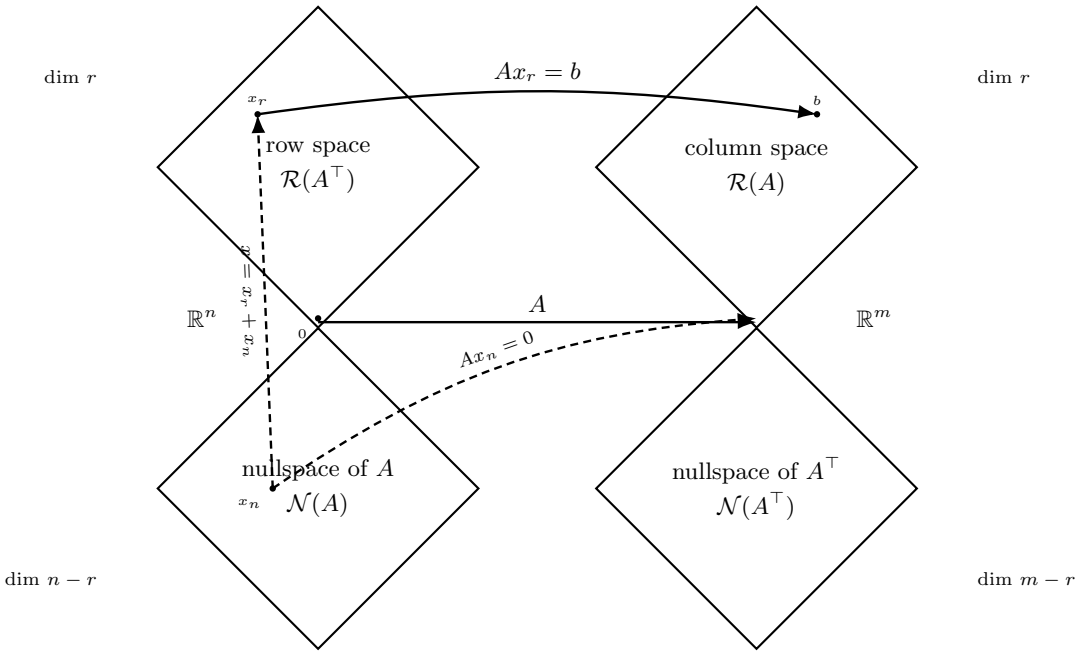
MATH 110

Lecture Notes

Semester Fall 2025



Alexander Lu

Instructor: Ken Ribet

Contents

Chapter 1

Vector Spaces

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

1.1.1 Complex Numbers

Complex numbers are an extension of the real number system, where the square-roots of negative numbers of possible.

Definition 1.1.1: Imaginary number, i

$i^2 = -1$

Definition 1.1.2: Complex Numbers, $\mathbb{C}$

- Every complex number $x \in \mathbb{C}$ is an ordered pair of two numbers (a, b) where $a, b \in \mathbb{R}$. a denotes the real component of x, while b denotes the imaginary component of x. We write $x = a + bi$
- The set of all complex numbers $\mathbb{C} = \{a + bi : a, b \in \mathbb{R}\}$
- multiplication and addition on C where $a, b, c, d \in \mathbb{R}$

$$(a + bi) + (c + di) = (a + c) + (b + d)i$$

$$(a + bi)(c + di) = (ac - bd) + (ad + bc)i$$

The familiar properties of real numbers apply to the complex system as well.

Theorem 1.1.3: Properties of complex arithmetic

$\forall \alpha, \beta, \lambda \in \mathbb{C}$

(a) **Commutativity:** $\alpha + \beta = \beta + \alpha$ and $\alpha\beta = \beta\alpha$.

(b) **Associativity:** $(\alpha + \beta) + \lambda = \alpha + (\beta + \lambda)$ and $(\alpha\beta)\lambda = \alpha(\beta\lambda)$

(c) **Identities:** $\lambda + 0 = \lambda$ and $\lambda 1 = \lambda$

(d) **Additive Inverse:** $\forall \alpha \in \mathbb{C}, \exists! \beta \in \mathbb{C}$ such that $\alpha + \beta = 0$

(e) **Multiplicative Inverse:** $\forall \alpha \in \mathbb{C}, \alpha \neq 0 \implies \exists! \beta \in \mathbb{C}, \alpha\beta = 1$

(f) **Distributive property:** $\lambda(\alpha + \beta) = \lambda\alpha + \lambda\beta$

Proof. Let $\alpha = a + bi$, $\beta = c + di$, $\lambda = e + fi$

(a)

$$\begin{aligned} \alpha + \beta &= a + bi + c + di \\ &= (a + c) + (b + d)i \\ &= (c + a) + (d + b)i \\ &= \beta + \alpha \end{aligned}$$

$$\begin{aligned} \alpha\beta &= (a + bi)(c + di) \\ &= ac + adi + cbi - bd \\ &= ca + cbi + adi - db \\ &= (c + di)(a + bi) \\ &= \beta\alpha \end{aligned}$$

(b)

$$\begin{aligned} (\alpha + \beta) + \lambda &= ((a + c) + (b + d)i) + (e + fi) \\ &= (a + c + e) + (b + d + f)i \\ &= a + (c + e) + (b + (d + f))i \\ &= \alpha + (\beta + \lambda) \end{aligned}$$

$$\begin{aligned} (\alpha\beta)\lambda &= (ac + adi + cbi - bd)(e + fi) \\ &= ace + acfi + adei - adf + cbei - cbf - bde - bdfi \end{aligned}$$

$$\begin{aligned} \alpha(\beta\lambda) &= (a + bi)(ce + cfi + dei - df) \\ &= ace + acfi + adei - adf + bcei - bcf - bde - bdfi \end{aligned}$$

$$(\alpha\beta)\lambda = \alpha(\beta\lambda)$$

(c)

$$\lambda + 0 = (e + fi) + (0 + 0i) = (e + 0) + (f + 0)i = e + fi = \lambda$$

$$\lambda 1 = (e + fi)(1) = (e \cdot 1 + fi \cdot 1) = e + fi = \lambda$$

(d) First prove existence of β. Let $\beta = a \cdot -1 + bi \cdot -1 = -a - bi$. Thus:

$$\alpha + \beta = (a + bi) + (-a - bi) = (a - a) + (b - b)i = 0 + 0i = 0$$

Then prove uniqueness of β. Suppose there exists $\beta_1 \neq \beta$ such that $\alpha + \beta_1 = 0$. We know $\alpha + \beta = 0$. Hence:

$$(\alpha + \beta_1) - (\alpha + \beta) = \beta_1 - \beta = 0$$

$$\beta_1 = \beta$$

A contradiction.

(e) If $\alpha \neq 0$, $a^2 + b^2 > 0$. We define $\beta = \frac{a}{a^2+b^2} + \frac{-b}{a^2+b^2}i$.

$$\begin{aligned}
\alpha\beta &= (a + bi)\left(\frac{a}{a^2 + b^2} + \frac{-b}{a^2 + b^2}i\right) \\
&= \frac{a^2}{a^2 + b^2} + \frac{-abi}{a^2 + b^2} + \frac{abi}{a^2 + b^2} + \frac{b^2}{(a^2 + b^2)} \\
&= \frac{a^2}{a^2 + b^2} + \frac{b^2}{a^2 + b^2} + 0i \\
&= \frac{a^2 + b^2}{a^2 + b^2} + 0i \\
&= 1 + 0i \\
&= 1
\end{aligned}$$

Now prove the uniqueness of β. Suppose there exists $\beta_1 \neq \beta$ such that $\alpha\beta_1 = 1$. We know $\alpha\beta = 1$. Hence:

$$\alpha\beta - \alpha\beta_1 = \alpha(\beta - \beta_1) = 0 \implies \beta - \beta_1 = 0$$

$$\beta = \beta_1$$

A contradiction.

(f)

$$\begin{aligned}
\lambda(\alpha + \beta) &= (e + fi)((a + bi) + (c + di)) \\
&= (e + fi)((a + c) + (b + d)i) \\
&= e(a + c) + e(b + d)i + f(a + c)i - f(b + d) \\
&= e(a + c) + f(a + c)i + e(b + d)i - f(b + d) \\
&= \lambda(a + c) + \lambda(bi + di) \\
&= \lambda a + \lambda bi + \lambda c + \lambda di \\
&= \lambda(a + bi) + \lambda(c + di) \\
&= \lambda\alpha + \lambda\beta
\end{aligned}$$

□

For rigor, division and subtraction operations must also be defined:

Definition 1.1.4: Subtraction ($-\alpha$), and Division ($\frac{1}{\alpha}$)

Let $\alpha, \beta \in \mathbb{C}$

- **Subtraction:** We define $\alpha - \beta$ as $\alpha + (-\beta)$ where $-\beta$ is the additive inverse of β.
- **Division:** For $\beta \neq 0$ We define $\frac{\alpha}{\beta}$ as $\alpha \cdot \frac{1}{\beta}$ where $\frac{1}{\beta}$ is the unique multiplicative inverse of β.

Both $\mathbb{R}$ and $\mathbb{C}$ are fields. For simplities sake, we define the following:

Definition 1.1.5: $\mathbb{F}$

$\mathbb{F}$ represents either $\mathbb{R}$ or $\mathbb{C}$

1.1.2 Lists

Definition 1.1.6: List and length

- Let n be a natural number. A list of length n is an ordered set of n objects.
- Two lists are equal if and only if they have the same length, the same elements, and the same order
- A list of length n is written as such:

$$(z_1, \ldots, z_n)$$

- lists have **finite** length

A list of length n is often called an n-**tuple**

1.1.3 $\mathbb{F}^n$

The $\mathbb{R}^2$ is the set of all ordered pairs of real numbers, which geometrically forms a plane:

$$\mathbb{R}^2 = \{(x, y) : x, y \in \mathbb{R}\}$$

Higher dimensional sets consists of elements that are n-tuples, where n is the dimensionality of the set.

Definition 1.1.7: $\mathbb{F}^n$ and coordinates

$\mathbb{F}^n$ is the set of all lists of length n of elements belonging to $\mathbb{F}$:

$$\mathbb{F}^n = \{(x_1 \ldots, x_n) : x_k \in \mathbb{F} \text{ for } k = 1, \ldots, n\}$$

x_k is denoted as the k^{th} coordinate of $(x_1, \ldots, x_n)$.

Definition 1.1.8: addition in $\mathbb{F}^n$

Addition between elements $\mathbf{x}, \mathbf{y} \in \mathbb{F}^n$ is defined by adding coordinates of $\mathbf{x}$ and $\mathbf{y}$ together element-wise.

$$\mathbf{x} + \mathbf{y} = (x_1, \ldots, x_n) + (y_1, \ldots, y_n) = (x_1 + y_1, \ldots, x_n + y_n)$$

Definition 1.1.9: 0

$\mathbf{0}$ is the list of length n whose coordinates are all 0

$$\mathbf{0} = (0, \ldots, 0)$$

An element of $\mathbb{F}^n$ can either be interpreted as a point $(x_1, \ldots, x_n)$, or as an arrow extending from the origin, a **vector**. The geometric way of adding two element of $\mathbb{R}^n$ together, is to simply shift the second vector ontop of the end of the first vector (the arrow tip), and draw a new vector extending from the base of the first vector, to the tip of the second vector. This is the vector sum.

Theorem 1.1.10: Properties of $\mathbb{F}^n$

for $\forall \mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{F}^n$ and $\forall a, b, \lambda \in \mathbb{F}$:

(a) **Commutativity of Addition:** $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$

(b) **Associativity of Addition:** $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$

(c) **Associativity of scalar multiplication** $(ab)\mathbf{x} = a(b\mathbf{x})$

(d) **Scalar multiplicative identity:** $1\mathbf{x} = \mathbf{x}$

(e) **Left distributivity of scalar multiplication over vector addition:** $\lambda(\mathbf{x} + \mathbf{y}) = \lambda\mathbf{x} + \lambda\mathbf{x}$

(f) **Right distributivity of scalar multiplication over vector addition:** $(a + b)\mathbf{x} = a\mathbf{x} + b\mathbf{x}$

Proof. $\forall \mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{F}^n$ and $\forall a, b, \lambda \in \mathbb{F}$

(a)

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

(b)

$$\begin{aligned}(\mathbf{x}+\mathbf{y})+\mathbf{z} &= ((x_1+y_1),\dots(x_n+y_n))+(z_1,\dots,z_n)\\ &= ((x_1+y_1)+z_1,\dots,(x_n+y_n)+z_n)\\ &= (x_1+(y_1+z_1),\dots,x_n+(y_n+z_n))\\ &= \mathbf{x}+(\mathbf{y}+\mathbf{z})\end{aligned}$$

(c)

$$\begin{aligned}(ab)\mathbf{x} &= (ab)(x_1,\dots,x_n)\\ &= (abx_1,\dots abx_n)\\ &= (a(bx_1),\dots,a(bx_n))\\ &= a(b\mathbf{x})\end{aligned}$$

(d)

$$1\mathbf{x} = 1(x_1,\dots,x_n) = (1x_1,\dots,1x_n) = (x_1,\dots,x_n) = \mathbf{x}$$

(e)

$$\begin{aligned}\lambda(\mathbf{x}+\mathbf{y}) &= \lambda(x_1+y_1,\dots,x_n+y_n)\\ &= (\lambda(x_1+y_1),\dots,\lambda(x_n+y_n))\\ &= (\lambda x_1+\lambda y_1,\dots\lambda x_n+\lambda y_n)\\ &= (\lambda x_1,\dots\lambda x_n)+(\lambda y_1,\dots\lambda y_n)\\ &= \lambda\mathbf{x}+\lambda\mathbf{y}\end{aligned}$$

(f)

$$\begin{aligned}(a+b)\mathbf{x} &= (a+b)(x_1,\dots,x_n)\\ &= ((a+b)x_1,\dots,(a+b)x_n)\\ &= (ax_1+bx_1,\dots,ax_n+bx_n)\\ &= (ax_1,\dots,ax_n)+(bx_1,\dots,bx_n)\\ &= a\mathbf{x}+b\mathbf{x}\end{aligned}$$

□

Definition 1.1.11: Additive inverse in $\mathbb{F}^n$, $-\mathbf{x}$

$\forall x \in \mathbb{F}^n$, the additive inverse of $\mathbf{x}$, $-\mathbf{x}$m is the vector $\mathbf{x} \in \mathbb{F}^n$ such that

$$\mathbf{x}+(-\mathbf{x}) = \mathbf{0}$$

Given that $\mathbf{x} = (x_1,\dots,x_n)$, then $-\mathbf{x} = (-x_1,\dots,-x_n)$.

The additive inverse of $\mathbf{x}$ is a vector with the same length but opposite direction of $\mathbf{x}$

Most multiplication in this course will not deal with vector-vector element multiplication. Instead, multiplication between a scalar and a vector will be crucial.

Definition 1.1.12: Scalar multiplication in $\mathbb{F}^n$

The product of a scalar $\lambda \in \mathbb{F}$ and a vector $\mathbf{x} \in \mathbb{F}^n$ is derived by multiplying each coordinate of the vector by λ

$$\lambda(x_1, \ldots, x_n) = (\lambda x_1, \ldots, \lambda x_n)$$

The scalar multiplication of a vector is geometrically a scaling of the vector by the magnitude of the scalar. An alternative vector product is the **dot product**, where the element-wise product of two vectors are summed up to return a scalar.

1.1.4 Exercises

Exercise 1.1.1. Show that

$$\frac{-1+\sqrt{3}i}{2}$$

is a cube root of 1 (meaning that its cube equals 1).

Solution

$$\begin{aligned}\left(\frac{-1+\sqrt{3}i}{2}\right)^3 &= \left(\frac{-2-2\sqrt{3}i}{4}\right)\left(\frac{-1+\sqrt{3}i}{2}\right) \\ &= \frac{2-2\sqrt{3}i+2\sqrt{3}i+6}{8} \\ &= \frac{8}{8} \\ &= \boxed{1}\end{aligned}$$

□

Exercise 1.1.2. Find two distinct square roots of i

Solution

We note that we want a complex number such that the real parts cancel out when the number if squared. This suggests that both the real and complex part of the number must be equal in magnitude. Trying $1+i$ yields:

$$(1+i)^2 = 1+2i-1 = 2i$$

We then divide by $\sqrt{2}$ to get the first root:

$$\boxed{\frac{1+i}{\sqrt{2}}}$$

The second root is just the additive inverse of the first root:

$$\boxed{-\frac{1+i}{\sqrt{2}}}$$

As expected, both roots evaluate to i:

$$\left[-\left(\frac{1+i}{2}\right)\right]^2 = \left(\frac{1+i}{\sqrt{2}}\right)^2 = \frac{1+2i-1}{2} = i$$

□

Exercise 1.1.3. Find $\mathbf{x} \in \mathbb{R}^4$ such that

$$(4,-3,1,7) + 2\mathbf{x} = (5,9,-6,8)$$

Solution

$$\begin{aligned}(4,-3,1,7) + 2\mathbf{x} &= (5,9,-6,8)\\ 2\mathbf{x} &= (1,12,-7,1)\\ \mathbf{x} &= \boxed{\left(\frac{1}{2}, 6, -\frac{7}{2}, \frac{1}{2}\right)}\end{aligned}$$

Exercise 1.1.4. Explain why there does not exist $\lambda \in \mathbb{C}$ such that

$$\lambda(2-3i, 5+4i, -6+7i) = (12-5i, 7+22i, -32-9i)$$

Solution

The vectors $(2-3i, 5+4i, -6+7i)$ and $(12-5i, 7+22i, -32-9i)$ do not point in the same direction. This can be verified by taking the element-wise ratios of these two vectors. For the first element:

$$\frac{2-3i}{12-5i} = \frac{2-3i}{12-5i}\cdot\frac{12+5i}{12+5i} = \frac{24+10i-36i+15}{144+25} = \frac{39-26i}{169} = \frac{3}{13} - \frac{2}{13}i$$

For the second element:

$$\frac{5+4i}{7+22i} = \frac{5+4i}{7+22i}\cdot\frac{7-22i}{7-22i} = \frac{35-110i+28i+88}{49+484} = \frac{123-82i}{533} = \frac{3}{13} - \frac{2}{13}i$$

For the third element:

$$\frac{-6+7i}{-32-9i} = \frac{-6+7i}{-32-9i}\cdot\frac{-32+9i}{-32+9i} = \frac{192-54i-324i-63}{1024+81} = \frac{129-378i}{1105}$$

Since the ratios are not equal, the vectors do not point in the same direction, so there does not exist a scalar λ that relates the two vectors.

Alternatively: For argument's sake, assume $\lambda \in \mathbb{C}$ DOES exist. This means that $\lambda = a+bi$ where $a, b \in \mathbb{R}$. Since $a, b \in \mathbb{R}$, we can use the definition of scalar multiplication to write $\lambda(2-3i, 5+4i, -6+7i)$ as $(a+b)(2-3i, 5+4i, -6+7i)$, which is equal to $(2a-3b)$ □

1.2 Definition of Vector Space

1.2.1 Vector Spaces

Before defining what a vector space is, we must formalize **vector addition** and **scalar multiplication**:

Definition 1.2.1: Addition, scalar multiplication

- A **vector addition** on a set $\mathcal{V}$ is a function that assigns an element $u+v \in \mathcal{V}$ to each pair of elements $u, v \in \mathcal{V}$.
- A **scalar multiplication** on a set $\mathcal{V}$ is a function that assigns an element $\lambda v \in \mathcal{V}$ to each $\lambda \in \mathbb{F}$ and each $v \in \mathcal{V}$.

A vector space is a set $\mathcal{V}$ that satisfies the following properties

Definition 1.2.2: Vector Space

A **vector space** is a set $\mathcal{V}$ along with an addition on $\mathcal{V}$ and a scalar multiplication on $\mathcal{V}$ such that the following properties hold:

- **Commutativity:** $u+v=v+u$ for all $u, v \in \mathcal{V}$.
- **Associativity:** $(u+v)+w=u+(v+w)$ and $(ab)v=a(bv)$ for all $u, v, w \in \mathcal{V}$ and for all $a, b \in \mathbb{F}$.
- **Additive identity:** There exists an element $0 \in \mathcal{V}$ such that $v+0=v$ for all $v \in \mathcal{V}$.
- **Additive inverse:** For every $v \in \mathcal{V}$, there exists $w \in \mathcal{V}$ such that $v+w=0$.
- **Multiplicative identity** $1v=v$ for all $v \in \mathcal{V}$.
- **Distributive Properties:** $a(u+v)=au+av$ and $(a+b)v=av+bv$ for all $a, b \in \mathbb{F}$ and all $u, v \in \mathcal{V}$.

The elements of a vector space are often referred to as **vectors** or **points**. Moreover, scalar multiplication in a vector space $\mathcal{V}$ depends on the underlying field $\mathbb{V}$. To be precise, we say that $\mathcal{V}$ is a **vector space over** $\mathbb{F}$. 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**.

Example.

> The simplest vector space is $\{0\}$

Example.

$\mathbb{F}^\infty$ is the set of all sequences of elements of $\mathbb{F}$:

$$\mathbb{F}^\infty = \{(x_1, x_2, \ldots) : x_k \in \mathbb{F} \text{ for } k = 1, 2, \ldots\}$$

. The properties of a vector space still holds with $\mathbb{F}^\infty$.

Vector spaces aren't only concerned with lists of elements. They could be applied to any arbitrary object, so long as the necessary operations are defined and the properties are satisfied. For example example, functions.

Definition 1.2.3: $\mathbb{F}^S$

- $\mathbb{F}^S$ denotes the set of functions from a set S to $\mathbb{F}$.
- For $f, g \in \mathbb{F}^S$, the sum $f + g \in \mathbb{F}^S$ is the function defined by

  $$(f + g)(x) = f(x) + g(x)$$

  for all $x \in S$
- For $\lambda \in \mathbb{F}$ and $f \in \mathbb{F}^S$, the product of $\lambda f \in \mathbb{F}^S$ is the function defined by

  $$(\lambda f)(x) = \lambda f(x)$$

  for all $x \in S$

Example.

$\mathbb{R}^n$ is the set of all real-valued functions on the interval $[0, 1]$

The vector space $\mathbb{F}^n$ is actually a special case of $\mathbb{F}^S$. Since each element in $\mathbb{F}^n$ can be thought of as a function (mapping) from the set $\{1, 2, \ldots, n\}$, where the function is:

$$x(k) \mapsto x_k$$

Remark.

$\mathbb{F}^n$ is also $\mathbb{F}^{\{1,\ldots,n\}}$

Theorem 1.2.4: Unique additive identity

Any vector space has a unique additive identity 0.

Proof. Suppose for the sake of contradiction that there exists distinct additive identities for vector space $\mathcal{V}$: 0 and 0^*. Since addition commutes:

$$0 = 0 + 0^* = 0^* + 0 = 0*$$

Since $0 = 0^*$, we have a contradiction, $\mathcal{V}$ has only one additive identity. □ □

Theorem 1.2.5: Unique additive inverse

Every element in a vector space $\mathcal{V}$ has a unique additive inverse.

Proof. Suppose for the sake of contradiction for an element $x \in \mathcal{V}$ there exists two distinct additive inverses, y and y^*. Then:

$$y = y + 0 = y + (x + y^*) = (y + x) + y^* = 0 + y^* = y^*$$

Since $y = y^*$, we have a contradiction, $\mathcal{V}$ has only one additive inverse. □

We now have the tools to define subtraction in a vector space:

Definition 1.2.6: Subtraction

Let $\mathbf{v}, \mathbf{w} \in \mathcal{V}$.

- $-\mathbf{v}$ denotes the additive inverse of $\mathbf{v}$.
- **Subtraction:** $\mathbf{w} - \mathbf{v}$ is defined as $\mathbf{w} + (-\mathbf{v})$.

The following theorem appears obvious but has a pretty slick proof:

Theorem 1.2.7: Any vector multiplied by 0 yields $\mathbf{0}$

$0\mathbf{v} = \mathbf{0}$ for every $\mathbf{v} \in \mathcal{V}$

Proof. $\forall \mathbf{v} \in \mathcal{V}$, we have:

$$0\mathbf{v} = (0 + 0)\mathbf{v} = 0\mathbf{v} + 0\mathbf{v}$$

This implies that $0\mathbf{v}$ is the additive inverse of itself, which means that $0\mathbf{v} = \mathbf{0}$. More rigorously, adding the additive inverse of $0\mathbf{v}$ to both sides would yield:

$$\mathbf{0} = 0\mathbf{v}$$

□ □

We have another different, yet similar theorem:

Theorem 1.2.8: A number times the vector 0 is 0

$a\mathbf{0} = \mathbf{0}$ for every $a \in \mathbb{F}$

Proof. $\forall a \in \mathbb{F}$:

$$a\mathbf{0} = \underbrace{a(\mathbf{0} + \mathbf{0})}_{\mathbf{0} \text{ is its own additive inverse}} = a\mathbf{0} + a\mathbf{0}$$

And then by adding the additive inverse of $a\mathbf{0}$ to both sides, we get:

$$\mathbf{0} = a\mathbf{0}$$

□

Whenever a vector $\mathbf{v} \in \mathcal{V}$ is multiplied by the scalar -1, then the result is the additive inverse of $\mathbf{v}$.

Theorem 1.2.9: The scalar -1 times a vector v yields $-\mathbf{v}$

$(-1)\mathbf{v} = -\mathbf{v}$ for every $v \in \mathcal{V}$

Proof. $\forall \mathbf{v} \in \mathcal{V}$, we have:

$$\mathbf{v} + (-1)\mathbf{v} = 1\mathbf{v} + (-1)\mathbf{v} = (1 + (-1))\mathbf{v} = 0\mathbf{v} = \mathbf{0}$$

By definition, $(-1)\mathbf{v} = -\mathbf{v}$. □

1.2.2 Exercises

Exercise 1.2.1. Prove that $-(-\mathbf{v}) = \mathbf{v}$ for every $\mathbf{v} \in \mathcal{V}$.

Solution

We note that:

$$-\mathbf{v} + (-(-\mathbf{v})) = \mathbf{0}$$

$$-\mathbf{v} + \mathbf{v} = \mathbf{0}$$

By the uniqueness of the additive inverse, it's necessary for $(-(-\mathbf{v})) = \mathbf{v}$. □

Exercise 1.2.2. Suppose $a \in \mathbb{F}, \mathbf{v} \in \mathcal{V}$, and $a\mathbf{v} = \mathbf{0}$. Prove that $a = 0$ or $v = 0$.

Solution

We prove by cases:

- **Case 1** $a = 0$**:** This is trivial
- **Case 2** $a \neq 0$**:** Then there exists a multiplicative inverse of a, $a^{-1} \in \mathbb{F}$. Multiplying both sides of $a\mathbf{v} = \mathbf{0}$ yields:

$$\mathbf{v} = \mathbf{0}$$

So no matter what, $a = 0$ or $\mathbf{v} = \mathbf{0}$ □

Exercise 1.2.3. Suppose $\mathbf{v}, \mathbf{w} \in \mathcal{V}$. Explain why there exists a unique $\mathbf{x} \in \mathcal{V}$ such that $\mathbf{v} + 3\mathbf{x} = \mathbf{w}$

Solution

By adding the additive inverse of v to both sides, we get:

$$3\mathbf{x} = \mathbf{w} - \mathbf{v}$$

This proves existence, as:

$$\mathbf{x} = \frac{1}{3}(\mathbf{w} - \mathbf{v})$$

Verify by:

$$\begin{aligned}\mathbf{v} + 3\mathbf{x} &= \mathbf{v} + 3\left(\frac{1}{3}(\mathbf{w} - \mathbf{v})\right) \\ &= \mathbf{v} + (\mathbf{w} - \mathbf{v}) \\ &= \mathbf{w}\end{aligned}$$

To prove existence, suppose that there exists some other solution $\mathbf{x}^*$:

$$\mathbf{v} + 3\mathbf{x} = \mathbf{v} + 3\mathbf{x}^* = \mathbf{w}$$

By adding the additive inverse of $\mathbf{v}$, we have:

$$3\mathbf{x} = 3\mathbf{x}^* \implies \mathbf{x} = \mathbf{x}^*$$

□

Exercise 1.2.4. The empty set is not a vector space. The empty set fails to satisfy only one of the requirements listed in the definition of a vector space. Which one?

Solution

Since the empty set has no elements, there does **not exist an additive identity element**. □

Exercise 1.2.5. Show that in the definition of a vector space, the additive inverse condition can be replaced with the condition that

$$0\mathbf{v} = \mathbf{0} \text{ for all } \mathbf{v} \in \mathcal{V}.$$

Solution

The statement for the additive inverse condition for a vector space $\mathcal{V}$ is:

$$\forall \mathbf{v} \in \mathcal{V}, \exists! - \mathbf{v} \in \mathcal{V}, \mathbf{v} + (-\mathbf{v}) = \mathbf{0}$$

We recall that $(-1)\mathbf{v} = -\mathbf{v}$ for every $\mathbf{v} \in \mathcal{V}$ from theorem 1.2.9. Hence:

$$\mathbf{v} + (-\mathbf{v}) = (1 + (-1))\mathbf{v} = 0\mathbf{v} = \mathbf{0}$$

To show the converse, suppose that we know every $\mathbf{v} \in \mathcal{V}$ has an additive inverse. Then we know:

$$0\mathbf{v} = (0 + 0)\mathbf{v} = 0\mathbf{v} + 0\mathbf{v}$$

If we add the additive inverse on both sides, we get:

$$0\mathbf{v} = \mathbf{0}$$

Which is what we wanted to show □ □

Exercise 1.2.6. Let ∞ and $-\infty$ denote two distinct objects, neither of which is in $\mathbb{R}$. Define an addition and scalar multiplication on $\mathbb{R} \cup \{\infty, -\infty\}$ as you could guess from the notation. Specifically the sum and product of two real numbers is as usual, and for $t \in \mathbb{R}$ define

$$t\infty = \begin{cases} -\infty & \text{if } t < 0, \\ 0 & \text{if } t = 0, \\ \infty & \text{if } t > 0, \end{cases} \quad t(-\infty) = \begin{cases} +\infty, & \text{if } t < 0, \\ 0, & \text{if } t = 0, \\ -\infty, & \text{if } t > 0. \end{cases}$$

and

$$\begin{aligned} t + \infty = \infty + t = \infty + \infty = \infty, \\ t + (-\infty) = (-\infty) + t = (-\infty) + (-\infty) = -\infty, \\ \infty + (-\infty) = (-\infty) + \infty = 0. \end{aligned}$$

Which these operations of addition and scalar multiplication, is $\mathbb{R} \cup \{\infty, -\infty\}$ a vector space over $\mathbb{R}$? Explain.

Solution

No $\mathbb{R} \cup \{\infty, -\infty\}$ is not a vector space. Take the definition:

$$\infty + t = \infty + \infty$$

Suppose for the sake of contradiction that $\mathbb{R} \cup \{\infty, -\infty\}$ was a vector space, we then add the additive inverse of ∞ to both sides and get:

$$t = \infty, t \in \mathbb{R}$$

But since $t \in \mathbb{R}$, and $\infty \notin \mathbb{R}$, this is a contradiction. □

Exercise 1.2.7. Suppose S is a nonempty set. Let $\mathcal{V}^S$ denote the set of functions from S to V. Define a natural addition and scalar multiplication on $\mathcal{V}^S$, and show that $\mathcal{V}^S$ is a vector space with these definitions.

Solution

Let $f, g, h \in \mathcal{V}^S$, $a \in S$, $\lambda, \gamma \in \mathcal{V}$. We define the natural addition of f and g on $\mathcal{V}^S$ as:

$$(f + g)(a) = f(a) + g(a)$$

We define the scalar multiplication on $\mathcal{V}^S$ as:

$$(\lambda f)(a) = \lambda\,(f(a))$$

To show that $\mathcal{V}^S$, we must prove it satisfies all the properties of theorem 1.1.3.

- **Commutativity of addition:**

$$(f+g)(a) = f(a) + g(a) = g(a) + f(a) = (g+f)(a)$$

- **Associativity:**

$$\begin{aligned}((f+g)+h)(a) &= (f+g)(a) + h(a)\\ &= f(a) + g(a) + h(a)\\ &= f(a) + (g(a) + h(a))\\ &= (f + (g+h))(a)\end{aligned}$$

$$\begin{aligned}((\lambda\gamma)f)(a) &= (\lambda\gamma)f(a)\\ &= \lambda(\gamma f(a))\\ &= (\lambda(\gamma f))(a)\end{aligned}$$

- **Additive identity:** Let 0 be the additive identity of $\mathcal{V}$. Define $z(a) \in \mathcal{V}^S$ as:

$$z(a) = 0$$

  Hence:

$$(f+z)(a) = f(a) + z(a) = f(a) + 0 = f(a)$$

- **Additive inverse:** We know that by the definition of vector space, $\forall\lambda \in \mathcal{V}, \exists(-\lambda) \in \mathcal{V}$. Since $\forall s \in S$, $\forall f \in \mathcal{V}^S$, $f(a) \in \mathcal{V}$, there exists some additive inverse $-f(a)$ for every $f(a)$. Hence, we can define the additive inverse g of f as:

$$g(a) = -f(a)$$

  Hence:

$$\begin{aligned}(g+f)(a) &= g(a) + f(a)\\ &= -f(a) + f(a)\\ &= 0\\ &= z(a) \quad \text{(the identity)}\end{aligned}$$

- **Multiplicative identity:**

$$1f(a) = f(a)$$

- **Distributive properties:**

$$\begin{aligned}(\lambda(f+g))(a) &= \lambda(f+g)(a)\\ &= \lambda(f(a) + g(a))\\ &= \lambda f(a) + \lambda g(a)\\ &= (\lambda f + \lambda g)(a)\end{aligned}$$

$$\begin{aligned}((\lambda+\gamma)f)(a) &= (\lambda f+\gamma f)(a)\\ &= \lambda f(a)+\gamma f(a)\\ &= (\lambda f+\gamma f)(a)\end{aligned}$$

Thus $\mathcal{V}^S$ is a vector field.

Exercise 1.2.8. Suppose $\mathcal{V}$ is a real vector space.

