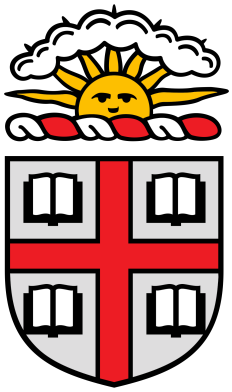
Honors Linear Algebra

MATH0540

Professor Melody Chan

Brown University



Edited by

Richard Tang

Contents

Chapter 1

Fundamentals of Linear Algebra

§1.1 Sets

Sets serve as a fundamental construct in higher-level mathematics. We start with a brief introduction to set theory.

Definition 1.1.1: Sets

A **set** is a collection of elements.

1. $x \in X$ means x is an element of X.
2. $x \notin$ means x is not an element of X.
3. $X \subset Y$ means X is a subset of Y (i.e. $\forall x \in X, x \in Y$.)
4. $X = Y \iff X \subset Y \wedge Y \subset X$.
5. $A \cap B := \{x \mid x \in A \wedge x \in B\}$ means set intersection.
6. $A \cup B := \{x \mid x \in A \vee x \in B\}$ means set union.
7. $A \setminus B := \{x \mid x \in A \wedge x \notin B\}$ means set difference.

Example 1. *Let*

$$\mathbb{Z} = \{\ldots, -3, -2, -1, 0, 1, 2, \ldots\}.$$

denote the set of integers, and let

$$\mathbb{Z}^+ = \{0, 1, \ldots\}.$$

denote the set of positive integers.

§1.1.1 Set Builder notation

Sets may be defined formally with set-builder notation:

$$X = \{\ expression \mid rule\}.$$

Example 2. *1. Let E represent the set of all even numbers. This set is expressed*

$$E = \{n \in \mathbb{Q} \mid \exists k \in \mathbb{Z} \text{ s.t. } n = 2k\}.$$

2. Let A represent the set of real numbers whose squares are rational numbers:

$$A = \{a \in \mathbb{R} \mid a^2 \in \mathbb{Q}\}.$$

§1.1.2 Cartesian Products

Definition 1.1.2: Ordered Tuples

An **ordered pair** is defined (x, y). An **n-ordered tuple** is an ordered list of n items

$$(x_1, \ldots, x_n).$$

Definition 1.1.3: Cartesian Products

Let A, B be sets. The **cartesian product** $A \times B$ is defined

$$A \times B := \{(a, b) \mid a \in A, b \in B\}.$$

Similarly, define the n-fold cartesian product

$$A^n := A \times A \times \cdots \times A.$$

Example 3. *$\mathbb{R}^2$ and $\mathbb{R}^3$ are examples of commonly known Cartesian products, which represent the 2D- and 3D-plane respectively.*

Example 4. *R^n is a first example of a **vector space.** Let $n \in \mathbb{Z}^+ \cup \{0\}$:*

1. *(Addition in R^n) We define an **addition operation** on $\mathbb{R}^n$ by adding coordinate-wise*
$$(x_1, \ldots, x_n) + (y_1, \ldots, y_n) = (x_1 + y_1, \ldots, x_n + y_n).$$
2. *(Scaling) Given $(x_1, \ldots, x_n) \in \mathbb{R}^n, \lambda \in \mathbb{R}$, we define*
$$\lambda \cdot (x_1, \ldots, x_n) = (\lambda x_1, \ldots, \lambda x_n).$$

Remark 1. $\mathbb{R}_0 = \{0\}$.

§1.1.3 Functions

Let A, B be sets. Informally, a function $f : A \to B$ deterministically returns an element $b \in B$ for each $a \in A$. We write $f(a) = b$.

Example 5. *The function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^2$ maps $\mathbb{R}$ to the subset*

$$S \subset \mathbb{R} = \{(x, x^2) \mid x \in \mathbb{R}\}.$$

Definition 1.1.4: Functions

Let A, B be sets. A function $f : A \to B$ is a subset $G_f \subset A \times B$ such that for all $a \in A$, there exists at most one $b \in B$ s.t. $(a, b) \in G_f$. We write $f(a) = b$ when $(a, b) \in G_f$.

Definition 1.1.5: Codomain

Given a function $f : A \to B$, A is the **domain** of f, and B is the **codomain** or **target** of f. Let the **range** of f be defined as

$$\{b \in B \mid f(a) = b, a \in A\}.$$

The range is the subset of B. Importantly, the number of elements in the range of f cannot be larger than the number of elements in A, as each $f(a)$ maps to at most one $b \in B$.

Definition 1.1.6: Bijectivity

Let $f : A \to B$ be a function.

1. f is **injective**, or an **injection**, if $a_1, a_2 \in A$ and $f(a_1) = f(a_2)$ implies $a_1 = a_2$.
2. f is **surjective**, or a **surjection**, if for any $b \in B$, there exists an $a \in A$ such that $f(a) = b$. Equivalently, the range is the whole codomain.
3. f is **bijective**, or a **bijection**, if it is both injective and surjective. Equivalently, for every $b \in B$, there is a unique $a \in A$ such that $f(a) = b$.

§1.2 Fields

Roughly speaking, a **field** is a set, together with operations addition and multiplication. Vector spaces may be defined *over* fields.

Definition 1.2.1: Fields

A **field** is a set $\mathbb{F}$ containing elements named 0 and 1, together with binary operations $+$ and $\cdot$ satisfying, for all $a, b, c \in \mathbb{F}$:

- **commutativity**: $a+b=b+a, a\cdot b=b\cdot a$
- **associativity**: $a+(b+c)=(a+b)+c, a\cdot(b\cdot c)=(a\cdot b)\cdot c$
- **identities**: $0+a=a, 1\cdot a=a$
- **additive inverse**: For any $a\in\mathbb{F}$, there exists a $b\in\mathbb{F}$ such that $a+b=0$. We denote this $b=-a$
- **multiplicative inverse**: For any $a\in\mathbb{F}$, $a\neq 0$, there exists a $b\in\mathbb{F}$ such that $ab=1$.
- **distributivity**: $a\cdot(b+c)=a\cdot b+a\cdot c$.

Example 6. $\mathbb{R}^+\setminus\{0\}$ *is* ***not*** *a field under* $+,\cdot$.

