem
Suppose $T \in \mathcal{L}(V)$. Let q denote the characteristic polynomial of T. Then, $q(T) = 0$.

Proof 13. We've shown Cayley-Hamilton holds on complex vector spaces. Assume V is a real vector space. Then, we know $q(T_{\mathbb{C}}) = 0$, which implies $q(T) = 0$. ■

Corollary 8.1.19 Suppose $T \in \mathcal{L}(V)$. Then,

1. the degree of the minimal polynomial of T is at most $\dim V$;
2. the characteristic polynomial of T is a polynomial multiple of the minimal polynomial of T.

8.2 Operators on Real Inner Product Spaces

Theorem 8.2.1 Normal but Not Self-Adjoint Operators
Suppose V is a 2-dimensional real inner product space and $T \in \mathcal{L}(V)$. Then, the following are equivalent:

1. T is normal but not self-adjoint;
2. The matrix of T with respect to every orthonormal basis of V has the form $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$, with $b \neq 0$.
3. The matrix of T with respect to some orthonormal basis of V has the form $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$, with $b > 0$.

Proof 1.
$(1) \implies (2)$: Suppose $TT^* = T^*T$ but $T \neq T^*$. Let e_1, e_2 be an orthonormal basis of V. Suppose

$$\mathcal{M}(T, (e_1, e_2)) = \begin{pmatrix} a & c \\ b & d \end{pmatrix}.$$

Then, $Te_1 = ae_1 + be_2$. So, $\|Te_1\|^2 = \|ae_1 + be_2\|^2 = a^2 + b^2$. Since T is normal $\iff \|Tv\| = \|T^*v\| \quad \forall\, v \in V$. So, $\|T^*e_1\|^2 = \|Te_1\|^2 = a^2 + b^2$. Note that

$$\mathcal{M}(T, (e_1, e_2)) = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

the conjugate transpose of $\mathcal{M}(T, (e_1, e_2))$. So, $\|T^*e_1\|^2 = \|ae_1 + ce_2\|^2 = a^2 + c^2$. Therefore, $a^2 + b^2 = a^2 + c^2$, or $b^2 = c^2$. Then, $b = c$ or $b = -c$.

1. If $c = b$, then
$$\mathcal{M}(T) = \begin{pmatrix} a & c \\ c & d \end{pmatrix} = \mathcal{M}(T^*).$$
That implies $T = T^*$, which contradicts with our assumption that $T \neq T^*$. So, this situation is omitted.

2. So, $c = -b$, and then $\mathcal{M}(T) = \begin{pmatrix} a & -b \\ b & d \end{pmatrix}$. Note if $b = 0$, then $\mathcal{M}(T) = \begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix} = \mathcal{M}(T^*)$, contradicting with our assumption that $T \neq T^*$. So, $b \neq 0$.

Finally, since T is normal, we have $\mathcal{M}(T)\mathcal{M}(T^*) = \mathcal{M}(T^*)\mathcal{M}(T)$. That is,

$$\begin{pmatrix} a & -b \\ b & d \end{pmatrix}\begin{pmatrix} a & b \\ -b & d \end{pmatrix} = \begin{pmatrix} a & b \\ -b & d \end{pmatrix}\begin{pmatrix} a & -b \\ b & d \end{pmatrix} \implies ab - bd = -ab + bd \implies ab = bd.$$

Since $b \neq 0$, we have $a = d$. So,

$$\mathcal{M}(T) = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}, \quad b \neq 0. \qquad \square$$

(2) $\implies$ (3): Choose an orthonormal basis e_1, e_2. Then,

$$\mathcal{M}(T, (e_1, e_2)) = \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \quad \text{with } b \neq 0.$$

If $b > 0$, then (3) holds. If $b < 0$, then

$$\mathcal{M}(T, (e_1, -e_2)) = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}.$$

Then, $-b > 0$, which implies (3) holds. $\square$

(3) $\implies$ (1): Suppose $\exists$ an orthonormal basis e_1, e_2 *s.t.*

$$\mathcal{M}(T, (e_1, e_2)) = \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \quad \text{with } b > 0.$$

Then, $\mathcal{M}(T, (e_1, e_2))^t = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}$. Since $b > 0$, $\mathcal{M}(T) \neq \mathcal{M}(T)^t$. So, T is not self-adjoint. Since $\mathcal{M}(T)\mathcal{M}(T^*) = \mathcal{M}(T^*)\mathcal{M}(T)$ is clear, we have shown T is normal. ■

Theorem 8.2.2
Suppose V is an inner product space, $T \in \mathcal{L}(V)$ is normal, and U is a subspace of V that is invariant under T. Then,

1. $U^\perp$ is invariant under T;
2. U is invariant under T^*;
3. $(T|_U)^* = (T^*)|_U$;
4. $T|_U \in \mathcal{L}(U)$ and $T|_{U^\perp} \in \mathcal{L}(U^\perp)$ are normal operators.

Proof 2.