- The **complexification** of $\mathcal{V}$ denoted by $\mathcal{V}_C$, equals $\mathcal{V}\times\mathcal{V}$. An element of $\mathcal{V}_C$ is an ordered pair $(\mathbf{u},\mathbf{v})$, where $\mathbf{u},\mathbf{v}\in\mathcal{V}$, but we write this as $\mathbf{u}+i\mathbf{v}$.
- Addition on $\mathcal{V}_C$ is defined by

$$(\mathbf{u}_1+i\mathbf{v}_1)+(\mathbf{u}_2+i\mathbf{v}_2)=(\mathbf{u}_1+\mathbf{u}_2)+i(\mathbf{v}_1+\mathbf{v}_2),\forall\mathbf{u}_1,\mathbf{u}_2,\mathbf{v}_1,\mathbf{v}_2\in\mathcal{V}$$

- Complex scalar multiplication on $\mathcal{V}_C$ is defined by

$$(a+bi)(\mathbf{u}+i\mathbf{v})=(a\mathbf{u}-b\mathbf{v})+i(a\mathbf{v}+b\mathbf{u})$$

  for all $a,b,\in\mathbb{R}$ and all $\mathbf{u},\mathbf{v}\in\mathcal{V}$.

Prove that with the definitions of addition and scalar multiplication as above, $\mathcal{V}_C$ is a complex vector space.

Solution

We must once again prove all axioms of a vector space hold for $\mathcal{V}_C$. Let $\lambda=a+bi$ and $\gamma=c+di$ where $a,b,c,d\in\mathbb{R}$. Let $\mathbf{x},\mathbf{y},\mathbf{z}\in\mathcal{V}_C$ where $\mathbf{x}=\mathbf{u}_1+i\mathbf{v}_1$, $\mathbf{y}=\mathbf{u}_2+i\mathbf{v}_2$, and $\mathbf{z}=\mathbf{u}_3+i\mathbf{v}_3$:

- **Commutativity:**

$$\begin{aligned}\mathbf{x}+\mathbf{y} &= (\mathbf{u}_1+i\mathbf{v}_1)+(\mathbf{u}_2+i\mathbf{v}_2)\\ &= (\mathbf{u}_1+\mathbf{u}_2)+i(\mathbf{v}_1+\mathbf{v}_2)\\ &= (\mathbf{u}_2+\mathbf{u}_1)+i(\mathbf{v}_2+\mathbf{v}_1)\\ &= (\mathbf{u}_2+i\mathbf{v}_2)+(\mathbf{u}_1+i\mathbf{v}_1)\\ &= \mathbf{y}+\mathbf{x}\end{aligned}$$

- **Associativity:**

$$\begin{aligned}(\mathbf{x}+\mathbf{y})+\mathbf{z} &= ((\mathbf{u}_1+i\mathbf{v}_1)+(\mathbf{u}_2+i\mathbf{v}_2))+(\mathbf{u}_3+i\mathbf{v}_3)\\ &= ((\mathbf{u}_1+\mathbf{u}_2)+i(\mathbf{v}_1+\mathbf{v}_2))+(\mathbf{u}_3+i\mathbf{v}_3)\\ &= ((\mathbf{u}_1+\mathbf{u}_2)+\mathbf{u}_3)+i((\mathbf{v}_1+\mathbf{v}_2)+\mathbf{v}_3)\\ &= (\mathbf{u}_1+(\mathbf{u}_2+\mathbf{u}_3))+i(\mathbf{v}_1+(\mathbf{v}_2+\mathbf{v}_3))\\ &= (\mathbf{u}_1+i\mathbf{v}_1)+((\mathbf{u}_2+\mathbf{u}_3)+i(\mathbf{v}_2+\mathbf{v}_3))\\ &= (\mathbf{u}_1+i\mathbf{v}_1)+((\mathbf{u}_2+i\mathbf{v}_2)+(\mathbf{u}_3+i\mathbf{v}_3))\\ &= \mathbf{x}+(\mathbf{y}+\mathbf{z})\end{aligned}$$

$$\begin{aligned}(\lambda\gamma)\mathbf{x} &= ((a+bi)(c+di))(\mathbf{u}_1 + i\mathbf{v}_1)\\ &= ((ac-bd)+i(ad+bc))(\mathbf{u}_1+i\mathbf{v}_1)\\ &= (ac\mathbf{u}_1 - ad\mathbf{v}_1 - bc\mathbf{v_1} - bd\mathbf{u}_1) + i(ac\mathbf{v}_1 + ad\mathbf{u}_1 + bc\mathbf{u}_1 - bdv_1)\end{aligned}$$

$$\begin{aligned}\lambda(\gamma\mathbf{x}) &= (a+bi)((c+di)(\mathbf{u}_1+i\mathbf{v}_1))\\ &= (a+bi)(c\mathbf{u}_1 + ic\mathbf{v}_1 + id\mathbf{u}_1 - d\mathbf{v}_1)\\ &= (ac\mathbf{u}_1 - ad\mathbf{v}_1 - bc\mathbf{v_1} - bd\mathbf{u}_1) + i(ac\mathbf{v}_1 + ad\mathbf{u}_1 + bc\mathbf{u}_1 - bdv_1)\end{aligned}$$

So...

$$(\lambda\gamma)\mathbf{x} = \lambda(\gamma\mathbf{x})$$

- **Additive Identity:** Assume that $\mathbf{0}$ is the additive identity of $\mathcal{V}$. We define $\mathbf{z} \in \mathcal{V}_C$ as $(\mathbf{0}, \mathbf{0})$, or:

  $$\mathbf{z} = \mathbf{0} + i\mathbf{0}$$

  We claim that $\mathbf{z}$ is the additive identity of $\mathcal{V}_C$. We can show this as such:

  $$\mathbf{x} + \mathbf{z} = (\mathbf{u}_1 + i\mathbf{v}_1) + (\mathbf{0} + i\mathbf{0}) = (\mathbf{u}_1 + \mathbf{0}) + i(\mathbf{v}_1 + \mathbf{0}) = \mathbf{u}_1 + i\mathbf{v}_1 = \mathbf{x}$$

- **Additive Inverse:** Imagine the additive inverses of $\mathbf{u}_1, \mathbf{v}_1 \in \mathcal{V}$ are $\mathbf{w}, \mathbf{t} \in \mathcal{V}$ respectively. Then, the additive inverse of any $\mathbf{x} \in \mathcal{V}_C$ is just $-\mathbf{x} = \mathbf{w} + i\mathbf{t}$. We can show this:

  $$\mathbf{x} + (-\mathbf{x}) = (\mathbf{u}_1 + i\mathbf{v}_1) + (\mathbf{w} + i\mathbf{t}) = (\mathbf{u}_1 + \mathbf{w}) + i(\mathbf{v}_1 + \mathbf{t}) = \mathbf{0} + i\mathbf{0} = \mathbf{z}$$

- **Multiplicative Identity:** Imagine that the multiplicative identity of $\mathcal{V}$ is $\mathbf{1}$, and the additive identity of $\mathcal{V}$ is $\mathbf{0}$. We can define $\mathbf{m}$ as:

  $$\mathbf{m} = \mathbf{1} + i\mathbf{0}$$

  We claim that $\mathbf{m}$ is the multiplicative identity, and we may verify this by:

  $$\mathbf{m}\mathbf{x} = (\mathbf{1} + i\mathbf{0})(\mathbf{u}_1 + i\mathbf{v}_1) = (\mathbf{u}_1 + i\mathbf{v}_1) + (\mathbf{0} + i\mathbf{0}) = \mathbf{u}_1 + i\mathbf{v}_1 = \mathbf{x}$$

- **Distributive properties:**

$$\begin{aligned}\lambda(\mathbf{x}+\mathbf{y}) &= (a+bi)((\mathbf{u}_1 + i\mathbf{v_1}) + (\mathbf{u}_2 + i\mathbf{v}_2))\\ &= (a+bi)((\mathbf{u}_1+\mathbf{u}_2) + i(\mathbf{v}_1+\mathbf{v}_2))\\ &= (a(\mathbf{u}_1+\mathbf{u}_2) - b(\mathbf{v}_1+\mathbf{v}_2)) + i(b(\mathbf{u}_1+\mathbf{u}_2) + a(\mathbf{v}_1+\mathbf{v}_2))\\ &= (a\mathbf{u_1} + ia\mathbf{v}_1 + ib\mathbf{u}_1 - b\mathbf{v}_1) + (a\mathbf{u_2} + ia\mathbf{v}_2 + ib\mathbf{u}_2 - b\mathbf{v}_2)\\ &= (a+bi)(\mathbf{u}_1 + i\mathbf{v}_1) + (a+bi)(\mathbf{u}_2 + i\mathbf{v}_2)\\ &= \lambda\mathbf{x} + \lambda\mathbf{y}\end{aligned}$$

$$\begin{aligned}(\lambda+\gamma)\mathbf{x} &= ((a+bi)+(c+di))(\mathbf{u}_1+i\mathbf{v_1}) \\ &= ((a+c)+i(b+d))(\mathbf{u}_1+i\mathbf{v}_1) \\ &= a\mathbf{u}_1+ia\mathbf{v_1}+c\mathbf{u}_1+ic\mathbf{v}_1+ib\mathbf{u}_1-b\mathbf{v}_1+id\mathbf{u}_1-d\mathbf{v}_1 \\ &= (a\mathbf{u}_1+ia\mathbf{v}_1+ib\mathbf{u}_1-b\mathbf{v_1})+(c\mathbf{u}_1+ic\mathbf{v}_1+id\mathbf{u_1}-d\mathbf{v_1}) \\ &= (a+bi)(\mathbf{u}_1+i\mathbf{v}_1)+(a+bi)(\mathbf{u}_1+i\mathbf{v}_1) \\ &= \lambda\mathbf{x}+\gamma\mathbf{x}\end{aligned}$$

Having satisfied all properties of a vector field, we have shown that $\mathcal{V}_C$ is a vector field. □

1.3 Subspaces

Definition 1.3.1: Subspace

A subset $\mathcal{U}$ of $\mathcal{V}$ is a **subspace** of $\mathcal{V}$ if $\mathcal{U}$ is also a vector space with the same additive identity, addition, and scalar multiplication as $\mathcal{V}$ does

However, the conditions for a subspace may be simplified to just three checks:

Theorem 1.3.2: Conditions for subspace

A subset $\mathcal{U}$ of $\mathcal{V}$ is a subspace if and only if $\mathcal{U}$ satisfies the following:

- **Additive identity:** $\mathbf{0} \in \mathcal{U}$
- **Additive Closure:** $\mathbf{u}, \mathbf{w} \in \mathcal{U} \implies \mathbf{u}+\mathbf{w} \in \mathcal{U}$
- **Closure under scalar multiplication:** For $a \in \mathbb{F}$ and $\mathbf{u} \in \mathcal{U}$ implies $a\mathbf{u} \in \mathcal{U}$

Proof. If $\mathcal{U}$ is a vector space itself, then the above checks are all trivial.

Conversely, the conditions for the subspaces already guarantees an additive identity. Furthermore, the second and third conditions ensures us the addition and scalar multiplication behave as expected on $\mathcal{U}$.

The third condition also guarantees that the additive inverse of any element of $\mathcal{U}$ also belongs in $\mathcal{U}$ as $\forall \mathbf{u} \in \mathcal{U}, (-1)\mathbf{u} \in \mathcal{U}$. Which satisfies the additive inverse property of a vector field.

Because $\mathcal{U} \in \mathcal{V}$, we know that all the other properties of a vector field, the commutativity, associativity, and distributive properties, are all satisfied on $\mathcal{U}$ because they hold for $\mathcal{V}$. With this, all properties now hold for $\mathcal{U}$ to be a vector space, and hence a subspace of $\mathcal{V}$. □

Corollary 1.3.3

The additive identity condition for a subspace may be substituted by showing that the subspace is not empty. These two statements are the same thing as we can take any element of a non-empty set and multiply it by 0 to get the additive inverse.

Subspaces allow us to illustrate the pattern in linearity in other branches of mathematics. For instance, consider the following examples:

Example.

(a) The set of continuous real-valued functions on the interval $[0,1]$ is a subspace of $\mathbb{R}^{[0,1]}$.

(b) The set of differentiable real-valued functions on $\mathbb{R}$ is a subspace of $\mathbb{R}^{\mathbb{R}}$.

(c) The set of differentiable real-valued functions f on the interval $(0,3)$ such that $f'(2) = b$ is a subspace of $\mathbb{R}^{(0,3)}$ if and only if $b = 0$.

(d) The set of all sequences of complex numbers with limit 0 is a subspace of $\mathbb{C}^{\infty}$

Example (a) relies on the fact that the sum of continuous real-valued functions is continuous, and the scalar multiple of a continuous function is continuous.

Example (b) requires that the sum and scalar multiplies of differentiable functions are differentiable.

Example (c) requires the fact that the derivative of a function cf where f is a differentiable function and c is a scalar is equal to cf'.

Examples (d) requires the fact that the sum of two convergent sequences produces a sequence whose limit is the sum of the limit of the two preceding sequences.

Additionally, there is an alternative view at the subspaces of $\mathbb{R}^n$, or **n-dimensional Euclidean Spaces**. The subspaces of $\mathbb{R}^2$ is either the set $\{0\}$, all lines in $\mathbb{R}^2$ containing the origin, or $\mathbb{R}$ itself. The same extends to $\mathbb{R}^3$, with the added subspaces that consists of all planes containing the origin, and $\mathbb{R}^3$ itself. In fact, these are the only subspaces of $\mathbb{R}^2$ and $\mathbb{R}^3$.

1.3.1 Sums of Subspaces

The union of two subspaces is rarely a subspace.

Definition 1.3.4: Sum of subspaces

Suppose $\mathcal{V}_1, \ldots, \mathcal{V}_m$ are subspaces of $\mathcal{V}$. The **sum** of $\mathcal{V}_1, \ldots, \mathcal{V}_m$, denoted by $\mathcal{V}_1 + \cdots + \mathcal{V}_m$ is the set of all possible sums of elements of $\mathcal{V}_1, \ldots, \mathcal{V}_m$.

$$\mathcal{V}_1 + \cdots + \mathcal{V}_m = \{v_1 + \cdots + v_m : v_1 \in \mathcal{V}_1, \ldots, v_m \in \mathcal{V}_m\}$$

Example.

A sum of subspaces of $\mathbb{F}^3$: Let $\mathcal{U}$ be the set of all elements of $\mathbb{F}^3$ whose second and third coordinates are 0. Let $\mathcal{W}$ be the set of all elements of $\mathbb{F}^3$ whose first and third coordinates equal 0. It should be clear that both $\mathcal{U}$ and $\mathcal{W}$ are subspaces. Then the sum of $\mathcal{U}$ and $\mathcal{W}$ is:

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

Theorem 1.3.5: Sum of subspaces is the smallest subspace containing all summands

Suppose $\mathcal{V}_1, \ldots, \mathcal{V}_m$ are subspaces of $\mathcal{V}$. Then $\mathcal{V}_1 + \cdots + \mathcal{V}_m$ is the smallest subspace of $\mathcal{V}$ containing $\mathcal{V}_1, \ldots, \mathcal{V}_m$

Proof. It is clear that the additive identity $\mathbf{0}$ is included in $\mathcal{V}_1 + \cdots + \mathcal{V}_m$ as each one of $\mathcal{V}_1, \ldots, \mathcal{V}_m$ contains $\mathbf{0}$. Furthermore, $\mathcal{V}_1 + \cdots + \mathcal{V}_m$ is closed under addition and scalar multiplication because any element in $\mathcal{V}_1 + \cdots + \mathcal{V}_m$ can be expressed as a sum of elements belonging to $\mathcal{V}_1, \ldots, \mathcal{V}_m$. Thus, the sum of any two elements of $\mathcal{V}_1 + \cdots + \mathcal{V}_m$ also is a sum of $\mathcal{V}_1, \ldots, \mathcal{V}_m$. Additionally, any scalar multiple of an element of $\mathcal{V}_1 + \cdots + \mathcal{V}_m$ can be written as a sum of scalar multiples of elements of $\mathcal{V}_1, \ldots, \mathcal{V}_m$. Thus, $\mathcal{V}_1 + \cdots + \mathcal{V}_m$ is a subspace of $\mathcal{V}$.

To show that $\mathcal{V}_1 + \cdots + \mathcal{V}_m$ is the smallest subspace of $\mathcal{V}$ containing all of $\mathcal{V}_1, \ldots, \mathcal{V}_m$, we see that $\mathcal{V}_1 + \cdots + \mathcal{V}_m$ contains all of $\mathcal{V}_1, \ldots, \mathcal{V}_m$, as every element from $\mathcal{V}_1, \ldots, \mathcal{V}_m$ may be expressed as the sum of itself and zero vectors. Conversely, we see that every subspace of V that contains $\mathcal{V}_1, \ldots, \mathcal{V}_m$ must necessarily contain $\mathcal{V}_1 + \cdots + \mathcal{V}_m$, as these subspaces must contain all finite sum of all elements within the space. Hence, we have proven that $\mathcal{V}_1 + \cdots + \mathcal{V}_m$ is the smallest subspace of $\mathcal{V}$ containing $\mathcal{V}_1 \ldots, \mathcal{V}_m$. □ □

Sums of subspaces are analogous to unions of sets. The smallest subspace that contains all subspaces of a set is the sum of the subspaces. The smallest set that contains all sets in a set is the union of the sets.

1.3.2 Direct Sums

A special case of sums of subspaces is the **direct sum**.

Definition 1.3.6: Direct sum

Suppose $\mathcal{V}_1, \ldots, \mathcal{V}_m$ are subspaces of $\mathcal{V}$.

- The sum $\mathcal{V}_1 + \cdots + \mathcal{V}_m$ is a **direct sum** if each element of $\mathcal{V}_1 + \cdots + \mathcal{V}_m$ can be written in only one way as a sum of the elements of $\mathcal{V}_1, \ldots, \mathcal{V}_m$.
- If $\mathcal{V}_1 + \cdots + \mathcal{V}_m$ is a direct sum, then we denote it by $\mathcal{V}_1 \oplus \cdots \oplus \mathcal{V}_m$.

Example.

> Suppose $\mathcal{V}_k$ is the subspace of $\mathbb{F}^n$ whose vector's coordinates are all zero except in the kth slot. Then $\mathbb{F}^n = \mathcal{V}_1 \oplus \cdots \oplus \mathcal{V}_n$.

Theorem 1.3.7: Condition for a Direct Sum

Suppose $\mathcal{V}_1, \ldots, \mathcal{V}_m$ are subspaces of $\mathcal{V}$. Then $\mathcal{V}_1 + \cdots + \mathcal{V}_m$ is a direct sum if and only if the only way to write 0 as a sum of elements of $\mathcal{V}_1, \ldots, \mathcal{V}_m$ is by taking all the elements to be 0.

Proof. Suppose that $\mathcal{V}_1 + \cdots + \mathcal{V}_m$ is a direct sum. Then, by definition, the only way to write 0 as a sum of elements of $\mathcal{V}_1, \ldots, \mathcal{V}_m$ is by taking all the elements to be 0.

Now suppose that the only way to write 0 as a sum of elements is $\mathbf{v}_1 + \cdots + \mathbf{v}_m$ where each $\mathbf{v}_k = \mathbf{0}$. Assume for the sake of contradiction that $\mathcal{V}_1 + \cdots + \mathcal{V}_m$ is not a direct sum. Then there exists some $\mathbf{u} \in \mathcal{V}_1 + \cdots + \mathcal{V}_m$ such that $\mathbf{u} = \mathbf{u}_1 + \cdots + \mathbf{u}_m = \mathbf{u}_1^* + \cdots + \mathbf{u}_m^*$ where $\mathbf{u}_k, \mathbf{u}_k^* \in \mathcal{V}_k$ and $\exists k$ such that $\mathbf{u}_k \neq \mathbf{u}_k^*$. Then:

$$\mathbf{0} = \mathbf{u} - \mathbf{u} = (\mathbf{u}_1 - \mathbf{u}_1^*) + \cdots + (\mathbf{u}_m - \mathbf{u}_m^*)$$

By construction, we know that at least one of $(\mathbf{u}_k - \mathbf{u}_k^x) \neq 0$. This is a contradiction, and hence $\mathcal{V}_1 + \cdots + \mathcal{V}_m$ is a direct sum. □

There is another way of testing if two subspaces are direct sums:

Theorem 1.3.8: Direct sum of two subspaces

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

Proof. Suppose that $\mathcal{U} + \mathcal{W}$ is a direct sum. Then, by theorem 1.3.7, the only way to write 0 as a sum of elements of $\mathcal{U}$ and $\mathcal{W}$ is by taking both elements to be 0. However, if there exists some $\mathbf{v} \neq 0 \in \mathcal{U} \cap \mathcal{W}$, we also know that $-\mathbf{v} \in \mathcal{U} \cap \mathcal{W}$ then we can write:

$$\mathbf{0} = \mathbf{v} + (-\mathbf{v})$$

This means that there are now two ways to express 0 as a sum of elements of $\mathcal{U}$ and $\mathcal{W}$. Thus, $\mathcal{U} + \mathcal{W}$ cannot possibly be a direct sum.

For the reverse direction, we assume $\mathcal{U} \cap \mathcal{W} = \{\mathbf{0}\}$. To prove that $\mathcal{U} + \mathcal{W}$ is a direct sum, let $\mathbf{u} \in \mathcal{U}$ and $\mathbf{w} \in \mathcal{W}$, and:

$$0 = \mathbf{u} + \mathbf{w}$$

By theorem 1.3.7, we just need to show that $\mathbf{u} = \mathbf{w} = 0$. The equation above implies that $\mathbf{u}$ and $\mathbf{w}$ are additive inverses of one another. Hence, $\mathbf{u} = -\mathbf{w} \in \mathcal{W}$. Since $\mathbf{u} \in \mathcal{U} \cap \mathcal{W}$, $\mathbf{u} = 0$, and $\mathbf{w} = 0$ also. Hence, the only way to express $\mathbf{0}$ as a sum of two vectors from $\mathcal{U}$ and $\mathcal{W}$ is by summing the two zero vectors, meaning $\mathcal{U} + \mathcal{W}$ is a direct sum. □

Intuitively, sums of subspaces are analogous to unions of subsets. So direct sums of subspaces are analogous to disjoint unions of subsets. However, no two

Chapter 2

Finite-Dimensional Vector Spaces

2.1 Span and Linear Independence

When talking about combiniations and lists of vectors, the notation we use is comma-separated n-tuples without wrapping parenthesis. For example: $(1, 2, 3), (4, 5, 6), (7, 8, 9)$ is a list of 3 3-tuples.

2.1.1 Linear Combinations and Span

A **linear combination** of a list vectors is a sum of scalar multiples of the vectors in the list.

Definition 2.1.1: Linear Combination

A linear combination of vectors $v_1, \ldots, v_n$ in a vector space $\mathcal{V}$ is another vector of the form:

$$\lambda_1 v_1 + \cdots + \lambda_n v_n$$

Where $\lambda_1, \ldots, \lambda_n \in \mathbb{F}$

Oftentimes, systems of linear equations are the same problem as finding a set of coefficients that allow for some given list of vectors to be scalar-multiplied and summed to some resulting vector.

We may also represent the set of all possible linear combinations of a list of vectors in a vector space as a **span**.

Definition 2.1.2: span

The set of all possible linear combininations of a list of vectors $v_1, \ldots, v_n \in \mathcal{V}$ is denoted as the **span** of those vectors, often written as $\text{span}(v_1, \ldots, v_n)$. Formally,

this may be expressed as:

$$\text{span}(v_1, \ldots, v_n) = \{\lambda_1 v_1 + \cdots + \lambda_n v_n \mid \lambda_1, \ldots, \lambda_n \in \mathbb{F}\}$$

The span of the empty list ( ) is defined to be $\{\mathbf{0}\}$

Theorem 2.1.3: Span is the smallest containing subspace

The span of a list of vectors in $\mathcal{V}$ is the smallest subspace of $\mathcal{V}$ containing all vectors in the list.

Proof. We first show that a span of vectors is a subspace. For starters, the span of any list of vectors always contains the zero element, as we may multiply every vector by the zero element and sum to get the zero element. The span of an empty list is also defined to be the set containing only the zero element. Furthermore, $\text{span}(v_1, \ldots, v_n)$ is closed under addition. For any two vectors $u_1, u_2 \in \text{span}(v_1, \ldots, v_n)$:

$$\begin{aligned} u_1 + u_2 &= (\lambda_1 v_1 + \cdots + \lambda_n v_n) + (\gamma_1 v_1 + \cdots + \gamma_n v_n) \\ &= (\lambda_1 + \gamma_1) v_1 + \cdots + (\lambda_n + \gamma_n) v_n \end{aligned}$$

Since each sum $\lambda_i + \gamma_i \in \mathbb{F}$ we know that $u_1 + u_2 \in \text{span}(v_1, \ldots, v_n)$ as it is also a linear combination of vectors $v_1, \ldots, v_n$. Similarly, $\text{span}(v_1, \ldots, v_n)$ is closed under scalar multiplication as:

$$\gamma(\lambda_1 v_1 + \cdots + \lambda_n v_n) = \gamma\lambda_1 v_1 + \cdots + \gamma\lambda_n v_n$$

Since every $\gamma\lambda_i \in \mathbb{F}$, we know that γu_1 is also a linear combination, and thus belongs in $\text{span}(v_1, \ldots, v_n)$.

Now we have to prove that the span is the smallest such subspace that contains all vectors in the list. First of all, for any potential subspace $\mathcal{U}$ that contains all vectors $v_1, \ldots, v_n$, we know that through closure under addition and scalar multiplication, we may freely multiply any vector by a scalar and stil get a vector present in $\mathcal{U}$, and we may freely add any vectors together and still be contained within $\mathcal{U}$. Hence, any linear combination of vectors $v_1, \ldots, v_n$ must also belong in $\mathcal{U}$, so $\text{span}(v_1, \ldots, v_n) \in \mathcal{U}$, thus, any subspace of $\mathcal{V}$ that contains $v_1, \ldots, v_n$ must also contain the span of these vectors, meaning the span is the smallest such subspace. □

Definition 2.1.4: Spans

If $\text{span}(v_1, \ldots, v_n)$ equals $\mathcal{V}$, then the list $v_1, \ldots, v_n$ **spans** $\mathcal{V}$.

Example.

The list of all canonical n-tuples of $\mathbb{F}^n$ spans $\mathbb{F}^n$. For $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{F}^n$:

$$(x_1, \ldots, x_n) = x_1(1, 0 \ldots, 0) + x_2(0, 1 \ldots, 0) + \cdots + x_n(0, 0 \ldots, 1)$$

Definition 2.1.5: Finite-dimensional vector space

A vector space is **finite-dimensional** if some list of vectors in it spans the space. By definition, every list has finite length.

Definition 2.1.6: Polynomial, $\mathcal{P}(\mathbb{F})$

- A function $p : \mathbb{F} \to \mathbb{F}$ is a polynomial with coefficients in $\mathbb{F}$ if there exists $a_0, \ldots, a_m \in \mathbb{F}$ such that:

  $$p(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_m z^m$$

  for all $z \in \mathbb{F}$

- $\mathcal{P}(\mathbb{F})$ is the set of all polynomials with coefficients in $\mathbb{F}$. $\mathcal{P}(\mathbb{F}) = \bigcup_m \mathcal{P}_m(\mathbb{F})$

It's easy to see that $\mathcal{P}(\mathbb{F})$ is a vector space over $\mathbb{F}$, there is a zero polynomial, the sum of two polynomials is a polynomial, and also the scalar multiple of a polynomial is still a polynomial. Thus, $\mathcal{P}(\mathbb{F})$ is a subspace of $\mathbb{F}^{\mathbb{F}}$, which is the vector space of functions from $\mathbb{F}$ to $\mathbb{F}$. Moreover, the coefficients of a polynomial are unique determined by the polynomial, no two different polynomials have the same coefficients.

Definition 2.1.7: Degree of polynomial $\deg p$ and $\mathcal{P}_m(\mathbb{F})$

- A polynomial $p \in \mathcal{P}(\mathbb{F})$ has degree m if there exists scalars $a_0, a_1, \ldots, a_m$ such that for every $z \in \mathcal{F}$, we have:

  $$p(z) = a_0 + a_1 z + \cdots + a_m z^m$$

- The polynomial that is identically 0 has degree $-\infty$
- The degree of a polynomial p is denoted by $\deg p$.
- For a non-negative integer m, $\mathcal{P}_m(\mathbb{F})$ denotes the set of all polynomials with coefficients in $\mathbb{F}$ and degree at most m.

If m is a nonnegative integer, then $\mathcal{P}_m(\mathbb{F}) = \text{span}(1, z, \ldots, z^m)$. Hence $\mathcal{P}_m(\mathbb{F})$ is a finite-dimensional vector space for each non-negative integer m.

Definition 2.1.8: Infinite-dimensional vector space

A vector space is **infinite-dimensional** if it is not finite-dimensional.

Example.

> $\mathcal{P}(\mathbb{F})$ is infinite dimensional. No list spans $\mathcal{P}(\mathbb{F})$, because we can always construct a polynomial that is degree 1 larger than the max degree in any list.

2.1.2 Linear Independence

Let $v_1, \ldots, v_n \in V$ and let $v \in \text{span}(v_1, \ldots, v_n)$. We know that there exists scalars $\lambda_1, \ldots, \lambda_n$ such that

$$v = \lambda_1 v_1 + \cdots + \lambda_n v_n$$

An interesting question that we can ask is if the set of scalars that satisfy the equation are unique. If the set of scalars are not unique, then there exists anot set of scalars $\mu_1, \ldots, \mu_n$ such that at least one such $\mu_i \neq \lambda_i$ and:

$$v = \mu_1 v_1 + \cdots + \mu_n v_n$$

Then we see that:

$$v - v = 0 = (\lambda_1 - \mu_1)v_1 + \cdots + (\lambda_n - \mu_n)v_n$$

Hence, we see that as long as the only way to write 0 as a linear combination of v_n is by setting all scalars to 0, then all vectors have a unique representation in terms of $v_1, \ldots, v_n$. Alternatively, if there was a non-zero combination that yielded 0, we may add that combination onto any linear combination and still get the same vector, giving us multiple representations.

Finally, in the context of linear maps and matrices, this is identical to saying that the null space of $v_1, \ldots, v_n$ is just the zero vector (if we were to arrange them as columns in a matrix). This brings us to the important concept of the linear independence of a list of vectors.

Definition 2.1.9: Linearly independent

- A list of vectors $v_1, \ldots, v_n$ in $\mathcal{V}$ is linear independent if the only choice of $\lambda_1, \ldots, \lambda_n \in \mathbb{F}$ that makes

  $$\lambda_1 v_1 + \cdots + \lambda_n v_n = 0$$

  is $a_1 = \cdots = a_m = 0$.
- The empty list ( ) is also linearly independent.

A list of length 1 is a linearly indepndent list if and only if the vector is not 0.

A list of length 2 is a linearly independent list if and only if the two vectors are not scalar multiples of each other.

If vectors are removed from a linearly independent list, the remaining list is also linearly independent.

Proof. Let $v_1, \ldots, v_n$ be a list of vectors of arbitrary length n that is linearly independent. Suppose we remove v_n and have a list $v_1, \ldots, v_{n-1}$. Assume for the sake of contradiction that the remaining list is not linearly independent. Then there exists scalars $\lambda_1, \ldots, \lambda_{n-1}$ that are all 0 such that

$$\lambda_1 v_1 + \cdots + \lambda_{n-1} v_{n-1} = 0$$

This is a contradiction as it implies $v_1, \ldots, v_n$ is linearly dependent, as we may easily form a non-trivial combination that results in 0 by setting λ_n to 0. □

2.1.3 Linear Dependence

Definition 2.1.10: Linearly Dependent

- A list of vectors in V is called **linearly dependent** if it is not linearly independent.
- A linearly dependent list has a non-trivial linear combination that results in 0.

Lemma 2.1.11:

If some vector in a list of vectors in V is a linear combination of the other vectors, then the list of vectors is linearly dependent.

Proof. Let $v_1, \ldots, v_n$ be a list of n vectors. Suppose:

$$v_k = \lambda_1 v_1 + \ldots \lambda_{k-1} v_{k-1} + \lambda_{k+1} v_{k+1} + \cdots + \lambda_n v_n$$

Then:

$$(\lambda_1 v_1 + \ldots \lambda_{k-1} v_{k-1} + \lambda_{k+1} v_{k+1} + \cdots + \lambda_n v_n) - v_k = 0$$

□

Lemma 2.1.12: Linear Dependence Lemma

Suppose that $v_1, \ldots, v_m$ are linearly independent in V. Then there exists $k \in \{1, \ldots, m\}$ such that:

$$k \in \text{span}(v_1, \ldots, v_{k-1})$$

If k satisfies the condition above and v_k is removed from the list, then the span of the remaining list equals $\text{span}(v_1, \ldots, v_m)$.

Proof. Linear dependence $\implies \exists \lambda_1, \ldots, \lambda_n$ not all zero such that $\lambda_1 v_1 + \cdots + \lambda_n v_n = 0$. Pick k to be the largest index where $\lambda_k \neq 0$. We see that:

$$v_k = \frac{1}{\lambda_k}(-\lambda_1 v_1 + \cdots + -\lambda_{k-1} v_{k-1})$$

Hence, $v_k \in \text{span}(v_1, \ldots, v_{k-1})$. To show the the span of the list with v_k removed is the same as the span of the original list, we simply substitute v_k back into the original expression just in terms of $v_1, \ldots, v_{k-1}$ We see that nothing about the span changed at all, except now v_k is not in the span.

Note, if $k = 1$, then $v_k \in \text{span}(v_1, \ldots, v_n)$ implies that $v_1 = 0$ as the only element in the empty span is 0. □

Theorem 2.1.13: Length of linearly independent list $\leq$ length of spanning list

In a finite-dimensional vector space, the length of every linearly independent list of vectors is less than or equal to the length of every spanning list of vectors.

Proof. Let $v_1, \ldots, v_n$ be a list of vectors in V and $u_1, \ldots, u_m$ be a spanning list of V. We want to prove that $n \leq m$. We do this by incrementally exchanging each u_i with a v_i until we have no v_is left to exchange.

Step 1: We take v_1 and append it to the start of the spanning list $u_1, \ldots, u_m$. We note that $v_1, u_1, \ldots, u_m$ is linearly dependent because $v_1 \in \text{span}(u_1, \ldots, u_m)$. We also note that by the linear dependence lemma, there exists some vector that may written as a linear combination of the previous vectors, and can be removed from $v_1, u_1, \ldots, u_m$ without changing the span of the list. This vector cannot be v_1, as v_1 cannot be written as a linear combination of the 0 vector. Hence, one of the u_i vectors must be removed. And so our new list of length m still spans V.