Example 7. *(Finite Fields) Let* p *prime (e.g.* $p=5$*). Define*

$$\mathbb{F}_p=\{0,\dots,p-1\},$$

with binary operations $+_p,\cdot_p$ *given by addition and multiplication modulo* p*. We claim (without proof) that* $\mathbb{F}_p$ *is a field.*

Example 8. *Let* $\mathbb{C}=\{a+bi\mid a,b\in\mathbb{R}\}$*. Elements of* $\mathbb{C}$ *are called* ***complex numbers****. Formally, a complex number is an ordered pair* $(a,b), a,b\in\mathbb{R}$*. We define addition as*

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

and multiplication as

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

Showing $\mathbb{C}$ *is a field is left as an exercise for the reader.*

Proposition 1.2.1: $\mathbb{C}$ Multiplicative Inverse

For every $\alpha\in\mathbb{C}\setminus\{0\}$, there exists $\beta\in\mathbb{C}$ with $\alpha\cdot\beta=1$.

Proof. Given $\alpha\in\mathbb{C}\setminus\{0\}$, let us write $\alpha=a+bi$. Then not both $a,b=0$. Let $\beta=\frac{a}{a^2+b^2}+-\frac{b}{a^2+b^2}i$. Then $\alpha\beta=(a+bi)\left(\frac{a}{a^2+b^2}+-\frac{b}{a^2+b^2}\right)=1$.
Thus $\forall\alpha\in\mathbb{C}\setminus\{0\},\exists\beta\in\mathbb{C}$ s.t. $\alpha\cdot\beta=1$. □

$\mathbb{R}^n$ and $\mathbb{C}^n$ are specific examples of fields, but by no means the only ones (for instance, $\mathbb{F}^2$ with addition and multiplication modulo 2 is a field). Fields serve as the underlying set of numbers and operations that vector spaces are built on. In this course, we focus primarily on $\mathbb{R}$ and $\mathbb{C}$; but many of the definitions, theorems, and proofs work interchangeably with abstract fields.

§1.3 Vector Spaces

Vector spaces serve as the fundamental abstract structure of linear algebra. All future topics will build on vector spaces. Roughly, a vector space V is a set of **vectors** with an addition operation and scalar multiplication, where scalars are drawn from a field $\mathbb{F}$. We now formalize this definition.

Definition 1.3.1: Vector Spaces

Given a field $\mathbb{F}$, A **vector space** over $\mathbb{F}$, denoted $V_{\mathbb{F}}$, is a set V, together with vector addition on V

$$+ : V \times V \longrightarrow V$$

and scalar multiplication on V

$$\cdot : \mathbb{F} \times V \longrightarrow V$$

satisfying the following properties:

- (additive associativity) For all $u, v, w \in V$, $u + (v + w) = (u + v) + w$.
- (additive identity) There exists an element $0 \in V$ such that $v+0 = 0+v = 0$.
- (additive inverse) For all $v \in V$, there exists $w \in V$ such that $v+w = w+v = 0$. We denote $w = -v$.
- (additive commutativity) For all $v, w \in V$, $v + w = w + v$.
- (scalar multiplicative associativity) For all $\alpha, \beta \in \mathbb{F}, v \in V$, $\alpha(\beta v) = (\alpha\beta)v$.
- (scalar multiplicative identity) There exists an element $1 \in \mathbb{F}$ such that $1v = v$ for all $v \in V$.
- (Distributive Law I) For every $\alpha \in \mathbb{F}$, $v, w \in V$, $a \cdot (v + w) = a \cdot v + a \cdot w$.
- (Distributive Law II) For every $\alpha, \beta \in \mathbb{F}, v \in V$, $(\alpha + \beta) \cdot v = \alpha \cdot v + \beta \cdot v$.

We call elements of $\mathbb{F}$ **scalars**, and elements of V **vectors**, or **points**.

Example 9. *We say 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 10. *Let $\mathbb{F}$ be a field.*

1. *For some integers $n \geq 0$, $\mathbb{F}^n = \{(a_1, \ldots, a_n) \mid a_i \in \mathbb{F}\}$ with vector addition defined*

$$(a_1, \ldots, a_n) + (b_1, \ldots, b_n) = (a_1 + b_1, \ldots, a_n + b_n)$$

and scalar multiplication defined

$$\lambda \cdot (v_1, v_2, \ldots, v_n) = (\lambda v_1, \lambda v_2, \ldots, \lambda v_n).$$

Note that $F^0 = \{0\}$.

2. $\mathbb{F}^\infty = P\{(a_1, a_2, a_3, \ldots) \mid a_j \in \mathbb{F}, j \in \mathbb{N}\}$ *with vector addition and scalar multiplication defined similarly.*

3. *Let* S *be any set; consider* $\{g : S \to \mathbb{F}\}$ *be the set of functions from* S *to* $\mathbb{F}$. *Given* $f, g : S \to \mathbb{F}$, $\lambda \in \mathbb{F}$, *define vector addition* $(f + g) : S \to \mathbb{F}$ *as*

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

and scalar multiplication $\lambda f : S \to \mathbb{F}$ *as*

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

Perhaps counterintuitively, example 3 subsumes example 1! For example, let $S = \{1, 2, \ldots, n\}$, and let $\mathbb{R}^{\{1,\ldots,n\}}$ be the set of all functions from $\{1, \ldots, n\} \to \mathbb{R}$. One such f may be

$$\begin{aligned} f : \{1, \ldots, n\} &\longrightarrow \mathbb{R} \\ x &\longmapsto f(x) = x^2 - 3. \end{aligned}$$

But f can also be thought of as an n-tuple. For instance, with $n = 3$, we can define a function

$$f = (-2, 1, 6) \in \mathbb{R}^3.$$

This is equivalent to $f(1) = -2, f(2) = 1, f(3) = 6$. Similarly, if $f(x) = e^x$, then $f \in \mathbb{R}^{\{1,2,3\}} = \left(e, e^2, e^3\right) \in \mathbb{R}^3$, since $f(1) = e, f(2) = e^2, f(3) = e^3$.

In other words, every n-tuple $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ could be represented as a function $f : \{1, 2, \ldots, n\} \to \mathbb{R}$, where $f(1) = x_1, f(2) = x_2, \ldots, f(n) = x_n$. The key insight here is that **the function f is the n-tuple**; the one function $f(x) = e^x$ is equivalent to the n-tuple $\left(e, e^2, \ldots, e^n\right)$.