1. Let $e_1, \ldots, e_m$ be an orthonormal basis of U. Then, extend it to an orthonormal basis $e_1, \ldots, e_m, f_1, \ldots, f_n$ of V. Since U is invariant under T, $Tu \in U$. Then, each $Te_j \in U$. That is, Te_j is a linear combination of $e_1, \ldots, e_m$. Thus, $\mathcal{M}(T, (e_1, \ldots, e_m, f_1, \ldots, f_n))$ is of the form

$$\mathcal{M}(T) = \begin{array}{c|c} & \begin{array}{cc} e_1 \cdots e_m & f_1 \cdots f_n \end{array} \\ \begin{array}{c} e_1 \\ \vdots \\ e_m \\ f_1 \\ \vdots \\ f_n \end{array} & \left(\begin{array}{c|c} A & B \\ \hline 0 & C \end{array}\right) \end{array}$$

For each $j \in \{1, \dots, m\}$, let $Te_j = a_{1,j}e_1 + \cdots + a_{m,j}e_m$. Then, $\|Te_j\|^2 = a_{1,j}^2 + \cdots + a_{m,j}^2$. Then,

$$\sum_{j=1}^{m} \|Te_j\|^2 = \sum_{j=1}^{m} \left(a_{1,j}^2 + \cdots + a_{m,j}^2\right).$$

Note that

$$\mathcal{M}(T^*) = \left(\begin{array}{c|c} A^t & 0 \\ \hline B^t & C^t \end{array}\right).$$

Then,

$$\sum_{j=1}^{m} \|T^*e_j\|^2 = \sum_{j=1}^{m} \left(a_{1,j}^2 + \cdots + a_{m,j}^2 + b_{j,1}^2 + \cdots + b_{j,n}^2\right).$$

Since $\displaystyle\sum_{j=1}^{m} \|Te_j\|^2 = \sum_{j=1}^{m} \|T^*e_j\|^2$, we have

$$\sum_{j=1}^{m} \left(a_{1,j}^2 + \cdots + a_{m,j}^2\right) = \sum_{j=1}^{m} \left(a_{1,j}^2 + \cdots + a_{m,j}^2 + b_{j,1}^2 + \cdots + b_{j,n}^2\right).$$

Then, each $b_{i,j} = 0$. So, $B = 0_{m\times n}$. That is,

$$\mathcal{M}(T) = \begin{array}{c|c} & \begin{array}{cc} e_1 \cdots e_m & f_1 \cdots f_n \end{array} \\ \begin{array}{c} e_1 \\ \vdots \\ e_m \\ f_1 \\ \vdots \\ f_n \end{array} & \left(\begin{array}{c|c} A & 0 \\ \hline 0 & C \end{array}\right) \end{array}$$

Then, for each $k \in \{1, \dots, n\}$, $Tf_k = 0e_1 + \cdots + 0e_m + c_{1,k}f_1 + \cdots + c_{n,k}f_n$. That is,

$$Tf_k \in \operatorname{span}(f_1, \dots, f_n) = U^\perp.$$

Therefore, $Tv \in U^\perp$ whenever $v \in U^\perp$. Hence, $U^\perp$ is invariant under T. $\square$

2. Note that

$$\mathcal{M}(T^*) = \begin{pmatrix} A^t & 0 \\ 0 & C^t \end{pmatrix}.$$

Then, $T^*e_j \in \operatorname{span}(e_1, \dots, e_m) = U$. So, U is invariant under T^*. $\square$

3. Let $S = T|_U \in \mathcal{L}(U)$. Fix $v \in U$. Then, $\forall\, u \in U$,

$$\langle Su, v\rangle = \langle Tu, v\rangle = \langle u, T^*v\rangle.$$

From (2), we know $T^*v \in U$. Then, we have

$$\langle u, S^*v\rangle = \langle Su, v\rangle = \langle u, T^*v\rangle.$$

So, $S^*v = T^*v$. That is, $(T|_U)^* = (T^*)|_U$. $\square$

4. Since T is normal, T commutes with T^*. By (3): $(T|_U)^* = (T^*)|_U$. So, we have $(T|_U)(T|_U)^* = (T|_U)^*(T|_U)$. That is, $T|_U$ is normal. Similarly, interchanging the roles of U and $U^\perp$,

$$(T|_{U^\perp})(T|_{U^\perp})^* = (T|_{U^\perp})^*(T|_{U^\perp}).$$

Then, $T_{U^\perp}$ is also normal.

$\blacksquare$

Lemma 8.2.3 Suppose $A = \begin{pmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_m \end{pmatrix}$ and $B = \begin{pmatrix} B_1 & & 0 \\ & \ddots & \\ 0 & & B_m \end{pmatrix}$, where A_j and B_j are matrices of the same size, then

$$AB = \begin{pmatrix} A_1B_1 & & 0 \\ & \ddots & \\ 0 & & A_mB_m \end{pmatrix}.$$

Theorem 8.2.4
Suppose V is a real inner product space and $T \in \mathcal{L}(V)$. Then, the following are equivalent:

1. T is normal;
2. $\exists$ an orthonormal basis of V with respect to which T has a block diagonal matrix *s.t.* each block is an 1×1 matrix or a 2×2 matrix of the form $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$ with $b > 0$.