Steps k for k = 2,. . . ,n: At step k, we have the list:

$$v_1, \ldots, v_{k-1}, u_{?_1}, \ldots, u_{?_{m-k+1}}$$

By adding v_k to the list, we see that the list becomes linearly dependent, so once again by the linear dependence lemma, there exists a vector that can be written as a linear combination of the previous vectors that we may remove from the list without changing the vector space spanned by the list. We see that this removable vectors cannot be any of the v_is because they are all linearly independent and cannot be written as a linear combination of each other. Thus, it must mean that there still exists a w vector from the original spanning list that can be removed.

By repeating these steps, right after we add v_n to the list, we see that we have removed at least n spanning list vectors. Hence m is at least n, so a linearly independent list is always smaller than or equal to in length to a spanning list. □

The implications of the theorem tells us that if we are ever presented with a list of vectors that are longer than a known spanning list of a vector space, we know that the list must be linearly dependent. Likewise, if we know a list of a certain size n is linearly independent in V, then the spanning list must be larger than or equal to the linearly independent list in length, otherwise it wouldn't span.

Theorem 2.1.14: Finite-dimensional Subspace

Every subspace of a finite-dimensional vector space is finite-dimensional

Proof. Let V be a finite dimensional vector space. Suppose that V is spanned by the list $v_1, \ldots, v_n$. Furthermore, suppose that $U \subseteq V$ is a subspace of V. Since every vector

$u \in U \implies u \in V$, every vector u may be written as a linear combination of $v_1, \ldots, v_n$. Hence, $v_1, \ldots, v_n$ spans U, and since the spanning list is of finite-length, U is also finite-dimensional. □

2.2 Bases

A basis of a vector space V is a list of vectors that span V and are linearly independent.

Definition 2.2.1: Basis

A *basis* of V is a list of vectors in V that is linearly independent and spans V.

Theorem 2.2.2: Criterion for Basis

A list $v_1, \ldots, v_n$ of vectors in V is a basis of V if and only if every $v \in V$ can be written uniquely in the form

$$v = \lambda_1 v_1 + \cdots \lambda_n v_n$$

For scalars $\lambda_1, \ldots, \lambda_n \in \mathbb{F}$

Proof. **Forward Direction:** Let $v_1, \ldots, v_n$ be a basis of V. Then we know that $v_1, \ldots, v_n$ is a spanning list and linearly independent. $v_1, \ldots, v_n$ being a spanning list means that every vector in V may be represented as a linear combination of $v_1, \ldots, v_n$. $v_1, \ldots, v_n$ being linearly independent means that every vector in $\text{span}(v_1, \ldots, v_n)$ has a unique representation as a linear combination of $v_1, \ldots, v_n$. (This is because otherwise, the list must be linearly dependent as there is a non-trivial linear combination that equals 0). Hence $\forall v \in V$, v can be uniquely written as a linear combination of $v_1, \ldots, v_n$.

Reverse Direction: If every $v \in V$ can be written unique in the form of $v = \lambda_1 v_1 + \cdots \lambda_n v_n$, then that means the list $v_1, \ldots, v_n$ spans V. Additionally, it also implies that 0 may be uniquely written in the form of $v = \lambda_1 v_1 + \cdots \lambda_n v_n$, since $0 = 0v_1 + \cdots + 0v_n$, that means the only way to express 0 as a linear combination of $v_1 \ldots, v_n$ is with a trivial combination, implying $v_1, \ldots, v_n$ is linearly independent. Hence $v_1, \ldots, v_n$ is a basis. □

A spanning list is not necessarily a basis, this is because the list may be linearly dependent. However, every spanning list contains a basis, which means we may keep removing vectors from the list until we have such a basis.

Theorem 2.2.3: Every Spanning List Contains a Basis

Every spanning list in a vector sapce can be reduced to a basis of the vector space.

Proof. We may prove this algorithm by procedurally showing how vectors may be removed from some spanning list. Let $B = v_1, \ldots, v_n$ be a spanning list of vector space V. To remove vectors from B such that the remaining vectors forms a basis of V, we may proceed with the steps below:

Step 1: If $v_1 = 0$. remove. v_1 from the B as $0\,\text{span}(\quad)$. Otherwise, do nothing.

step k: If $v_k \in \text{span}(v_1, \ldots, v_{k-1})$, then remove v_k from B. Othewise, do nothing.

The resulting list B after n steps is a basis as it fully spans V because we only removed vectors that were in the of previous vectors. For the same reason, B is also linearly independent, so B is now a basis of V. □

Corollary 2.2.4

Basis of Finite-Dimensional Vector Space

Every finite-dimensional vector space has a basis.

Proof. By definition, there exists some list B that spans any finite-dimensional vector space V. Hence, B is a spanning list that may or may not be a basis already. If B is not a basis, then it may be reduced to a basis of V by the previous theorem. □

A dual theorem to Theorem 2.2.3 is the fact that any linearly independent list may be extended to a basis of V.

Theorem 2.2.5: Every Linearly Independent List Extends to a Basis

Every linearly independent list of vectors in a finite-dimensional vector space can be extended to a basis of the vector space.

Proof. Let $u_1, \ldots, u_m$ be a linearly independent list in a finite-dimensional vector space V. Let $w_1, \ldots, w_n$ be a list of vectors in V that spans V. The following list must span V:

$$u_1, \ldots, u_m, w_1, \ldots, w_n$$

However, the list is not necessarily linearly independent. If the list is linearly dependent, then we may reduce this list to a basis containing $u_1, \ldots, u_m$ (As all us are linearly independent with each other, none get removed), and some of the w_is by removing the ones that are in the span of previous vector. □

The previous result is highly useful to prove the concept that every subspace of a finite-dimensional vector space can be paired with a complement subspace to form a direct sum of the entire space.

Theorem 2.2.6: Every Subspace of V is Part of a Direct Sum Equal to V

Suppose V is finite-dimensional and U is a subspace of V. Then there is a subspace W of V such that $V = U \oplus W$.

Proof. Since V is finite-dimensional, U must be finite dimensional too. Hence, there is a list of vectors $w_1, \ldots, w_n$ that spans V and and a list of vectors $u_1 \ldots, u_m$ that is the basis of U. We can concatenate the two lists together to form another spanning list of V: $u_1, \ldots, u_m, w_1, \ldots, w_n$. We show that this list of vectors may be reduced to be a basis of V, containg all $u_1, \ldots, u_m$ as they are linearly independent, and some remaining $v_i s$. $u_1, \ldots, u_m$ is still the basis of U while we claim the remaining w_is (call them $w_1, \ldots, w_k$) form the basis of the complement subspace W.

By our construction, it's obvious that $U + W = V$, since the concatenation of their basis vectors forms a basis of V. For any $v \in V$:

$$v = \lambda_1 u_1 + \cdots + \lambda_n u_n + \mu_1 w_1 + \cdots + \mu_k w_k$$

We see that $\lambda_1 u_1 + \cdots + \lambda_n u_n \in U$, and $\mu_1 w_1 + \cdots + \mu_k w_k \in W$, so any vector from V can be written as a sum of vectors from U and W.

To show that $U + V$ is a direct sum, we just need to show $U \cap V = \{0\}$. 0 belongs to both spaces already. To show that no other vectors belong on this intersection, suppose that some vector $v \in U \cap W$. Then there exists scalars $\lambda_1, \ldots, \lambda_n; \mu_1, \ldots, \mu_k \in \mathbb{F}$ such that:

$$v = \lambda_1 u_1 + \cdots + \lambda_n u_n = \mu_1 w_1 + \cdots + \mu_k w_k$$

Heence:

$$\lambda_1 u_1 + \cdots + \lambda_n u_n - \mu_1 w_1 - \cdots - \mu_k w_k = 0$$

Since the vectors $u_1, \ldots, u_n, w_1, \ldots, w_k$ are linearly independent, all scalars must be 0, so any $v \in U \cap W$ must be the 0 vector. □

2.3 Dimension

We may formally define the dimension of a space to be the length of the basis of that space. However, for this definition to make sense, all basis must have the same length.

Theorem 2.3.1: Every Basis of a Vector Space is the same length

Any two bases of a finite-dimensional vector space have the same length.

Proof. Take two bases of vector space V: A and B. Say A has length m and B has length n. Both A and B are linearly independent, and are spanning lists. Hence, $n \leq m$ and $m \leq n$ implying that $n = m$. □

Definition 2.3.2: Dimension $\dim V$

- The *dimension* of a finite-dimensional vector space is the length of any basis of the vector space.

- The dimension of a finite-dimensional vector space V is denoted by $\dim V$.

Theorem 2.3.3: Dimension of a subspace

If V is finite-dimensional and U is a subspace of V, then $\dim U \leq \dim V$.

Proof. Any basis A of U is a linearly independent list in V. Any basis B of V is a spanning list of V. Hence the length of $A \leq B$, so $\dim U \leq \dim V$. □

A basis of V is defined as a list that spans and is linearly independent. We can show that any linearly independent list spans V as long as it has length $\dim V$.

Theorem 2.3.4: Linearly Independent List of the Right Length is a Basis

Suppose V is finite-dimensional. Then every linearly independent list of vectors in V of length $\dim V$ is a basis of V.

Proof. Suppose $\dim V = n$ and $v_1, \ldots, v_n$ is linearly independent in V. The list $v_1, \ldots, v_n$ can be extended to a basis of V. However, since $n = \dim V$, the extension is trivial. This means any vectors we extend to $v_1, \ldots, v_n$ can be written as a linearly combination of previous vectors, meaning that $v_1, \ldots, v_n$ is already a spanning list. □

We draw an important corollary from this:

Corollary 2.3.5

Suppose that V is finite-dimensional and U is a subspace of V such that $\dim U = \dim V$. Then $U = V$.

Proof. Let $u_1, \ldots, u_n$ be a basis of U. We see that $n = \dim U = \dim V$. Hence $u_1, \ldots, u_n$ is also a basis of V, as any vectors in V can be written as a linear combination of the basis of U. Hence $U = V$. □

Additionally, we can also say that any spanning list of the right length is also a basis:

Theorem 2.3.6: Spanning list of the right length is a basis

Suppose V is finite-dimensional. Then every spanning list of vectors in V of length $\dim V$ is a basis of V.

Proof. Let $\dim V = n$ and let $v_1, \ldots, v_n$ span V. We know that any spanning list of a vector space may be reduced down to a basis of V. However, since every basis has length n, the reduction is trivial, so $v_1, \ldots, v_n$ is a basis of V. □

2.3.1 Dimension of sums

We may derive a formula for the sum of two subspaces of a finite-dimensional vector space. The formula is analogous to the inclusion exclusion principle in counting and probability. Intuitively, the number of elements in the union of two sets is equal ot the number of elements in the first set plus the number of elements in the second set, and minus the number of elements in the intersection.

Theorem 2.3.7: Dimension of a sum

If V_1 and V_2 are subspaces of a finite-dimensional vector space, then:

$$\dim(V_1 + V_2) = \dim V_1 + \dim V_2 - \dim(V_1 \cap V_2)$$

Proof. Let $v_1, \ldots, v_m$ be a basis of $V_1 \cap V_2$, so $\dim(V_1 \cap V_2) = m$. Since $v_1, \ldots, v_m$ is a basis of $V_1 \cap V_2$, it is linearly independent in V_1. Hence, $v_1, \ldots, v_m$ may be extended to be a basis $v_1, \ldots, v_m; u_1, \ldots, u_j$ of V_1. Similarly, $v_1, \ldots, v_m$ may be extended to a basis $v_1, \ldots, v_m; w_1, \ldots, w_k$ of V_2. Hence:

$$\dim V_1 = m + j \quad \dim V_2 = m + k$$

We assert that the following list:

$$v_1, \ldots, v_m; u_1, \ldots, u_j; w_1 \ldots, w_k$$

is a basis of $V_1 + V_2$. To show this, we note that for any $v \in V_1 + V_2$, $v = u + w$ where $u \in V_1$ and $w \in V_2$. Additionally, we know there exists scalars $\lambda_1, \ldots \lambda_{m+j}$ and $\mu_1, \ldots, \mu_{m+k}$ such that:

$$u = \lambda_1 v_1 + \cdots + \lambda_m v_m + \lambda_{m+1} u_1 + \cdots + \lambda_{m+j} u_j$$

$$w = \mu_1 v_1 + \cdots + \mu_m v_m + \mu_{m+1} w_1 + \cdots + \mu_{m+k} w_k$$

Hence:

$$v = u+w = (\lambda_1+\mu_1)v_1+\cdots+(\lambda_m+\mu_m)v_m+\lambda_{m+1}u_1+\cdots+\lambda_{m+j}u_j+\mu_{m+1}w_1+\cdots+\mu_{m+k}w_k$$

Hence, $v_1, \ldots, v_m; u_1, \ldots, u_j; w_1 \ldots, w_k$ is a spanning list of $V_1 + V_2$. To show that this list is linearly independent, we need to show that if a linear combination of the vectors is 0, then all scalars must be 0. Suppose:

$$a_1 v_1 + \cdots + a_m v_m + b_1 u_1 + \cdots + b_m u_j + c_1 w_1 + \cdots + c_k w_k = 0$$

Where all a, b, c's are scalars. We must show that all these scalars are 0. We can show that equation may be rewritten as:

$$c_1 w_1 + \cdots + c_k w_k = -a_1 v_1 + \cdots - a_m v_m - b_1 u_1 - \cdots - b_m u_j$$

Since $v_1, \ldots v_m; u_1, \ldots, u_j$ is a basis of v_1, we see that $c_1 w_1 + \cdots + c_k w_k \in V_1$, so $c_1 w_1 +$

$\cdots + c_k w_k \in V_1 \cap V_2$. Since we know that $v_1, \ldots, v_m$ is a basis of $V_1 \cap V_2$, we know that:

$$c_1 w_1 + \cdots + c_k w_k = d_1 v_1 + \cdots + d_m v_m$$

for scalars c's and d's. However, we know that $v_1, \ldots, v_m, w_1, \ldots, w_k$ is linearly independent, so that implies that all the c's and d's are 0. Hence $c_1 w_1 + \cdots + c_k w_k = 0$, so we have:

$$-a_1 v_1 + \cdots - a_m v_m - b_1 u_1 - \cdots - b_m u_j = 0$$

Which is only possible when the a's and b's are all 0 as $v_1, \ldots, v_m, u_1, \ldots, u_j$ are linearly independent. Hence, all a's, b's, and c's are 0, which implies $v_1, \ldots, v_m; u_1, \ldots, u_j; w_1 \ldots, w_k$ is linearly independent. Hence, $v_1, \ldots, v_m; u_1, \ldots, u_j; w_1 \ldots, w_k$ is a basis of $V_1 + V_2$.
We also know that $\dim U + V = m + j + k$. Hence:

$$\begin{aligned} \dim V_1 + V_2 &= m + j + k \\ &= (m + j) + (m + j) - m \\ &= \dim V_1 + \dim V_2 - \dim(V_1 \cap V_2) \end{aligned}$$

□

Corollary 2.3.8

Direct sums preserve dimensionality. If $X \oplus Y = V$, then $\dim V = \dim X + \dim Y$

Proof. $X \oplus Y = V$, then $X \cap Y = \{0\}$, so $\dim(X \cap Y) = 0$. Hence:

$$\dim V = \dim(X + Y) = \dim X + \dim Y - \dim(X \cap Y) = \dim X + \dim Y$$

□

Proposition 2.3.9

Suppose that V and W are finite-dimensional vector spaces. Then $\dim V \times W = \dim V + \dim W$.

Proof. $V \times W$ is the vector space of all pairs of vectors from V and W. This can be expressed as:

$$V \times W = V \times \{0\} \oplus \{0\} \times W$$

We see that $\dim V \times \{0\} = \dim V$, and likewise, $\dim W \times \{0\} = \dim W$. So $\dim V \times W = \dim V + \dim W$. □

Chapter 3

Linear Maps

In this chapter, we will cover a powerful fundamental of linear maps, (also known as the **Rank-Nullity Theorem**), which states that the dimension of the domain of a linear map is equal to the dimension of the subspace that gets sent to 0 plus the dimension of the range.

Additionally, we explore how connected linear maps and matricies are, and how a linear map corresponds with a matrix. This matrix also corresponds with the basis of the domain space, and also a basis of the target space.

We denote U, V, and W as vector spaces.

3.1 Vector Spaces of Linear Maps

3.1.1 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:** $T(u + v) = Tu + Tv$ for all $u, v \in V$.
- **Homogeneity:** $T(\lambda v) = \lambda T(v)$ for all $\lambda \in \mathbb{F}$ and all $v \in V$.

For a linear map $T : V \to W$, we call the space V the domain, or source of T.

Definition 3.1.2: $\mathcal{L}(V, W)$, $\mathcal{L}(V)$

- The set of linear maps from V to W is denoted by $\mathcal{L}(V, W)$.
- The set of linear maps from V to V is denoted by $\mathcal{L}(V)$. In other words, $\mathcal{L}(V) = \mathcal{L}(V, V)$

In particular, we can define a **zero** linear map that takes every element of some vector space

to the additive identity of another (or the same) vector space. $0 \in \mathcal{L}(V, W)$ is defined as:

$$0(v) = 0$$

Additionally, the **identity operator** I is the linear map on a vector space that maps each element to itself. $I \in \mathcal{L}(V)$ is defined as:

$$I(v) = v$$

A lot of mathematical objects are linear in nature. For instance, the derivative and integral operations are linear. Moreover, the composition of two polynomials p and q from $\mathcal{P}(\mathbb{F})$ is also a linear transformation.

The following lemma proves that a linear map from one vector space to another solely depends on how it maps vectors in the basis (This is intuitively true because a linear map has all the linear properties that would ensure linear combinations of the basis vectors would be preserved under the map)

Lemma 3.1.3: Linear map lemma

Suppose $v_1, \ldots, v_n$ is a basis of V and $w_1, \ldots, w_n \in W$. Then there exists a unique linear map $T : V \to W$ such that

$$Tv_k = w_k$$

for each $k = 1, \ldots, n$. (Note V and W are the same dimension)

Proof. We define $T : V \to W$ as:

$$T(c_1v_1 + \cdots + c_nv_n) = c_1w_1 + \cdots + c_nw_n$$

Where $c_1, \ldots, c_n$ are arbitrary elements of $\mathbb{F}$. By this definition, for each k take $c_k = 1$ and set all other c_i to 0 to verify that $T(v_1) = w_1$. We may show that this map is linear as follows:

If $u, v \in V$ with $u = a_1v_1 + \cdots + a_nv_n$ and $v = c_1v_1 + \cdots + c_nv_n$, then we see:

$$\begin{aligned} T(u+v) &= T(a_1v_1 + \cdots + a_nv_n + c_1v_1 + \cdots + c_nv_n) \\ &= T((a_1 + c_1)v_1 + \cdots + (a_n + c_n)v_n) \\ &= (a_1 + c_1)w_1 + \cdots + (a_n + c_n)w_n \\ &= (a_1w_1 + \cdots + a_nw_n) + (c_1w_1 + \cdots + c_nw_n) \\ &= T(u) + T(v) \end{aligned}$$

Likewise, if $\lambda \in \mathbb{F}$ was a scalar, and $v = a_1v_1 + \cdots + a_nv_n$, then:

$$\begin{aligned} T(\lambda v) &= T(\lambda a_1 v_1 + \cdots + \lambda a_n v_n) \\ &= \lambda a_1 w_1 + \cdots + \lambda a_n w_n \\ &= \lambda(a_1 w_1 + \cdots a_n w_n) \\ &= \lambda T(v) \end{aligned}$$

Hence, we have shown that T is linear. It remains to show that T is unique. By our definition of T, we know that $T(c_1v_1 + \cdots + c_nv_n) = c_1T(v_1) + \cdots + c_nT(v_n) = c_1w_1 + \cdots + c_nw_n$. This tells us that the value of T on any $v \in V$ is fixed once we know how the basis vectors map. If any other linear map satisfied this constraint, then it must be the same as T. □

3.1.2 Algebraic Operations on $\mathcal{L}(V, W)$

Definition 3.1.4: Addition and Scalar Multiplication on $\mathcal{L}(V, W)$

Suppose $S, T \in \mathcal{L}(V, W)$ and $\lambda \in \mathbb{F}$. The sum $S + T$ and the product λT are the linear maps from $V \mapsto W$ defined by:

$$(S + T)(v) = Sv + Tv \quad (\lambda T)(v) = \lambda(Tv)$$

for all $v \in V$.

Because we may define addition and scalar multiplication on $\mathcal{L}(V, W)$, we may show that $\mathcal{L}(V, W)$ satisfies the properties of a vector space. Unlike normal vector spaces, a product of linear maps may be defined as well:

Definition 3.1.5: 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

$$(ST)(u) = S(Tu)$$

for all $u \in U$

ST is intuitively the composition of S and T. T first maps from U to V, and S the maps from V to W, meaning that ST is a map from U to W. However, ST for S and T is only defined when T maps into the domain of S.

Definition 3.1.6: Algebraic Properties of products of linear maps

- **Associativity:** $(T_1T_2)T_3 = T_1(T_2T_3)$ whenever when the domains and codomains of T_1, T_2, and T_3 make sense.
- **Identity:** $TI_V = I_WT = T$ whenever $T \in \mathcal{L}(V, W)$. I_V denotes the identity on V and I_W denotes the identity on W.

- **Distributive Properties:** $(S_1 + S_2)T = S_1T + S_2T$ and $S(T_1 + T_2) = ST_1 + ST_2$ whenever, T, T_1, T_2 $in(L)U, V$, and $S, S_1, S_2 \in \mathcal{L}(V, W)$.

Finally, we may show that all linear maps map 0 to 0.

Theorem 3.1.7: Linear maps take 0 to 0

Suppose T is a linear map from V to W. Then $T(0) = 0$.

Proof. By the additivity of linearity:

$$T(0) = T(0 + 0) = T(0) + T(0)$$

Add the addtive inverse of $T(0)$ to both sides to get:

$$0 = T(0)$$

□

This conclusion tells us that any linear function is only linear if and only if the constant term is 0, since otherwise, the function won't map 0 to 0.

3.2 Null Spaces and Ranges

Every linear map has two related subspaces: its **Null Space** and its **Column Space**.

Definition 3.2.1: Null Space null T

For $T \in \mathcal{L}(V, W)$, the **null space** of T, denoted by $\operatorname{null} T$, is the subset of V consisting of vectors that T maps to 0:

$$\operatorname{null} T = \{v \in V : Tv = 0\}$$

In fact, the null space of a linear map is a subspace of the domain of the linear map. 0 is in the null space of every linear map.

Definition 3.2.2: The null space is a subspace

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

Proof. We show the three subspace properties:
Existance of 0: Obviously $0 \in \operatorname{null} T$. A linear map always maps 0 to 0.

Closure under addition: Given $v \in V$ such that $Tv = 0$, and $w \in V$ such that $Tw = 0$. We see that for $v + w$, $T(v + w) = T(v) + T(w) = 0$. So $v + w \in \operatorname{null} T$.

Closure under scalar multiplication: Fiven $v \in V$ such that $Tv = 0$. We see that for scalar $\lambda \in \mathbb{F}$, $T(\lambda v) = \lambda T(v) = \lambda 0 = 0$. So $\lambda v \in V$. □

From real analysis and discrete math, we learned that an **injective** map is one that is one-to-one, while a **surjective** map is one that is unto. We may define these properties for linear maps as well:

Definition 3.2.3: Injective

A function $T : V \to W$ is called **injective** if $Tu = Tv$ implies $u = v$

In other words, a linear map T is injective if and only if it maps distinct inputs to distinct outputs. Similar to linear independence, we can check if a linear map is injective just by checking if 0 is the only vector that is mapped to 0. This is because if there exists some other element in W that has multiple inputs that map to it from V under T, then there exists more than one way to map to 0.

Theorem 3.2.4

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

Proof. Suppose T is injective, then by definition, it must only map 0 to 0. Suppose that $T(v) = 0 = T(0)$, we see $v = 0$.

Now suppose $\operatorname{null} T = \{0\}$. Take $u, v \in V, u \neq v$ such that:

$$T(u) - T(v) = T(u - v) = 0$$

Hence, $u - v = 0$, which implies $u = v$. Hence, T must be injective. □

3.2.1 Range and Surjectivity

Definition 3.2.5: Range

For $T \in \mathcal{L}(V, W)$ the **range** of T is the subset of W consisting of vectors that are equal to Tv for some $v \in V$:

$$\operatorname{range} T = \{Tv : v \in V\}$$

Similar to the null-space of linear maps, range T is a subspace of W.

Theorem 3.2.6: The range is a subspace

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

Proof. We may prove the subspace proofs again:
Existence of zero element: We know that $T(0) = 0$, so $0 \in \operatorname{range} T$.

Closure under addition: Let $u, v \in V$ and $w_1, w_2 \in \text{range } T$. We know that $T(u) \in \text{range } T$ and $T(v) \in \text{range } T$. Hence: $T(u+v) = T(u) + T(v) = w_1 + w_2 \in \text{range } T$.

Closure under scalar multiplication: Let $w \in \text{range } T$. Let $\lambda \in \mathbb{F}$. There exists $v \in V$ such that $T(v) = w$. Hence, $T(\lambda v) = \lambda T(v) = \lambda w \in \text{range } T$. □

We see now how the range of a linear map relates to the range of a linear map.

Definition 3.2.7: Surjective

A function $T : V \to W$ is **surjective** if its range is equal to W.

Alternatively, this means that every vector in $w \in W$ has at least one pre-image in V that maps to itself. The surjectivity of a linear map depends on the space that it maps to. *The differentiation map $D \in \mathcal{L}(\mathcal{P}_5(\mathbb{R}))$ is not surjective as any functions of degree 5 does not have an antiderivative in $\mathcal{P}_5(\mathbb{R})$. However, the differentiation map $S \in \mathcal{L}(\mathcal{P}_4(\mathbb{R}), \mathcal{P}_5(\mathbb{R}))$ is surjective as the range is $\mathcal{P}_4(\mathbb{R})$.*

3.2.2 Fundamental Theorem of Linear Maps

Theorem 3.2.8: Fundamental theorem of linear maps

Suppose V is finite-dimensional and $T \in \mathcal{L}(V, W)$. Then range T is finite-dimensional and

$$\dim V = \dim \text{null } T + \dim \text{range } T$$

Proof. Let $u_1, \ldots, u_m$ be a basis of null T, hence dim null $T = m$We can extend this basis to a basis of V:

$$u_1, \ldots, u_m; v_1, \ldots, v_n$$

So $\dim V = m + n$. We must show that $\dim \text{range } T = n$. To do this, we show that $Tv_1, \ldots, Tv_n$ is a basis of range T. For any $v \in V$, we have:

$$v = \lambda_1 u_1 + \cdots + \lambda_m u_m + \mu_1 v_1 + \cdots + \mu_n u_n$$

We see then that:

$$\begin{aligned} Tv &= T(\lambda_1 u_1 + \cdots + \lambda_m u_m + \mu_1 v_1 + \cdots + \mu_n v_n) \\ &= \underbrace{T(\lambda_1 u_1) + \cdots + T(\lambda_m u_m)}_{\text{All} \in \text{null } T} + T(\mu_1 v_1) + \cdots + T(\mu_n v_n) \\ &= T(\mu_1 v_1) + \cdots + T(\mu_n v_n) \\ &= \mu_1 T(v_1) + \cdots + \mu_n T(v_n) \end{aligned}$$

Hence, we see that $T(v_1), \ldots, T(v_n)$ is a spanning list of range T, implying that range T is finite dimensional as well. All that remains is to show that $T(v_1), \ldots, T(v_n)$ is linearly

independent. Suppose that:

$$\mu_1 T(v_1) + \cdots + \mu_n T(v_n) = 0$$

This implies that

$$T(\mu_1 v_1 + \cdots + \mu_n v_n) = 0$$

Hence, $\mu_1 v_1 + \cdots + \mu_n v_n \in \operatorname{null} T$. Hence, we may write:

$$\mu_1 v_1 + \cdots + \mu_n v_n = \gamma_1 u_1 + \cdots + \gamma_m u_m$$

Since $u_1, \ldots, u_m; v_1, \ldots, v_n$ is a basis of V, the v's and the u's are all linearly independent, which implies the μ's are all 0, so $T(v_1), \ldots, T(v_n)$ is linearly independent. Hence, $T(v_1), \ldots, T(v_n)$ is a basis for $\operatorname{range} T$, so $\dim \operatorname{range} T = n$. Hence:

$$\begin{aligned} \dim V &= m + n \\ &= \dim \operatorname{null} T + \dim \operatorname{range} T \end{aligned}$$

□

3.2.3 Dimensionality Constraints

Intuitively, it makes sense that no linear map from a finite-dimensional vector space to a smaller vector space can be injective, and also no linear map from a finite-dimensional vector space to a "larger" vector space can be surjective.

We have the following bounds for $\dim \operatorname{null} T$ and $\dim \operatorname{range} T$:

$$\dim \operatorname{range} T \leq \min\{\dim V, \dim W\}$$

$$\dim \operatorname{null} T \leq \dim V$$

Theorem 3.2.9: Linear map to a Lower-dimensional space is not injective

Let V and W be finite-dimensional vector spaces such that $\dim V > \dim W$. Then no linear map from V to W is injective.

Proof. Any dimensionality argument like this one can be simply proved with rank nullity. We know that:

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

Let V be dimension n and W be dimension m, where $n > m$. We know that:

$$\dim \operatorname{range} T \leq m < n$$

Hence:

$$\dim \operatorname{null} T \geq n - m > 0$$

Since $\dim \operatorname{null} T > 0$, the null space of T is not trivial, so T cannot be injective. □

Theorem 3.2.10: Linear map to a higher-dimensional space is not surjective

Let V and W be finite-dimensional vector spaces such that $\dim V < \dim W$. Then no linear map from V to W is surjective.

Proof. Suppose for the sake of contradiction that there existed some map $T : V \to W$ such that T was surjective. This means that $\dim \operatorname{range} T \geq \dim W$, which implies that $\dim \operatorname{range} T > \dim V$, which is a contradiction.

This can also be proven using rank nullity, as setting $\dim \operatorname{null} T = 0$ tells us that $\dim \operatorname{range} T \leq \dim V < \dim W$, meaning that $\operatorname{range} T \neq W$, so T is not surjective. □

The dimensionality constraints provide important consequences when dealing with systems of linear equations. Linear maps can be used to represent systems of linear equations. We may think of linear systems as combinations of n unknowns and m equations. In the context of homogeneous systems (systems where each equation equals 0), we're interested in finding out if there exists non-trivial solutions. We see that the ideas we've seen above directly translate to the ideas of:

- A homogenous system of linear equations with more variables than equations has nonzero solutions. (Not injective)
- A homogenous system of linear equations with more equations than variables has no solution for some choice of constant terms. (Not surjective)

3.3 Matrices

3.3.1 Representing a Linear Map by a Matrix

An isomorphism (as defined in lecture), is a linear map that is also bijective. We see that there exists an isomophism between $\mathcal{L}(V, W)$ and W^n (n-tuples of elements of W). This means that for a given basis $v_1, \ldots, v_n$, every linear map is unique deterined by some $w_1, \ldots, w_n$ where $Tv_1 = w_1$. Hence, every linear map from $V \to W$ is structurally identical to n vectors from W. This give rise to the concept of a Matrix.

Definition 3.3.1: Matrix, $A_{j,k}$

Let m and n be nonnegative integers. An m-by-n matrix A is a rectangular array of elements of $\mathbb{F}$ with m rows and n columns:

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

With this definition, we can formulize the concept that every linear map is just a matrix.

Definition 3.3.2: Matrix of a linear map, $\mathcal{M}(T)$

Let $T \in \mathcal{L}(V, W)$ and $v_1, \ldots, v_n$ is a basis of V and $w_1, \ldots, w_m$ is a basis of W. The matrix of T with respect to these bases is the m-by-n matrix $\mathcal{M}(T)$ whose entries $A_{j,k}$ are defined by:

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

For constructing a basis, remember this diagram:

$$\mathcal{M}(T) = \begin{array}{c|cccccc} & v_1 & \cdots & v_k & \cdots & v_n \\ \hline w_1 & A_{1,1} & \cdots & A_{1,k} & \cdots & A_{1,n} \\ \vdots & \vdots & & \vdots & & \vdots \\ w_m & A_{m,1} & \cdots & A_{m,k} & \cdots & A_{m,n} \end{array}$$

Where the kth column denotes how to map v_k to a linear combination of the w's.

3.3.2 Addition and Scalar Multiplication of Matrices

Definition 3.3.3: Matrix Addition

The sume of two matrices of the same size is the matrix obtained b adding corresponding entries in the matrices:

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

Hence, we may define the matrix of the sum of linear maps.

Theorem 3.3.4: Matrix of the sum of linear spaces

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

Proof. We know that:

$$(S + T)v = Sv + Tv$$

Spplying the definition, we see that:

$$\mathcal{M}(S + T) = \mathcal{M}(S) + \mathcal{M}(T)$$

□

Similarly, we can also define scalar multiplication for a matric as well:

Definition 3.3.5: 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:

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

Theorem 3.3.6: Matrix of the scalar times a linear spaces

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

Since we may define addition and scalar multiplication for matricies, we may also define a vector space of matrices.

Definition 3.3.7: $\mathbb{F}^{m,n}$

For positive integers m and n, the set of all m-by-n matrices with entires in $\mathbb{F}$ is denoted by $\mathbb{F}^{m,n}$ and is also a vector space.

Note that the dimension of such a vector space is mn, as we may form a basis for the vector space out of all matrices whose elements are all 0 except for one. There are mn "degrees of freedom". Moreover, since $L(\mathbb{F}^{\times}, \mathbb{F}^{\geq})$ is an isomorphism with $\mathbb{F}^{m,n}$, both vector spaces have dimension mn.