From this, we get that the set of functions $\mathbb{R}^{\{1,\ldots,n\}} = \mathbb{R}^n$, the set of n-tuples.

Remark 2. *Reinterpret* $\mathbb{F}^0 = \{$*functions* $f : \varnothing \longrightarrow \mathbb{F}\}$. *How many functions are there from* $\varnothing \longrightarrow \mathbb{F}$*?*
One function $\varnothing = \varnothing \times \mathbb{F}$.

Example 11. *The set of continuous functions* $f : \mathbb{R} \to \mathbb{R}$ *forms a vector space over* $\mathbb{R}$. *In particular, the sum of two continuous functions is continuous; and* $a \cdot f$ *is continuous for any* $a \in \mathbb{R}$, *and* f *continuous.*

But what about fields over fields? Are these vector spaces?

Example 12. *Let $\mathbb{K}$ be a field, and say $\mathbb{F} \subseteq \mathbb{K}$ ($\mathbb{F}$ is a subfield of $\mathbb{K}$). Then $\mathbb{K}$ is a vector space over $\mathbb{F}$, with addition defined as in $\mathbb{K}$, and with scalar multiplication defined*

$$\lambda \cdot x = \lambda x, \text{ where } \lambda \in \mathbb{F}, x \in \mathbb{K}.$$

Thus $\mathbb{C}$ is a real vector space (this is why we draw the complex plane like $\mathbb{R}^2$!).

§1.3.1 Properties of Vector Spaces

We now observe some fascinating properties of vector spaces. Let V be a vector space over a field $\mathbb{F}$.

Proposition 1.3.1: Unique Additive Identity

V has a unique additive identity.

Proof. Suppose $e, e' \in V$ are both additive identities. Then

$$\begin{aligned} e &= e + e' \\ &= e'. \end{aligned}$$

Thus $e = e'$. □

Proposition 1.3.2: Unique Additive Inverse

Every vector $v \in V$ has a unique additive inverse.

Proof. Let $v \in V$, and suppose $w, w' \in V$ are both additive inverses of v. Then

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

Thus $w = w'$. □

Let us also define a notion of subtraction: we say $v - w = v + (-w)$.

Proposition 1.3.3: -v

For any $v \in V$,

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

Proof. Let $v, -v \in V$ where $-v$ is the inverse of v. Then

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

Since every $v \in V$ has a unique additive inverse, $-v = (-1) \cdot v$. □

Proposition 1.3.4: 0 Times a Vector

For every $v \in V$, $0v = 0$.

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

$$0v = (0+0)v = 0v + 0v.$$

Adding the additive inverse of $0v$ to both sides, we get $0v = 0$. □

Proposition 1.3.5: Scalar Times 0

For every $a \in \mathbb{F}$, $a\mathbf{0} = \mathbf{0}$.

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

$$a\mathbf{0} = a(\mathbf{0} + \mathbf{0}) = a\mathbf{0} + a\mathbf{0}.$$

Adding the additive inverse to both sides yields $a\mathbf{0} = \mathbf{0}$. □

§1.4 Subspaces

Subspaces can greatly expand our examples of vector spaces.

Definition 1.4.1: Subspaces

A subset $U \subseteq V$ is a **subspace** (or a **linear subspace**) of V if U is also a vector space.
U is a subspace of V if and only if

1. $\mathbf{0} \in U$.
2. For all $u, w \in U$, $u + w \in U$.
3. For all $u \in U$, $\lambda \in \mathbb{F}$, $\lambda \cdot u \in U$.

That is, addition and scalar multiplication are **closed** in U, and the identity element exists.
We see that these three properties are enough for U to satisfy the six properties of vector spaces: associativity, commutativity, and distributivity are automatically satisfied, as they hold on the larger space V (and so also hold on the subspace U); addition and scalar multiplication make sense in U, and the additive identity exists; the third condition guarantees the additive inverse ($-v = -1v$).

Example 13. *What are the subspaces of* $\mathbb{R}^2$ *and* $\mathbb{R}^3$*?*

Solution: It turns out that there are only three valid types of subspaces of $\mathbb{R}^2$:

1. The zero vector $\mathbf{0} = (0, 0)$.

2. All lines through the origin ($y = \alpha x$).
3. $\mathbb{R}^2$ itself.

Similarly, there are only four valid types of subspaces of $\mathbb{R}^2$:

1. The zero vector $\mathbf{0} = (0, 0, 0)$.
2. All lines through through the origin.
3. All planes through the origin.
4. $\mathbb{R}^3$ itself.

Let us now do a rough sketch of a proof that the list of subspaces of $\mathbb{R}^2$ is complete.

Proof. Let W be a subspace of R^2. If W has no nonzero vectors, then $W = \{\mathbf{0}\}$. If W has a non-zero vector $v \in V \setminus \{\mathbf{0}\}$, then W must contain the line through v passing through $\mathbf{0}$.
Moreover, if W contains some $w \in V$ not on the line, we have the ability to "turn" the coordinate plane, such that any $u \in V$ can be formed by $\alpha v + \beta w$. □

§1.4.1 Sums of Subspaces

With vector spaces, we are primarily only interested in subspaces, not arbitrary subsets. Thus, the notion of the sum of subspaces is useful.

Definition 1.4.2: Sum of Subsets

Suppose $U_1, \ldots, U_m$ are subsets of V. The **sum** of $U_1, \ldots, U_m$, denoted $U_1 + \ldots + U_m$, is the set of all possible sums of elements of $U_1, \ldots, U_m$. Precisely,

$$U_1 + \ldots + U_m = \{u_1 + \ldots + u_m \mid u_1 \in U_1, \ldots, u_m \in U_m\}.$$

Example 14. *Suppose $V = \mathbb{R}^3$. Let $U_1 = \{(x, 0, 0) \in \mathbb{R}^3 \mid x \in \mathbb{R}\}$ be the subspace containing elements with only x components, and $U_2 = \{(0, y, 0) \in \mathbb{R}^3 \mid y \in \mathbb{R}\}$ be the subspace containing elements with only y components. Then*

$$U_1 + U_2 = \{(x, y, 0)\mathbb{R}^3 \mid x, y \in \mathbb{R}\},$$

or the xy-plane.

Are these sums of subspaces actually subspaces themselves? Indeed, it is the smallest subspace containing all of the individual subspaces.