Proof 3.
(2) $\implies$ (1): With respect to the basis given by (2),

$$\mathcal{M}(t)\mathcal{M}(t^*) = \mathcal{M}(T^*)\mathcal{M}(T).$$

Note that

$$\mathcal{M}(T) = \begin{pmatrix} \ddots & & \\ & \begin{pmatrix} a & -b \\ b & a \end{pmatrix} & \\ & & \ddots \end{pmatrix} \quad \text{and} \quad \mathcal{M}(T^*) = \begin{pmatrix} \ddots & & \\ & \begin{pmatrix} a & b \\ -b & a \end{pmatrix} & \\ & & \ddots \end{pmatrix}.$$

Since

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix}\begin{pmatrix} a & b \\ -b & a \end{pmatrix} = \begin{pmatrix} a^2+b^2 & 0 \\ 0 & a^2+b^2 \end{pmatrix} = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}\begin{pmatrix} a & -b \\ b & a \end{pmatrix},$$

we have $TT^* = T^*T$. So, T is normal. $\square$

(1) $\implies$ (2): We will use induction on $\dim V$. When $\dim V = 1$, the desired results hold. When $\dim V = 2$, if T is self-adjoint, then use the Real Spectrum Theorem, the desired results hold. If $\dim V = 2$ and T is not self-adjoint, by Theorem 8.2.1, the desired results also hold.

Now, assume that $\dim V > 2$ and the desired result holds on vector spaces of dimension smaller than $\dim V$. Let U be a subspace of V with $\dim U = 1$, and U is invariant under T. If such a subspace exists, (i.e., if T has an eigenvector v, then let $U = \operatorname{span}(v)$). If no such subspace exists, let U be a subspace of V of dimension 2 that is invariant under T.

If $\dim U = 1$, choose a vector u with $\|u\| = 1$. Then, u is an orthonormal basis of U, and $\mathcal{M}(T|_U)$ is 1×1. If $\dim U = 2$, then $T|_U \in \mathcal{L}(U)$ is normal by Theorem 8.2.2, but $T|_U$ is not self-adjoint (otherwise $T|_U$ would have an eigenvector). Thus, we can choose an orthonormal basis of U, say, $e_1, e_2,$ *s.t.*

$$\mathcal{M}(T|_U, (e_1, e_2)) = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}.$$

Now, $U^\perp$ is invariant under T and $T|_{U^\perp}$ is normal by Theorem 8.2.2. Then, $\dim U^\perp < \dim V$. By our inductive hypothesis, $\exists$ an orthonormal basis $f_1, \ldots, f_n$ of $U^\perp$ *s.t.*

$$\mathcal{M}(T|_{U^\perp}, (f_1, \ldots, f_n)) = \begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & \begin{pmatrix} a & -b \\ b & a \end{pmatrix} & & \\ & & & \ddots & \\ & & & & 1 \end{pmatrix}$$

Since $V = U \oplus U^\perp$, we finally have

$$\mathcal{M}(T) = \begin{array}{c} \\ e_1 \\ \vdots \\ e_m \\ f_1 \\ \vdots \\ f_n \end{array} \begin{array}{c} \begin{array}{c|c} e_1 \cdots e_m & f_1 \cdots f_n \end{array} \\ \left(\begin{array}{c|c} \begin{matrix} a & -b \\ b & a \end{matrix} & 0 \\ \hline 0 & \text{Desired Form} \end{array}\right) \end{array}$$

which is in the desired form. $\blacksquare$

Example 8.2.5 Let $\theta \in \mathbb{R}$. Then, the operator on $\mathbb{R}^2$ of counter-clockwise rotation centered at the origin by θ is an isometry. The matrix of this operator with respect to the standard basis is

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

Remark. *If θ is not an integer multiple of π, then no non-zero vector of $\mathbb{R}^2$ gets mapped to a scalar multiple of itself, and have the operator has no eigenvalues.*

Theorem 8.2.6
Suppose V is a real inner product space and $S \in \mathcal{L}(V)$. Then, the following are equivalent:

1. S is an isometry;
2. $\exists$ an orthonormal basis of V with respect to which S has a block diagonal matrix *s.t.* each block on the diagonal is an 1×1 matrix containing 1 or -1 or is a 2×2 matrix of the form

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix},$$

with $\theta \in (0, \pi)$.

Proof 4.
$(1) \implies (2)$: Suppose S is an isometry. Then, S is normal. So, $\exists$ an orthonormal basis $e_1, \dots, e_n$ *s.t.*

$$\mathcal{M}(S, (e_1, \dots, e_n)) = \begin{pmatrix} \ddots & & & & \\ & \lambda & & & \\ & & \ddots & & \\ & & & \begin{pmatrix} a & -b \\ b & a \end{pmatrix} & \\ & & & & \ddots \end{pmatrix},$$

with $b > 0$. If λ is an entry in a 1×1 matrix along the diagonal, then $\exists$ a basis vector e_j *s.t.* $Se_j = \lambda e_j$. So, $\|Se_j\| = \|\lambda e_j\| = |\lambda|\|e_j\| = \|e_j\|$. So, $|\lambda| = 1$, or $\lambda = \pm 1$.