3.3.3 Matrix Multiplication

Previously we have considered the composition of linear maps $T : U \to V$ and $S : V \to W$. The composition ST is thus another linear map from U to W. It makes sense that we could write a matrix that corresponds to ST, but does $\mathcal{M}(ST) = \mathcal{M}(S)\mathcal{M}(T)$? We may define matrix multiplication such that this is true:

Theorem 3.3.8: Matrix of Products of Linear Maps

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

Proof. The proof of this depends on how we define matrix multiplication. Suppose $\mathcal{M}(S) = A$ and $\mathcal{M}(T) = B$. Let A have dimension $m \times n$ and let B have dimension $n \times p$, where $p = \dim U$, $n = \dim V$, and $m = \dim W$. Let $u_1, \ldots, u_p$ be a basis of U, $v_1, \ldots, v_n$ be a

basis of V and $w_1, \ldots, w_m$ be a basis of W. For $1 \leq k \leq p$, we have:

$$\begin{aligned}(ST)u_k &= S\left(\sum_{r=1}^{n} B_{r,k} v_r\right) \\ &= \sum_{r=1}^{n} B_{r,k} S v_r \\ &= \sum_{r=1}^{n} B_{r,k} \sum_{j=1}^{m} A_{j,r} w_r \\ &= \sum_{r=1}^{n} \sum_{j=1}^{m} B_{r,k} A_{j,r} w_r \\ &= \sum_{j=1}^{m} \left(\sum_{r=1}^{n} A_{j,r} B_{r,k}\right) w_r\end{aligned}$$

Hence, we see that $\mathcal{M}(ST)$ is the $m \times p$ matrix C whose value $C_{j,k} = \sum_{r=1}^{n} A_{j,r} B_{r,k}$ □

Definition 3.3.9: Matrix Multiplication

Suppose A is an m-by-n matrix and B is an n-by-p matrix. Then AB is defined to be the m-by-p matric whose entry in row j, column k, is given by the equation

$$(AB)_{j,k} = \sum_{r=1}^{n} A_{j,r} B_{r,k}$$

Hence, the entry in row j, column k of AB is computed by taking row j of A and column k of B, multiplying together corresponding entries, and then summing.

Matrix multiplication only makes sense when the number of columns in the first matrix A matches the number of rows in the second matrix B. This corresponds to the idea that when we compose linear maps together, the map in the "middle" is the same, and must have the same dimension.

3.3.4 Column-Row Factorization and Rank of a Matrix

Definition 3.3.10: Column rank, row rank

Suppose A is an $m \times n$ matrix with entries in $\mathbb{F}$.

- The **column rank** of A is the dimension of the span of the column of A in $\mathbb{F}^{m,1}$. (column vectors)
- The **row rank** of A is the dimension of the span of the rows of A in $\mathbb{F}^{1,n}$. (row vectors)
- The column and row rank of A are at most $\min\{n, m\}$. Each column vector is at most m elements, and there are only n of them.

Definition 3.3.11: Transpose, A^T

The **transpose** of a matrix A, denoted by A^T, is the matrix obtained from A by interchanging rows and columns. If A is $m \times n$, then A^T is $n \times m$ whose entries are given by:

$$A^T_{j,k} = A_{k,j}$$

Theorem 3.3.12: Properties of transpose

- $(A+B)^T = A^T + B^T$
- $(\lambda A)^T = \lambda A^T$
- $(AC)^T = C^T A^T$

Proof. We prove the following:

- Let $C = A + B$, we know that $C_{j,k} = A_{j,k} + B_{j,k}$. $C^T_{j,k} = C_{k,j} = A_{k,j} + B_{k,j}$. Hence, $C^T_{j,k} = A^T_{j,k} + B^T_{j,k}$, so $C^T = A^T + B^T$.
- Let $C = \lambda A$, we know that $C_{j,k} = \lambda A_{j,k}$. Additionally, $C^T_{j,k} = C_{k,j} = \lambda A_{k,j} = \lambda A^T_{j,k}$. Hence, $C^T = \lambda A^T$.
- Let $D = AC$. We know that $D_{j,k} = \sum_{r=1}^{n} A_{j,r} C_{r,k}$. Hence $D^T_{j,k} = D_{k,j} = \sum_{r=1}^{n} A_{k,r} C_{r,j}$. Note that:

  $$A_{k,r} = A^T_{r,k} \quad C_{r,j} = C^T_{j,r}$$

  Hence, $A_{k,r} C_{r,j} = C^T_{j,r} A^T_{r,k}$, so $D^T = C^T A^T$.

□

Theorem 3.3.13: Column-Row factorization

Suppose A is an $m \times n$ matrix with entries in $\mathbb{F}$ and column rank $c \geq 1$. Then there exist an $m \times c$ matrix C and a $c \times n$ matrix R both with entries in $\mathbb{F}$, such that $A = CR$

Proof. The column vectors of A may be reduced to a basis of the span of the columns of A. This list has length c, and serves as the columns for C. Each column of R then, can be constructed as a column vector in F^c, that denotes the coefficients of the linear combination of the columns of C that will yield each column in A. We see that C is $n \times c$ while R is $c \times m$ and $A = CR$. □

Theorem 3.3.14: Column rank equals row rank

Suppose $A \in \mathbb{F}^{m,n}$. Then the column rank of A equals the row rank of A.

Proof. Let C denote the column rank of A. Let $A = CR$ be the column-row factorization of A. Every row of A is a linear combination of the rows of R (By the row-view of matrix

multiplication). Since R has c rows, we know that rowrank $A \leq c =$ columnrank A. To show they are equal, we must show that columnrank $A \leq$ rowrank A. We take the transpose of A, note that:

$$\text{columnrank } A = \text{rowrank } A \leq \text{columnrank } A^T = \text{rowrank } A$$

hence, we have shown that columnrank $A =$ rowrank A. □

Definition 3.3.15: Rank

The **rank** of a matrix $A \in \mathbb{F}^{m,n}$ is just the column rank of A.

3.4 Invertibility and Isomorphisms

3.4.1 Invertible linear maps

Definition 3.4.1: invertible, inverse

- A linear map $T \in \mathcal{L}(V, W)$ is invertible if there exists some linear map $S \in \mathcal{W}, \mathcal{V}$ such that ST equals the identity operator on V and TS equals the identity operator on W.
- A linear map $S \in \mathcal{L}W, V$ satisfying $ST = I_V$ and $TS = I_W$ is known as the **inverse** of T.

Intuitively, it makes sense that the inverse of a linear map should be unique. We prove this now:

Theorem 3.4.2: Inverse is unique

An invertible linear map has a unique inverse, and is denoted by T^{-1}

Proof. Let $T \in \mathcal{L}V, W$ and let $S_1, S_2 \in \mathcal{L}(W, V)$ be inverses of T. Note that:

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

So $S_1 = S_2$ □

Sometimes, it may be difficult to find an exact inverse of a map. Maybe we don't know the map itself. Regardless, we can still prove it's invertibility with the following theorem:

Theorem 3.4.3: invertibility $\iff$ injectivity and surjectivity

A linear map is invertible if and only if it is injective and surjective.

Proof. ($\rightarrow$) Suppose that T is invertible. To show that T is injective, we want to show

that if $Tv = 0$, then $v = 0$:

$$0 = Tv \implies T^{-1}0 = (T^{-1}T)v \implies v = 0$$

Hence, T is injective. Now to prove that T is surjective, we show that for all $w \in W$, we have some $v \in V$ such that $Tv = w$. We know that $T^{-1}w \in V$, and if apply T to this vector, we see that $TT^{-1}w = I_W w = w$. Hence, every vector in W has some pre-image under T in V, so T is surjective.

($\leftarrow$) Now suppose T is injective and surjective. We show that T is invertible by constructing the inverse of T. Take any basis $v_1, \ldots, v_n$ of V. Since T is bijective, we know that $Tv_1, \ldots, Tv_n = w_1, \ldots, w_n$ is an linearly independent list of vectors in W that is also a basis. This is because by rank-nullity, $\dim V = \dim W$. Create the map $S \in \mathcal{L}(W, V)$ such that $Sw_k = v_k$ for all $k = 1, \ldots, n$. Now, we see that for any $v \in V$, we have:

$$\begin{aligned} ST(v) &= ST(a_1v_1 + \cdots + a_nv_n) \\ &= S(a_1Tv_1 + \cdots + a_nTv_n) \\ &= S(a_1w_1 + \cdots + a_nw_n) \\ &= a_1Sw_1 + \cdots + a_nSw_n \\ &= a_1v_1 + \cdots + a_nv_n \\ &= v \end{aligned}$$

Hence $ST = I_V$. Similarly, we have:

$$\begin{aligned} TS(w) &= TS(a_1w_1 + \cdots + a_nw_n) \\ &= T(a_1Sw_1 + \cdots + a_nSw_n) \\ &= T(a_1v_1 + \cdots + a_nv_n) \\ &= a_1Tv_1 + \cdots + a_nTv_n \\ &= a_1w_1 + \cdots + a_nw_n \\ &= w \end{aligned}$$

Hence $TS = I_V$. We have shown that $S = T^{-1}$, so T is invertible. □

We now introduce a powerful corollary of the rank-nullity theorem. Given a linear transformation between two vector spaces of the same dimension, we may show that one of injectivity/surjectivity implies the other.

Theorem 3.4.4: Injectivity is equivalent to surjectivity (if $\dim V = \dim W < \infty$)

Suppose that V and W are finite-dimensional vector sapces, $\dim V = \dim W$, and $T \in \mathcal{L}(V, W)$. Then:

$$T \text{ is invertible} \iff T \text{ is injective} \iff T \text{ is surjective}$$

Proof. The rank nullity theorem states that:

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

We see that if T is injective, then $\operatorname{null} T = \{0\}$, so $\dim \operatorname{null} T = 0$, so we have $\dim V = \dim \operatorname{range} T = \dim W$, so $\operatorname{range} T = W$. If T is surjective, then $\operatorname{range} T = W$, so $\dim \operatorname{range} T = \dim W$, which implies that $\dim \operatorname{null} T = 0$, T is injective. Hence, if T is either injective or surjective, T must also be invertible. □

If T happens to be a linear operator on V, that is, T maps from V to V, then the fact that $ST = I$ implies that $TS = ST$.

Theorem 3.4.5: $ST = I \iff TS = I$ (on vector sapces of the same dimension)

Suppose V and W are finite-dimensional vector sapces of the same dimension, $S \in \mathcal{L}(W, V)$, and $T \in \mathcal{L}(V, W)$. Then $ST = I$ if and only if $TS = I$

Proof. First suppose $ST = I$. If $v \in V$ and $Tv = 0$, then:

$$v = Iv = (ST)v = S(Tv) = S0 = 0$$

which implies that T is injective, and by our previous theorem, T is surjective so T is invertible. Multiply the equation $ST = I$ on both sides by T^{-1}, and we get $ST(T^{-1}) = S = T^{-1}$, so $TS = TT^{-1} = I$. The proof for the other direction is identical. □

3.4.2 Isomorphic Vector Spaces

We can have two vector spaces that are structurally the same but have different representations. The two spaces are denoted as **isomorphic** to each other.

Definition 3.4.6: Isomorphsim, isomorphic

- An **isomorphism** is an invertible linear map.
- Two vector spaces are called **isomorphic** if there is an isomorphism from one vector space onto the other one.

An isormorphism $T : V \to W$ essentially relables every vector $v \in V$ as a vector $Tv \in W$. To easily and reliably determine if two vector spaces are isomorphic, we only need to check their dimension.

Theorem 3.4.7: Dimension shows whether vector spaces are isomorphic

Two finite-dimensional vector spaces over $\mathbb{F}$ are isomorphic if and only if they have the same dimension

Proof. ($\rightarrow$) Suppose that V and W are isomorphic, this means that there is some isomorphism $T : V \to W$. Since T is invertible, we have that T is injective and surjective, so $\operatorname{null} T = \{0\}$, and $\operatorname{range} T = W$. By rank-nullity, we see that $\dim V = \dim W$.

($\leftarrow$) Suppose that V and W have the same dimension. Let $v_1, \ldots, v_n$ be a basis of V and $w_1, \ldots, w_n$ be a basis of W. We define $T : V \to W$ as:

$$T(v) = T(a_1 v_1 + \cdots + a_n v_n) = a_1 w_1 + \cdots a_n w_n$$

We see that T is linear as it obeys additivity and homogeneity. Additionally, T is surjective because it maps to every basis vector of W, and since $\dim V = \dim W$, we have that T is also injective. Hence, T is an isomorphism, so V and W are isomorphic. □

Theorem 3.4.7 has many useful consequences. For one, any finite-dimensional vector space V is isomorphic to $\mathbb{F}^n$ where $n = \dim V$. This holds true for spaces like the polynomial space, for which a valid isomorphism between $\mathcal{P}_n(\mathbb{F})$ and $\mathbb{F}^{n+1}$ is mapping a polynomial to a vector containing the coefficients of each power in the polynomial.

$\mathbb{F}^{m,n}$ denotes the vector space of $m \times n$ matrices whose entries lie in $\mathbb{F}$. Let $v_1, \ldots, v_n$ be a basis of V, and $w_1, \ldots, w_m$ be a basis of W. For each $T \in \mathcal{L}(V, W)$, there is a matrix $\mathcal{M}(T) \in \mathbb{F}^{m,n}$. $\mathcal{M}$ is actually a linear map from the space of linear transformations to the space of matrices. It actually serves as an isomorphism.

Theorem 3.4.8: $\mathcal{L}(V, W)$ and $\mathbb{F}^{m,n}$ are isomorphic

Suppose $v_1, \ldots, v_n$ is a basis of V and $w_1, \ldots, w_n$ is a basis of W. Then $\mathcal{M}$ is an isomorphism between $\mathcal{L}(V, W)$ and $\mathbb{F}^{m,n}$.

Proof. $\mathcal{M}$ is linear, so it remains to show that $\mathcal{M}$ is injective and surjective. To prove that $\mathcal{M}$ is injective, we show that $\mathcal{M}(T) = 0$ only when $T = 0$. If $\mathcal{M}(T) = 0$, then $Tv_k = 0$ for each $k = 1, \ldots, n$. Since $v_1, \ldots, v_n$ is a basis of V, $T = 0$, so $\mathcal{M}$ is injective.

To show that $\mathcal{M}$ is surjective, take $A \in \mathcal{F}^{m,n}$. There exists some $T \in \mathcal{V}, \mathcal{W}$ such that:

$$Tv_k = \sum_{j=1}^{m} A_{j,k} w_j$$

We see that $\mathcal{M}(T) = A$, so the range of $\mathcal{M}$ is $\mathbb{F}^{m,n}$, so the two spaces are isomorphic. □

A corollary of the previous result informs of how we could deterine the dimension of a particular linear map.

Corollary 3.4.9

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

Suppose that V and W are finite-dimensional. Then $\mathcal{L}(V, W)$ is finite-dimensional and

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

Solution

We know that $\dim \mathcal{L}(V,W) = \dim \mathbb{F}^{m,n}$ as they are isomorphic. We know that $\dim \mathbb{F}^{m,n} = mn$, so $\dim \mathcal{L}(V,W) = (\dim V)(\dim W)$

3.4.3 Linear maps thought of as Matrix Multiplication

Definition 3.4.10: Matrix of a vector, $\mathcal{M}(v)$

Suppose $v \in V$ and $v_1, \ldots, v_n$ is a basis of V. The **matrix of** v with respect to this basis is the $n \times 1$ matrix:

$$\dim \mathcal{M}(v) = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$$

where $b_1, \ldots, b_n$ are scalars such that $v = b_1 v_1 + \cdots + b_n v_n$.

This matrix of a vector depends on what the underlying basis we've chosen are. Once again, the $\mathcal{M}$ transformation is an isomorphism between V and $\mathbb{F}^{\dim V}$. Suppose that $A \in \mathbb{F}^{m,n}$, then $A_{\cdot,k}$ denotes the kth column of A that is a m by 1 matrix.

Theorem 3.4.11: $\mathcal{M}(T)_{\cdot,k} = \mathcal{M}(Tv_k)$

Suppose $T \in \mathcal{L}(V,W)$, $v_1, \ldots, v_n$ is a basis of V, and $w_1, \ldots, w_m$ is a basis of W. Then the kth column of $\mathcal{M}(T)$ is equal to $\mathcal{M}(Tv_k)$.

Proof. By definition, this is true. □

Theorem 3.4.12: Linear maps act like matrix multiplication

Suppose $T \in \mathcal{L}(V,W)$ and $v \in V$. Suppose $v_1, \ldots, v_n$ is a basis of V and $w_1, \ldots, w_m$ is a basis of W. Then:

$$\mathcal{M}(Tv) = \mathcal{M}(T)\mathcal{M}(v)$$

Proof. We know that $v = a_1 v_1 + \cdots + a_n v_n$ for some $a_1, \ldots, a_n \in \mathbb{F}$. Hence:

$$Tv = a_1 Tv_1 + \cdots + a_n Tv_n$$

We apply the $\mathcal{M}$ transformation, and see that:

$$\begin{aligned} \mathcal{M}(Tv) &= a_1 \mathcal{M}(v_1) + \cdots + a_n \mathcal{M}(v_n) \\ &= a_1 \mathcal{M}(T_{\cdot,1}) + \cdots + a_n \mathcal{M}(T_{\cdot,n}) \\ &= \mathcal{M}\sqsubseteq \end{aligned}$$

□

The specific matrix A of a linear map depends on the underlying choice of bases that we have. However, no matter what choice of basis we use, the column rank of $\mathcal{M}(T)$ is always

the same, as range T does not depend on the choice of a basis.

Theorem 3.4.13: dimension of range T equals column rank of $\mathcal{M}(T)$

Suppose V and W were finite-dimensional and $T \in \mathcal{L}(V, W)$, Then $\dim \operatorname{range} T = \operatorname{rank}(\mathcal{M}(T))$

Proof. Let $v_1, \ldots, v_n$ be a basis of V, and $w_1, \ldots, w_n$ be a basis of W. The linear map $\mathcal{M}$ that takes $w \in W$ to $\mathcal{M}(w)$ is an isomorphism from W to $\mathbb{F}^{m,1}$. If we restrict this isomorphism onto range T, the restriction is an isomorphism from range T to $\operatorname{span}(\mathcal{M}(Tv_1), \ldots, \mathcal{M}(Tv_n))$. Each $\mathcal{M}(Tv_k)$ is the k-th column of $\mathcal{M}(T)$, so:

$$\dim \operatorname{range} T = \dim \operatorname{range} \operatorname{span}(\mathcal{M}(Tv_1), \ldots, \mathcal{M}(Tv_n)) = \operatorname{columnrank}(\mathcal{M}(T))$$

□

3.4.4 Change of Basis

Definition 3.4.14: Invertible, inverse, A^{-1}

A swuare matrix A is **invertible** if there is a square matrix B of the same size such that $AB = BA = I$. B is the inverse of A and is denoted by A^{-1}.

Theorem 3.4.15: Change of basis formula

Suppose $T \in \mathcal{L}(V)$. Suppose $u_1, \ldots, u_n$ and $v_1, \ldots, v_n$ are basis of V. Let:

$$A = \mathcal{M}(T, (u_1, \ldots, u_n)) \quad \text{and} \quad B = \mathcal{M}(T, (v_1, \ldots, v_n))$$

and $C = \mathcal{M}(I, (u_1, \ldots, u_n), (v_1, \ldots, v_n))$. Then

$$A = C^{-1}BC$$

Proof. We see that:

$$A = C^{-1}\mathcal{M}(T, (u_1, \ldots, u_n), (v_1, \ldots, v_n))$$

Note that C^{-1} is the identity map that sends vectors in the basis $v_1, \ldots, v_n$ to the basis $u_1, \ldots, u_n$. Now, we may derive an expression for $\mathcal{M}(T, (u_1, \ldots, u_n), (v_1, \ldots, v_n)) = BC$. C essentially casts any vector in basis $u_1, \ldots, u_n$ to the v basis. B then performs the transformation. Finally, we take the transformed vector and convert it back to the dimension we started with through C^{-1}. Giving us:

$$A = C^{-1}BC$$

□

Theorem 3.4.16: Matrix of inverse equals inverse of matrix

Suppose that $v_1, \ldots, v_n$ is a basis of V and $T \in \mathcal{L}(V)$ is invertible. Then $\mathcal{M}(T^{-1}) = (\mathcal{M}(T))^{-1}$, where both matrices are with respect to the basis $v_1, \ldots, v_n$.

Proof. By definition, we know that:

$$T^{-1}T = I$$

Hence:

$$\mathcal{M}(T^{-1}T) = \mathcal{M}(I) = I$$

Thus we have:

$$\mathcal{M}(T^{-1})\mathcal{T} = I$$

which shows that $\mathcal{M}(T^{-1}) = (\mathcal{M}(T))^{-1}$. □

3.5 Products and Quotients of Vector Spaces

3.5.1 Products of Vector Spaces

Definition 3.5.1: Product of Vector Spaces

Suppose $V_1, \ldots, V_m$ are vector spaces over $\mathbb{F}$.

- The **product** $V_1, \ldots, V_m$ is defined by:

$$V_1 \times \cdots \times V_m = \{(v_1, \ldots, v_m) : v_1 \in V_1, \ldots, v_m \in V_m\}.$$

- Addition on $V_1 \times \cdots \times V_m$ is defined by

$$(u_1, \ldots, u_m) + (v_1, \ldots, v_m) = (u_1 + v_1, \ldots, u_m + v_m)$$

- Scalar multiplication on $V_1 \times \cdots \times V_m$ is defined by

$$\lambda(v_1, \ldots, v_m) = (\lambda v_1, \ldots, \lambda v_m)$$

Theorem 3.5.2: Product of vector spaces is a vector space

Suppose $V_1, \ldots, V_m$ are vect or spaces over $\mathbb{F}$. Then $V_1 \times \cdots \times V_m$ is a vector space over $\mathbb{F}$.

Theorem 3.5.3: Dimension of a product is the sum of dimensions

Suppose $V_1, \ldots, V_m$ are finite-dimensional vector spaces. Then $V_1 \times \cdots \times V_m$ is finite-dimensional and

$$\dim(V_1 \times \cdots \times V_m) = \dim V_1 + \cdots + \dim V_m$$

Proof. For all V_k, choose a basis. For each basis vector of V_k, we may form a basis vector of $V_1 \times \cdots \times V_m$ such that the kth element is the said basis vector of V_k, with all other elements 0. If we take all such vectors and form a list, we see that they are linearly independent and span $V_1 \times \ldots \times V_m$. Hence, it is a basis with length $\dim V_1 + \cdots + \dim V_m$, so it has dimension $\dim V_1 + \cdots + \dim V_m$. □

We may relate the idea of product spaces to a direct sum.

Theorem 3.5.4: products and direct sums

Suppose that $V_1, \ldots, V_m$ are subspaces of V. We define $\Gamma : V_1 \times \cdots \times V_m \to V_1 + \cdots + V_m$ by:

$$\Gamma(v_1, \ldots, v_m) = v_1 + \cdots + v_m$$

Then $V_1 + \cdots + V_m$ is a direct sum if and only if Γ is injective.

Proof. Γ is injective if and only if the only way for $\Gamma(v_1, \ldots, v_m) = 0$ is if all $v_1, \ldots, v_m$ were 0. However, this is just the definition of a direct sum. □

3.5.2 Quotient Spaces

The crux of a quotient space lies in the idea of the sum of a vector and a subset. We call such a set a **coset**

Definition 3.5.5: $v + U$, translate

Let $v \in V$ and $U \subseteq V$. Then $v + U$ is the subset of V defined by:

$$v + U = \{v + u : u \in U\}$$

This is also called a **translate** of U.

We're now ready to define a quotient space of V for some subspace U.

Definition 3.5.6: quotient space, V/U

Suppose U is a subspace of V. Then the **quotient space** V/U is the set of all translates of U.

$$V/U = \{v + U : \forall v \in V\}$$

Theorem 3.5.7: Two translates of a subspace are equal or disjoint

Suppose U is a subspace of V and $v, w \in V$. Then:

$$v - w \in U \iff v + U = w + U \iff (v + U) \cap (w + U) \neq \emptyset$$

Proof. Suppose $v-w \in U$. Then $w = v+u$ for some $u \in U$. Hence $w+U = v+u+U = v+U$ as $u + U = U$. The fact that $v + U = w + U$ proves the fact that $(v + U) \cap (w + U) \neq \emptyset$. If $(v + U) \cap (w + U) \neq \emptyset$, then we know that there is some vector $u \in V$ such that $u = v + u_1 = w + u_2$. Now, see that:

$$v + u_1 - w - u_2 = 0 \implies v - w = u_2 - u_1 \in U$$

□

To show that V/U is a vector space, we must define addition and scalar multiplication on it.

Definition 3.5.8: addition and scalar multiplication on V/U

Suppose U is a subspace of V. We define addition and scalar multiplication on V/U as:

$$(v_1 + U) + (v_2 + U) = v_1 + v_2 + U + U = (v_1 + v_2) + U$$

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

The only issue with addition and scalar multiplication, is that in a quotient space, we are not working with distinct elements, but rather equivalent classes.

Theorem 3.5.9: Quotient Space operations are well defined

Across different representations of equivalence classes, the operations hold.

Proof. Take $v_1, v_2, w_1, w_2 \in V$ such that

$$v_1 + U = v_2 + U \quad \text{and} \quad w_1 + U = w_2 + U$$

We show that $(v_1 + w_1) + U = (v_2 + w_2) + U$. We know that:

$$v_1 - v_2 \in U \quad \text{and} \quad w_1 - w_2 \in U$$

This implies that $(v_1 - v_2) + (w_1 - w_2) \in U$, which means that $v_1 + w_1 - (v_2 + w_2) \in U$. Applying our properties, we get that:

$$v_1 + w_1 + U = v_2 + w_2 + U$$

So addition is well-defined. Now we show that if $v_1+U = w_1+U$, then $\lambda(v_1+U) = \lambda(w_1+U)$. Note:

$$v_1 - w_1 \in U \implies \lambda(v_1 - w_1) \in U$$

So we have $\lambda v_1 + U = \lambda w_1 + U$, so scalar multiplication is also well-defined. The vector space proof is just applying the axioms, and the additive identity of V/U is $0 + U$, and the additive inverse of any element $v + U$ is just $-v + U$. □

Since V/U is a vector space, we may assign a dimension to it as well. We show that the $\dim V/U = \dim V - \dim U$.

Definition 3.5.10: Quotient map, π

Suppose U is a subspace of V. Then the quotient map $\pi : V \to V/U$ is the linear map defined by:

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

for all $v \in V$

We may use the quotient map and the fundamental theorem of linear algebra to derive an expression for the dimension of $\dim V/U$.

Theorem 3.5.11: Dimension of quotient space

Suppose V is finite-dimensional and U is a subspace of V. Then:

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

Proof. Let π be the quotient map from V to V/U. We see that $\operatorname{range}\pi = V/U$, so $\dim \operatorname{range}\pi = \dim V/U$. Additionally, by the fundamental theorem of linear algebra, we know that:

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

We see that $\operatorname{null}\pi = U$, so the fundamental theorem just gives us:

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

□

We see that quotient spaces essentially takes a subspace U of V and collapses it to a single point. Hence, for every linear map T on V, we map create an induced linear map $\tilde{T}$ on $V/(\operatorname{null}T)$. This is a powerful idea, as we essentially collapse the null-space of T into a single point: 0.

Definition 3.5.12: $\tilde{T}$

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. We show that this map is well-defined, that is, for different representations of the same vector, is their image under $\tilde{T}$ equal? Take $v, w \in V$ such that $v + \operatorname{null} T = w + \operatorname{null} T$. We see that $v - w \in \operatorname{null} T$. We see that:

$$T(v - w) = 0 \implies T(v) - T(w) = 0 \implies T(v) = T(w)$$

Hence, we have that:

$$\tilde{T}(v + \operatorname{null} T) = T(v) = T(w) = \tilde{T}(w + \operatorname{null} T)$$

□

This map $\tilde{T}$ has the following properties:

Theorem 3.5.13: Null space and range of $\tilde{T}$

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

1. $\tilde{T} \circ \pi = T$ where π is the quotient map of V onto $V/\operatorname{null} T$
2. $\tilde{T}$ is injective
3. $\operatorname{range} \tilde{T} = \operatorname{range} T$
4. $V/(\operatorname{null} T)$ and $\operatorname{range} T$ are isomorphic

Proof.

1. For every vector $v \in V$, we have: $(\tilde{T} \circ \pi)v = \tilde{T}(\pi(v)) = \tilde{T}(v + \operatorname{null} T) = Tv$
2. Suppose $\tilde{T}(v + \operatorname{null} T) = Tv = 0$. Then $Tv = 0$, which implies $v \in \operatorname{null} T$. Hence, $v + \operatorname{null} T \in \operatorname{null} T$, which is the zero element of $V/\operatorname{null} T$, so $\tilde{T}$ is injective.
3. By definition, we see that $\operatorname{range} \tilde{T} = \operatorname{range} T$
4. To show that they are isomorphic, we must show that there exists an isomoprhism between the two spaces. We see that $\tilde{T}$ is the isomorphism, as it is injective and surjective.

□

3.6 Duality

3.6.1 Dual Space and Dual Map

We introduce the concept of linear maps that map elements of a vector space V into a scalar field $\mathbb{F}$. Such a mapping is called a **linear functional**.

Definition 3.6.1: Linear Functional

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

Definition 3.6.2: Dual space, V'

The vector space $\mathcal{L}(V, \mathbb{F})$ is known as the **dual space** of V, denoted by V'

We can make statements about the dimension of the dual space as well:

Theorem 3.6.3: $\dim V' = \dim V$

Suppose V is finite-dimensional. Then V' is also finite-dimensional and

$$\dim V = \dim V'$$

Proof.

$$\dim V' = \dim \mathcal{L}(V, \mathbb{F}) = (\dim V)(\dim \mathbb{F}) = \dim V$$

□

Similar to basis on a regular vector space V, we may define a basis of V' as well:

Definition 3.6.4: Dual Basis

If $v_1, \ldots, v_n$ is a basis of V, then the **dual basis** of $v_1, \ldots, v_n$ is the list $\phi_1, \ldots, \phi_n$ of elements of V', where each ϕ_j is the linear functional on V such that:

$$\phi_j(v_k) = \begin{cases} 1 & \text{if } j = k \\ 0 & \text{o.w.} \end{cases}$$

One way of thinking about the dual basis, is that each of the dual basis functionals ϕ_j applied onto a vector $v \in V$ gives the coefficient of the j-th basis vector of V in the linear combination of v in terms of v_ks.

Theorem 3.6.5: Dual Basis gives coefficients for linear combination

Suppose $v_1, \ldots, v_n$ is basis of V and $\phi_1, \ldots, \phi_n$ is the dual basis. Then

$$v = \phi_1(v)v_1 + \cdots + \phi_n(v)(v_n)$$

for each $v \in V$.

Proof. let $v = a_1v_1 + \cdots + a_nv_n$, we see that $\phi_j(v) = a_1\phi_j(v_1) + \cdots + a_n\phi_j(v_n)$. By the definition of the dual basis, we get that $\phi_j(v) = a_j$. Hence, $v = \phi_1(v)v_1 + \cdots + \phi_n(v)v_n$.

□

We now prove taht the dual space is in fact a basis of the dual space.

Theorem 3.6.6: dual basis is a basis of the dual space

Suppose V is finite-dimensional. Then the dual basis of a basis of V is a basis of V'.

Proof. Suppose $v_1, \ldots, v_n$ is a basis of V. Let $\phi_1, \ldots, \phi_n$ denote the dual basis. We show that $\phi_1, \ldots, \phi_n$ is linearly independent. Suppose that for $a_1, \ldots, a_n \in \mathbb{F}$, we have:

$$a_1\phi_1 + \cdots + a_n\phi_n = 0$$

We then have:

$$(a_1\phi_1 + \cdots + a_n\phi_n)(v_k) = a_k$$

Hence, we see that for the linear combination to hold true, $a_k = 0$ for all k, so $\phi_1, \ldots, \phi_n$ is a linearly independent list with $\dim V = \dim V'$ vectors, so it's a basis of the dual space. □

For any linear map $T : V \to W$, we can create a corresponding map $T' : W' \to V'$

Definition 3.6.7: dual map

Let $T \in \mathcal{L}(V, W)$. The **dual map** of T is the linear map $T' \in \mathcal{L}(W', V')$ defined for each $\phi \in W'$ as:

$$T'(\phi) = \phi \circ T$$

To verify this, we see that the transformation first converts a given vector $v \in V$ to a vector $w \in W$. Then, ϕ converts $w \in W$ to $\lambda \in \mathbb{F}$. We see that the composition then sends any vector in V to $\mathbb{F}$ which tells us that it is a linear functional of V. It could be verified that T' is linear:

- Let $\phi, \psi \in W'$, then:

$$T'(\phi + \psi) = (\phi + \psi) \circ T = \phi \circ T + \psi \circ T = T'(\phi) + T'(\psi)$$

- Let $\lambda \in \mathbb{F}$, $\phi \in W'$, then:

$$T'(\lambda\phi) = (\lambda\phi) \circ T = \lambda(\phi \circ T) + \lambda T'(\phi)$$

Theorem 3.6.8: Algebraic properties of dual maps

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

(a) $(S + T)' = S' + T'$ for all $S \in \mathcal{L}(V, W)$

(b) $(\lambda T)' = \lambda T'$ for all $\lambda \in \mathbb{F}$

(c) $(ST)' = T'S'$ for all $S \in \mathcal{L}(W, U)$.

Proof. (a) Note that for all $\phi \in W'$:

$$(S+T)'(\phi) = \phi \circ (S+T) = \phi \circ S + \phi \circ T = S' + T'$$

(b) For all $\phi \in W'$:

$$(\lambda T)'(\phi) = \phi \circ (\lambda T) = \lambda(\phi \circ T) = \lambda T'(\phi)$$

(c) For all $\phi \in U'$:

$$(ST)'\phi = \phi \circ S \circ T = T'(\phi \circ S) = T'S'\phi$$

□

3.6.2 Null Space and Range of Dual of Linear Map

Since T and T' are closely related, it should be no surprise that we can express $\operatorname{null} T'$ and $\operatorname{range} T'$ in terms of $\operatorname{null} T$ and $\operatorname{range} T$.

Theorem 3.6.9: Annihilator, U^0

For $U \subseteq V$ the **Annihilator** of U, denoted by U^0, is defined by:

$$U^0 = \{\phi \in V : \phi(u) = 0 \text{ for } u \in U\}$$

The annihlator of a subspace is a subset of V', more specfiically, it is a subspace of the dual space.

Theorem 3.6.10: The annihilator is a subspace

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

Proof. We must prove the subspace properties for U^0. For closure under addition, we take $\phi, \psi \in U^0$, we know that for all $u \in U$:

$$\phi(u) = 0 \quad \text{and} \quad \psi(u) = 0$$

Take $\phi + \psi$, we have:

$$(\phi + \psi)u = \phi(u) + \psi(u) = 0 + 0 = 0$$

Hence, $(\phi + \psi) \in U^0$, so U^0 is closed under addition. To show closure under scalar multiplication, note that:

$$(\lambda\phi)u = \lambda\phi(u) = \lambda(0) = 0$$

for all $\lambda \in \mathbb{F}$, so $\lambda\phi \in U^0$, so U^0 is closed under scalar multiplication. To verify that the zero element exists, note that $\phi = 0$ belongs in U^0, as it sends all vectors in V to 0, so it must also send all vectors in U to 0. □

Now we may define the dimension of the annhilator based on the dimension of the parent space and the subspace.