Proposition 1.4.1: Sum of Subspaces

Suppose $U_1, \ldots, U_m$ are subspaces of V. Then $U_1 + \ldots + U_m$ is the smallest subspace of V containing $U_1, \ldots, U_m$.

Proof. Clearly, $0 \in U_1+\ldots+U_m$ and addition and scalar multiplication in $U_1+\ldots+U_m$ is closed. Thus $U_1 + \ldots + U_m$ is a subspace of V.

To show that it is the smallest, observe first that $U_1, \ldots, U_m$ are all contained in $U_1+\ldots+U_m$ (for U_j, simply set $u_i = 0$ for any $i \neq j$). Additionally, every subspace of V containing $U_1, \ldots, U_m$ contains $U_1+\ldots+U_m$ as well, since subspaces must contain all finite sums of their elements (in this case, $u_i \in U_i$). Thus, since $U_1+\ldots+U_m$ contains every individual subspace, and any subspace containing $U_1, \ldots, U_m$ also contains $U_1 + \ldots + U_m$, we have that $U_1 + \ldots + U_m$ is the smallest subspace containing $U_1, \ldots, U_m$. □

§1.4.2 Direct Sums

Suppose $U_1, \ldots, U_m$ are subspaces of V. Every element of $U_1 + \ldots + U_m$ can be written as

$$u_1 + \ldots + u_m,$$

where each u_j is in U_j. Like the concept of injectivity, we are interested in the case when each vector in $U_1 + \ldots + U_m$ can only be written in one way. We call these **direct sums**.

Definition 1.4.3: Direct Sum

Suppose $U_1, \ldots, U_m$ are subspaces of V. The sum $U_1+\ldots+U_m$ is a **direct sum** if each element of $U_1+\ldots+U_m$ can be written in only one way as a sum $u_1+\ldots+u_m$, where $u_j \in U_j$. We denote this sum

$$U_1 \oplus \ldots \oplus U_m.$$

Two theorems are useful in determining if a sum of subspaces is a direct sum. Their proofs are left as an exercise for the reader.

Theorem 1.4.1: Condition for a Direct Sum

Suppose $U_1, \ldots, U_m$ are subspaces of V. Then $U_1 + \ldots + U_m$ is a direct sum if and only if the only way to write

$$0 = u_1 + \ldots + u_m$$

is by setting each $u_j = 0$.

Proof. One direction is easy. To show the other direction, assume there are multiple ways to write a vector v, and perform arithmetic $0 = v - v$ to arrive at $u_j = 0$. □

Theorem 1.4.2: Direct Sum of Two Subspaces

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

Proof. If we know direct sum, then there is only one way to write $0 = v + -v$ ($v \in U \cap W$). For the other direction, try writing $0 = u + w$ for some $u \in U$, $w \in W$, and showing that $u = w = 0$ necessarily. □

Chapter 2

Finite-Dimensional Vector Spaces

§2.1 Span and Linear Independence

Suppose a friend imagines a subspace $W \subseteq \mathbb{R}^3$. You know that $(1,0,0), (0,1,0) \in W$. What else do you know must be in W? Well, first, $\mathbf{0} = (0,0,0) \in W$ by definition. But moreover, anything in the form $\{(a,b,0) \mid a,b \in \mathbb{R}\}$ (the xy-plane) must be in W, since any point on the plane can be made by $\alpha \cdot a + \beta \cdot b$ (we will later see that $(1,0)$ and $(0,1)$ are **basis vectors** of $\mathbb{R}^2$).

Definition 2.1.1: Linear Combination and Span

A **linear combination** of a list of vectors $v_1, \ldots, v_n \in V$ is a vector of the form

$$\lambda_1 v_1 + \ldots + \lambda_n v_n, \text{ where } \lambda_i \in \mathbb{F}.$$

The **span** (or **linear span**) of $v_1, \ldots, v_n$, is the set of all linear combinations of $v_1, \ldots, v_n$:

$$\text{span}(v_1, \ldots, v_n) = \{a_1 v_1 + \ldots + a_m v_m \mid a_i \in \mathbb{F}\}.$$

The span of no vectors is $\{0\}$.

Proposition 2.1.1: Span is Smallest Subspace

The span of $v_1, \ldots, v_m$ is the smallest subspace of V containing $v_1, \ldots, v_m$. Precisely:

1. $\text{span}(v_1, \ldots, v_m)$ is a subspace of V.
2. Any subspace W of V containing $v_1, \ldots, v_m$ also contains $\text{span}(v_1, \ldots, v_m)$.

Proof. Let $v_1, \ldots, v_m$ be a list of vectors in V.

$\text{span}(v_1, \ldots, v_m)$ is clearly a subspace of V: achieve $\mathbf{0} \in \text{span}(v_1, \ldots, v_m)$ by setting each $a_j = 0$, and since $a_j + b_j, \lambda a_j \in \mathbb{F}$, $\text{span}(v_1, \ldots, v_m)$ is closed under addition and scalar multiplication.

Now, we show that $\text{span}(v_1, \ldots, v_m)$ is the smallest subspace containing $v_1, \ldots, v_m$. Every vector v_j is a linear combination of $v_1, \ldots, v_m$ (take, for $i \neq j$, $a_i = 0$); thus $\text{span}(v_1, \ldots, v_m)$ contains each v_j. Additionally, every subspace U of V that contains each $v_j \in U$ is closed under addition and scalar multiplication, so U contains every linear combination of $v_1, \ldots, v_m$; thus U contains $\text{span}(v_1, \ldots, v_m)$. So, since $\text{span}(v_1, \ldots, v_m)$

contains every vector v_j, and any subspace U of V that contains every vector v_j also contains $\text{span}(v_1, \ldots, v_m)$, the span is the smallest subspace containing every v_j. □

Definition 2.1.2: Spanning a Vector Space

If $\text{span}(v_1, \ldots, v_m) = V$, then $v_1, \ldots, v_m$ **spans** V, and $v_1, \ldots, v_m$ are a **spanning set**.

We now make one of the key definitions of linear algebra.

Definition 2.1.3: Finite Dimensional Vector Spaces

If V is spanned by a **finite** list of vectors $v_1, \ldots, v_m$ then V is **finite-dimensional**.

If V is not finite-dimensional, then V is **infinite-dimensional**.