Now, consider a 2×2 matrix of the form $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$ with $b > 0$ along the diagonal. Then, $\exists$ a basis e_i, e_{i+1} *s.t.* $Se_i = ae_i + be_{i+1}$. So,

$$\begin{aligned} 1 = \|e_i\|^2 = \|Se_i\|^2 &= \|ae_i + be_{i+1}\|^2 \\ &= \|ae_i\|^2 + \|be_{i+1}\|^2 \\ &= a^2 + b^2. \end{aligned}$$

So, $\exists \theta \in (0, \pi)$ *s.t.* $a = \cos\theta$ and $b = \sin\theta$, given $b > 0$. Therefore, this direction holds. $\square$

$(2) \implies (1)$: Suppose $\exists$ an orthonormal basis of V with respect to which the matrix of S has the desired form. Thus, we have a direct sum decomposition: $V = U_1 \oplus \cdots \oplus U_m$, where each U_j is a subspace of V of dimension 1 or 2. Furthermore, any two vectors belonging to distinct U's are orthogonal, and each $S|_{U_j}$ is an isometry mapping U_j into U_j. If $v \in V$, we can write $v = u_1 + \cdots + u_m$, where each $u_j \in U_j$. Applying S to the equation:

$$\begin{aligned} \|Sv\|^2 &= \|Su_1 + \cdots + Su_m\|^2 \\ &= \|Su_1\|^2 + \cdots + \|Su_m\|^2 \\ &= \|u_1\|^2 + \cdots + \|u_m\|^2 = \|v\|^2. \end{aligned}$$

Thus, S is an isometry. ∎

9 Trace and Determinant

9.1 Trace

Remark. *With respect to every basis of V, the matrix of the identity operator $I \in \mathcal{L}(V)$ is the diagonal matrix with 1's on the diagonal and 0's elsewhere.*

Definition 9.1.1 (Identity Matrix/I). Suppose n is a positive integer. The $n \times n$ diagonal matrix

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

is called the *identity matrix* and is denoted I.

Definition 9.1.2 (Invertible/Inverse/A^{-1}). A square matrix A is called *invertible* if there is a square matrix B of the same size such that $AB = BA = I$; we call B the *inverse* of A and denote it by A^{-1}.

> **Theorem 9.1.3**
> If A is an invertible square matrix, then $\exists$ a unique matrix B *s.t.* $AB = BA = I$.

Proof 1. Suppose $\exists$ two matrices B, B' *s.t.*

$$AB = BA = I \quad \text{and} \quad AB' = B'A = I.$$

Then, we have $AB = AB'$. So, $BAB = BAB'$. Therefore, $IB = IB'$, or $B = B'$. ∎

> **Theorem 9.1.4**
> Suppose $T \in \mathcal{L}(V)$ and $v_1, \ldots, v_n$ is a basis of V. Then, $\mathcal{M}(T, (v_1, \ldots, v_n))$ is invertible if and only if T is invertible.

Proof 2.

($\Rightarrow$) Suppose T is invertible, so $\exists S \in \mathcal{L}(V)$, $ST = TS = I$. Then, $\mathcal{M}(ST) = \mathcal{M}(TS) = \mathcal{M}(I)$. That is,

$$\mathcal{M}(S)\mathcal{M}(T) = \mathcal{M}(T)\mathcal{M}(S) = I.$$

So, $\mathcal{M}(T)$ is invertible. □

($\Leftarrow$) Let $A = \mathcal{M}(T)$ is invertible. Then, $\exists$ a matrix B *s.t.* $AB = BA = I$. Let $S \in \mathcal{L}(V)$ *s.t.* $B = \mathcal{M}(S)$. So,

$$\mathcal{M}(T)\mathcal{M}(S) = \mathcal{M}(S)\mathcal{M}(T) = \mathcal{M}(I).$$

That is, $\mathcal{M}(TS) = \mathcal{M}(ST) = I$, or $TS = ST = I$. Then, by definition, T is invertible. ∎

> **Theorem 9.1.5**
> Suppose $u_1, \ldots, u_n$ and $v_1, \ldots, v_n$ and $w_1, \ldots, w_m$ are all bases of V. Suppose $S, T \in \mathcal{L}(V)$. Then,
>
> $$\mathcal{M}(ST, (u_1, \ldots, u_n), (w_1, \ldots, w_n)) = \mathcal{M}(S, (v_1, \ldots, v_n), (w_1, \ldots, w_n))\mathcal{M}(T, (u_1, \ldots, u_n), (v_1, \ldots, v_n)).$$

Theorem 9.1.6
Suppose $u_1, \ldots, u_n$ and $v_1, \ldots, v_n$ are bases of V. Then, the matrices $\mathcal{M}(I, (u_1, \ldots, u_n), (v_1, \ldots, v_n))$ and $\mathcal{M}(I, (u_1, \ldots, u_n), (v_1, \ldots, v_n))$ are invertible, and each is the inverse of the other.