Theorem 3.6.11: dimension of the annihilator

Suppose V is finite-dimensional and U is a subspace of V. Then

$$\dim U^0 = \dim V - \dim U$$

Proof. We begin with some basis $v_1, \ldots, v_n$ of U, now we may extend this basis to a basis $v_1, \ldots, v_m$ of V. There is a corresponding dual basis $\phi_1, \ldots, \phi_m$ of V' such that $\phi_j(v_j) = 1, \phi_{j \neq k}(v_k) = 0$. Note that for all $\phi \in U^0$:

$$\phi(u) = (a_1\phi_1 + \cdots + a_m\phi_m)(u) = 0$$

for all $u \in U$. For this to hold true, $\phi \in \text{span}(\phi_{n+1}, \ldots, \phi_m)$, as if ϕ had a non-zero component of some dual basis of U, it wouldn't map all $u \in U$ to 0. Hence, $U^0 = \text{span}(\phi_{n+1}, \ldots, \phi_m)$, so $\dim U^0 = \dim V - \dim U$.

Alternative proof: Let $i \in \mathcal{L}(U, V)$ be the map defined by $i(u) = u$ for all $u \in U$. Take i', a linear map from V' to U'. By the fundamental theorem of linear algebra:

$$\dim V' = \dim \text{null}\, i' + \dim \text{range}\, i'$$

Note that $\text{null}\, i' = \dim U^0$, $\dim V' = \dim V$ and $\dim \text{range}\, i' = \dim U' = \dim U$. Hence, we have:

$$\dim U^0 = \dim V - \dim U$$

as desired. □

It is useful for proofs to know the behaviors of the annihilator when it is the full dual space, or when it is the set containing only 0.

Theorem 3.6.12: Condition for the anihilator to equal $\{0\}$ or the whole space

Suppose V is finite-dimensional and U is a subspace of V. Then

(a) $U^0 = \{0\} \iff U = V$

(b) $U^0 = V' \iff U = \{0\}$

Proof. (a) Suppose that $U^0 = \{0\}$. Then no $\phi \in V'$ where $\phi \neq 0$ sends all vectors in U to 0 except for 0. This means that for all ϕ_k in the dual basis of V', we can find some element $u \in U$ such that $\phi u \neq 0$. Hence, all the basis vectors of V are contained within U, so $U = V$. For the other direction, note that if $U = V$, then all linear functionals map some non-zero vectors to some non-zero value, with the exception for the 0 map, so $U^0 = \{0\}$.

(b) Suppose that $U^0 = V'$. This means that all linear functionals in V' map all vectors in $U = 0$. Take the dual basis $\phi_1, \ldots, \phi_n$ of V'. For each of dual basis functionals, $\phi_k u = 0$ for all $u \in U$. This means that no basis vector may be in the span of U, so

$U = \{0\}$. For the other direction, note that if $U = \{0\}$, then for all linear functionals ϕ in V', $\phi 0 = 0$, so $U^0 = V'$.

□

Theorem 3.6.13: The null space of T'

Suppose V and W are finite-dimensional and $T \in \mathcal{L}(V, W)$. Then

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

(b) $\dim \operatorname{null} T' = \dim \operatorname{null} T + \dim W - \dim V$

Proof. (a) For all $\phi \in \operatorname{null} T'$, note that $\phi \circ T = 0$ for all $v \in V$. Hence, for all $u \in \operatorname{range} T$, $\phi u = 0$, so $\phi \in (\operatorname{range} T)^0$. For all $\phi \in (\operatorname{range} T)^0$, note that $\phi \circ Tv = 0$ for all $v \in V$. Hence, $\phi \circ T = 0$, so $\phi \in \operatorname{null} T'$. Since $\operatorname{null} T' \subseteq (\operatorname{range} T)^0$ and $(\operatorname{range} T)^0 \in \operatorname{null} T'$, $\operatorname{null} T' = (\operatorname{range} T)^0$.

(b) Note that $\dim \operatorname{null} T' = \dim(\operatorname{range} T)^0$. By the previous theoresm, we know that:

$$\dim(\operatorname{range} T)^0 = \dim W - \dim \operatorname{range} T$$

Since $\dim \operatorname{range} T = \dim V - \dim \operatorname{null} T$, we have:

$$\dim(\operatorname{range} T)^0 = \dim \operatorname{null} T + \dim W - \dim V$$

□

Because the null space and ranges of a vector space and it's dual space are connected, we can prove the surjectivity of T through the injectivity of T'.

Theorem 3.6.14: T surjective is equivalent to T' injective

Suppose V and W are finite-dimensional and $T \in \mathcal{L}(V, W)$. Then

$$T \text{ is surjective} \iff T' \text{ is injective}$$

Proof.

$$\begin{aligned} T \text{ is surjective} &\iff \operatorname{range} T = W \\ &\iff (\operatorname{range} T)^0 = \{0\} \\ &\iff \operatorname{null} T' = \{0\} \\ &iff T' \text{ is injective} \end{aligned}$$

□

We many show something similar for $\operatorname{range} T'$:

Theorem 3.6.15: the range of T'

Suppose V and W are finite-dimensional and $T \in \mathcal{L}(V, W)$. Then

(a) $\dim \operatorname{range} T' = \dim \operatorname{range} T$

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

Proof. (a) Note:

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

(b) Take $\phi \in \operatorname{range} T'$, we know that $\phi = \psi \circ T$ for some $\psi \in W'$. If $v \in \operatorname{null} T$, then:

$$\phi(v) = \psi \circ Tv = \psi(0) = 0$$

Hence, $\phi \in (\operatorname{null} T)^0$, so $\operatorname{range} T' \subseteq (\operatorname{null} T)^0$. To show equality, we show that the dimension o the two spaces are the same:

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

□

And like before, we can show that T is injective if T' is surjective.

Theorem 3.6.16: T injective is equivalent to T' surjective

Suppose V and W are finite-dimensional and $T \in \mathcal{L}(V, W)$. Then

$$\text{is injective} \iff T' \text{ is surjective}$$

Proof.

$$\begin{aligned}T \text{ is injective} &\iff \operatorname{null} T = 0 \\ &\iff \dim \operatorname{null} T = 0 \\ &\iff \dim(\operatorname{null} T)^0 = \dim V = \dim V' \iff \dim \operatorname{range} T' = \dim V' \iff T' \text{ is surjective}\end{aligned}$$

□

3.6.3 Matrix of Dual of Linear Map

We can actually derive the matrix of the dual map with respect to the matrix of the original map:

Theorem 3.6.17: Matrix of T' is transpose of matrix of T

Suppose V and W are finite-dimensional and $T \in \mathcal{L}(V, W)$. Then:

$$\mathcal{M}(T') = (\mathcal{M}(T))^T$$

Proof. Let $A = \mathcal{M}(T)$ and $C = \mathcal{M}(T')$. Let $1 \leq j \leq m$ and $1 \leq k \leq n$, where $m = \dim W$ and $n = \dim V$. We know that:

$$T'(\psi_j) = \sum_{r=0}^{n} C_{r,j}\phi_r = \psi_j \circ T$$

Hence, we have:

$$\psi_j \circ T(v_k) = \sum_{r=0}^{n} C_{r,j}\phi_r(v_k) = C_{k,j}$$

Additionally,

$$\begin{aligned}\psi_j \circ T(v_k) &= \psi_j \left(\sum_{r=1}^{m} A_{r,k} w_r \right) \\ &= A_{j,k}\end{aligned}$$

Since $C_{k,j} = A_{j,k}$, we see that the two matrices are transposes of each other. □

The duality of the two spaces can actually be used to prove that column rank is equal to row rank.

Theorem 3.6.18: Column rank equals row rank

Suppose $A \in \mathbb{F}^{m,n}$. Then the column rank of A equals the row rank of A.

Proof. Let $T : \mathbb{F}^{\cdot,n} \to \mathbb{F}^{\cdot,m}$ where $\mathcal{M}(T) = A$. Then we have that:

$$\begin{aligned}\text{column-rank}(A) &= \text{column-rank}(\mathcal{M}(T)) \\ &= \dim \text{range}\, T \\ &= \dim \text{range}\, T' \\ &= \text{column-rank}\, \mathcal{M}(T') \\ &= \text{column-rank}\, A^T \\ &= \text{row-rank}(A)\end{aligned}$$

□

Chapter 4

Polynomials

We will explore polynomials over both the field of real numbers $\mathbb{R}$ and the field of complex numbers $\mathbb{C}$.

4.1 Complex Numbers

Definition 4.1.1: Real part, $\Re z$, imaginary part, $\Im z$

Suppose $z = a + bi$, where a and b are real numbers.

- The **real part** of z, denoted by $\Re z$ is defined by $\Re z = a$
- The **imaginary part** of z, denoted by $\Im z$ is defined by $\Im z = b$

For every complex number z, we have $z = \Re z + i\Im z$. Over the field of complex numbers, there is also the concept of a complex conjugate, denoted as $\bar{z}$.

Definition 4.1.2: Complex conjugate, $\bar{z}$, absolute value, $|z|$

Suppose $z \in \mathbb{C}$.

- The **complex conjugate** of $z \in \mathbb{C}$, denoted by $\bar{z}$, is defined by $\bar{z} = \Re z - \Im z i$
- The **absolute value** of z, denoted by $|z|$, is defined by $|z| = \sqrt{(\Re z)^2 + (\Im z)^2}$

We note that a scalar in $\mathbb{C}$ is analogous to a vector in $\mathbb{R}^2$, as the real and imaginary parts are real values. Hence $\mathbb{C}$ is the two-dimensional real-vector space over $\mathbb{R}$, but the one-dimensional complex vector space over $\mathbb{C}$. There are various properties of the complex numbers:

Theorem 4.1.3: Properties of Complex Numbers

- **Sum of z and $\bar{z}$:** $z + \bar{z} = 2\Re z$
- **Difference of z and $\bar{z}$:** $z - \bar{z} = 2(\Im z)i$
- **Product of z and $\bar{z}$:** $z\bar{z} = (\Re z)^2 + (\Im z)^2 = |z|^2$
- **Additivity and multiplicativity of complex conjugate:** $\overline{z+w} = \bar{z} + \bar{w}$ and $\overline{zw} = \bar{z}\bar{w}$
- **Double comlex conjugate:** $\bar{\bar{z}} = z$
- **Real and imaginary parts are bounded by $|z|$:** $\Re z| \leq |z|$ and $|\Im z| \leq |z|$.
- **Absolute value of the complex conjugate:** $|\overline{z}| = |z|$
- **Multiplicativity of absolute value:** $|zw| = |z||w|$
- **Triangle inequality:** $|z + w| \leq |z| + |w|$

Proof. We show how to prove the triangle inequality. Note that:

$$\begin{aligned}
|w+z|^2 &= (w+z)(\overline{w}+\overline{z}) \\
&= w\overline{w} + w\overline{z} + z\overline{w} + z\overline{z} \\
&= |w|^2 + |z|^2 + w\overline{z} + z\overline{w} \\
&= |w|^2 + |z|^2 + w\overline{z} + \overline{w\overline{z}} \\
&= |w|^2 + |z|^2 + 2\Re(w\overline{z}) \\
&\leq |w|^2 + |z|^2 + 2|w\overline{z}| \\
&= |w|^2 + |z|^2 + 2|w||z| \\
&= (|w| + |z|)^2
\end{aligned}$$

Taking the square root of both sides gives the triangle inequality:

$$|w + z| \leq |w| + |z|$$

□

4.2 Zeros of Polynomials

A function $p : \mathbb{F} \to \mathbb{F}$ is called a polynomial of degree m if there exists $a_0, \ldots, a_m \in \mathbb{F}$ with $a_m \neq 0$ such that:

$$p(z) = a_0 + a_1 z + \cdots + a_m z^m$$

for all $z \in \mathbb{F}$. We must show that the degree of a polynomial is unique, which is the same thing as showing if the representation of p as a linear combination of powers of z is unique. We first define solutions to the equation $p(z) = 0$.

Definition 4.2.1: zero of a polynomial

A number $\lambda \in \mathbb{F}$ is called a **zero** of the polynomial $p \in \mathbb{P}(\mathbb{F})$ if $p(\lambda) = 0$.

In fact, we can show that a polynomial p can be factored by monomials corresponding to each zero.

Theorem 4.2.2: each zero of a polynomial corresponds to a degree-one factor

Suppose m is a positive integer and $p \in \mathcal{P}(\mathbb{F})$ is a polynomial of degree m. Then $\lambda \in \mathbb{F}$ is a zero of p if and only if there exists some polynomial $q \in \mathcal{P}(\mathbb{F})$ of degree $m-1$ such that

$$p(z) = (z - \lambda)q(z)$$

for every $\lambda \in \mathbb{F}$

Proof. Suppose that $p(\lambda) = 0$. Let $a_0, a_1, \ldots, a_m \in \mathbb{F}$ be such that:

$$p(z) = a_0 + a_1 z + \cdots + a_m z^m$$

for all $z \in \mathbb{F}$. We see that:

$$p(z) = p(z) - p(\lambda) = a_1(z - \lambda) + a_2(z^2 - \lambda^2) + \cdots + a_m(z^m - \lambda^m)$$

We can use a clever factorization of the difference of powers, where:

$$z^n - \lambda^n = (z - \lambda)(z^{n-1} + z^{n-2}\lambda + \cdots + \lambda^{n-1}) = \sum_{k=0}^{n-1} z^{n-1-k}\lambda^k$$

Hence, we see that we can factor out a $(z - \lambda)$ out of every term in $p(z) = p(z) - p(\lambda)$, we see that $p(z)$ can be written as the product of $(z - \lambda)$ and another polynomial $q(z)$, where

$$q(z) = \sum_{k=1}^{m} a_k \sum_{j=0}^{k-1} z^{k-1-j}\lambda^j,$$

which is a polynomial of degree at most $m - 1$, we obtain

$$p(z) = (z - \lambda)q(z),$$

as required. □ □

There is a limit on the number of zeros that a polynomial can have. This limit is the degree of the polynomial.

Theorem 4.2.3: degree m implies at most m zeros

Suppose m is an positive integer and $p \in \mathcal{P}(\mathbb{F})$ is a polynomial of degree m. Then p has at most m zeros in $\mathbb{F}$.

Proof. We prove via induction on m. Our base case is $m = 0$, which means p is a constant polynomial. If p is non-zero, then p has no roots, so our claim is true. However, if p is 0, then every scalar value is a root of p, so we say $p = 0$ has degree $-\infty$. Assume that our claim holds for polynomials of degree $m = k$. Consider a polynomial p of degree $m = k + 1$. If m has no roots, then we are done. Otherwise, $\exists \lambda \in \mathbb{F}$ such that $p(\lambda) = 0$. We may factor out $(z - \lambda)$ from p to get $p = (z - \lambda)q$, where q is a polynomial of degree k, which by our induction hypothesis has at most k roots. Hence p has at most $k + 1$ roots. □

The above theorem gives us the following corrolary:

Theorem 4.2.4: Polynomial has unique coefficients

The coefficients of a polynomial uniquely determine the polynomial.

Proof. Suppose p and q are polynomials of degree m such that $p(z) = q(z)$ for all $z \in \mathbb{F}$, and p and q have different coefficients. We see that $p - q$ would yield a non-zero polynomial, but this polynomial would have infinitely many 0s, which is a contradiction. So a polynomial must be uniquely identified by its coefficients. □

4.2.1 Division Algorithm for Polynomials

If p and s are nonnegative integers and $s \neq 0$, then there are non-negative q and r such that $p = sq + r$. The same idea holds for polynomials. We invent a basis of $\mathcal{P}_n\mathbb{F}$

Theorem 4.2.5: Division algorithm for polynomials

Suppose that $p, s \in \mathcal{P}(\mathbb{F})$, with $s \neq 0$. Then there exist unique polyomials $q, r \in \mathcal{P}\mathbb{F}$ such that:

$$p = sq + r$$

and $\deg r < \deg s$.

Proof. Let $n = \deg p$ and let $m = \deg s$. If $n \leq m$, then take $q = 0$ and $r = p$ to get the equation $p = 0 + p$ where $\deg r < \deg s$. Now assume $n \geq m$. We assert that the following list is linearly independent $\mathcal{P}_n(\mathbb{F})$:

$$1, z, \ldots, z^{m-1}, s, sz, \ldots, sz^{n-m}$$

This is because each polynomial in this list has a different degree. Moreover, this list has an overall length of $n + 1$, so it serves as a basis of $\mathcal{P}_n(\mathbb{F})$. Because $p \in \mathcal{P}_n(\mathbb{F})$, there exist some set of constants $a_0, a_1, \ldots, a_{m-1} \in \mathbb{F}$ and $b_0, b_1, \ldots, b_{n-m} \in \mathbb{F}$ such that:

$$p = a_0 + a_1 z + \cdots + a_{m-1} z^{m-1} + b_0 s + b_1 sz + \cdots + b_{n-m} sz^{n-m}$$

We now pick:

$$r = a_0 + a_1 z + \cdots + a_{m-1} z^{m-1}$$

and

$$q = b_0 + b_1 z + \cdots + b_{n-m} z^{n-m}$$

We see that $p = sq + r$, where $\deg r < \deg s$. The uniqueness properties follows from the fact that q and r have unique representations in the basis twe constructed. □

4.2.2 Factorization of Polynomials over $\mathbb{C}$

The behavior of polynomial factorization is different over $\mathbb{R}$ and $\mathbb{C}$. We can use the results over the complex plane the prove similar properties in the real plane. To do this, we introduce the **Fundamental Theorem of Algebra**. This theorem is an existance theorem, which does not provide a means for finding zeros of a polynomial.

Theorem 4.2.6: Fundamental theorem of algebra

Every nonconstant polynomial with complex coefficients has a zero in $\mathbb{C}$.

Proof. See advanced textbooks on complex analysis for various proofs. □

Using the fundamental theorem of algebra, we can prove that every polynomial over $\mathbb{C}$ can be factored into degree-one polynomials. Now, we define the concept of irreducible polynomials over $\mathbb{F}$.

Definition 4.2.7: Irreducible polynomial

A nonconstant polynomial $p \in \mathcal{P}(\mathbb{F})$ is called **irreducible** over $\mathbb{F}$ if whenever there are polynomials $q, r \in \mathcal{P}(\mathbb{F})$ such that $p = qr$, then either $\deg q = 0$ or $\deg r = 0$. They do not factor into two non-constant polynomials. They are analogous to prime numbers in the integers.

More specficially, when we work with irreducible polyomials in the context of factoring, we are interested in **monic** irreducible polynomials, which are irreducible polynomials with leading coefficients 1. The following theorem shows that the only irreducible polynomials over $\mathbb{C}$ are degree-one polynomials.

Theorem 4.2.8: Irreducible polynomials over $\mathbb{C}$ are degree one

If $p \in \mathcal{P}(\mathbb{C})$ is a nonconstant polynomial, then p has a unique factorization (except for the order of the factors) of the form:

$$p(z) = c(z - \lambda_1)(z - \lambda_2) \cdots (z - \lambda_m),$$

where $c, \lambda_1, \ldots, \lambda_m \in \mathbb{C}$ and $m = \deg p$.

Proof. We prove this with induction on $m = \deg p$. For the base case, $m = 1$, which is trivial as we already have a degree-one polynomial. Assume that our claim holds for

polynomials of degree $m = k$. Now consider a polynomial p of degree $m = k + 1$. By the fundamental theorem of algebra, we know that p has at least one root $\lambda_1 \in \mathbb{C}$. We may factor out this root to get:

$$p(z) = (z - \lambda_1)q(z)$$

where q is a polynomial of degree k. By our induction hypothesis, we know that q can be factored as:

$$q(z) = c(z - \lambda_2)(z - \lambda_3) \cdots (z - \lambda_{k+1})$$

for some $\lambda_2, \ldots, \lambda_{k+1} \in \mathbb{C}$. Hence, we have:

$$p(z) = c(z - \lambda_1)(z - \lambda_2) \cdots (z - \lambda_{k+1})$$

as desired. The constant term in the factorization comes from the leading coefficient of p. Now to show that this factorization is unique, suppose there was some other factorization with roots $\tau_1, \ldots, \tau_{k+1} \in \mathbb{C}$. For each λ_k, we have that $p(\lambda_k) = 0$, so one of the τ_j must be equal to λ_k. By repeating this argument for all λ_k, we see that the two factorizations are the same, except for the order of the factors. □

4.3 Factorization of Polynomials over $\mathbb{R}$

A polynomial with real coefficients may have no real zeros (Consider $x^2 + 1$). We derive the following theorem:

Theorem 4.3.1: Polynomials with real coefficients have nonreal zeros in pairs

Suppose $p \in \mathcal{P}(\mathbb{C})$ is a polynomial with real coefficients. If $\lambda \in \mathbb{C}$ is a zero of p, then we know $\overline{\lambda}$ is also a zero.

Proof. By our properties of complex numbers, we know that:

$$\overline{p(\lambda)} = \overline{a_0 + a_1\lambda + \cdots + a_m\lambda^m} = a_0 + a_1\overline{\lambda} + \cdots + a_m\overline{\lambda}^m = p(\overline{\lambda})$$

Which is the desired result. □

The property of polynomials over $\mathbb{R}$ differs from those over $\mathbb{C}$ because of the existence of non-real zeros. Intuitively though, since non-real zeros come in complex-conjugate pairs, their product is always a quadratic polynomial with real coefficients. This hints at the property that the monic irreducible polynomials over $\mathbb{R}$ consists of degree-one terms like before, and certain degree-two terms. We formalize this in the following theorem:

Theorem 4.3.2: Factorization of a quadratic polynomial

Suppose $b, c \in \mathbb{R}$. Then there is a polynomial factorization of the form:

$$x^2 + bx + c = (x - \lambda_1)(x - \lambda_2)$$

with $\lambda_1, \lambda_2 \in \mathbb{R}$ if and only if $b^2 \geq 4c$.

Proof. This follows from the quadratic formula. Note that for any quadratic polynomial $x^2 + bx + c$:

$$x^2 + bx + c = \left(x + \frac{b}{2}\right)^2 + \left(c - \frac{b^2}{4}\right)$$

Suppose that $b^2 < 4c$, then the right side of the equation is positive for all $x \in \mathbb{R}$, so there are no real zeros of the polynomial, so it cannot be factored into two degree-one polynomials. Conversely, suppose that $b^2 \geq 4c$. Then the term $\frac{b^2}{4} - c$ is now positive, which implies that there exists some real number d such that $d^2 = \frac{b^2}{4} - c$. Hence, we have:

$$\begin{aligned} x^2 + bx + c &= \left(x + \frac{b}{2}\right)^2 - d^2 \\ &= \left(x + \frac{b}{2} + d\right) + \left(x + \frac{b}{2} - d\right) \end{aligned}$$

Which provides the factorization we wanted □

To show that the only irreducible polynomials over the reals are monic linear terms and certain quadratics, we prove that all polynomials over the reals have a unique factorization into the irreducible polynomials.

Theorem 4.3.3: factorization of a polynomial over $\mathbb{R}$

Suppose $p \in \mathcal{P}(\mathbb{R})$ is a nonconstant polynomial. Then p has a unique factorization (except for the order of the factors) of the form

$$p(x) = c(x - \lambda_1) \cdots (x - \lambda_m)(x^2 + b_1 x + c_1) \cdots (x^2 + b_M x + c_M)$$

where $c, \lambda_1, \ldots, \lambda_m, b_1, \ldots, b_M, c_1, \ldots, c_M \in \mathbb{R}$, with $b_k^2 < 4c_k$ for each k.

Proof. First consider the polynomial as an element of $\mathcal{P}(()\mathbb{C})$. In the case that all complex roots are real, then we have what we want. Otherwise, suppose p has some complex root λ, then by the complex conjugate root theorem, we have that $\overline{\lambda}$ is also a root of p. Hence:

$$p(x) = (x - \lambda)(x + \lambda)q(x) = (x^2 - 2(\Re\lambda)x + |\lambda|^2)q(x)$$

Where $\deg q = \deg p - 2$. Proving that q has real-coefficients would complete this inductive proof. q may be written as:

$$q(x) = \frac{p(x)}{x^2 - 2(\Re\lambda)x + |\lambda|^2}$$

for all $x \in \mathbb{R}$. Since the numerator and the denominator are both real, this implies that $q(x) \in \mathbb{R}$ for all $x \in \mathbb{R}$. We may write q also as:

$$q(x) = a_0 + a_1 x + \cdots + a_{n-2} x^{n-2}$$

where $n = \deg p$ and $a_0, \ldots, a_{n-2} \in \mathbb{C}$. Combining the two, we see that for all $x \in \mathbb{R}$.

$$\Im q(x) = 0 = \Im a_0 + (\Im a_1)x + \cdots + (\Im a_2)x^{n-2}$$

Hence $\Im q(x)$ must be the zero polynomial, so all coefficients of q must be real.

Now to show that q is unique, we just need to note that an irreducible quadratic in the reals can be uniquely factored as $(x - \lambda_k)(x + \lambda_k)$. The rest follows... $\square$

Chapter 5

Eigenvalues and Eigenvectors

5.1 Invariant Subspaces

5.1.1 Eigenvalues

Definition 5.1.1: Operator

A linear map from a vector sapce to itself is an **operator**

Definition 5.1.2: Restriction of a linear operator, $T_{|U}$

Suppose $T \in \mathcal{L}(V)$, and U is a subspace of V. Then the restriction of T on U is the map $T_{|U} \in \mathcal{L}(U)$ such that $T_{|U}u = Tu$ for all $u \in U$

The restriction of a linear operator may make it easier to analyze a map as we work with a smaller subspace. However, the results tend not to be helpful if the restriction of the operator may map outside of the subspace. Things are much easier with subspaces such that T maps all elements of the subspace back into the subspace.

Definition 5.1.3: Invariant Subspace

Suppose $T \in \mathcal{L}(V)$. A subspace U of V is **invariant** under T if $Tu \in U$ for all $u \in U$. That is, $T_{|U}$ is an operator on U

There are four major invariant subspaces that are nice to know. For any $T \in \mathcal{L}(V)$, we have the following invariant subspaces:

$\{0\}$	Invariant under T as $T0 = 0$
V	Invariant under T as $Tv = V$ for all $v \in V$
$\operatorname{null} T$	Invariant under T as $Tv = 0$ for all $v \in \operatorname{null} T$ and $0 \in \operatorname{null} T$
$\operatorname{range} T$	Invariant under T as $Tv \in \operatorname{range} T$ for all v

One of the simplest possible invariant subspaces we could have is an invariant subspace of

dimension 1. This can be constructed by taking any nonzero $v \in V$ and construct U as:

$$U = \{\lambda v : \lambda \in \mathbb{F}\} = \text{span}(v)$$

We see that U is an one-dimensional subspace of V. If U is a T invariant subspace for some $T \in \mathcal{L}(V)$, then there exists some scalar such that for all $u \in U$:

$$Tu = \lambda u$$

This gives us the concept of an **eigenvalue**

Definition 5.1.4: eigenvalue

Suppose $T \in \mathcal{L}(V)$. A number $\lambda \in \mathbb{F}$ is an **eigenvalue** of T if there exists $v \in V$ such that $v \neq 0$ and $Tv = \lambda v$.

The aforementioned equation gives us a lot of linear-algebraic methods to verify if any scalar is a eigenvalue of an operator T.

Theorem 5.1.5: Equivalent conditions to be an eigenvalue

Suppose V is finite-dimensional, $T \in \mathcal{L}(V)$, and $\lambda \in \mathbb{F}$. Then the following are equivalent.

(a) λ is an eigenvalue of T.

(b) $T - \lambda I$ is not injective.

(c) $T - \lambda I$ is not surjective.

(d) $T - \lambda I$ is not invertible.

Proof. Suppose λ was an eigenvalue of T, then by definition, there exists some scalar $\lambda \in \mathbb{F}$ such that $Tv = \lambda v$. Note that $v = Iv$, so $Tv = \lambda Iv$. Rearranding gives us the desired quantity of:

$$Tv - \lambda Iv = 0 \implies (T - \lambda I)v = 0$$

Hence, $v \in \text{null}\, T - \lambda I$, so $T - \lambda I$ is not injective. Now $T - \lambda I$ is also an operator on V, so non-injectivity also implies non-surjectivity, which finally implies non-invertibility. □

Definition 5.1.6: eigenvector

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

The following is an important property of eigenvectors that will enable us to make claims and statements about spaces spanned by eigenvectors.

Theorem 5.1.7: Linearly independent eigenvectors

Suppose $T \in \mathcal{L}(V)$. Then every list of eigenvectors of T corresponding to distinct eigenvalues of T is linearly independent.

Proof. Suppose that the claim is false. Then there exists some smallest integer m such that there is a linearly dependent list of eigenvectors $v_1, \ldots, v_n$ corresponding to eigenvalues $\lambda_1, \ldots, \lambda_n$ of T. Now suppose:

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

where none of the a_ks are 0 (as m is minimal). Now apply the operator $(T - \lambda_m I)$ onto both sides of the equality, and we get:

$$a_1(\lambda_1 - \lambda_m)v_1 + \cdots + a_{m-1}(\lambda_1 - \lambda_m)v_{m-1} = 0$$

Which is a linear combination of $m - 1$ vectors such that all coefficients are non-zero (as eigenvalues are distinct) and equals 0. This is a contradiction to the minimality of m, so our claim must have been true. □

A direct corrolary of the above theorem provides an upperbound on the number of eigenvalues an operator may have:

Theorem 5.1.8: Operator cannot have more eigenvalues than dimension

Suppose V is finite-dimensional, then any operator $T \in \mathcal{L}(V)$ may have at most $\dim V$ distinct eigenvalues.

Proof. Suppose T had n distinct eigenvalues where $n > \dim V$. Then this implies there is some list of n eigenvalues that are linearly independent, which is a contradiction. □

5.2 Polynomials Applied to Operators

Because an operator always maps back to the original vector space, we can consider powers of operators, and combined with linearity, this allows us to consider applying polynomials to operators.

Definition 5.2.1: T^m

Suppose $T \in \mathcal{L}(V)$ and M is a positive integer.

- $T^m \in \mathcal{L}(V)$ is deinfed by $T^m = \underbrace{T \cdots T}_{m \text{ times}}$
- T^0 is defined to be the identity operator I on V.

- If T is invertible with inverse T^{-1}, then $T^{-m} \in \mathcal{L}(V)$ is defined by

$$T^{-m} = (T^{-1})^m$$

Definition 5.2.2: $p(T)$

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

$$p(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_m z^m$$

for all $z \in \mathbb{F}$. Then $p(T)$ is the operator on V defined by:

$$p(T) = a_0 I + a_1 T + \cdots + a_m T^m$$

Polynomials are nice in that the map from p to $p(T)$ is a linear map. In addition, the product of polynomials also behave nicely:

Definition 5.2.3: 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 all $z \in \mathbb{F}$.

The following gives theorem gives us an important yet powerful idea of the invariant subspaces of T:

Theorem 5.2.4: Null space and range or $p(T)$ are invariant under T

Suppose $T \in \mathcal{L}(V)$ and $p \in \mathcal{P}(\mathbb{F})$. THen $\operatorname{null} p(T)$ and $\operatorname{range} p(T)$ are invariant under T.

Proof. Suppose $u \in \operatorname{null} p(T)$. Then $p(T)u = 0$. Hence:

$$(p(T))\,(Tu) = (p(T)T)\,(u) = (Tp(T))\,(u) = T\,(p(T)u) = T(0) = 0$$

Hence $Tu \in \operatorname{null} p(T)$. Thus $\operatorname{null} p(T)$ is invariant under T, as desired. Suppose $u \in \operatorname{range} p(T)$. Then there is $v \in V$ such that $u = p(T)v$:

$$Tu = T(p(T)v) = p(T)(Tv)$$

Hence $Tu \in \operatorname{range} p(T)$. Hence, $\operatorname{range} p(T)$ is T invariant. □

5.3 The Minimal Polynomial

5.3.1 Existence of Eigenvalues on Complex Vector Spaces

The core theorem of this section is

Theorem 5.3.1: Existence of Eigenvalues

Every operator on a finite-dimensional nonzero complex vector space has an eigenvalue

Proof. Suppose V is a finite-dimensional complex vector space of dimension $n > 0$ and $T \in \mathcal{L}(V)$. Pick any nonzero $v \in V$, we have:

$$v, Tv, T^2v, \ldots, T^nv$$

This list is not linearly independent as there $n+1$ vectors. This means that some non-trivial linear combination of vectors above would yield 0. Let p be the nonconstant polynomial of the smallest such degree:

$$p(T)v = 0$$

By the fundamental theorem of algebra, $p(z)$ must have some root λ such that $p(z) = 0$. Hence, p may be factored into some:

$$p(z) = (z - \lambda)q(z)$$

This implies that:

$$p(T)v = (T - \lambda I)q(T)v = 0$$

Since $p(T)$ is minimal, $q(T)v$ must be non-zero, which implies that $(T-\lambda I)$ must send $q(T)v$ to 0, so λ is an eigenvalue of T with eigenvector $q(T)v$ □

A counter example to this would be in an inifite dimensional space, we may define a map such that it always maps a vector to some other vector that is not a scalar multiple of itself.

5.3.2 Eigenvalues and the Minimal Polynomial

Definition 5.3.2: Monic Polynomial

A **monic polynomial** is a polynomial whose highest-degree coefficient equals 1.

Consider a map α that sends vectors from $\mathcal{P}(\mathbb{F}) \to \mathcal{L}(V)$ such that $\alpha(p) = p(T)$. we know that $\mathcal{P}(\mathbb{F})$ is an infinite-dimensional vector space, while $\mathcal{L}(V)$ is a finite-dimensional vector space with dimension $(\dim V)^2$. Hence the null space of this map is non-trivial (it is very big, infinitely big!). What can we say about this null space?

For any polynomial $p \in \operatorname{null} \alpha$, we have $p(T) = 0$. Let m be the minimal degree polynomial such that $m(T) = 0$. Since the field of polynomials is homomorphic, we know that for

any polynomial $q(z)$, $(mq)(T) = m(T)q(T) = 0$. Now take p. We can divide p by m to get:

$$p(z) = m(z)q(z) + r(z)$$

for some $q(z)$. Note that $m(T)q(T) = 0$, which implies that $r(T) = 0$. Hence, all polynomials that T satisfies is a multiple of this "minimal-degree" polynomial. We call this polynomial, the **minimal polynomial**

Definition 5.3.3: minimal polynomial

Suppose V is finite-dimensional and $T \in \mathcal{L}(V)$. Then the **minimal polynomial** of T is the unique monic polynomial $p \in \mathcal{P}(\mathbb{F})$ of smallest degree such that $p(T) = 0$.