Example 15. *Let $\mathcal{P}(\mathbb{F})$ be the set (indeed, vector space) of polynomials over a field $\mathbb{F}$. Show $\mathcal{P}(\mathbb{F})$ is infinite-dimensional.*

Solution: Let $p \in \mathcal{P}(\mathbb{F})$, and let m denote the highest degree polynomial in $\mathcal{P}(\mathbb{F})$. Then p has at most degree m; thus a polynomial p^{m+1} is not spanned by any list of vectors in $\mathcal{P}(\mathbb{F})$; thus $\mathcal{P}(\mathbb{F})$ is finite-dimensional.

§2.1.1 Linear Independence

As with sums/direct sums, we are interested if a vector has a unique linear combination; that is, given a list $v_1, \ldots, v_m \in V$, and $v \in \text{span}(v_1, \ldots, v_m)$, are there unique $a1, \ldots, a_m \in \mathbb{F}$ such that

$$v = a_1 v_1 + \ldots + a_m v_m?$$

In other words, is there only one way to create a certain vector given a span? Suppose there's more than one way; then there exists $b_1, \ldots, b_m \in \mathbb{F}$ such that

$$v = b_1 v_1 + \ldots + b_m v_m;$$

then

$$0 = (a_1 - b_1)v_1 + \ldots + (a_m b_m)v_m.$$

If the only way to do this is the obvious way, where $a_i - b_i = 0$, then the representation is unique. We call this **linear independence**.

Definition 2.1.4: Linear Independence

A list of vectors $v_1, \ldots, v_m \in V$ is **linearly independent** if the only choice of $a_1, \ldots, a_m \in \mathbb{F}$ that makes $a_1 v_1 + \ldots + a_m v_m$ equal 0 is $a_i = 0$.

A list of vectors in V is **linearly dependent** if it is not linearly independent.

That is, there exist non-zero $a_i \in \mathbb{F}$ such that

$$0 = \sum_{i=1}^{m} a_i v_i.$$

An empty list of vectors () is linearly independent.

Example 16. *1. A list of one vector $v \in V$ is linearly independent if and only if v is non-zero.*

2. A list of two vectors $v1, v_2 \in V$ is linearly independent if and only if one vector is not a scalar combination of the other vector; that is, $v_1 \neq \lambda v_2$ for some $\lambda \in \mathbb{F}$.

3. $(1,0,0),\ (0,1,0) \in \mathbb{R}^3$ is linearly independent.

4. $(1,-1,0),(-1,0,1),(0,1,-1) \in \mathbb{R}^3$ is linearly dependent. In particular, $(1,-1,0)+(-1,0,1)+(0,1,-1) = \boldsymbol{0}$. Alternatively, we can write $(-1,0,1)$ as a linear combination of the other two:

$$(-1,0,1) = -1 \cdot (1,-1,0) - (0,1,-1).$$

Intuitively, a list of vectors is linearly independent if none of its vectors are a linear combination of the other vectors; each vector is "independent" of the other vectors. In other words, a vector is linearly independent if it is not in the span of the other vectors. Its negation is also important, and is arguably more intuitive: a list of vectors is linearly dependent **iff** it is in the span of the other vectors (it is "dependent" on the other vectors). Formally, this gives rise to an important lemma, and theorem.

Lemma 2.1.1: Linear Dependence Lemma

Suppose that $v_1, \ldots, v_m \in V$ is a linearly dependent list of vectors. Then there exists some $j \in \{1, \ldots, m\}$ such that:

1. $v_j \in \text{span}(v_1, ..., v_{j-1})$
2. If the j^{th} term is removed from the list, the span of the remaining vectors $v_1, \ldots, \hat{v_j}^{a}, \ldots, v_m$ equals $\text{span}(v_1, \ldots, v_m)$.

In other words, removing the linearly dependent vector has no effect on the overall span of the vectors.

[a] here, hat means "with v_j removed"

Proof. Because the list $v_1, \ldots, v_m$ is linearly dependent, there exist $a_1, \ldots, a_m \in \mathbb{F}$ not all 0 such that

$$a_1 v_1 + \ldots + a_m v_m = 0.$$

Let j be the *largest element* of $\{1, \ldots, m\}$ such that $a_j \neq 0$. Then

$$v_j = -\frac{a_1}{a_j} v_1 - \ldots - \frac{a_{j-1}}{a_j};$$

hence v_j is in the span of $v_1, \ldots, v_{j-1}$.

Now, suppose $u \in \text{span}(v_1, \ldots, v_m)$. Then there exist $b_1, \ldots, b_m \in \mathbb{F}$ such that

$$u = b_1 v_1 + \ldots + b_m v_m.$$

If we replace v_j with 2.1.1, the resulting list consists only of $v_1, \ldots, \hat{v_j}, \ldots, v_m$; thus we see that u is in the span of the list. □

Theorem 2.1.1: Length of Linearly Independent List and Span

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. Left as an exercise for the reader. Try starting with $u_1, \ldots, u_m \in V$ a list of linearly independent vectors, and $v_1, \ldots, v_n \in V$ a spanning list of V, and show that $m \leq n$. Use the Linear Dependence Lemma to iteratively add u_i and remove w_j; eventually, we are left with a list with all u_i, and optionally some w_j.
To see why we cannot have more u than w, if that were the case, then $u_1, \ldots, u_n$ would span V, but $u_{n+1}, \ldots, u_m$ would be linearly independent, a contradiction. Thus $m \leq n$. □

Intuitively, every subspace of a finite-dimensional vector space is also finite-dimensional.

Proposition 2.1.2: Finite-Dimensional Subspaces

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

Proof. Suppose V is finite-dimensional and U is a subspace of V. We construct a spanning list of U:

- If $U = \{0\}$, then U is finite-dimensional and we are done, so choose a non-zero $v_1 \in U$.
- If $U = \text{span}(v_1, \ldots, v_{j-1})$, then U is finite-dimensional and we are done; otherwise, if $U \neq \text{span}(v_1, \ldots, v_{j-1})$, choose a vector $v_j \in U$ such that $v_j \notin \text{span}(v_1, \ldots, v_{j-1})$ (equivalently, $v_1, \ldots, v_j$ is linearly independent).

After the process, we are left with a linearly independent spanning list of U. Since U is a subspace of V, this linearly independent list cannot be longer than the length of V's basis (aka spanning list), and so U is finite-dimensional as well. □

§2.2 Bases

Spanning lists and linearly independent lists go hand in hand; now, we bring these concepts together.