Proof 3. By Theorem 9.1.5, replacing w_j with u_j, we have

$$I = \mathcal{M}(I, (v_1, \ldots, v_n), (u_1, \ldots, u_n))\mathcal{M}(I, (u_1, \ldots, u_n), (v_1, \ldots, v_n)).$$

Now, interchanging the roles of u's and v's, we get

$$I = \mathcal{M}(I, (u_1, \ldots, u_n), (v_1, \ldots, v_n))\mathcal{M}(I, (v_1, \ldots, v_n), (u_1, \ldots, u_n)).$$

So, by definition, the desired result holds. ■

Example 9.1.7 Consider the bases $(4, 2), (5, 3)$ and $(1, 0), (0, 1)$ of $\mathbb{F}^2$. Then,

$$\mathcal{M}(I, ((4, 2), (5, 3)), ((1, 0), (0, 1))) = \begin{pmatrix} 4 & 5 \\ 2 & 3 \end{pmatrix}$$

because $I(4, 2) = 4(1, 0) + 2(0, 1)$ and $I(5, 3) = 5(1, 0) + 3(0, 1)$. Find the inverse of it.

Solution 4.
Suppose $I(1, 0) = a(4, 2) + b(5, 3)$ and $I(0, 1) = c(4, 2) + d(5, 3)$. Then, solve for

$$\begin{cases} 4a + 5b = 1 \\ 2a + 3b = 0 \end{cases} \quad \text{and} \quad \begin{cases} 4c + 5d = 0 \\ 2c + 3d = 1 \end{cases},$$

we have

$$\begin{cases} a = 3/2 \\ b = -1 \end{cases} \quad \text{and} \quad \begin{cases} c = -5/2 \\ d = 2 \end{cases}.$$

So, the inverse is

$$\begin{pmatrix} 3/2 & -5/2 \\ -1 & 2 \end{pmatrix}.$$

□

Theorem 9.1.8 Change of Basis Formula
Suppose $T \in \mathcal{L}(V)$. Let $u_1, \ldots, u_n$ and $v_1, \ldots, v_n$ be bases of V. Let

$$A = \mathcal{M}(I, (u_1, \ldots, u_n), (v_1, \ldots, v_n)).$$

Then

$$\mathcal{M}(T, (u_1, \ldots, u_n)) = A^{-1}\mathcal{M}(T, (v_1, \ldots, v_n))A.$$

Proof 5. By Theorem 9.1.5, replacing w_j with u_j and replace S with I, we have $\mathcal{M}(T, (u_1, \ldots, u_n)) = A^{-1}\mathcal{M}(T, (u_1, \ldots, u_n), (v_1, \ldots, v_n))$. Again, by Theorem 9.1.5, replacing w_j with v_j, T with I, and S with

T, we get

$$\mathcal{M}(T,(u_1,\ldots,u_n),(v_1,\ldots,v_n)) = \mathcal{M}(T,(v_1,\ldots,v_n))A.$$

Therefore, we've shown

$$\mathcal{M}(T,(u_1,\ldots,u_n)) = A^{-1}\mathcal{M}(T,(v_1,\ldots,v_n))A.$$

■

Definition 9.1.9 (Trace of an Operator). Suppose $T \in \mathcal{L}(V)$

- If $\mathbb{F} = \mathbb{C}$, then the trace of T is the sum of the eigenvalues of T, with each eigenvalue repeated according to its multiplicity.
- If $\mathbb{F} = \mathbb{R}$, then the trace of T is the sum of the eigenvalues of $T_{\mathbb{C}}$, with each eigenvalue repeated according to its multiplicity.

The trace of T is denoted $\operatorname{tr} T$.

Theorem 9.1.10
Suppose $T \in \mathcal{L}(V)$. Let $n = \dim V$. Then, $\operatorname{tr} T$ equals the negative of the coefficient of z^{n-1} in the characteristic polynomial of T.

Proof 6. Suppose $\lambda_1,\ldots,\lambda_n$ are eigenvalues of T with each eigenvalue repeated according to its multiplicity. Then, $(z-\lambda_1)\cdots(z-\lambda_n) = z^n - (\lambda_1+\cdots+\lambda_n)z^{n-1} + \cdots + (-1)^n(\lambda_!\cdots\lambda_n)$. Hence, we complete the proof. ■

Definition 9.1.11 (Trace of a Matrix). The *trace* of a square matrix A, denoted $\operatorname{tr} A$, is defined to be the sum of the diagonal entries of A.

Lemma 9.1.12 If A and B are square matrices of the same size, then $\operatorname{tr}(AB) = \operatorname{tr}(BA)$.

Proof 7. Suppose

$$A = \begin{pmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & \ddots & \vdots \\ A_{n1} & \cdots & A_{nn} \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} B_{11} & \cdots & B_{1n} \\ \vdots & \ddots & \vdots \\ B_{n1} & \cdots & B_{nn} \end{pmatrix}.$$

Then, $\left(AB\right)_{jj} = \sum_{k=1}^{n} A_{jk}B_{kj}$. So,

$$\begin{aligned} \operatorname{tr}(AB) = \sum_{j=1}^{n}\left(AB\right)_{jj} &= \sum_{j=1}^{n}\sum_{k=1}^{n} A_{jk}B_{kj} \\ &= \sum_{k=1}^{n}\sum_{j=1}^{n} B_{kj}A_{jk} \\ &= \sum_{k=1}^{n}\left(BA\right)_{kk} \\ &= \operatorname{tr}(BA). \end{aligned}$$