Now we have a proposition. We argue that if we take the minimal polynomial to be monic, then it must be **unique**.

Proposition 5.3.4

The minimal polynomial is unique

For $T \in \mathcal{L}(V)$, there exist a unique monic minimal polynomial of the lowest degree such that $m(T) = 0$.

Proof. Take any minimal polynomial m and divide it by its lead coefficient. We know that it must then be unique, as a polynomial is uniquely defined by its coefficients. □

A key property of this polynomial is that its degree is bounded.

Theorem 5.3.5: $\deg m \leq \dim V$

For any $T \in \mathcal{L}(V)$, the minimal polynomial m of T has degree at most $\dim V$

Proof. We prove this with induction on $\dim V$. When V is zero-dimensional, it must be the zero vector space, hence, the minimal polynomial is $p(z) = 1$, which is a 0 degree polynomial. Now suppose our claim was true for $\dim V < n$. Consider a vector space V with dimension n. Consider the list of poynomials:

$$v, Tv, T^2v, \ldots, T^nv$$

This list is linearly dependent as V is n-dimensional, but the list has length $n + 1$. Hence, there exists some minimal index m such that T^mv can be written as a linear combination of $v, Tv, \ldots, T^{m-1}v$. Note that $U = \text{span}(v, Tv, \ldots, T^{m-1}v)$ is a T-invariant subspace of V, as applying T to any vector spanning vector $v, Tv, \ldots, T^{m-1}v$ would result in another vector in U, and $TT^{m-1}v = T^mv$ which is in the span as well. Now consider the restriction of T on U, $T_{|U}$. We claim that we can find a polynomial p such that $p(T_{|U}) = 0$. We know that for some $a_0, \ldots, a_{m-1}$:

$$a_0I + a_1T + a_2T^2 + \cdots + a_{m-1}T^{m-1} - T^m = 0$$

Hence, we can take $p(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_{m-1} z^{m-1} - z^m$. Hence $p(z)$ is a degree $m = \dim U$ polynomial that satisfies T, so the minimal polynomial of $T_{|U}$ has degree at most $\dim U$. Now consider the map $T_{V/U} : v + U \to Tv + U$. We know that V/U has dimension $\dim V - \dim U$, which by our induction hypothesis has some satisfying polynomial of T, q, such that $q(T)(v + U) = U$ for all $v \in V$, where deg $\leq \dim V - \dim V/U$.

Finally, we claim that $(pq)(T) = 0$, and has degree of $\leq \dim U + \dim V - \dim U = \dim V$. We know that for all vectors $v \in V$, $q(T)v = U$. Additionally, $p(u) = 0$ for any $u \in U$. Hence, $p(q(z))$ is a polynomial that safiesfies T for all v, and has degree at most $\dim V$. Hence, the minimal polynomial of T must also have degree of at most $\dim V$. □

THe following result is one of the most important theorems concerning the minimal polynomial. It will enable us to make statements about the very forms of the matrices of linear transformations. We observe that any root of a minimal polynomial must be an eigenvalue of the linear map of the minimal polynomial.

Theorem 5.3.6: Eigenvalues are the zeros of the minimal polynomial

Suppose V is finite-dimensinoal and $T \in \mathcal{L}(V)$.

(a) The zeros of the minimal polynomial of T are the eigenvalues of T.

(b) If V is a complex vector space, then the minimal polynomial of T has the form

$$(z - \lambda_1) \cdots (z - \lambda_m)$$

Where $\lambda_1, \ldots, \lambda_m$ is a list of all eigenvalues of T, possibly with repititions

Proof. Suppose p was the minimal polynomial of T. Furthermore, suppose $\lambda \in \mathbb{F}$ was a root of T. Then we may factor p as:

$$p(z) = (z - \lambda)q(z)$$

We know that $q(T) = 0$, so $(T - \lambda I)q(T)v$ for all $v \in V$. However, since we know that deg $q <$ deg p, there exists some vector $w \in V$ such that $q(T)v \neq 0$. Hence, $q(T)w \in \text{null}\, T - \lambda I$. This implies that $q(T)w$ is an eigenvector of T with eigenvalue λ.

We now show that every eigenvalue of T is a root of the minimal polynomial. Suppose $\lambda \in \mathbb{F}$ was an eigenvalue of T. There exists a nonzero $v \in V$ such that $Tv = \lambda v$. We see that:

$$p(T)v = p(\lambda)v = 0$$

Hence, $p(\lambda) = 0$, which implies λ is a root of p. To verify the factoring of p when V is a complex vector space, note that every polynomial in $\mathcal{P}(\mathbb{C})$ may be factored into linear irreducible terms. □

We are now ready to prove that all polynomials that is satisfied by T must be a multiple of the minimal polynomial of T.

Theorem 5.3.7: $q(T) = 0 \iff q$ is a polynomial multiple of the minimal polynomial

Suppose V is finite-dimensional, $T \in \mathcal{L}(V)$, 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. Let p be the minimal polynomial of T. Let q be any polynomial such that $q(T) = 0$. We know that we may divide q by p to yield:

$$q = ps + r$$

where $\deg r < \deg p$. We know that:

$$0 = q(T) = p(T)s(T) + r(T) = r(T)$$

Since p is the minimal polynomial, we know that r must be the zero polynomial as otherwise it would contradict the minimality of p. The other direction is easier to prove. If a polynomial q was a multiple of p. Then $q(T) = p(T)s(T) = 0$. □

From the above theorem, we derive a nice corollary:

Corollary 5.3.8

Suppose V is finite-dimensional, $T \in \mathcal{L}(V)$, and U is a subspace of V that is invariant under T. Then the minimal polynomial of T is a polynomial multiple of the minimal polynomial of $T_{|U}$ and is a polynomial multiple of the minimal polynomial of $T_{V/U}$.

Proof. If p was the minimial polynomial of T and q was the minimal polynomial of $T_{|U}$, then we know that $p(T_{|U}) = 0$. By our theorem above, p must be a mutliple of q. Moreover, take s to be the minimal polynomial of $T_{V/U}$ such that $s(T_{V/U}) = U$. This means that $s(T)v = 0$ for $v \in V - U$. Since $p(T)v = 0$, we know that p must be a product of s. □

Another property of the minimal polynomial is that it provides another way of telling the invertibility of a linear map:

Theorem 5.3.9: T not invertible $\iff$ the constant term of minimal polynomial is 0

Suppose $T \in \mathcal{L}(V)$ and p is the minimal polynomial of T. Then:

$$\begin{aligned} T \text{ is not invertible} &\iff 0 \text{ is an eigenvalue of } T \\ &\iff 0 \text{ is a zero of } p \\ &\iff \text{the constant term of } p \text{ is } 0 \end{aligned}$$

Finally,

5.3.3 Eigenvalues on Odd-Dimensional Real Vector Spaces

We cannot make general statements anything about the existence of eigenvectors on real vector spaces, but if we consider the restriction of odd-dimensional vector spaces, we can show that every such space has an eigenvalue.

Theorem 5.3.10: Even-dimensional Null Space

Suppose V is a finite-dimensional real vector space and $T \in \mathcal{L}(V)$ and $b, c \in \mathbb{R}$ such that $b^2 < 4c$. Then $\dim \operatorname{null} T^2 + bT + cI$ is an even number.

Proof. We know that $\operatorname{null} T^2 + bT + cI$ is a T-invariant subspace. Now, let us replace V with $\operatorname{null} T^2 + bT + cI$. We prove a stronger statement to help us prove the original theorem:

Claim:

If $T \in \mathcal{L}(V)$ where V is a finite-dimensional real vector space, and $p(z)$ is an irreducible quadratic polynomial over $\mathbb{R}$. If $p(T) = 0$, then $\dim V$ is even

Proof. Consider the map:

$$S := \frac{2T + b}{\sqrt{4c - b^2}}$$

Then:

$$S^2 = \frac{4T^2 + 4bT + b^2 I}{(4c - b^2)I}$$

Note that $T^2 + bT = -cI$, Hence, we have:

$$S^2 = \frac{-4cI + b^2 I}{(4c - b^2)I} = -I$$

Then, we have $S^2 + I = 0$, which means that S satisfies $x^2 + 1$. □

Now to finish up the proof, we supply another claim:

Claim:

Suppose that S is an operator on a finite-dimensional real vector space V such that $S^2 + I = 0$. Then $\dim V$ is even

Proof. On a real vector space V, the map S would behave as the complex number i. We can view V as a complex vector space by using S, by defining scalar multiplication of vectors such as:

$$(a + bi)v = av + biv = av + bSv$$

where a and b are integers. To finally show that V must be even dimension, we assert that $\dim_{\mathbb{R}} V = 2 \dim_{\mathbb{C}} V$. This is because for a basis $v_1, \ldots, v_d$ of V over $\mathbb{C}$, we have that $v_1, \ldots, v_d, iv_1, \ldots, iv_n$ is a basis of V over $\mathbb{R}$, so the dimension of V must be even. □

□

The following theorem gaurantees us eigenvalues on odd-dimensional vector spaces.

Theorem 5.3.11: Operators on odd-dimensional vector spaces have eigenvalues

Every operator on an odd-dimensional vector space has an eigenvalue

Proof. We prove this by induction on $\dim V$. Our base case is when $\dim V = 1$. In this case, for any $T \in \mathcal{L}(V)$, the singular basis vector of V is an eigenvalue of T. Now suppose that $\dim V = n \geq 3$ and that the claim holds for all odd dimensional spaces with dimension less than n. Let p be the minimal polynomial of T. Suppose that p can be factored by some linear irreducible $(x - \lambda)$ where $\lambda \in R$, then λ is an eigenvalue of T and we're done. Otherwise, p must be factored by irreducible quadratic $z^2 + bz + c$. Hence:

$$0 = p(T) = q(T)(T^2 + bT + cI)$$

We know that since p is the minimal polynomial of T, then $q(T) \neq 0$, so there must exist some vectors v such that $v \in \operatorname{null} T^2 + bT + cI$. Moreover, from the previous result, we know that $\operatorname{null} T^2 + bT + cI$ has even dimension, which implies that $\operatorname{range} T^2 + bT + cI$ must have odd dimension by rank nullity. We know that $\operatorname{range} T^2 + bT + cI$ must be a T invariant subspace of odd dimension smaller than V, so the restriction of T to $\operatorname{range} T^2 + bT + cI$ must have an eigenvalue by our hypothsis, so T must have an eigenvalue as well. □

5.4 Upper-Triangular Matrices

We typically refer to linear maps that map from a space to itself as **operators**. All operators have square matrices as the dimension of the domain and codomain are the same. A core concept in linear algebra is to always find a matrix for an operator that have a simple and nice form, often including many zeroes. One such form is the upper-triangular matrix, where all values below the diagonal of the matrix are 0.

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

There are a set of equivalent conditions that guarantees an operator has a upper-triangular matrix in some basis.

Theorem 5.4.1: Equivalent conditions for upper-triangular matrix

Suppose $T \in \mathcal{L}(V)$ and $v_1, \ldots, v_n$ is a basis of V. Then the following are equivalent:

(a) The matrix of T with respect to $v_1, \ldots, v_n$ is upper-triangular.

(b) $\text{span}(v_1, \ldots, v_k)$ is invariant under T for each $k = 1, \ldots, n$.

(c) $Tv_k \in \text{span}(v_1, \ldots, v_k)$ for each $k = 1, \ldots, n$.

Proof. To prove these are equivalent, we just need to show that each condition implies the next. Suppose that the matrix of T with respect to $v_1, \ldots, v_n$ is upper triangular. Now for any $k \in \{1, \ldots, m\}$, consider any vector $u \in \text{span}(v_1, \ldots, v_k)$. We see that:

$$u = a_1 v_1 + \cdots + a_k v_k$$

Now when we apply T onto u, we see that:

$$Tu = a_1 Tv_1 + \cdots + a_k Tv_k$$

by linearity. Note that since the matrix of T is upper-triangular, each basis vector is mapped to a linear combination of itself and all previous basis vectors, with no components in the subsequent basis vectors. Hence, $Tu \in \text{span}(v_1, \ldots, v_k)$, so $\text{span}(v_1, \ldots, v_k)$ is invariant under T.

Now suppose that $\text{span}(v_1, \ldots, v_k)$ is invariant under T for each $k = 1, \ldots, n$. Then trivially, we have that $v_k \in \text{span}(v_1, \ldots, v_k)$, so $Tv_k \in \text{span}(v_1, \ldots, v_k)$ for each $k = 1, \ldots, n$.

Finally, suppose that $Tv_k \in \text{span}(v_1, \ldots, v_k)$ for each $k = 1, \ldots, n$. Then we may write:

$$Tv_k = a_1 v_1 + \cdots + a_k v_k$$

for some scalars $a_1, \ldots, a_k \in \mathbb{F}$. Hence, the matrix of T with respect to $v_1, \ldots, v_n$ must be upper-triangular. □

The nice property of upper-triangular matrices is that their eigenvalues are nicely found on the diagonal. To show this, we first need to prove that an upper-triangular operator satisfies a certain polynomial composed of linear factors of the diagonal entries.

Theorem 5.4.2: Equation Satisfied by Operator with Upper-triangular Matrix

Suppose $T \in \mathcal{L}(V)$ and V has a basis with respect to which T has an upper-triangular matrix with diagonal entries $\lambda_1, \ldots, \lambda_n$. Then

$$(T - \lambda_1 I)(T - \lambda_2 I) \cdots (T - \lambda_n I) = 0$$

Proof. We prove this by indunction on n. When $n = 1$, we have that $T = \lambda_1 I$, so $(T - \lambda_1 T) = 0$. Now suppose our claim was true for $n - 1$. Consider the subspace $U =$

$\text{span}(v_1, \ldots, v_{n-1})$. We know that U is invariant under T, so consider the restriction of T on U, $T_{|U}$. By our induction hypothesis, we have:

$$(T_{|U} - \lambda_1 I)(T_{|U} - \lambda_2 I) \cdots (T_{|U} - \lambda_{n-1} I) = 0$$

Now consider any $v \in V$. We may write $v = u + av_n$ for some $u \in U$ and $a \in \mathbb{F}$. Note that:

$$(T - \lambda_n I)v = (T - \lambda_n I)(u + av_n) = (T - \lambda_n I)u + a(T - \lambda_n I)v_n$$

Note that $(T - \lambda_n I)u \in U$ as U is invariant under T. Additionally, $(T - \lambda_n I)v_n \in U$ as well, as the matrix of T is upper-triangular. Hence, $(T - \lambda_n I)v \in U$. Now applying the previous result, we have:

$$\begin{aligned}
&(T - \lambda_1 I)(T - \lambda_2 I) \cdots (T - \lambda_n I)v \\
&= (T - \lambda_1 I)(T - \lambda_2 I) \cdots (T - \lambda_{n-1} I)((T - \lambda_n I)v) \\
&= 0
\end{aligned}$$

as $(T - \lambda_n I)v \in U$. Hence, our claim holds. □

Intuitively, we may think of each term in the product polynomial as "eliminating" one dimension of the vector space until we reach the zero vector space. This is because each term $(T - \lambda_k I)$ would map v_k into the span of the previous basis vectors ($\text{span}(v_1, \ldots, v_{k-1})$), so after applying all n terms, we would reach the zero vector space. The map $(T - \lambda_k I)$ essentially zeros out the component of v_k in the direction of itself (the kth diagonal value).

From this, we may conclude that the eigenvalues of an upper-triangular matrix are exactly the diagonal entries.

Theorem 5.4.3: Eigenvalues of Upper-triangular Matrix

Suppose $T \in \mathcal{L}(V)$ has an upper-triangular matrix with respect to some basis of V. Then the eigenvalues of T are precisely the entries on the diagonal of that upper-triangular matrix.

Proof. Suppose that T was upper-triangular with respect to some basis $v_1, \ldots, v_n$ of V with diagonal entries $\lambda_1, \ldots, \lambda_n$. We see that $Tv_1 = \lambda_1 v_1$, so λ_1 is an eigenvalue of T. Now consider $k \in \{2, \ldots, n\}$. We know by the previous part that:

$$(T - \lambda_k I)v_k \in \text{span}(v_1, \ldots, v_{k_1})$$

Hence, the map $(T - \lambda_k I)$ maps $\text{span}(v_1, \ldots, v_k)$ to $\text{span}(v_1, \ldots, v_{k-1})$. Since $(T - \lambda_k I)$ maps from a space of dimension k to a space of dimension $k - 1$, it cannot be injective, so there exists some v such that:

$$(T - \lambda_k I)v = 0$$

Hence, λ_k is an eigenvalue of T with eigenvector v. Thus, all diagonal entries of T are eigenvalues of T. To show that T has no other eigenvalues, note that since T satisfies the polynomial:

$$p(z) = (z - \lambda_1)(z - \lambda_2) \cdots (z - \lambda_n)$$

p must also satisfy the minimal polynomial of T. Hence, any eigenvalue λ of T must be a root of p, so λ must be one of the λ_ks. □

With this theorem, we may now prove a core condition for an operator to have an upper-triangular matrix.

Theorem 5.4.4: Necessary and Sufficient Condition for Upper-triangular Matrix

Suppose V is finite-dimensional and $T \in \mathcal{L}(V)$. Then T has an upper-triangular matrix with respect to some basis of V if and only if the minimal polynomial of T equals $(z-\lambda_1)(z-\lambda_2)\cdots(z-\lambda_m)$ for some $\lambda_1, \ldots, \lambda_m \in \mathbb{F}$.

Proof. ($\Longrightarrow$) Suppose that T is upper-triangular with respect to some basis of V with diagonals $\alpha_1, \ldots, \alpha_n$. By the previous theorem, T must satisfy the polynomial $p(z) = \prod_{k=1}^{n}(z-\alpha_k)$. Since p must be a polynomial divisible by the minimal polynomial of T, the minimal polynomial of T must also be a product of linear factors.

($\Longleftarrow$) Suppose that the minimal polynomial of T is $p(z) = \prod_{k=1}^{m}(z-\lambda_k)$ for some $\lambda_1, \ldots, \lambda_m \in \mathbb{F}$. We prove T has an upper-triangular matrix with induction on m. When $m = 1$, we have that $T = \lambda_1 I$, so the matrix of T with respect to any basis of V is upper-triangular. Now suppose that for an arbitrary positive m, our claim was true. Let $U = \operatorname{range} T - \lambda_{m+1} I$. We know that U is invariant under T (as the kernel and range of any polynomial of an operator is invariant under that operator). Hence, the restriction of T to U, $T_{|U}$ is an operator on U.

Now consider $u \in U$, then we know that $u = (T - \lambda_{m+1}I)v$ for some $v \in V$. Hence, we hace:

$$\begin{aligned}&(T-\lambda_1 I)(T-\lambda_2 I)\cdots(T-\lambda_m I)u\\ &\quad = (T-\lambda_1 I)(T-\lambda_2 I)\cdots(T-\lambda_{m+1} I)v\\ &\quad = 0\end{aligned}$$

Hence, we see that $(T-\lambda_1 I)(T-\lambda_2 I)\cdots(T-\lambda_m I)$ is a polynomial that satisfies $T_{|U}$, so it must be a product of the minimal polynomial of $T_{|U}$. Hence, the minimal polynomial of $T_{|U}$ is the product of at most $m-1$ linear terms, so by our induction hypothesis, $T_{|U}$ has an upper-triangular matrix with respect to some basis $u_1, \ldots, u_M$ of U. This means that for $k = 1, \ldots, M$:

$$Tu_k \in \operatorname{span}(u_1, \ldots, u_k)$$

Now, we can extend the basis of U with some vectors $v_1, \ldots, v_N$ to form a basis of V. We claim that with respect to the basis $u_1, \ldots, u_M, v_1, \ldots, v_N$, T has an upper-triangular matrix. Note that for any $j \in \{1, \ldots, N\}$:

$$(T - \lambda_{m+1}I)v_j \in U = \operatorname{span}(u_1, \ldots, u_M)$$

Hence, we have:

$$Tv_j \in \operatorname{span}(u_1, \ldots, u_M, v_j)$$

Thus, for any $k \in \{1, \dots, N\}$, $Tv_k \in \text{span}(u_1, \dots, U_M, v_1, \dots, v_k)$, so by our earlier theorem, the matrix of T with respect to this basis is upper-triangular. □ □

Finally, combined with the result from the fundamental theorem of algebra, we have the following corollary:

Corollary 5.4.5

If $\mathbb{F} = \mathbb{C}$, then every operator on a finite-dimensional nonzero complex vector space has an upper-triangular matrix with respect to some basis of V.

Proof. For any $T \in \mathcal{L}(V)$, the minimal polynomial of T may be factored into linear terms over $\mathbb{C}$ by the fundamental theorem of algebra. Hence, by the previous theorem, T has an upper-triangular matrix with respect to some basis of V. □

5.5 Diagonalizable Operators

An operator T is said to be **diagonalizable** if there exists a basis of V with respect to which the matrix of T is diagonal.

Definition 5.5.1: Diagonalizable Operator

- An operator $T \in \mathcal{L}(V)$ is **diagonalizable** if there exists a basis of V with respect to which the matrix of T is diagonal.
- A matrix is **diagonal** if it is a square matrix that is 0 everywhere except on the main diagonal.

$$\begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

Following from the prevous chapter, any diagonal matrix is also an upper-triangular matrix, so the diagonals of an diagonal matrix are also the eigenvalues of the operator. To work with diagonal operators, we introduce the concept of a **eigenspace**.

Definition 5.5.2: Eigenspace, $E(\lambda, T)$

Suppose $T \in \mathcal{L}(V)$ and $\lambda \in \mathbb{F}$. The eigenspace of T corresponding to λ is the subspace $E(\lambda, T)$ of V defined by:

$$E(\lambda T) = (T - \lambda I) = \{v \in V : Tv = \lambda v\}$$

Hence, the eigenspace corresponding to λ is the set of all eigenvectors of T with eigenvalue λ, along with the zero vector.

Using the fact that eigenvectors of distinct eigenvalues are linearly independent, we may now prove a special property of eigenspaces:

Theorem 5.5.3: sum of eigenspaces is direct

Suppose $T \in \mathcal{L}(V)$ and $\lambda_1, \ldots, \lambda_m$ are distinct eigenvalues of T. Then

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

is a direct sum. Furthermore, if V is finite-dimensional, then

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

Proof. To prove that the sum is direct, suppose that $v_1 + \cdots + v_m = 0$ where $v_k \in E(\lambda_k, T)$ for each $k = 1, \ldots, m$. Since the eigenvectors of distinct eigenvalues are linearly independent, we know that each $v_k = 0$, so the sum is direct.

To show that the dimension is bounded, we suppose for the sake of contradiction that:

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

Suppose $E(\lambda_k, T)$ has dimension d_k, such that we may pick a basis of eigenvectors $v_1, \ldots, v_{d_k}$ for $E(\lambda_k, T)$. Then the list of all of these basis vectors for all eigenspaces would have length:

$$\sum_{k=1}^{m} d_k > \dim V$$

But this is a contradiction, as these vectors are linearly independent and cannot have length greater than $\dim V$. Hence, our claim holds. □

5.5.1 Conditions for Diagonalizability

Theorem 5.5.4: Equivalent conditions for diagonalizability

Suppose V is finite-dimensional and $T \in \mathcal{L}(V)$. Let $\lambda_1, \ldots, \lambda_m$ be distinct eigenvalues of T. Then the following are equivalent.

- T is diagonalizable.
- V has a basis consisting of eigenvectors of T.
- $V = \bigoplus_{k=1}^{m} E(\lambda_k, T)$.
- $\dim V = \sum_{k=1}^{m} \dim E(\lambda_k, T)$.

Proof. We prove that each condition implies the next.

(1 $\implies$ 2) Suppose that T is diagonal with respect to a basis $v_1, \ldots, v_n$ of V. This is true if and only if $Tv_k = \lambda_k v_k$ for each $k = 1, \ldots, n$ where each λ_k is the kth diagonal entry. Hence, each v_k is an eigenvector of T, so V has a basis consisting of eigenvectors of T.

(2 $\implies$ 3) Suppose that V has a basis consisting of eigenvectors of T. We may group these basis by their. corresponding eigenvalues to get:

$$V = E(\lambda_1, T) + E(\lambda_2, T) + \cdots + E(\lambda_m, T)$$

Since eigenvectors of distinct eigenvalues are linearly independent, we know that this sum is direct, so:

$$V = \bigoplus_{k=1}^{m} E(\lambda_k, T)$$

(3 $\implies$ 4) Suppose that $V = \bigoplus_{k=1}^{m} E(\lambda_k, T)$. Then by the dimension formula for direct sums, we have:

$$\dim V = \sum_{k=1}^{m} \dim E(\lambda_k, T)$$

(4 $\implies$ 1) Suppose that $\dim V = \sum_{k=1}^{m} \dim E(\lambda_k, T)$. We may pick a basis of eigenvectors for each eigenspace $E(\lambda_k, T)$, and combine all of these basis to form a basis of V. With respect to this basis, the matrix of T is diagonal, so T is diagonalizable. □

We have the following corollary from the previous result:

Corollary 5.5.5

Suppose V is finite-dimensional and $T \in \mathcal{L}(V)$ has $\dim V$ distinct eigenvalues, then T is diagonalizable.

Proof. If T has $\dim V$ distinct eigenvalues, then T has $\dim V$ linearly independent eigenvectors, so V has a basis consisting of eigenvectors of T, so by the previous theorem, T is diagonalizable. □

The result above is sufficient for diagonalizability, but NOT necessary, it's important to note the distinction that it's not an if and only if relation. There are more general conditions for diagonalizablility, such as the spectral theorem for real and complex inner product spaces, but we will cover them in a future chapter.

An advantage of diagonal matrices is that they allow us to efficiently compute high powers of operators. This is because for a diagonal matrix with diagonal entries $\lambda_1, \ldots, \lambda_n$, we have:

$$\begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}^k = \begin{pmatrix} \lambda_1^k & 0 & \cdots & 0 \\ 0 & \lambda_2^k & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n^k \end{pmatrix}$$

We now introduce a necessary and sufficient condition for diagonalizability in terms of the minimal polynomial of an operator.

Theorem 5.5.6: Necessary and Sufficient Condition for Diagonalizability

Suppose V is finite-dimensional and $T \in \mathcal{L}(V)$. Then T is diagonalzable if and only if the minimal polynomial of T has the form:

$$(z - \lambda_1)(z - \lambda_2) \cdots (z - \lambda_m)$$

For some distinct $\lambda_1, \ldots, \lambda_m \in \mathbb{F}$.

Proof. ($\Longrightarrow$) Suppose that T is diagonalizable, then V has a basis consisting of eigenvectors of T. Let $\lambda_1, \ldots, \lambda_m$ be the distinct eigenvalues of T, we know that for each vector v_j in the eigenbasis $v_1, \ldots, v_n$ of V, we have:

$$(T - \lambda_1 I)(T - \lambda_2 I) \cdots (T\lambda_m I)v_j = 0$$

Hence, the minimal polynomial of T is the product of linear factors corresponding to the distinct eigenvalues of T.

($\Longleftarrow$) Suppose that the minimal polynomial of T equals $(z-\lambda_1)(z-\lambda_2)\cdots(z-\lambda_m)$ for some distinct $\lambda_1, \ldots, \lambda_m \in \mathbb{F}$. We will show that T is diagonal by induction on m. When $m-1$, we have that $(T - \lambda_1 I) = 0$, so $T = \lambda_1 I$, which is diagonalizable. Now suppose that the claim was true for $m-1$ and all smaller values. Consider the subspace $U = \operatorname{range} T - \lambda_m I$, which is invariant under T. Like before, we see that for any $u \in U$, there is a $v \in V$ such that $u = (T - \lambda_m I)v$, so we have:

$$\begin{aligned}
&(T - \lambda_1 I)(T - \lambda_2 I) \cdots (T - \lambda_{m-1} I)u \\
&= (T - \lambda_1 I)(T - \lambda_2 I) \cdots (T - \lambda_m I)v \\
&= 0
\end{aligned}$$

So we know that $(T - \lambda_1 I)(T - \lambda_2 I) \cdots (T - \lambda_{m-1} I)$ is a polynomial that satisfies $T_{|U}$, so it must be a multiple of the minimal polynomial of $T_{|U}$. Hence, the minimal polynomial of $T_{|U}$ is the product of at most $m-1$ linear terms, so by our induction hypothesis, $T_{|U}$ is diagonalizable. This means that there exists a basis $u_1, \ldots, u_M$ of U consisting of eigenvectors of $T_{|U}$, which are also eigenvectors of T. Now consider the subspace $W = \operatorname{null} T - \lambda_m I$. W contains all the eigenvectors of T corresponding to eigenvalue λ_m. We claim that the sum $U + W$ is direct and spans V. To see that the sum is direct, suppose some vector $v \in U \cap W$. Since $v \in W$, we have $Tv = \lambda_m v$. Since $v \in U$, we have:

$$\begin{aligned}
0 &= (T - \lambda_1 I) \cdots (T - \lambda_{m-1} I)v \\
&= (\lambda_m - \lambda_1) \cdots (\lambda_m - \lambda_{m-1})v
\end{aligned}$$

Since the λ_ks are distinct, we see that $v = 0$, so the sum is direct. To see that the sum spans V, consider any $v \in V$. We may write:

$$u = (T - \lambda_m I)v \in U$$

Since U has a basis of eigenvectors of T, and W has a basis of eigenvectors of T, concatenating these two basis gives us a basis of V with eigenvectors of T, so T is diagonalizable. □

The diagonalization property carries onto invariant subspaces as well, which will be important when considering the simltaneous diagonalization of operators.

Corollary 5.5.7

Suppose V is finite-dimensional and $T \in \mathcal{L}(V)$ is diagonalizable. If U is a subspace of V that is invariant under T, then the restriction of T to U, $T_{|U}$ is also diagonalizable.

Proof. Since T is diagonalizable, the minimal polynomial of T is a product of distinct linear factors. Since U is invariant under T, the minimal polynomial of $T_{|U}$ must be a product of distinct linear factors as well, so by the previous theorem, $T_{|U}$ is diagonalizable. □

5.6 Commuting Operators

It's rare that two arbitrary operators commute, however, when they do, we may derive some nice properties from them.

Definition 5.6.1: Commute

- Two operators $S, T \in \mathcal{L}(V)$ are said to **commute** if $ST = TS$.
- Two square matrices A, B of the same size commute if $AB = BA$.

It should be no surprise that if two operators commute, then their matrices also commute:

Theorem 5.6.2: Commuting Operators have Commuting Matrices

Suppose V is finite-dimensional and $S, T \in \mathcal{L}(V)$. $ST = TS$ iff for any basis of V, the matrices of S and T with respect to that basis also commute.

Proof.

$$\begin{aligned} S \text{ and } T \text{ commute} &\iff ST = TS \\ &\iff \mathcal{M}(ST) = \mathcal{M}(TS) \\ &\iff \mathcal{M}(S)\mathcal{M}(T) = \mathcal{M}(T)\mathcal{M}(S) \\ &\iff \text{the matrices of } S \text{ and } T \text{ commute} \end{aligned}$$

□

A core property of commuting operators is that they preserve each other's eigenspaces. The eigenspace of an operator is invariant uner any operator that commutes with it.

Theorem 5.6.3: Eigenspaces are Invariant under Commuting Operators

Suppose V is finite-dimensional and $S, T \in \mathcal{L}(V)$ such that $ST = TS$ and $\lambda \in \mathbb{F}$. Then $E(\lambda, S)$ is invariant under T.

Proof. Take any vector $v \in E(\lambda, S)$, then we have:

$$Sv = \lambda v$$

Now consider Tv, we see that:

$$STv = TSv = T(\lambda v) = \lambda Tv$$

Hence, $Tv \in E(\lambda, S)$, so $E(\lambda, S)$ is invariant under T. □

This property allows us to prove a powerful result about commuting diagonalizable operators.

Theorem 5.6.4: Commuting Diagonalizable Operators are Simultaneously Diagonalizable

Two diagonalizable operators on the same vector space have diagonal matrices with respect to the same basis if and only if the two operators commute.

Proof. ($\Longrightarrow$) Suppose that $S, T \in \mathcal{L}(V)$ are diagonalizable with respect to the same basis of V. Then the matrices of S and T with respect to this basis are diagonal matrices, which commute. Hence, by our previous theorem, $ST = TS$.

($\Longleftarrow$) Suppose that $S, T \in \mathcal{L}(V)$ are diagonalizable and commute. Let $\lambda_1, \ldots, \lambda_m$ be the distinct eigenvalues of S. We know that since S is diagonalzable:

$$V = \bigoplus_{k=1}^{m} E(\lambda_k, S)$$

By the previous result, we know that each eigenspace $E(\lambda_k, S)$ is invariant under T. Hence, the restriction of T to each eigenspace $E(\lambda_k, S)$, $T_{|E(\lambda_k, S)}$ is diagonalizable as well. We may pick a basis of eigenvectors of $T_{|E(\lambda_k, S)}$ for each eigenspace $E(\lambda_k, S)$, and concatenating all of these basis gives us a basis of V with respect to which both S and T are diagonal. □

In the case of a complex vector space, commuting operators guarantee a common eigenvector.

Theorem 5.6.5: Common Eigenvector for Commuting Operators

Every pair of commuting operators on a finite-dimensional nonzero complex vector space has a common eigenvector.

Proof. Suppose V is a finited-imensional nonzero complex vector space and $S, T \in \mathcal{L}(V)$ commute. Let λ be an eigenvalue of S, which exists by the fundamental theorem of algebra. Consider the eigenspace $E(\lambda, S)$, which is nonzero, which is invariant under T. Now consider the restriction of T to $E(\lambda, S)$, $T_{|E(\lambda,S)}$. Since $E(\lambda, S)$ is a nonzero complex vector space, $T_{|E(\lambda,S)}$ must have an eigenvalue μ. Hence, there exists some nonzero vector $v \in E(\lambda, S)$ such that:

$$T_{|E(\lambda,S)}v = \mu v$$

Since $v \in E(\lambda, S)$, we have $Sv = \lambda v$. Hence, v is a common eigenvector of S and T □

Commutativity is so powerful, that it even provides us with a sufficient condition for simultaneous upper-triangularizability in complex vector spaces.