Definition 2.2.1: Basis

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

Example 17. *1.* $e_1, e_2, \ldots, e_n = \{(1,0,\ldots,0),(0,1,\ldots,0),\ldots,(0,\ldots,1)\}$ *is the* ***standard basis*** *for* $\mathbb{F}^n$.

2. $(1,2)$, $(3,5)$ *is another basis for* $\mathbb{R}^2$*; however,* $(1,2)$, $(3,5)$, $(4,13)$ *spans* $\mathbb{R}^2$ *but is not linearly independent.*

3. $1, x, \ldots, x^m$ *is a basis of* $\mathcal{P}_m(\mathbb{R})$.

Bases are incredibly useful in constructing unique vectors in a vector space.

Proposition 2.2.1: Criterion for Bases

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

$$v = a_1 v_1 + \ldots + a_n v_n,$$

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

Proof. First, suppose $v_1, \ldots, v_n$ is a basis of V. Let $v \in V$. Since $V = \text{span}(v_1, \ldots, v_n)$, we have, for $a_i \in \mathbb{F}$

$$v = a_1 v_1 + \ldots + a_n v_n.$$

Suppose, for $c_i \in \mathbb{F}$, that

$$v = c_1 v_1 + \ldots + c_n v_n.$$

This implies

$$(v - v) = 0 = \sum_{i=1}^{n} (a_i - c_i) v_i.$$

Since $v_1, ..., v_n$ is linearly independent, all $a_i - c_i = 0$, and so $a_i = c_i$.
Now, suppose $v \in V$ can be represented uniquely by $a_1 v_1 + \ldots + a_n v_n$. Clearly, $v_1, \ldots, v_n$ spans V; and given

$$0 = a_1 v_1 + \ldots + a_n v_n,$$

since the representation is unique, we must have $a_i = 0$ (since otherwise, if any $a_i \neq 0$, then the representation wouldn't be unique); hence $v_1, \ldots, v_n$ is linearly independent as well, and so is a basis of V. □

Spanning lists may not be bases of V due to linear independence, while linearly independent lists may not be bases due to spanning. Thus, we look for ways to create bases from spanning/linearly independent lists.
For spanning lists, we get the idea that we can discard "useless" vectors while maintaining span.

Proposition 2.2.2: Spanning List contains a Basis

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

Proof. Let $B = v_1, \ldots, v_n$ span V; we iteratively remove "useless" vectors until left with a basis.

- If $v_1 = 0$, delete v_1 from B; otherwise, leave B unchanged.
- If $v_j \in \text{span}(v_1, \ldots, v_{j-1})$, delete v_j from B. Otherwise, leave B unchanged.

After iterating n times (through the entire list), we are left with a list B that spans V (since we only removed linearly dependent vectors in the span of the other vectors). Moreover, B is linearly independent, since no vector $v_i \in B$ is in the span of the previous vectors. Hence B is a basis of V. □

An easy corollary follows:

Corollary 2.2.1: Basis of Finite-Dimensional Vector Space

Every finite-dimensional vector space has a basis.

Proof. By definition, a finite-dimensional vector space has a spanning list; using the previous result, we reduce this list B to a basis. □

Now, we work with linearly independent lists; we can add "uncovered" vectors until such a list spans V while maintaining linear independence.

Proposition 2.2.3: Linearly Independent List Extends to a Basis

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

Proof. Let $u_1, \ldots, u_m \in V$ be a linearly independent list, and let $w_1, \ldots, w_n \in V$ be a basis for V. Thus the list

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

spans V. Using the procedure before, we reduce this list to a basis of V; this basis has all of the u's, since $u_1, \ldots, u_m$ is linearly independent, and some of the w's. □

For example, suppose we have $(2, 3, 4), (9, 6, 8) \in \mathbb{R}^3$. Using $e_1, e_2, e_3 \in \mathbb{R}^3$ as the standard basis, the procedure results in a basis $(2, 3, 4), (9, 6, 8), (0, 1, 0)$.

We finish with some subspaces of V; intuitively, two subspaces can be combined to form V.

Proposition 2.2.4: 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. Because V is finite-dimensional, so is U; so let $u_1, \ldots, u_m$ be a basis of U. Since $u_1, \ldots, u_m \in V$ is linearly independent, extend the list to a basis $u_1, \ldots, u_m, w_1, \ldots, w_n$ of V, and let $W = \text{span}(w_1, \ldots, w_n)$. To prove $V = U \oplus W$, we need to show

$$V = U + W \text{ and } U \cap W = \{0\}.$$

For any $v \in V$, since $u_1, \ldots, u_m, w_1, \ldots, w_n$ is a basis for v, we have $a_i, b_i \in \mathbb{F}$ such that

$$v = \sum_{i=1}^{m} a_i u_i + \sum_{j=1}^{n} b_j w_j.$$

Since $\sum_{i=1}^{m} a_i u_i \in U$, $\sum_{j=1}^{n} b_j w_j \in W$, we have $v = u + w$, $u \in U$, $w \in W$. Thus $v \in U + W$, and so $V = U + W$.
Now, suppose $v \in U \cap W$. Then we have, for $a_i, b_i \in \mathbb{F}$,

$$v = \sum_{i=1}^{m} a_i u_i = \sum_{j=1}^{n} b_j w_j,$$

and so

$$\sum_{i=1}^{m} a_i u_i - \sum_{j=1}^{n} b_j w_j = 0.$$

Since $u_1, \ldots, u_m, w_1, \ldots, w_n$ is a basis and so linearly independent, every $a_i, b_j = 0$, and so $v = 0(u_1 + \ldots + u_m) = 0(w_1 + \ldots + w_n) = 0$. Hence $U \cap W = \{0\}$, and so

$$V = U \oplus W.$$

□

§2.3 Dimension

With a space such as $\mathbb{R}^2$ or $\mathbb{R}^3$, we get the notion of two or three dimensions. For each, we see that their bases are that length (e.g. $(1,0), (0,1)$ is length two, $(1,0,0), (0,1,0), (0,0,1)$ is length three); hence, it seems that dimension is dependent on length of basis. However, in a finite-dimensional vector space, this would only make sense if every basis had the same length. Fortunately, this is the case:

Proposition 2.3.1: Basis Length does not Depend on Basis

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

Proof. Let B_1, B_2 be bases of a finite-dimensional vector space V. Then B_1 is linearly independent, and B_2 spans V. Hence the length of B_1 is less than or equal to the length of B_2. Swapping roles (e.g. B_1 spans V, B_2 linearly independent), we see that the length of B_1 is greater than or equal to the length of B_2. Hence their lengths are equal. □

Now, we can formally define dimension:

Definition 2.3.1: Dimension

The **dimension** of a finite-dimensional vector space V, denoted $\dim V$, is the length of any basis B of V.