■

Lemma 9.1.13 Let $T \in \mathcal{L}(V)$. Suppose $u_1, \ldots, u_n$ and $v_1, \ldots, v_n$ are bases of V. Then,

$$\operatorname{tr} \mathcal{M}(T, (u_1, \ldots, u_n)) = \operatorname{tr} \mathcal{M}(T, (v_1, \ldots, v_n)).$$

Proof 8. Let $A = \mathcal{M}(I, (u_1, \ldots, u_n), (v_1, \ldots, v_n))$. Then,

$$\begin{aligned}
\operatorname{tr} \mathcal{M}(T, (u_1, \ldots, u_n)) &= \operatorname{tr} \left(A^{-1} \mathcal{M}(T, (v_1, \ldots, v_n))A\right) \\
&= \operatorname{tr} \left((\mathcal{M}(T, (v_1, \ldots, v_n))A)A^{-1}\right) \\
&= \operatorname{tr} \left(\mathcal{M}(T, (v_1, \ldots, v_n))(AA^{-1})\right) \\
&= \operatorname{tr} \mathcal{M}(T, (v_1, \ldots, v_n)).
\end{aligned}$$

■

Theorem 9.1.14
Suppose $T \in \mathcal{L}(V)$. Then, $\operatorname{tr} T = \operatorname{tr} \mathcal{M}(T)$.

Proof 9. By Lemma 9.1.13, we know $\operatorname{tr} \mathcal{M}(T)$ is independent of the choice of basis. Use the basis introduced by block diagonal matrix with upper-triangular blocks in previous Chapters, we have the desired result. If T is defined on a real vector space, then consider $\operatorname{tr} \mathcal{M}(T)$ on $T_{\mathbb{C}}$. ■

Theorem 9.1.15
Suppose $S, T \in \mathcal{L}(V)$. Then, $\operatorname{tr}(S + T) = \operatorname{tr} S + \operatorname{tr} T$.

Proof 10. Choose a basis of V. Then,

$$\begin{aligned}
\operatorname{tr}(S + T) &= \operatorname{tr} \mathcal{M}(S + T) \\
&= \operatorname{tr}(\mathcal{M}(S) + \mathcal{M}(T)) \\
&= \operatorname{tr} \mathcal{M}(S) + \operatorname{tr} \mathcal{M}(T) \\
&= \operatorname{tr} S + \operatorname{tr} T.
\end{aligned}$$

■

Theorem 9.1.16
$\nexists$ operators $S, T \in \mathcal{L}(V)$ *s.t.* $ST - TS = I$.

Proof 11. Let $S, T \in \mathcal{L}(V)$. Then,

$$\begin{aligned}
\operatorname{tr}(ST - TS) &= \operatorname{tr}(ST) - \operatorname{tr}(TS) \\
&= \operatorname{tr} \mathcal{M}(ST) - \operatorname{tr} \mathcal{M}(TS) \\
&= \operatorname{tr}(\mathcal{M}(S)\mathcal{M}(T)) - \operatorname{tr}(\mathcal{M}(T)\mathcal{M}(S)) \\
&= 0.
\end{aligned}$$

Since $\operatorname{tr} I = \dim V \neq 0$, $\operatorname{tr}(I) \neq \operatorname{tr}(ST - TS)$. So, it must be that $\nexists S, T \in \mathcal{L}(V)$ *s.t.* $ST - TS = I$. ■

9.2 Determinant

Definition 9.2.1 (Determinant of an Operator/$\det T$). Suppose $T \in \mathcal{L}(V)$.

- If $\mathbb{F} = \mathbb{C}$, then the *determinant* of T is the product of the eigenvalues of T, with each eigenvalue repeated according to its multiplicity.
- If $\mathbb{F} = \mathbb{R}$, then the *determinant* of T is the product of the eigenvalues of $T_{\mathbb{C}}$, with each eigenvalue repeated according to its multiplicity.

The determinant of T is denoted by $\det T$.

Theorem 9.2.2
Suppose $T \in \mathcal{L}(V)$. Let $n = \dim V$. Then, $\det T$ equals $(-1)^n$ times the constant term of the characteristic polynomial of T.

Proof 1. Suppose $\lambda 1, \ldots, \lambda_n$ are eigenvalues of T with each eigenvalue repeated according to its multiplicity. Then,

$$(z-\lambda_1)\cdots(z-\lambda_n) = z^n - (\lambda_1 + \cdots + \lambda_n)z^{n-1} + \cdots + (-1)^n(\lambda 1 \cdots \lambda_n).$$

Hence, we complete the proof. ■

Theorem 9.2.3
Suppose $T \in \mathcal{L}(V)$. Then, the characteristic polynomial of T can be written as

$$z^n - (\operatorname{tr} T)z^{n-1} + \cdots + (-1)^n(\det T).$$

Proof 2. By Theorem 9.1.10 and Theorem 9.2.2, we complete the proof. ■

Theorem 9.2.4
An operator on V is invertible if and only if its determinant is non-zero.