Theorem 5.6.6: Commuting Operators are Simultaneously Upper-triangularizable

Suppose V is a finite-dimensional complex vector space and S, T are commuting operators on V. Then there exists a basis of V with respect to which both S and T have upper-triangular matrices.

Proof. Let $n = \dim V$, we proceed with induction on n. When $n = 1$, the result is trivial as all 1×1 matrices are upper-triangular. Now suppose that our claim was true for $n - 1$. By the previous result, we know that S and T have some common eigenvector v_1. Hence, we have:

$$Sv_1 = \lambda v_1 \in \text{span}(v_1), \quad Tv_1 = \mu v_1 \in \text{span}(v_1)$$

for some $\lambda, \mu \in \mathbb{C}$. Now consider the subspace $W = \text{span}(v_1)^\perp$ such that $V = \text{span}(v_1) \oplus W$. Consider a **projection map** $P : V \to W$ such that:

$$P(av_1 + w) = w$$

for any $a \in \mathbb{C}$ and $w \in W$. Now consider the maps $\hat{S}, \hat{T} \in \mathcal{L}(W)$ such that:

$$\hat{S} = PS, \quad \hat{T} = PT$$

We claim that $\hat{S}$ and $\hat{T}$ commute. To see this, consider any $w \in W$, there exists some a such that:

$$(\hat{S}\hat{T})w = \hat{S}(PTw) = \hat{S}(Tw - av_1) = P(S(Tw - av_1)) = P(STw)$$

Likewise, we have:

$$(\hat{T}\hat{S})w = P(TSw)$$

Since S and T commute, we have $P(STw) = P(TSw)$, so $\hat{S}$ and $\hat{T}$ commute. By our induction hypothesis, there is some basis $v_2, \ldots, v_n$ of W such that $\hat{S}$ and $\hat{T}$ both have upper-triangular matrices with respect to this basis.We see that $v_1, \ldots, v_n$ is a basis of V. Now for any $k \in \{2, \ldots, n\}$, we have:

$$Sv_k = \hat{S}v_k + a_k v_1 \in \text{span}(v_1, \ldots, v_k), \quad Tv_k = \hat{T}v_k + b_k v_1 \in \text{span}(v_1, \ldots, v_k)$$

Since $\hat{S}$ and $\hat{T}$ are upper-triangular with respect to $v_2, \ldots, v_n$, we have that $Sv_k, Tv_k \in \text{span}(v_1, \ldots, v_k)$ for each $k = 1, \ldots, k$, Hence, the matrices of S and T with respect to the basis $v_1, \ldots, v_n$ are both upper-triangular. □

We have a final corollary that relates the eigenvalues of commuting operators.

Corollary 5.6.7

Suppose V is a finite-dimensional complex vector space and S, T are commuting operaors on V. Then:

- Every eigenvalue of $S + T$ is an eigenvalue of S plus an eigenvalue of T
- Every eigenvalue of ST is a product of an eigenvalue of S and an eigenvalue of T

Proof. By the previous theorem, there exists a basis $v_1, \ldots, v_n$ of V with respect to which both S and T have upper-triangular matrices. Let the diagonal entries of the matrix of S be $\lambda_1, \ldots, \lambda_n$ and the diagonal entries of the matrix of T be $\mu_1, \ldots, \mu_n$. Then the matrix of $S + T$ with respect to this basis has diagonal entries $\lambda_1 + \mu_1, \ldots, \lambda_n + \mu_n$, so by our earlier result, the eigenvalues of $S + T$ are precisely $\{\lambda_k + \mu_k : k = 1, \ldots, n\}$. Similarly, the matrix of ST with respect to this basis has diagonal entries $\lambda_1\mu_1, \ldots, \lambda_n\mu_n$, so the eigenvalues of ST are precisely $\{\lambda_k\mu_k : k = 1, \ldots, n\}$. □

Chapter 6

Inner Product Spaces

Inner product spaces connects the algebraic nature of vector spaces to the geometric nature of Eucliean space. The inner product spaces induces the idea of lengths, angles, and orthogonality in vector spaces. The properties of an inner product space are fundamental to optimization problems and machine learning.

6.1 Inner Products and Norms

To geometrically think of vectors, they may represent a direction and magnitude in space. The length of a vector is called its **norm** and is denoted as $||v||$ For an n-dimensional real vector space, the norm of a vector $v = (v_1, \ldots, v_n)$ is defined as:

$$||v|| = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}$$

The norm itself is not a linear operator, but dot product of two vectors is a bilinear operator:

Definition 6.1.1: Dot Product

For $x, y \in \mathbb{R}^n$, the **dot product** of x and y $(x \cdot y)$ is the map $\cdot : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ defined as:

$$x \cdot y = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n$$

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

The dot product has several important properties:

Theorem 6.1.2: Properties of the Dot Product

- $x \cdot x \geq 0$ for all $x \in \mathbb{R}^n$.
- $x \cdot x = 0$ if and only if $x = 0$.
- For a fixed $y \in \mathbb{R}^n$, the map $x \mapsto x \cdot y$ is linear.
- $x \cdot y = y \cdot x$ for all $x, y \in \mathbb{R}^n$.

The dot product motivates us to define a more general concept of a such a map that maps two vectors to a scalar.

Definition 6.1.3: 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:

- **positivity:** $\langle v, v\rangle \geq 0$ for all $v \in V$.
- **Definiteness:** $\langle v, v\rangle = 0$ if and only if $v = 0$.
- **Linearity in the first slot:** For all $u, v, w \in V$ and $a, b \in \mathbb{F}$, $\langle au + bv, w\rangle = a\langle u, w\rangle + b\langle v, w\rangle$.
- **Conjugate Symmetry:** For all $u, v \in V$, $\langle u, v\rangle = \overline{\langle v, u\rangle}$.

Definition 6.1.4: Inner Product Space

An **inner product space** is a vector space V along with an inner product on V.

We may derive the following properties from the definition of an inner product:

Theorem 6.1.5: Additional Properties of an inner product

For $u, v, w \in V$ and $\lambda \in \mathbb{F}$:

(a) $\langle 0, v\rangle = \langle v, 0\rangle = 0$ for every $v \in V$

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

(c) $\langle u, \lambda v\rangle = \overline{\lambda}\langle u, v\rangle$

Proof. (a) $\langle 0, v\rangle$ is a inner product whose second slot is fixed, so it is a linear map. Hence, $\langle 0, v\rangle = 0$. By conjugate symmetry, we have $\langle v, 0\rangle = \overline{\langle 0, v\rangle} = 0$.

(b) We have:

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

(c) We have:

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

□

6.1.1 Norms

Using the inner product, we may define the norm of a vector in an inner product space.

Definition 6.1.6: Norm, $||v||$

For $v \in V$, the *norm* of v, denoted by $||v||$ is defined by:

$$||v|| = \sqrt{\langle v, v \rangle}$$

Once again, we may derive some properties of the norm from the definition:

Theorem 6.1.7: Properties of the Norm

Let $v \in V$ and $\lambda \in \mathbb{F}$. Then:

(a) $||v|| = 0$ if and only if $v = 0$.

(b) $||\lambda v|| = |\lambda| \, ||v||$.

Proof. (a) $\langle v, v \rangle = 0$ if and only if $v = 0$, so $||v|| = 0$ if and only if $v = 0$.

(b) We have:

$$\begin{aligned} ||\lambda v|| &= \sqrt{\langle \lambda v, \lambda v \rangle} \\ &= \sqrt{\lambda \overline{\lambda} \, \langle v, v \rangle} \\ &= \sqrt{|\lambda|^2 \, \langle v, v \rangle} \\ &= |\lambda| \, ||v|| \end{aligned}$$

□

6.1.2 Orthogonality

A crucial concept in inner product spaces occurs when two vectors are perpendicular to each other, or **orthogonal**.

Definition 6.1.8: Orthogonal Vector

Two vectors $u, v \in V$ are said to be **orthogonal** if $\langle u, v \rangle = 0$.

We can actually relate the angle θ in $\mathbb{R}^2$ between two vectors to their inner product:

$$\langle u, v \rangle = ||u|| \, ||v|| \cos \theta$$

Hence, orthogonality results when two vectors have an angle of $90°$ between them. We note that the zero vector is orthogonal to every vector in V since $\langle 0, v \rangle = 0$ for all $v \in V$, and zero is the only vector orthogonal to itself. We now revisit a middle school theorem:

Theorem 6.1.9: Pythagorean Theorem

Suppose $u, v \in V$. If u and v are orthogonal, then:

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

Proof. Suppose $\langle u, v\rangle = 0$, then:

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

□

Orthogonality introduces an important way of decomposing a vector into two orthogonal components. Given a vector u and a nonzero vector v, we may decompose u into a component parallel to v and a component orthogonal to v. Doing so, we may write u as:

$$u = pv + (u - pv)$$

for some scalar $p \in FF$. We want p such that $u - pv$ is orthogonal to v, so we have:

$$\begin{aligned} \langle u - pv, v\rangle &= 0 \\ \langle u, v\rangle - p\langle v, v\rangle &= 0 \\ p &= \frac{\langle u, v\rangle}{\langle v, v\rangle} \\ p &= \frac{\langle u, v\rangle}{||v||^2} \end{aligned}$$

Theorem 6.1.10: Orthogonal Projections

Suppose $u, v \in V$ with $v \neq 0$. Take $p = \frac{\langle u,v\rangle}{||v||^2}$. Then u may be uniquely written as:

$$u = pv + (u - pv)$$

where pv is parallel to v and $u - pv$ is orthogonal to v such that $\langle u - pv, v\rangle = 0$.

The vector pv is called the **orthogonal projection** of u onto v. This decomposition is often useful in optimization problems, where we want to minimize the distance between two vectors. For now though, we will use it to prove the Cauchy-Schwarz inequality, which allows us to bound the inner product of two vectors in terms of their norms.

Theorem 6.1.11: Cauchy-Schwarz Inequality

Suppose $u, v \in V$. Then:

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

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

Proof. If we have $v = 0$, then both sides of the inequality is 0, so the inequality stands. Now consider the orthogonal decomposition of u:

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

Now by the pythagorean theorem, we have:

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

Hence, we get:

$$|\langle u, v\rangle|^2 \leq ||u||\,||v||^2$$

and taking the square root yields us the Cauchy-Schwarz inequality. Furthermore, equality holds if and only if $||w|| = 0$, which happens if and only if u is a scalar multiple of v. □

Using the Cauchy-Schwarz inequality, we proceed to prove another important inequality called the triangle inequality, whose importance allows it to be treated as an axiom for analysis and metric spaces.

Theorem 6.1.12: Triangle Inequality

Suppose $u, v \in V$. Then:

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

The inequality is an equality if and only if one of u, v is a nonegative real mutliple of the other.

Proof. We have:

$$\begin{aligned}
||u+v||^2 &= \langle u+v, u+v\rangle \\
&= \langle u,u\rangle + \langle u,v\rangle + \langle v,u\rangle + \langle v,v\rangle \\
&= ||u||^2 + 2\Re(\langle u,v\rangle) + ||v||^2 \\
&\leq ||u||^2 + 2|inneruv| + ||v||^2 \\
&\leq ||u||^2 + 2\,||u||\,||v|| + ||v||^2 \\
&= (||u|| + ||v||)^2
\end{aligned}$$

Taking the square root of both sides yields us the triangle inequality. Furthermore, equality holds if and only if $\Re(\langle u,v\rangle) = |\,\langle u,v\rangle\,| = ||u||\,||v||$, which happens if and only if one of u, v is a nonnegative real multiple of the other. □

The triangle inequality is fundamental to analysis, as it allows us to define the distance between two vectors in an inner product space. For us, it gives us a tool to prove a final inquality, called the parallelogram law.

Theorem 6.1.13: Parallelogram Law

Suppose $u, v \in V$. Then:

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

Proof. Note that:

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

□

6.2 Orthonormal Bases

In inner product spaces, we may define a special type of basis called an orthonormal basis. It is often nice to work with orthonormal basis because they simplify operations for inner products and norms.

Definition 6.2.1: Orthonormal Basis

- A set of vectors $v_1, \ldots, v_n$ in V is said to be **orthonormal** if $||v_k|| = 1$ for all $k = 1, \ldots, n$ and $\langle v_j, v_k\rangle = 0$ for all $j \neq k$.
- A basis of V is said to be an **orthonormal basis** if it is an orthonormal set.

We now go through several properties of orthonormal bases that make them useful.

Theorem 6.2.2: Norm of an Orthonormal Linear Combination

Suppose $e_1, \ldots, e_n$ is an orthonormal list of vectors in V. Then

$$||a_1 e_1 + \cdots + a_m e_m||^2 = |a_1|^2 + \cdots + |a_m|^2$$

Proof.

$$\begin{aligned} ||a_1 e_1 + \cdots + a_m e_m||^2 &= \langle a_1 e_1 + \cdots + a_m e_m, a_1 e_1 + \cdots + a_m e_m \rangle \\ &= \sum_{i,j} a_i \overline{a_j} \langle e_i, e_j \rangle \\ &= \sum_{k=1}^{m} a_k \overline{a_k} \langle e_k, e_k \rangle \\ &= \sum_{k=1}^{m} |a_k|^2 \end{aligned}$$

□

Theorem 6.2.3: Orthonormal lists are linearly independent

Every orthonormal list of vectors is linearly independent

Proof. Suppose $e_1, \ldots, e_m$ is an orthonormal list of vectors in V and $a_1, \ldots, a_m \in \mathbb{F}$ such that:

$$a_1 e_1 + \cdots + a_m e_m = 0$$

Then we have:

$$0 = ||a_1 e_1 + \cdots + a_m e_m||^2 = |a_1|^2 + \cdots + |a_m|^2$$

Hence, all $a_k = 0$, so $e_1, \ldots, e_m$ is linearly independent. □

We now introduce an inequality that bounds the sum of squares of projections of a vector onto an orthonormal list.

Theorem 6.2.4: Bessel's Inequality

Suppose $e_1, \ldots, e_m$ is an orthonormal list of vectors in V. If $v \in V$ then:

$$|\langle v, e_1 \rangle|^2 + \cdots + |\langle v, e_m \rangle|^2 \leq ||v||^2$$

Proof. Suppose $v \in V$. THen:

$$v = \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_m \rangle e_m + v - (\langle v, e_1 \rangle e_1 + \cdots + \langle v, e_m \rangle e_m)$$

Now let $u = \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_m \rangle e_m$ and $w = v - u$. We want to show that $||u||^2 \leq ||v||^2$. Note that u and w are orthogonal, as w is essentially the component of v orthogonal to the span of $e_1, \ldots, e_m$. We can verify this by checking that $\langle w, e_k \rangle = 0$ for all $k = 1, \ldots, m$.

Hence, by the Pythagorean theorem, we have:

$$\begin{aligned} ||u||^2 + ||w||^2 &= ||v||^2 \\ ||u||^2 &\leq ||v||^2 \\ |\langle v, e_1 \rangle|^2 + \cdots + |\langle v, e_m \rangle|^2 &\leq ||v||^2 \end{aligned}$$

□

It cold be useful to treat an orthonormal basis as the standard basis in $\mathbb{F}^n$. The following theorem allows us to do so, by defining what the coordinates of a vector are with respect to an orthonormal basis:

Theorem 6.2.5: Writing a vector as a linear combination of an orthonormal basis

Let $e_1, \ldots, e_n$ be an orthonormal basis of V and $u, v \in V$. Then:

- $v = \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_n \rangle e_n$
- $||v||^2 = |\langle v, e_1 \rangle|^2 + \cdots + |\langle v, e_n \rangle|^2$
- $\langle u, v \rangle = \overline{\langle v, e_1 \rangle} \langle u, e_1 \rangle + \cdots + \overline{\langle v, e_n \rangle} \langle u, e_n \rangle$

Proof. Since $e_1, \ldots, e_n$ is a basis of V, we may write:

$$v = a_1 e_1 + \cdots + a_n e_n$$

for some scalars $a_1, \ldots, a_n \in \mathbb{F}$. Now consider $\langle v, e_k \rangle$ for some $k \in \{1, \ldots, n\}$, we have:

$$\begin{aligned} \langle v, e_k \rangle &= \langle a_1 e_1 + \cdots + a_n e_n, e_k \rangle \\ &= a_1 \langle e_1, e_k \rangle + \cdots + a_n \langle e_n, e_k \rangle \\ &= a_k \end{aligned}$$

Hence, we have:

$$v = \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_n \rangle e_n$$

The other two results follow from the previous two theorems. □

We now introduce an important algorithm for converting any linearly independent list of vectors into an orthonormal list of vectors that span the same subspace.

Theorem 6.2.6: Gram-Schmidt Process

Suppose $v_1, \ldots, v_m$ is a linearly independent list of vectors in V. Let $f_1 = v_1$. For $k = 2, \ldots, m$, define f_k inductively by:

$$f_k = v_k - \frac{\langle v_k, f_1 \rangle}{||f_1||^2} f_1 - \cdots - \frac{\langle v_k, f_{k-1} \rangle}{||f_{k-1}||^2} f_{k-1}$$

For each $k-1, \ldots, m$, let $e_k = \frac{f_k}{||f_k||}$. Then $e_1, \ldots, e_m$ is an orthonormal list of vectors in V such that:

$$\text{span}(v_1, \ldots, v_k) = \text{span}(e_1, ldots, e_k)$$

Proof. We will prove the result by induction on k. When $k = 1$. we have that $f_1 = v_1$, so $e_1 = \frac{v_1}{||v_1||}$ is a unit vector, and also $\text{span}(v_1) = \text{span}(e_1)$. Now suppose $1 < k \leq m$ and $e_1, \ldots, e_{k-1}$ is an orthonormal list such that:

$$\text{span}(v_1, \ldots, _{k-1}) = \text{span}(e_1, \ldots, e_{k-1})$$

We want to show that e_k is orthogonal to each of $e_1, \ldots, e_{k-1}$ and that

$$\text{span}(v_1, \ldots, v_k) = \text{span}(e_1, \ldots, e_k)$$

For starters, we have:

$$f_k = v_k - \frac{\langle v_k, f_1 \rangle}{||f_1||^2} f_1 - \cdots - \frac{\langle v_k, f_{k-1} \rangle}{||f_{k-1}||^2} f_{k-1}$$

$$e_k = \frac{f_k}{||f_k||}$$

Now consider any $j \in \{1, \ldots, k-1\}$, we have:

$$\begin{aligned}
\langle e_k, e_j \rangle &= \left\langle \frac{f_k}{||f_k||}, \frac{f_j}{||f_j||} \right\rangle \\
&= \frac{1}{||f_k|| \, ||f_j||} \langle f_k, f_j \rangle \\
&= \frac{1}{||f_k|| \, ||f_j||} \left(\langle v_k, f_j \rangle - \sum_{i=1}^{k-1} \frac{\langle v_k, f_i \rangle}{||f_i||^2} \langle f_i, f_j \rangle \right) \\
&= \frac{1}{||f_k|| \, ||f_j||} \left(\langle v_k, f_j \rangle - \frac{\langle v_k, f_j \rangle}{||f_k||^2} \langle f_j, f_j \rangle \right) \\
&= \frac{1}{||f_k|| \, ||f_j||} \left(\langle v_k, f_j \rangle - \langle v_k, f_j \rangle \right) \\
&= 0
\end{aligned}$$

Hence f_k is orthogonal to each of $f_1, \ldots, f_{k-1}$, so e_k is orthogonal to each of $e_1, \ldots, e_{k-1}$. Now we want to show that:

$$\text{span}(v_1, \ldots, v_k) = \text{span}(e_1, \ldots, e_k)$$

Note that for any vector in $\text{span}(v_1, \ldots, v_k)$, we may express it as a sum of two vectors u and w where $u \in \text{span}(v_1, \ldots, v_{k-1})$ and $w \in \text{span}(v_k)$. By our induction hypothesis, we have $u \in \text{span}(e_1, \ldots, e_{k-1})$. Now consider $w \in \text{span}(v_k)$, we have:

$$v_k = f_k + \frac{\langle v_k, f_1 \rangle}{||f_1||^2} f_1 + \cdots + \frac{\langle v_k, f_{k-1} \rangle}{||f_{k-1}||^2} f_{k-1}$$

Hence $v_k \in \text{span}(e_1, \ldots, e_k)$, so $w \in \text{span}(e_1, \ldots, e_k)$. Therefore, we have:

$$\text{span}(v_1, \ldots, v_k) \subseteq \text{span}(e_1, \ldots, e_k)$$

Since the two spans have the same dimension, they must be equal. □

A direct corollary of the Gram-Schmidt process is that every finite-dimensional inner product space has an orthonormal basis.

Corollary 6.2.7

Every finite-dimensional inner product space has an orthonormal basis.

Proof. Take any basis of V and apply the Gram-Schmidt process to obtain an orthonormal basis. □

Another corollary is that any linearly independent list of vectors may be extended to an orthonormal basis.

Corollary 6.2.8

Any linearly independent list of vectors in a finite-dimensional inner product space may be extended to an orthonormal basis.

Proof. Take any linearly independent list of vectors and apply the Gram-Schmidt process to obtain an orthonormal list. Then, extend this orthonormal list to an orthonormal basis using the previous corollary. □

It's often useful to use the gram-schmidt process to determine if a list of vectors is linearly independent. Should the list be linearly dependent, upon hitting the first dependent vector, the gram-schmidt process with yield a zero vector, which cannot be normalized, so the algorithm terminates.
Previously, we proved a necessary and sufficient condition for an operator to be upper-triangularizable on some basis. We now provide a stronger result for orthonormal bases.

Lemma 6.2.9: Upper-triangular matrix with respect to some orthonormal basis

Suppose V is finite-dimensional and $T \in \mathcal{L}(V)$. Then T has an upper-triangular matrix with respect to some orthonormal basis of V if and only if the minimal polynomial of T equals $(z - \lambda_1) \cdots (z - \lambda_m)$ for some $\lambda_1, \ldots, \lambda_m \in \mathbb{F}$.

Proof. We can actually show that any upper-triangular matrix may be converted into an upper-triangular matrix with respect to some orthonormal basis using the Gram-Schmidt process. Suppose T has an upper-triangular matrix with respect to some basis $v_1, \ldots, v_n$ of V. Applying the Gram-Schmidt process to this basis yields us an orthonormal basis $e_1, \ldots, e_n$ of V. We claim that the matrix of T with respect to this basis is also upper-triangular. In our proof of the Gram-schmidt process, we showed that for each $k = 1, \ldots, n$, we have:

$$\text{span}(v_1, \ldots, v_k) = \text{span}(e_1, \ldots, e_k)$$

Hence, for each $k = 1, \ldots, n$, we have:

$$Te_k \in \text{span}(v_1, \ldots, v_k) = \text{span}(e_1, \ldots, e_k)$$

so the matrix of T with respect to the basis $e_1, \ldots, e_n$ is upper-triangular. We just apply our previous upper-triangularization condition to finish the proof. □

Finally, we introduce **Schur's Theorem**, which is a corollary of our previous theorem:

Theorem 6.2.10: Schur's Theorem

Every operator on a finite-dimensional complex innner product space has an upper-triangular matrix with respect to some orthonormal basis.

Proof. We apply the fundamental theorem of algebra to see that the minimal polynomial of any operator on a finite-dimensional complex vector space splits into linear factors. Hence, by the previous lemma, every operator on a finite-dimensional complex inner product space has an upper-triangular matrix with respect to some orthonormal basis. □

6.2.1 Riesz Representation Theorem

If we fix v in an inner product space V, then the map $u \mapsto \langle u, v \rangle$ is a linear functional on V. Surprisingly, every linear functional on V may be represented as the inner product with a fixed vector in V.

Theorem 6.2.11: Riesz Representation Theorem

Suppose V is a finite-dimensional inner product space and ϕ is a linear functional on V. Then there is a unique vector $v \in V$ such that:

$$\phi(u) = \langle u, v \rangle$$

for every $u \in V$.

Proof. We first prove existence. Take an orthonormal basis $e_1, \ldots, e_n$ of V. We see that:

$$u = \langle u, e_1 \rangle e_1 + \cdots + \langle u, e_n \rangle e_n$$

Hence, we have:

$$\begin{aligned}\phi(u) &= \phi(\langle u, e_1\rangle\, e_1 + \cdots + \langle u, e_n\rangle\, e_n) \\ &\quad + \langle u, e_1\rangle\, \phi(e_1) + \cdots + \langle u, e_n\rangle\, \phi(e_n) \\ &= \left\langle u, \overline{\phi(e_1)}e_1 + \cdots + \overline{\phi(e_n)}e_n \right\rangle\end{aligned}$$

Where the last equality follows from the additivity in the second slot and the conjugate symmetry of the inner product. Hence, we may take:

$$v = \overline{\phi(e_1)}e_1 + \cdots + \overline{\phi(e_n)}e_n$$

to obtain the desired result. Now we prove uniqueness. Suppose there are two vectors $v, w \in V$ such that:

$$\phi(u) = \langle u, v\rangle = \langle u, w\rangle$$

for every $u \in V$. Then we have:

$$\langle u, v - w\rangle = 0$$

for every $u \in V$. Taking $u = v - w$, we have:

$$\langle v - w, v - w\rangle = 0$$

so $v - w = 0$, or $v = w$. Additionally, we may also note that a vector is uniquely determined by its components in an orthonormal basis, so the representation is unique. □

6.3 Orthogonal Complements and Minimization Problems

6.3.1 Orthogonal Complements

Given a subset U of an inner product space V, we may define the orthogonal complement of U in V.

Definition 6.3.1: Orthogonal Complement, $U^\perp$

If U is a subset of V, then the **orthogonal complement** of U, denoted by $U^\perp$, is the set of all vectors in V that are orthogonal to every vector in U:

$$U^\perp = \{v \in V : \langle v, u\rangle = 0 \text{ for all } u \in U\}$$

Proposition 6.3.2

I

t's often times useful to consider orthogonal complements because of the nice properties they have:

Theorem 6.3.3: Properties of Orthogonal Complement

(a) If U is a subset of V, then $U^\perp$ is a subspace of V.

(b) $\{0\}^\perp = V$

(c) $V^\perp = \{0\}$

(d) If U is a subset of V, then $U \cap U^\perp \subseteq \{0\}$

(e) If G and H are subsets of V and $G \subseteq H$, then $H^\perp \subseteq G^\perp$

Proof. (a) Note that $U^\perp$ is closed under scalar multiplication and addition, and the zero vector is in $U^\perp$ because 0 is orthogonal to all vectors in V. for all $v, w \in U^\perp$ and $u \in U$, we have:

$$\langle u, v + w\rangle = \langle u, v\rangle + \langle u, w\rangle = 0$$

So closed under addition. Similarly for scalar mutliplication:

$$\langle \lambda v, u\rangle = \lambda \langle v, u\rangle = 0$$

So closed under scalar multiplication

(b) Note that all vectors are orthogonal to 0, as $\langle v, 0\rangle = 0$ for all $v \in V$.

(c) Note that the only vector orthogonal to all vectors is 0.

(d) Suppose $u \in U \cap U^\perp$. Then we have:

$$\langle u, u\rangle = 0$$

The only vector orthogonal to itself is 0. So $U \cap U^\perp \subseteq \{0\}$.

(e) Suppose $G \subseteq H$, then for all $h \in H^\perp$, $\langle g, h\rangle = 0$ for all $g \in G$. Hence, $H^\perp \subseteq G^\perp$

□

A powerful consequence of orthogonal subspaces are that they form direct sums of the vector space:

Theorem 6.3.4: Direct sum of a subspace and its orthogonal complement

Suppose U is a finite-dimensional subspace of V. Then

$$V = U \oplus U^|perp$$

Proof. First we show that:

$$V = U + U^\perp$$

Let $v \in V$, and take an orthonormal basis $e_1, \ldots, e_m$ of U. We may extend this basis to an orthonormal basis $e_1, \ldots, e_n$ of V. We see that:

$$v = a_1 e_1 + \cdots + a_n e_n$$

for some scalars $a_1, \ldots, a_n \in \mathbb{F}$. Now we see that:

$$u = a_1 e_1 + \cdots + a_m e_m \in U$$

and

$$w = a_{m+1} e_{m+1} + \cdots + a_n e_n \in U^\perp$$

so we have $v = u + w \in U + U^\perp$. Hence, we have $V = U + U^\perp$. Now we show that:

$$U \cap U^\perp = \{0\}$$

Suppose $v \in U \cap U^\perp$. Then we have:

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

so $v = 0$. Therefore, we have:

$$V = U \oplus U^\perp$$

□

A direct corollary of this theorem is that we may easily compute the dimension of the orthogonal complement of a subspace.

Corollary 6.3.5

Let V be finite-dimensional and U is a subspace of V. Then

$$\dim U^\perp = \dim V = \dim U$$

Proof. This follows directly from the previous theorem and the dimension formula for direct sums. □

Another important consequence is:

Corollary 6.3.6

Let U be a finite-dimensional subspace of V. Then:

$$U = (U^\perp)^\perp$$

Proof. We first show $U \subseteq (U^\perp)^\perp$. We know that for every $u \in U$ and $v \in U^\perp$, we have $\langle u, v \rangle = 0$, so $u \in (U^\perp)^\perp$. Hence $U \subseteq (U^\perp)^\perp$. However, by the previous corollary, we have:

$$\dim (U^\perp)^\perp = \dim V - \dim U^\perp = \dim V - (\dim V - \dim U) = \dim U$$

So we must have $U = (U^\perp)^\perp$. □

We will now consider the main application of orthogonal complements: solving minimization problems in inner product spaces.

6.3.2 Orthogonal Projections and Minimization Problems

Definition 6.3.7: Orthogonal projection, P_U

Let U be a finite-dimensional subspace of V. The **orthogonal projection** of V onto U is the operator $P_U \in \mathcal{L}(V)$ defined as: For each $v \in V$, $v = u + w$, where $u \in U$ and $w \in U^\perp$. Let $P_U v = u$.

The orthogonal projection has several important properties:

Theorem 6.3.8: Properties of Orthogonal Projections

Let U be a finite-dimensional subspace of V. Then:

(a) $P_U \in \mathcal{L}(V)$.

(b) $P_U u = u$ for all $u \in U$.

(c) $P_U w = 0$ for all $w \in U^\perp$.

(d) $\text{range}\, P_U = U$.

(e) $\text{null}\, P_U = U^\perp$.

(f) $v - P_U v \in U^\perp$ for all $v \in V$.

(g) $P_U^2 = P_U$.

(h) $||P_U v|| \leq ||v||$ for all $v \in V$.

(i) If $e_1, \ldots, e_m$ is an orthonormal basis of U, and $v \in V$, then:

$$P_U v = \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_m \rangle e_m$$

Proof. (a) For any $v, w \in V$ and $\lambda \in \mathbb{F}$, we have:

$$P_U(v + w) = P_U v + P_U w$$

and

$$P_U(\lambda v) = \lambda P_U v$$

so P_U is linear.

(b) For any $u \in U$, we have $u = u + 0$, so $P_U u = u$.

(c) For any $w \in U^\perp$, we have $w = 0 + w$, so $P_U w = 0$.

(d) By definition, the range of P_U is U.

(e) Note that $\operatorname{null} P_U = \{v \in V : P_U v = 0\} = U^\perp$.

(f) We have $v = u + w$ for some $u \in U$ and $w \in U^\perp$. The definition of P_U gives us $P_U v = u$, so we have:

$$v - P_U v = w \in U^\perp$$

(g) For any $v \in V$, we have $P_U v = u$ for some $u \in U$. Then we have:

$$P_U^2 v = P_U u = u = P_U v$$

(h) For any $v \in V$, we have $v = u + w$ for some $u \in U$ and $w \in U^\perp$. Then we have:

$$||v||^2 = ||u||^2 + ||w||^2$$

so $||u|| \leq ||v||$. Since $P_U v = u$, we have $||P_U v|| \leq ||v||$.

(i) Let $v \in V$. We may write:

$$v = P_U v + (v - P_U v)$$

where $P_U v \in U$ and $v - P_U v \in U^\perp$. Now, since $e_1, \ldots, e_m$ is an orthonormal basis of U, we have:

$$P_U v = \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_m \rangle e_m$$

□

Similar to before, we may actually define the projection matrix to any subspace U with respect to a basis of U. Before we do so, we need to cover a few lemmas and definitions that would be useful in the proof. We first introduce what is called an adjoint matrix.

Definition 6.3.9: Adjoint Matrix

Suppose B is a matrix of real numbers of dimension $n \times d$. For $x \in \mathbb{F}^d$, we have $Bx \in \mathbb{F}^n$. The **adjoint** of B, denoted by B^* is the $d \times n$ matrix defined such that for all $x \in \mathbb{F}^d$ and $y \in \mathbb{R}^n$, we have:

$$\langle Bx, y \rangle = \langle x, B^* y \rangle$$

Where the inner product defined are the standard inner products on $\mathbb{F}^n$ and $\mathbb{F}^d$ respectively, and B^* is the conjugate transpose of B.

Proof. Note that by the definition of the standard inner product, we have:

$$\langle Bx, y \rangle = (Bx)^* y = x^* B^* y = \langle x, B^* y \rangle$$

□

Let U be a finite-dimensional subspace of V, and let $u_1, \ldots, u_d$ be a basis of U. Then the $n \times d$ matrix whose columns are $u_1, \ldots, u_d$ is denoted by A. Now we are ready to prove that the special matrix $A^\perp A$ is invertible.

Lemma 6.3.10: Invertibility of $A^\perp A$

The $d \times d$ matrix $A^\perp A$ is invertible.

Proof. Since $A^\perp A$ is a square matrix, it is an operator from $\mathbb{F}^d$ to $\mathbb{F}^d$. Hence, to show that $A^\perp A$ is invertible, it sufficies to show injectivity. Suppose there exists $x \in \mathbb{F}^d$ such that $A^\perp Ax = 0$. Then we have:

$$0 = \langle A^\perp Ax, x \rangle = \langle Ax, Ax \rangle = ||Ax||^2$$

Which we know to be true if and only if $Ax = 0$. But since the columns of A are linearly independent by construction, we must have $x = 0$. Hence, $A^\perp A$ is injective, so it is invertible. □

Theorem 6.3.11: Orthogonal Projection Matrix

Let U be a finite-dimensional subspace of V, and let $u_1, \ldots, u_d$ be a basis of U. Then the $n \times d$ matrix whose columns are $u_1, \ldots, u_d$ is denoted by A. Then for any $v \in V$, the orthogonal projection of v onto U is given by:

$$P_U v = A(A^\perp A)^{-1} A^\perp v$$

Proof. Let $v \in V$ such that $v = u + w$ where $u \in U$ and $w \in U^\perp$. Let $P = A(A^\perp A)^{-1} A^\perp$. We want to show that:

$$Pv = u$$

Which is equivalent to showing that:

$$Pu = u, \quad Pw = 0$$

Since $u \in U$, there exists $x \in \mathbb{F}^d$ such that $u = Ax$ as columns of A span U. Then we have:

$$Pu = PAx = A(A^\perp A)^{-1} A^\perp Ax = A(A^\perp A)^{-1}(A^\perp A)x = Ax = u$$

For $w \in U^\perp$, we have $A^\perp w = 0$ by defnition of the orthogonal complement (each column of A is orthogonal to w). Hence, we have:

$$Pw = A(A^\perp A)^{-1} A^\perp w = A(A^\perp A)^{-1} 0 = 0$$

□

Chapter 7

Operators on Inner Product Spaces

7.1 Adjoint Operators

In inner product spaces, we may define a special type of operator called the adjoint operator. We saw this earlier when we defined the adjoint matrix. We now generalize this to operators on inner product spaces.