For example, $\mathcal{P}_m\mathbb{F}$ has dimension $m+1$, because the basis $1, x, \ldots, x^m \in \mathcal{P}_m(\mathbb{F})$ has length $m+1$.
As expected, the dimension of a subspace is less than or equal to the dimension of the vector space.

Proposition 2.3.2: Dimension of a Subspace

Let V be a finite-dimensional vector space, and let U be a subspace of V. Then any basis $u_1, \ldots, u_m \in U$ is a linearly independent list in V, and any basis $v1, \ldots, v_n \in$

V is a spanning list of V, so $m \leq n$, or $\dim U \leq \dim V$.

Bases require two properties: linearly independent and spanning. It turns out that given any two out of three properties from length, linearly independent, and spanning, we can deduce whether a list is a basis. Clearly, if a list is linearly independent and spanning, it has the right length (e.g. $\dim V$), but the other two conditions (right length + lin. ind., or right length + spanning) may not be as obvious.

Proposition 2.3.3: Linearly Independent List of Right Length is a Basis

Given a finite-dimensional vector space V, every linearly independent list of vectors in V with length $\dim V$ is a basis for V.

Proof. Let $n = \dim V$ and $v_1, \ldots, v_n \in V$ be a linearly independent list in V, and suppose $v_1, \ldots, v_n$ does not span V. Then there exists some vector $v \in V$ such that $v \notin \text{span}(v_1, \ldots, v_n)$, so $v_1, \ldots, v_n, v$ is linearly independent. However, the length of every linearly independent list in V is less than or equal to the length of any spanning set of V; and since $n + 1 > n$ (V has a basis a.k.a. spanning list of length n), this is a contradiction. Thus $v_1, \ldots, v_n$ spans V, and therefore is a basis.

See Axler for an alternative proof; their logic makes more mental leaps. □

Proposition 2.3.4: Spanning List of Right Length is a Basis

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

Proof. Let $v_1, \ldots, v_n$ span V, and suppose $v_1, \ldots, v_n$ is linearly dependent. Then there exists $v_j \in \text{span}(v_1, \ldots, v_{j-1})$, so $v_1, \ldots, \hat{v}_j, \ldots, v_n$ spans V. However, again the length of every linearly independent list is less than or equal to the length of any spanning list of V; and since $n-1 < n$ (V has a basis a.k.a. linearly independent list of length n), this is a contradiction. Thus $v_1, \ldots, v_n$ is linearly independent, and therefore is a basis. □

Finally, we find the dimension of the sum of two subspaces of V. Intuitively, we keep all basis vectors of U_1, U_2, while discarding any "duplicates" (imagine a Venn Diagram!).

Proposition 2.3.5: Dimension of a Sum

If $U_1,\ U_2$ are subspaces of a finite-dimensional vector space, then

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

Proof. Let $u_1, \ldots, u_m$ be a basis for $U_1 \cap U_2$; thus $\dim U_1 \cap U_2 = m$. Because $u_1, \ldots, u_m$ is a basis, it is linearly independent in both U_1 and U_2; thus extend the list to a basis of U_1, $u_1, \ldots, u_m, v_1, \ldots, v_j$, and a basis of U_2, $u_1, \ldots, u_m, w_1, \ldots, w_k$. Thus $\dim U_1 = m + j$, and $\dim U_2 = m + k$.
We now show that

$$u_1, \ldots, u_m, v_1, \ldots, v_j, w_1, \ldots, w_k$$

is a basis for $U_1 + U_2$; this will show that $\dim(U_1 + U_2) = m + j + k = \dim U_1 + \dim U_2 - \dim U_1 \cap U_2$.

Clearly span$(u_1, \ldots, u_m, v_1, \ldots, v_j, w_1, \ldots, w_k)$ contains U_1 and U_2 (since any vector in either U_1 or U_2 could be made with a combination), and hence equals $U_1 + U_2$. To show linear independence, suppose

$$a_1u_1 + \ldots + a_mu_m + b_1v_1 + \ldots + b_jv_j + c_1w_1 + \ldots + c_kw_k = 0.$$

Rewriting, we get

$$c_1w_1 + \ldots + c_kw_k = -a_1u_1 - \ldots - a_mu_m - b_1v_1 - \ldots - b_jv_j,$$

and so $c_iw_i \in U_1$. Since all w_i are in U_2, this implies $c_1w_1 + \ldots + c_kw_k \in U_1 \cap U_2$. Since $u_1, ..., u_m$ is a basis for $U_1 \cap U_2$, we can rewrite as

$$c_1w_1 + \ldots + c_kw_k = d_1u_1 + \ldots + d_mu_m.$$

However, $u_1, \ldots, u_m, w_1, \ldots, w_k$ is linearly independent, so all $c_i, d_i = 0$. Thus we get, from the original equation,

$$a_1u_1 + \ldots + a_mu_m + b_1v_1 + \ldots + b_jv_j = 0.$$

Since this list is a basis, all $a_i, b_i = 0$. Thus all $a, b, c = 0$, and so the list is linearly independent.
Thus $u_1, \ldots, u_m, v_1, \ldots, v_j, w_1, \ldots, w_k$ is a basis for $U_1 + U_2$. □

Chapter 3

Linear Maps

§3.1 Linear Maps

Definition 3.1.1: Linear Maps

Let V, W be vector spaces over a field $\mathbb{F}$. A function

$$\begin{aligned} T : V &\longrightarrow W \\ v &\longmapsto T(v) \in W. \end{aligned}$$

is a **linear map** if it satisfies, given $v_1, v_2 \in V,\ \lambda \in \mathbb{F}$:

1. **Linearity**: $T(v_1 + v_2) = T(v_1) + T(v_2) \in W$.
2. **Homogeneity**: $T(\lambda v) = \lambda T(v)$.