Proof 3. First, suppose V is complex and $T \in \mathcal{L}(V)$. Note that

$$\begin{aligned} T \text{ is invertible} &\iff T \text{ is bijective} \\ &\iff T \text{ is injective} \\ &\iff \operatorname{null} T = \{0\} \\ &\iff Tv \neq 0 \text{ whenever } v \neq 0 \\ &\iff 0 \text{ is not an eigenvalue of } T \\ &\iff \det T \neq 0. \end{aligned}$$

Now, consider the case where V is real, then

$$\begin{aligned} T \text{ is invertible} &\iff 0 \text{ is not an eigenvalue of } T \\ &\iff 0 \text{ is not an eigenvalue of } T_{\mathbb{C}} \\ &\iff \det T \neq 0. \end{aligned}$$

■

Theorem 9.2.5
Suppose $T \in \mathcal{L}(V)$. Then, the characteristic polynomial of T equals $\det(zI - T)$.

Proof 4. Suppose V is a complex vector space. If $\lambda, z \in \mathbb{C}$, then λ is an eigenvalue of T if and only if $\exists\, v \neq 0$ *s.t.* $Tv = \lambda v$. Then, $zIv - Tv = zv - \lambda v$. So,

$$(zI - T)v = (z - \lambda)v.$$

Therefore, we have $z - \lambda$ is an eigenvalue of $zI - T$. Let d be the multiplicity of λ, then

$$d = \dim G(\lambda, T) = \text{null}\,(T - \lambda I)^{\dim V}.$$

Note that $(T - \lambda I) = (z - \lambda)I - (zI - T)$. Then,

$$(T - \lambda I)^{\dim V} = [(z - \lambda)I - (zI - T)]^{\dim V}.$$

So, we have

$$\text{null}\,(T - \lambda I)^{\dim V} = \text{null}\,[(z - \lambda)I - (zI - T)]^{\dim V}.$$

That is, $G(\lambda, T) = G(z - \lambda, zI - T)$. So, $\dim G(\lambda, T) = G(z - \lambda, zI - T)$. Then, the multiplicity of $z - \lambda$ is also d.

Let $\lambda_1, \ldots, \lambda_n$ denote the eigenvalues of T. Then, $z - \lambda_1, \ldots, z - \lambda_n$ are precisely the eigenvalues of $zI - T$. So, $\det(zI - T) = (z - \lambda_1) \cdots (z - \lambda_n)$, the characteristic polynomial of T.

Now, consider the case if V is a real vector space. Then, apply the proof above to $T_{\mathbb{C}}$, and then we complete the proof. ■

Definition 9.2.6 (Permutation/$\text{perm}\, n$). A *permutation* of $(1, \ldots, n)$ is a list $(m_1, \ldots, m_n)$ that contains each of the numbers $1, \ldots, n$ exactly once. The set of all permutations of $(1, \ldots, n)$ is denoted $\text{perm}\, n$.

Definition 9.2.7 (Sign of a Permutation). The *sign of a permutation* $(m_1, \ldots, m_n)$ is defined to be 1 if the number of pairs of integers (j, k) with $1 \leq j < k \leq n$ *s.t.* j appears after k in the list $(m_1, \ldots, m_n)$ is even, and -1 if the number of such pairs is odd. In other words, the sign of a permutation is 1 if the natural order has been changed an even number of times, and is -1 if the natural order has been changed an odd number of times.

Example 9.2.8 For the permutation $(2, 4, 5, 3)$, we have the following pairs of integers: $(2, 4), (2, 5), (2, 3), (4, 5), (4, 3), (5, 3)$, among which $(4, 3)$ and $(5, 3)$ are of unnatural order. So, $\text{sign}(2, 4, 5, 3) = 1$.

Theorem 9.2.9
Interchanging two entries in a permutation multiplies the sign of the permutation by -1.

Proof 5. Suppose we have $m_1, \ldots, m_i, \ldots, m_j, \ldots, m_n$ and we want the interchange m_i and m_j to get $m_1, \ldots, m_j, \ldots, m_i, \ldots, m_n$.

1. Adjacent Case: m_i and m_j are adjacent to each other.

Let number of pairs of reverse order from the original permutation to be N. Then

$$\mathrm{sign}(\text{original permutation}) = (-1)^N.$$

(a) If $m_i < m_j$m then after the interchange, we get one more reverse order, and so

$$\mathrm{sign}(\text{interchanged permutation}) = (-1)^{N+1} = (-1)(-1)^N.$$

(b) If $m_i > m_j$, then after the interchange, we get one less reverse order. So,

$$\mathrm{sign}(\text{interchanged permutation}) = (-1)^{N-1} = \frac{(-1)^N}{(-1)} = (-1)(-1)^N.$$

2. General Case: m_i and m_j are not adjacent.

   Then, suppose we need k times to move m_i to the position right after m_j. We need $k-1$ times to move m_j to the position m_i initially at. So,

$$\mathrm{sign}(interchangedpermutation) = (-1)^{N+2k-1} = (-1)(-1)^N.$$

■

Definition 9.2.10 (Determinant of a Matrix, $\det A$). Suppose A is an $n \times n$ matrix such that

$$A = \begin{pmatrix} A_{1,1} & \cdots & A_{1,n} \\ \vdots & \ddots & \vdots \\ A_{n,1} & \cdots & A_{n,n} \end{pmatrix}.$$