Definition 7.1.1: Adjoint, T^*

Let $T \in \mathcal{L}(V, W)$. The **adjoint** of T is the operator $T^* \in \mathcal{L}(W, V)$ such that:

$$\langle Tv, w \rangle = \langle v, T^* w \rangle$$

for every $v \in V$ and every $w \in W$.

It remains to prove that the adjoint operator actually exists and is unique.

Theorem 7.1.2: Existence and Uniqueness of the Adjoint Operator

Let $T \in \mathcal{L}(V, W)$. Then there exists a unique operator $T^* \in \mathcal{L}(W, V)$ such that:

$$\langle Tv, w \rangle = \langle v, T^* w \rangle$$

for every $v \in V$ and every $w \in W$.

Proof. We first prove existence. For each $w \in W$, define the linear functional phi_w on V by:

$$\phi_w(v) = \langle Tv, w \rangle$$

for every $v \in V$. By the Riesz representation theorem, there exists a unique vector $u_w \in V$

such that:

$$\phi_w(v) = \langle v, u_w \rangle$$

for every $v \in V$. Now define the operator $T^* \in \mathcal{L}(W, V)$ by:

$$T^* w = u_w$$

for every $w \in W$. Then we have:

$$\langle Tv, w \rangle = \phi_w(v) = \langle v, T^* w \rangle$$

for every $v \in V$ and every $w \in W$. Now we prove uniqueness. Suppose there exists another operator $S \in \mathcal{L}(W, V)$ such that:

$$\langle Tv, w \rangle = \langle v, Sw \rangle$$

for every $v \in V$ and every $w \in W$. Then we have:

$$\langle v, T^* w \rangle = \langle v, Sw \rangle$$

for every $v \in V$ and every $w \in W$. Taking $v = T^* w - Sw$, we have:

$$\langle T^* w - Sw, T^* w - Sw \rangle = 0$$

so $T^* w - Sw = 0$, or $T^* w = Sw$ for every $w \in W$. Hence, we have $T^* = S$. □

We must also show that the adjoint operator is linear.

Theorem 7.1.3: Linearity of the Adjoint Operator

Let $T \in \mathcal{L}(V, W)$. Then the adjoint operator $T^* \in \mathcal{L}(W, V)$ is linear.

Proof. For any $v \in V$, $w_1, w_2 \in W$ and $\lambda \in FF$, we have that:

$$\langle v, T^*(w_1 + w_2) \rangle = \langle Tv, w_1 + w_2 \rangle = \langle Tv, w_1 \rangle + \langle Tv, w_2 \rangle = \langle v, T^* w_1 \rangle + \langle v, T^* w_2 \rangle = \langle v, T^* w_1 + T^* w_2 \rangle$$

So $T^*(w_1 + w_2) = T^* w_1 + T^* w_2$. Similarly, we have:

$$\langle v, T^*(\lambda w) \rangle = \langle Tv, \lambda w \rangle = \overline{\lambda} \langle Tv, w \rangle = \overline{\lambda} \langle v, T^* w \rangle = \langle v, \lambda T^* w \rangle$$

So $T^*(\lambda w) = \lambda T^* w$. Hence, T^* is linear. □

The adjoint also has several important properties:

Theorem 7.1.4: Properties of the Adjoint Operator

Suppose $T \in \mathcal{L}VW$. Then:

(a) $(S+T)^* = S^* + T^*$ for all $S, T \in \mathcal{L}VW$.

(b) $(\lambda T)^* = \overline{\lambda} T^*$ for all $\lambda \in \mathbb{F}$.

(c) $(T^*)^* = T$.

(d) $(ST)^* = T^* S^*$ for all $S \in \mathcal{L}WU$ and $T \in \mathcal{L}VW$.

(e) $I^* = I$.

(f) T is invertible if and only if T^* is invertible. In this case, $(T^*)^{-1} = (T^{-1})^*$.

Proof. (a) By definition, we know that:

$$\begin{aligned} \langle (S+T)v, w \rangle &= \langle v, (S+T)^* w \rangle \\ &= \langle Sv, w \rangle + \langle Tv, w \rangle \\ &= \langle v, S^* w \rangle + \langle v, T^* w \rangle \\ &= \langle v, S^* w + T^* w \rangle \end{aligned}$$

So we have $(S+T)^* w = S^* w + T^* w$ for every $w \in W$. Hence, we have $(S+T)^* = S^* + T^*$.

(b) By defninition, we know that:

$$\langle \lambda T v, w \rangle = \langle v, (\lambda T)^* w \rangle = \lambda \langle Tv, w \rangle = \lambda \langle v, T^* w \rangle = \langle v, \overline{\lambda} T^* w \rangle$$

So we have $(\lambda T)^* w = \overline{\lambda} T^* w$ for every $w \in W$. Hence, we have $(\lambda T)^* = \overline{\lambda} T^*$.

(c) By definition, we know that:

$$\langle Tv, w \rangle = \langle v, T^* w \rangle = \langle T^* w, v \rangle = \langle w, (T^*)^* v \rangle = \langle (T^*)^* v, w \rangle$$

So we have $(T^*)^* v = Tv$ for every $v \in V$. Hence, we have $(T^*)^* = T$.

(d) By definition, we know that:

$$\langle STv, u \rangle = \langle Tv, S^* u \rangle = \langle v, T^* S^* u \rangle = \langle v, (ST)^* u \rangle$$

So we have $(ST)^* u = T^* S^* u$ for every $u \in U$. Hence, we have $(ST)^* = T^* S^*$.

(e) By definition, we know that:

$$\langle Iv, w \rangle = \langle v, I^* w \rangle = \langle v, w \rangle$$

So we have $I^* w = w$ for every $w \in V$. Hence, we have $I^* = I$.

(f) Suppose T is invertible. Then for every $v \in V$ and every $w \in W$, we have:

$$\langle v, w \rangle = \langle TT^{-1}v, w \rangle = \langle T^{-1}v, T^*w \rangle = \langle v, (T^*)^{-1}T^*w \rangle$$

So we have $(T^*)^{-1}T^*w = w$ for every $w \in W$. Hence, we have $(T^*)^{-1} = (T^{-1})^*$. The converse follows similarly.

□

We may also make statements about the null space and range of the adjoint operator.

Theorem 7.1.5: Null space and range of T^*

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

(a) $\operatorname{null} T^* = (\operatorname{range} T)^\perp$.

(b) $\operatorname{range} T^* = (\operatorname{null} T)^\perp$.

(c) $\operatorname{null} T = (\operatorname{range} T^*)^\perp$.

(d) $\operatorname{range} T = (\operatorname{null} T^*)^\perp$.

Proof. We prove (a) first. Let $w \in \operatorname{null} T^*$. Then we have $T^*w = 0$. so for every $v \in V$, we have:

$$\langle Tv, w \rangle = \langle v, T^*w \rangle = 0$$

Hence, we have $w \in (\operatorname{range} T)^\perp$. The reverse direction follows similarly. Once we have (a), we may take the orthogonal complements on both sides to derive (d). To derive (c), we simply swap the T with T^* in (a) and use the fact that $(T^*)^* = T$. Finally, we may take the orthogonal complements on both sides of (c) to derive (b). □

It shouldn't be surprising that the matrices of the adjoint operator and the operator itself are related. We introduce the conjugate transpose of a matrix to formalize this relationship

Definition 7.1.6: Conjugate Transpose, A^*

The *conjugate transpose* of a m-by-n matrix A is the n-by-m matrix A^* obtained by taking the transpose then taking the complex conjugate of each entry. For each entry, we have:

$$A^*_{i,j} = \overline{A_{j,i}}$$

we now show that with respect to orthonormal bases, the matrix of the adjoint operator is the conjugate transpose of the matrix of the operator itself.

Theorem 7.1.7: Matrix of the Adjoint Operator

Let $T \in \mathcal{L}(V, W)$ and $e_1, \ldots, e_n$ be an orthonormal basis of V and $f_1, \ldots, d_m$ be an orthonormal basis of W. Then $\mathcal{M}(T^*, (f_1, \ldots, f_m), (e_1, \ldots, e_n))$ is equal to the conjugate transpose of $\mathcal{M}(T, (e_1, \ldots, e_n), (f_1, \ldots, f_m))$.

$$\mathcal{M}(T^*) = (\mathcal{M}(T))^*$$

Proof. Note that the kth column of $\mathcal{M}(T)$ is given by writing Te_k as a linear combination of $f_1, \ldots, f_m$, where the scalars used in the linear combination become the entries of the kth column. Since $f_1, \ldots, f_m$ is an orthonormal basis of W, we have:

$$Te_k = \langle Te_k, f_1 \rangle f_1 + \cdots + \langle Te_k, f_m \rangle f_m$$

Hence, the (i, k)th entry of $\mathcal{M}(T)$ is given by $\langle Te_k, f_i \rangle$. Now we replace T with T^* and switch $e_1, \ldots, e_n$ with $f_1, \ldots, f_m$. We see that the jth entry in the kth row of $\mathcal{M}(T^*)$ is given by $\langle T^* f_k, e_j \rangle$. However, by the definition of the adjoint operator, we have:

$$\langle T^* f_k, e_j \rangle = \langle e_j, T^* f_k \rangle = \langle Te_j, f_k \rangle = \overline{\langle f_k, Te_j \rangle}$$

Hence, the (k, j)th entry of $\mathcal{M}(T^*)$ is equal to the conjugate of the (j, k)th entry of $\mathcal{M}(T)$. Therefore, we have:

$$\mathcal{M}(T^*) = (\mathcal{M}(T))^*$$

□

A close observation shows that this is highly similar to the relationship between a matrix and it's dual matrix. The two principles are the same idea applied in different contexts, as the orthogonal complement of a subset corresponds to the annhilator of the subset in the dual space, via the Riesz representation theorem.

7.1.1 Self-Adjoint Operators

When we focus on operators on inner product spaces to themselves, we may define a special type of operator called a self-adjoint operator.

Definition 7.1.8: Self-Adjoint Operator

An operator $T \in \mathcal{L}(V)$ is called self-adjoint if $T = T^*$.

Note that if a matrix is equal to it's complex-conjugate transpose, the complex-conjugate part guarantees that each entry is real, and the transpose part guarantees that the matrix is symmetric. Hence, self-adjoint operators are often called **Hermitian** operators (symmetric and real). Additionally, a nice property of self-adjoint operators is that their eigenvalues are always real.

Theorem 7.1.9: Eigenvalues

Every eigenvalue of a self-adjoint operator is real.

Proof. Let T be a self-adjoint operator on V. Let λ be an eigenvalue of T and let v be a nonzero eigenvector corresponding to λ. Then we have:

$$\lambda \left|\left|v\right|\right|^2 = \langle \lambda v, v\rangle = \langle Tv, v\rangle = \langle v, T^*v\rangle = \langle v, Tv\rangle = \langle v, \lambda v\rangle = \overline{\lambda} \left|\left|v\right|\right|^2$$

Since $\left|\left|v\right|\right|^2 > 0$, we have $\lambda = \overline{\lambda}$, so λ is real. □

The following theorem applies only on complex inner product spaces:

Theorem 7.1.10: Tv is orthogonal to v for all $v \iff T = 0$

Suppose V is a complex inner product space and $T \in \mathcal{L}(V)$. Then:

$$\langle Tv, v\rangle = 0 \text{ for every } v \in V \iff T = 0$$

Proof. Let $u, w \in V$. Note that:

$$\langle Tu, w\rangle = \frac{\langle T(u+w), u+w\rangle - \langle T(u-w), u-w\rangle}{4} + \frac{\langle T(u+iw), u+iw\rangle + \langle T(u-iw), u-iw\rangle}{4} i$$

Suppose that $\langle Tv, v\rangle = 0$ for every $v \in V$. Note that each term on the right hand side is zero by assumption, so we have $\langle Tu, w\rangle = 0$ for every $u, w \in V$, which is only true if $T = 0$. The converse is trivial. □

we have a similar result for self-adjoint operators:

Corollary 7.1.11

Suppose V is a complex inner product space and $T \in \mathcal{L}(V)$. Then:

$$\langle Tv, v\rangle \in \mathbb{R} \text{ for every } v \in V \iff T \text{ is self-adjoint}$$

Proof. If $v \in V$, then:

$$\langle T^*v, v\rangle = \overline{\langle v, T^*v\rangle} = \overline{\langle Tv, v\rangle}$$

Now:

$$\begin{aligned} T \text{ is self-adjoint} &\iff T - T^* = 0 \\ &\iff \langle (T - T^*)v, v\rangle = 0 \text{ for every } v \in V \\ &\iff \langle Tv, v\rangle - \overline{\langle T^*v, v\rangle} = 0 \text{ for every } v \in V \\ &\iff \langle Tv, v\rangle = \overline{\langle Tv, v\rangle} \text{ for every } v \in V \\ &\iff \langle Tv, v\rangle \in \mathbb{R} \text{ for every } v \in V \end{aligned}$$

□

We also prove that if Tv is orthogonal to v for all v, and T is self-adjoint, then $T = 0$.

Corollary 7.1.12

Suppose V is a complex inner product space and $T \in \mathcal{L}(V)$ is self-adjoint. Then:

$$\langle Tv, v\rangle = 0 \text{ for every } v \in V \iff T = 0$$

Proof. We have already proved this for the complex case. Now we assume V is real. Let $u, w \in V$. We have:

$$\langle Tu, w\rangle = \frac{\langle T(u+w), u+w\rangle - \langle T(u-w), u-w\rangle}{4}$$

This is because we have:

$$\langle Tw, u\rangle = \langle w, Tu\rangle = \langle Tu, w\rangle$$

by the self-adjoint property.

Now suppose $\langle Tv, v\rangle = 0$ for every $v \in V$. Note that each term on the right hand side is zero by assumption, so we have $\langle Tu, w\rangle = 0$ for every $u, w \in V$, which is only true if $T = 0$. □

The final thing we want to prove is a remarkable property that all real symmetric operators have: a guaranteed eigenvalue.

Theorem 7.1.13: Existence of Eigenvalues for Real Symmetric Operators

Suppose V is a real inner product space and $T \in \mathcal{L}(V)$ is self-adjoint. Then T has an eigenvalue.

Proof. To prove this statement, we first prove a lemma.

Lemma 7.1.14: Irreducible quadratic of a self-adjoint operator is injective

Suppose T is a self-adjoint operator on a real inner product space V, and $x^2 + bx + c$ is an irreducible quadratic polynomial with real coefficients. Then the operator $T^2 + bT + cI$ is injective.

Proof. The core of this lemma is by completing the square. Note that:

$$\begin{aligned} T^2 + bT + cI &= T^2 + bT + \frac{b^2}{4}I - \frac{b^2}{4}I + cI \\ &= (T + \frac{b}{2}I)^2 + (c - \frac{b^2}{4})I \\ &= S^2 + a^2 I \end{aligned}$$

Where $S = T + \frac{b}{2}I$ and $a = \sqrt{c - \frac{b^2}{4}} > 0$. Note that S is self-adjoint since T and a multiple of the identity operator are self-adjoint. Now suppose there exists $v \in V$ such that $(T^2 + bT + cI)v = 0$. Then we have:

$$S^2 v = -a^2 v$$

So we have:

$$\left|\left|S^2 v\right|\right| = \langle Sv, Sv\rangle = \left\langle S^2 v, v\right\rangle = \left\langle -a^2 v, v\right\rangle = -a^2 \left|\left|v\right|\right|^2$$

Which is only possible if $v = 0$. Hence, $T^2 + bT + cI$ is injective. □

Now we are ready to prove the theorem. Choose a vector $v \in V$, there exists some monic polynomial $p(x)$ of lowest degree such that $p(T)v = 0$. From here we have two cases:

- **Case 1:** $p(x)$ has a real root λ. Then we may factor $p(x)$ as:

  $$p(x) = (x - \lambda)q(x)$$

  for some polynomial $q(x)$. Hence, we have:

  $$0 = (T - \lambda I)q(T)v$$

  Which implies that $q(T)v$ is an eigenvector of T corresponding to the eigenvalue λ.

- **Case 2:** $p(x)$ has no real roots. The $p(x)$ must then be factored into irreducible quadratics with real coefficients such that $p(x) = q_1(x)q_2(x)\cdots q_m(x)$. Note that since $p(T)v = 0$, we have:

  $$0 = p(T)v = q_1(T)q_2(T)\cdots q_m(T)v$$

  Hence, there exists some i such that $q_i(T)$ is not injective. However, this contradicts the lemma we just proved. Hence, this case is impossible.

Thus, T must have an eigenvalue. □

7.1.2 Normal Operators

Previously, we have seen that operators tend to have nice properties when they commute. We now formalize this idea with the definition of normal operators for adjoints that commute.

Definition 7.1.15: Normal

An operator on an inner product space is **normal** if it commutes with it's adjoint, i.e. $TT^* = T^*T$.

Note that self-adjoint operators are normal by definition. Normal operators are interesting in that the magnitudes of the operator and it's adjoint are the same.

Theorem 7.1.16: T is normal if and only if Tv and T^*v have the same norm

Suppose $T \in \mathcal{L}(V)$. Then:

$$T \text{ is normal} \iff ||Tv|| = ||T^*v|| \text{ for every } v \in V$$

Proof. Note that:

$$\begin{aligned}
T \text{ is normal} &\iff TT^* = T^*T \\
&\iff TT^* - T^*T = 0 \\
&\iff \langle (TT^* - T^*T)v, v\rangle = 0 \text{ for every } v \in V \\
&\iff \langle TT^*v, v\rangle - \langle T^*Tv, v\rangle = 0 \text{ for every } v \in V \\
&\iff \langle T^*v, T^*v\rangle - \langle Tv, Tv\rangle = 0 \text{ for every } v \in V \\
&\iff ||T^*v||^2 = ||Tv||^2 \text{ for every } v \in V \\
&\iff ||T^*v|| = ||Tv|| \text{ for every } v \in V
\end{aligned}$$

□

Like self-ajoint operators, normal operators also have nice properties:

Theorem 7.1.17: Range, Nullspace, and Eigenvectors of Normal Operators

Suppose $T \in \mathcal{L}(V)$ is normal. Then:

(a) $\operatorname{null} T = \operatorname{null} T^*$.

(b) $\operatorname{range} T = \operatorname{range} T^*$.

(c) $V = \operatorname{null} T \oplus \operatorname{range} T$.

(d) $T - \lambda I$ is normal for every $\lambda \in \mathbb{F}$.

(e) If $v \in V$ and $\lambda \in \mathbb{F}$, then v is an eigenvector of T corresponding to λ if and only if v is an eigenvector of T^* corresponding to $\overline{\lambda}$. $Tv = \lambda v \iff T^*v = \overline{\lambda}v$.

Proof. (a) Suppose $v \in V$. Then:

$$v \in \operatorname{null} T \iff Tv = 0 \iff ||Tv|| = 0 \iff ||T^*v|| = 0 \iff T^*v = 0 \iff v \in \operatorname{null} T^*$$

(b) Note that:

$$\operatorname{range} T = (\operatorname{null} T^*)^\perp = (\operatorname{null} T)^\perp = \operatorname{range} T^*$$

(c) This follows directly from (a), (b), and the fundamental theorem of linear algebra.

$$V = \operatorname{null} T \oplus \operatorname{null} T^\perp = \operatorname{null} T \oplus \operatorname{range} T^* = \operatorname{null} T \oplus \operatorname{range} T$$

(d) Let $\lambda \in \mathbb{F}$. Note that:

$$\begin{aligned}(T-\lambda I)(T-\lambda I)^* &= (T-\lambda I)(T^*-\overline{\lambda}I)\\ &= TT^* - \overline{\lambda}T - \lambda T^* + |\lambda|^2 I\\ &= T^*T - \lambda T^* - \overline{\lambda}T + |\lambda|^2 I\\ &= (T^*-\overline{\lambda}I)(T-\lambda I)\\ &= (T-\lambda I)^*(T-\lambda I)\end{aligned}$$

(e) Suppose $v \in V$ and $\lambda \in \mathbb{F}$. Then:

$$||(T-\lambda I)v|| = ||(T-\lambda I)^*v|| = \big|\big|(T^*-\overline{\lambda}I)v\big|\big|$$

So we have:

$$(T-\lambda I)v = 0 \iff (T^*-\overline{\lambda}I)v = 0$$

So v is an eigenvector of T corresponding to λ if and only if v is an eigenvector of T^* corresponding to $\overline{\lambda}$.

□

we will prove an important property of normal operators right now, which also applies to self-adjoint operators:

Theorem 7.1.18: Orthogonal eigenvectors for normal operators

Suppose $T \in \mathcal{L}(V)$ is normal. The eigenvectors of T corresponding to distinct eigenvalues are orthogonal.

Proof. Let α, β be distinct eigenvalues of T with corresponding eigenvalues u, v. We want to show that $\langle u, v\rangle = 0$. Note that:

$$\begin{aligned}\alpha\langle u, v\rangle &= \langle Tu, v\rangle\\ &= \langle u, T^*v\rangle\\ &= \langle u, \overline{\beta}v\rangle\\ &= \beta\langle u, v\rangle\end{aligned}$$

Hence, we have $(\alpha-\beta)\langle u, v\rangle = 0$. Since $\alpha \neq \beta$, we must have $\langle u, v\rangle = 0$. □

7.2 The Spectral Theorem

Recall that a matrix is diagonalizable only when there exists a basis of V consisting of eigenvectors of the operator. An even better form of diagonalization is when the basis of eigenvectors are orthonormal. The spectral theorem is the result that characterizes these operators as self-adjoint operators when $\mathbb{F} = \mathbb{R}$ and normal operators when $\mathbb{F} = \mathbb{C}$. Hence, we will split the spectral theorem into two cases.

7.2.1 Real Spectral Theorem

To prove the real spectral theorem, we will need to first prove a lemma:

Lemma 7.2.1: Minimal polynomial of a self-adjoint operator splits into linear terms

Let $T \in \mathcal{L}(V)$ be self-adjoint. Then the minimal polynomial of T splits into linear terms over $\mathbb{R}$.

Proof. The simple case to consider is when $\mathbb{F} = \mathbb{C}$. We know that for a self-adjoint operator, all the eigenvalues are real. Moreover, over the complex field, every polynomial splits into linear factors. Hence, the minimal polynomial of T splits into linear factors over $\mathbb{C}$. However, since all the eigenvalues are real, the minimal polynomial must split into linear factors over $\mathbb{R}$ as well.

Now suppose $\mathbb{F} = \mathbb{R}$. For the minimal polynomial of T, we know there exists $\lambda_1, \dots, \lambda_m \in \mathbb{R}$ and $b_1, \dots, b_N, c_1, \dots, c_N \in \mathbb{R}$ such that that the minimal polynomial T may be factored as:

$$(x - \lambda_1)(x - \lambda_2) \cdots (x - \lambda_m)(x^2 + b_1 x + c_1)(x^2 + b_2 x + c_2) \cdots (x^2 + b_N x + c_N)$$

Hence, we have:

$$(T - \lambda_1 I)(T - \lambda_2 I) \cdots (T - \lambda_m I)(T^2 + b_1 T + c_1 I)(T^2 + b_2 T + c_2 I) \cdots (T^2 + b_N T + c_N I) = 0$$

Now suppose for the sake of contradiction that $N > 0$. Then by the lemma we proved earlier, each $T^2 + b_i T + c_i I$ is injective, and hence invertible. Hence, we may multiply both sides of the equation by the inverses of each $T^2 + b_i T + c_i I$ to obtain:

$$(T - \lambda_1 I)(T - \lambda_2 I) \cdots (T - \lambda_m I) = 0$$

But this contradicts the minimality of the minimal polynomial. Hence, we must have $N = 0$, so the minimal polynomial of T splits into linear factors over $\mathbb{R}$. □

Using this lemma, it is not possible to prove the real spectral theorem:

Theorem 7.2.2: Real Spectral Theorem

Suppose $\mathbb{F} = \mathbb{R}$ and $T \in \mathcal{L}(V)$. Then the following are equivalent:

(a) T is self-adjoint

(b) T has a diagonal matrix with respect to some orthonormal basis of V.

(c) V has an orthonormal basis consisting of eigenvectors of T.

Proof. Suppose that (a) holds. This means that T is self adjoint, so by the lemma we just proved, we know that the minimal polynomial of T splits into linear factors over $\mathbb{R}$. Hence, this means that T is upper-triangular with respect to some orthonormal basis of V by the

real Schur's theorem. With respect to this basis, the matrix of T^* is the transpose of T. But since $T^* = T$ as T is self-adjoint, this means that the matrix of T is equal to it's transpose, so the matrix of T is diagonal. Hence, we have (b).
Now suppose that (b) holds. Then there exists an orthonormal basis of V such that the matrix of T with respect to this basis is diagonal. Note that diagonal matrices are equal to their transposes, so the matrix of T is equal to the matrix of T^*. Hence, we have $T = T^*$, so T is self-adjoint. Thus, we have (a).
Finally, (b) and (c) are equivalent by the definition of diagonalizability. □

7.2.2 Complex Spectral Theorem

The complex spectral theorem is similar in language to the real spectral theorem, except it only requires the operator to be normal.

Theorem 7.2.3: Complex Spectral Theorem

Suppose $\mathbb{F} = \mathbb{C}$ and $T \in \mathcal{L}(V)$. THen the following are equivalent.

(a) T is normal.

(b) T has a diagonal matrix with respect to some orthonormal basis of V.

(c) V has an orthonormal basis consisting of eigenvectors of T.

Proof. We start with (a), which means T is normal. By the complex Schur's theorem, we know that T is upper-triangular with respect to some orthonormal basis of V:

$$\mathcal{M}(T) = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ 0 & a_{2,2} & \cdots & a_{2,2} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{n,n} \end{bmatrix}$$

$$\mathcal{M}(T^*) = \begin{bmatrix} \overline{a_{1,1}} & 0 & \cdots & 0 \\ \overline{a_{1,2}} & \overline{a_{2,2}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \overline{a_{1,n}} & \overline{a_{2,n}} & \cdots & \overline{a_{n,n}} \end{bmatrix}$$

Note that from the matrix, we have:

$$||Te_1||^2 = |a_{1,1}|^2$$
$$||T^*e_1||^2 = |a_{1,1}|^2 + |a_{1,2}|^2 + \cdots + |a_{1,n}|^2$$

However, since T is normal, we have $||Te_1|| = ||T^*e_1||$, so we must have $a_{1,2} = a_{1,3} = \cdots = a_{1,n} = 0$. Similarly, we may show that all the entries above the diagonal are zero. Hence, the matrix of T is diagonal with respect to this orthonormal basis. Thus, we have (b).
Now suppose that (b) holds. Then there exists an orthonormal basis of V such that the matrix of T with respect to this basis is diagonal. Note that diagonal matrices commute with their conjugate transposes, so the matrix of T commutes with the matrix of T^*. Hence,

we have $TT^* = T^*T$, so T is normal. Thus, we have (a).
Finally, (b) and (c) are equivalent by the definition of diagonalizability. □

7.3 Positive Operators

Definition 7.3.1: Positive Operator

An operator $T \in \mathcal{L}(V)$ is called *positive* if it is self-adjoint and $\langle Tv, v\rangle \geq 0$ for all $v \in V$.

If V is complex, then the condition that T is self-adjoint may be omitted, as it follows from the second condition. Moreover, we define the square root R of an operator as the map such that $R^2 = T$. We now go over some characterizations of positive operators.

Theorem 7.3.2: Characterizations of Positive Operators

Let $T \in \mathcal{L}(V)$. THen the following equivalent:

(a) T is a positive operator.

(b) T is self-adjoint and all eigenvalues of T are nonnegative.

(c) With respect to some orthonormal basis of V, the matrix of T is a diagonal matrix with only nonegative numbers on the diagonal.

(d) T has a positive square root.

(e) T has a self-adjoint square root.

(f) $T = R^*R$

Proof. We begin with the fact that T is positive, so by definition, we know T is self-adjoint and $\langle Tv, v\rangle \geq 0$ for all $v \in V$. Let λ value of T with eigenvector v. THen we have:

$$\lambda \left|\left|v\right|\right|^2 = \langle \lambda v, v\rangle = \langle Tv, v\rangle \geq 0$$

This is only possible if $\lambda \geq 0$. Hence, we have (b).
Now suppose (b) holds. Then by the real spectral theorem (if $\mathbb{F} = \mathbb{R}$) or the complex spectral theorem (if $\mathbb{F} = \mathbb{C}$), we know that there exists an orthonormal basis of V consisting of eigenvectors of T. With respect to this basis, the matrix of T is diagonal with the eigenvalues on the diagonal. Since all eigenvalues are nonnegative, the matrix of T is a diagonal matrix with only nonegative numbers on the diagonal. Hence, we have (c).
Now suppose (v) holds, since the matrix of T is a diagonal matrix with only positive numbers on the diagonal, we may define a diagonal matrix R with respect to the same orthonormal basis of V such that each diagonal entry of R is the square root of the corresponding diagonal entry of T. Hence, we have $R^2 = T$, so T has a positive square root. Thus, we have (d).
Now suppose (d) holds. Then there exists a positive square root R of T. Note that positive operators are self-adjoint, so R is self-adjoint. Hence, we have (e).

Now supppose that (e) and (f) are equivalent. Suppose (e) holds. Then there exists a self-adjoint square root R of T. Note that since R is self-adjoint, we have $R^* = R$. Hence, we have $T = R^2 = R^*R$, so we have (f).
Now suppose (f) holds. Then there exists an operator R such that $T = R^*R$. Note that:

$$\langle Tv, v\rangle = \langle R^*Rv, v\rangle = \langle Rv, Rv\rangle = ||Rv||^2 \geq 0$$

So T must be positive. Hence, we have (a). □

We can make a much stronger statement about positive operators, they don't just have a positive square root, they have a unique positive square root.

Theorem 7.3.3: Unique Positive Square Root

Suppose $T \in \mathcal{L}(V)$ is positive. Then there exists a unique positive operator $R \in \mathcal{L}(V)$ such that $R^2 = T$.

Proof. Suppose $T \in \mathcal{L}(V)$ is positive, and $v \in V$ is an eigenvector of T. Thus we know there is some real number $\lambda > 0$ such that $Tv = \lambda v$. Let R be a positive square root of T. We show that the behavior of T on all eigenvectors of T is uniquely determined, as there is a basis of V consisting of eigenvectors of T. Since R is a positive operator, we know that R is self-adjoint and has nonnegative eigenvalues. Moreover, the spectral theorem gaurantees an orthonormal set of basis vectors $e_1, \ldots, e_n$ that are eigenvectors of R. There are a set of nonnegative real numbers $\lambda_1, \ldots, \lambda_n$ such that $Rv = \sqrt{\lambda_k}e_k$ for each $k = 1, \ldots, n$. Note that:

$$v = a_1e_1 + \cdots + a_ne_nA$$

for some scalars $a_1, \ldots, a_n \in \mathbb{R}$. Hence, we have:

$$Rv = a_1\sqrt{\lambda_1}e_1 + \cdots + a_m\sqrt{\lambda_n}e_n$$

So we have:

$$\lambda v = Tv = R^2v = a_1\lambda_1e_1 + \cdots + a_n\lambda_ne_n = \lambda(a_1e_1 + \cdots + a_ne_n)$$

Subtracting both sides gives that $a_k(\lambda_k - \lambda) = 0$ for each $k = 1, \ldots, n$. So we have:

$$v = \sum_{k:\lambda_k=\lambda} a_ke_k$$

So $Rv = \sum_{k:\lambda_k=\lambda} a_k\sqrt{\lambda}e_k = \sqrt{\lambda}v$. Hence, the behavior of R on all eigenvectors of T is uniquely determined. Since there exists a basis of V consisting of eigenvectors of T, this means that R is uniquely determined on all of V. Thus, the positive square root of T is unique. We denote this unique positive square root of T as $\sqrt{T}$. □

The last corollary we will prove is that the inner product of the image of a positive operator and any vector is 0 if and only if the vector is in the nullspace of the operator.

Corollary 7.3.4

Suppose $T \in \mathcal{L}(V)$ is positive. Then for every $v \in V$:

$$\langle Tv, v\rangle = 0 \iff v \in \operatorname{null} T$$

Proof. Suppose:

$$0 = \langle Tv, v\rangle = \left\langle \sqrt{T}\sqrt{T}v, v \right\rangle = \left\langle \sqrt{T}v, \sqrt{T}v \right\rangle = \left|\left| \sqrt{T}v \right|\right|^2$$

Which is only possible if $\sqrt{T}v = 0$. Note that:

$$Tv = \sqrt{T}\sqrt{T}v = \sqrt{T}0 = 0$$

So we have $v \in \operatorname{null} T$. The converse is trivial. □

7.4 Isometries, Unitary Operators, and Matrix Factorization

7.4.1 Isometries

We define a special name for operators that preserve norms under transformation.

Definition 7.4.1: Isometry

A linear map $S \in \mathcal{L}(V, W)$ is called an **isometry** if

$$||Sv|| = ||v|| \text{ for every } v \in V$$

for every $v \in V$. An isometry is a linear map that preserves norms

Note that isometries are always injective, since if $Sv = 0$, then we have $||v|| = ||Sv|| = 0$, so $v = 0$. We will now provide a set of equivalent conditions for isometries.

Theorem 7.4.2: Characterizations of Isometries

Suppose $S \in \mathcal{L}(V, W)$. Let $e_1, \ldots, e_n$ is an orthonormal basis of V and $f_1, \ldots, f_m$ is an orthonormal basis of W. Then the following are equivalent:

(a) S is an isometryy.

(b) $S^*S = I$