The *determinant* of A, denoted $\det A$, is defined by

$$\det A = \sum_{(m_1,\ldots,m_n)\in \mathrm{perm}\, n} (\mathrm{sign}(m_1,\ldots,m_n)) A_{m_1,1}\cdots A_{m_n,n}.$$

Example 9.2.11 Compute determinant of an upper triangular matrix

$$A = \begin{pmatrix} A_{1,1} & & * \\ & \ddots & \\ 0 & & A_{n,n} \end{pmatrix}.$$

Solution 6.
By definition,

$$\det A = \sum_{(m_1,\ldots,m_n)\in \mathrm{perm}\, n} (\mathrm{sign}(m_1,\ldots,m_n)) A_{m_1,1}\cdots A_{m_n,n}.$$

Note that

$$A_{ij} \begin{cases} \neq 0 & i \le j \\ = 0 & i > j \end{cases}.$$

Consider $(1, \dots, n) \in \operatorname{perm} n$, $\operatorname{sign}(1, \dots, , n) = 1$, and $A_{m_1,1} \cdots A_{m_n,n}$ becomes $a_{1,1} \cdots A_{n,n}$. Now, if $(m_1, \dots, m_n) \neq (1, \dots, n)$, we can find some $A_{i,j} = 0$ with $i > j$. So,

$$\det A = (\operatorname{sign}(1, \dots, n))A_{1,1} \cdots A_{n,n} = A_{1,1} \cdots A_{n,n}.$$

□

Theorem 9.2.12
Suppose A is a square matrix and B is the matrix obtained from A by interchanging two columns. Then,

$$\det A = -\det B.$$

Proof 7. Suppose $A \in \mathbb{F}n \times n$ and $A = \begin{pmatrix} A_1 & \cdots & A_i & \cdots & A_j & \cdots & A_n \end{pmatrix}$. Then, by construction, we know $B = \begin{pmatrix} A_1 & \cdots & A_j & \cdots & A_i & \cdots & A_n \end{pmatrix}$. So,

$$\det A = \sum_{(m_1,\dots,m_n) \in \operatorname{perm} n} (\operatorname{sign}(m_1, \dots, m_n))A_{m_1,1} \cdots A_{m_i,i} \cdots A_{m_j,j} \cdots A_{m_n,n}$$

and

$$\det B = \sum_{(m_1,\dots,m_n) \in \operatorname{perm} n} (\operatorname{sign}(m_1, \dots, m_n))A_{m_1,1} \cdots A_{m_j,j} \cdots A_{m_i,i} \cdots A_{m_n,n}$$

Note that

$$\operatorname{sign}(m_1, \dots, m_i, \dots, m_j, \dots, m_n) = (-1)\operatorname{sign}(m_1, \dots, m_j, \dots, m_i, \dots, m_n).$$

So, by the linear properties of summation, we have

$$\det A = -\det B.$$

■

Theorem 9.2.13
If A is a square matrix that has two equal columns, then $\det A = 0$.

Proof 8. Interchanging the two equal columns, we still get the same matrix, A. Further, by Theorem 9.2.12, we have

$$\det A = -\det A,$$

suggesting $\det A = 0$. ■

Theorem 9.2.14
Suppose $A = \begin{pmatrix} A_{\cdot,1} & \cdots & A_{\cdot,n} \end{pmatrix}$ is an $n \times n$ matrix and $(m_1, \dots, m_n)$ is a permutation. Then,

$$\det \begin{pmatrix} A_{\cdot,m_1} & \cdots & A_{\cdot,m_n} \end{pmatrix} = (\operatorname{sign}(m_1, \dots, m_n)) \det A.$$

Theorem 9.2.15 Determinant is a Linear Function of Each Column
Suppose k, n are positive integers with $1 \leq k \leq n$. Fix $n \times 1$ matrices $A_{\cdot,1}, \cdots, A_{\cdot,n}$ except $A_{\cdot,k}$. Then, the function that takes an $n \times 1$ column vector $A_{\cdot,k}$ to $\det \begin{pmatrix} A_{\cdot,1} & \cdots & A_{\cdot,k} & \cdots & A_{\cdot,n} \end{pmatrix}$ is a linear map.

Theorem 9.2.16 Determinant is Multiplicative
Suppose A and B are square matrices of the same size. Then,

$$\det(AB) = \det(BA) = (\det A)(\det B).$$

Theorem 9.2.17
Let $T \in \mathcal{L}(V)$. Suppose $u_1, \ldots, u_n$ and $v_1, \ldots, v_n$ are bases of V. Then,

$$\det \mathcal{M}(T, (u_1, \ldots, u_n)) = \det \mathcal{M}(T, (v_1, \ldots, v_n)).$$

Theorem 9.2.18
Suppose $T \in \mathcal{L}(V)$. Then, $\det T = \det \mathcal{M}(T)$.

Theorem 9.2.19
Suppose $S, T \in \mathcal{L}(V)$. Then,

$$\det(ST) = \det(TS) = (\det T)(\det S).$$