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

Proposition 3.1.1: Linear Maps Preserve 0

If $T : V \to W$ is a linear map, then $T(\mathbf{0}) = \mathbf{0}$.

Proof. We have

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

Adding the additive inverse of $T(0)$ (which exists since $T(0) \in W$, and W is a vector space) to both sides, we have

$$0 = T(0).$$

□

Proposition 3.1.2: Combination of Linearity Properties

A function $T : V \to W$ is linear if and only if

$$T(\alpha v_1 + \beta v_2) = \alpha T(v_1) + \beta T(v_2)$$

for all $v_1, v_2 \in V,\ \alpha, \beta \in \mathbb{F}$.

Example 18. *Let V, W be any vector spaces over $\mathbb{F}$.*

1. *The **zero map***

$$\begin{aligned} 0 : V &\longrightarrow W \\ v &\longmapsto 0(v) = 0 \end{aligned}$$

   is a linear map.

2. *The **identity map***

$$\begin{aligned} I : V &\longrightarrow V \\ v &\longmapsto I(v) = v \end{aligned}$$

   is a linear map.

3. *Any linear map*

$$\begin{aligned} T : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto T(x) = ax \end{aligned}$$

   is a linear map.

The properties of linear maps are quite powerful; if we know how basis vectors are mapped, we can actually determine, completely the map, as seen below.

Example 19. *Say $T : \mathbb{R}^2 \to \mathbb{R}^2$ is a linear map such that $T(1,0) = (2,1)$ and $T(0,1) = (1,-1)$. What else do we know?*

- $T(0,0) = (0,0)$
- $T(1,1) = T((1,0) + (0,1)) = (2,1) + (1,-1) = (3,0)$
- $T(2,0) = (4,2)$
- $T(x,y) = (2x + y, x - y)$. *Wow!*

The following proposition asserts an intuitive understanding of linear functions in $\mathbb{R}$:

Proposition 3.1.3: Linear Maps in $\mathbb{R}$

Let $T : \mathbb{R} \to \mathbb{R}$ be a linear map. Then there is some $a \in \mathbb{R}$ such that $T(x) = ax$ for all $x \in \mathbb{R}$.

Proof. Let $a = T(1)$. Then for any $x \in \mathbb{R}$,

$$T(x) = T(x \cdot 1) = x \cdot T(1) = ax.$$

□

From the previous example and proposition, we get that knowing how basis vectors are mapped is incredibly important; this leads to the following theorem.
indeed, it turns out that given a basis in V, we can actually find a unique mapping to any list of vectors in W; that is, **linear maps are freely and uniquely determined by what they do to a basis**.

Theorem 3.1.1: Linear Maps and Basis of Domain

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

$$T(v_j) = w_j,\ j \in \{1, \ldots, n\}.$$

This statement is incredibly powerful; essentially, **linear maps are freely and uniquely determined by what they do to a basis**. That is, given any basis $v_1, \ldots, v_n \in V$, anyone can select **any** $w_i \in W$ (there are **no constraints** on the w_i), and there exists a unique map $T : V \to W$ that ensures v_i is mapped to w_i.

Proof. First, we show the existence of a linear map T with the desired property. Define $T : V \to W$ by

$$T(c_1v_1 + \ldots + c_nv_n) = c_1w_1 + \ldots + c_nw_n,\ c_i \in \mathbb{F}.$$

Since the list $v_1, \ldots, v_n$ is a basis, the function T is well defined; that is, since every element $v \in V$ has one unique representation $c_1 + \ldots + c_n$, every $T(v)$ only has one possible output. (If $v_1, \ldots, v_n$ were not a basis, then the map would not be a function; given two different representations of v, the map would produce two different output values for one input value).
For each j, taking $c_j = 1$ and $c_i = 0$ for the other c's shows that $T(v_j) = w_j$. One can easily verify that $T : V \to W$ is a linear map (linearity and homogeneity are trivial due to additive associativity and distributivity in W).
To prove uniqueness, now suppose $T \in \mathcal{L}(V, W)$, and that $T(v_j) = w_j$ for $j = 1, \ldots, n$. Let $c_i \in \mathbb{F}$. Homogeneity implies $T(c_jv_j) = c_jw_j$, and linearity implies

$$T(c_1v_1 + \ldots + c_nv_n) = c_1w_1 + \ldots + c_nw_n.$$

From this, because any $v = c_1v_1 + \ldots + c_nv_n$ is uniquely constructed from the basis, we see that any map $T \in \mathcal{L}(V, W)$ that sends v_j to w_j is the same map; that is, T is a unique linear map. □

§3.2 Null Spaces and Ranges

§3.2.1 Null Spaces

Now, we will explore two subspaces that are intimately connected with each linear map. First, we look at vectors that get sent to 0.

Definition 3.2.1: Null Space

For $T \in \mathcal{L}(V, W)$, the **null space of** T, denoted $\operatorname{null} T$, is the subset of V consisting

of vectors that T maps to 0:

$$\operatorname{null} T = \{v \in V \mid T(v) = 0\}.$$

Remark 3. *Advanced students may also know* $\operatorname{null} T$ *by its alternative name, the* ***kernel****. In fact, a vector space is a commutative ring over a field* $\mathbb{F}$*, and any linear map is actually a homomorphism (where isomorphism is given with bijective maps).*

Example 20. *Some examples of null spaces:*

- *If* T *is the zero map, then every* $v \in V$ *is in* $\operatorname{null} T$*; that is,* $\operatorname{null} T = V$*.*
- *If* $T \in \mathcal{L}(F^\infty, F^\infty)$ *is the backward shift*

$$T(x_1, x_2, \ldots) = (x_2, x_3, \ldots),$$

then $\operatorname{null} T = \{(a, 0, 0, \ldots) \mid a \in \mathbb{F}\}$*, since* x_1 *can be anything, and* $x_{i\geq 2}$ *must be* 0*.*

Now, we will see that the null space is a subspace of V, and discover an easier check for injectivity.

Proposition 3.2.1: Null Space is Subspace of V

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

Proof. Since T is a linear map, we know that $T(0) = 0$. Thus $0 \in \operatorname{null} T$. Now, suppose $u, v \in \operatorname{null} T$. Then

$$T(u + v) = T(u) + T(v) = 0 + 0 = 0,$$

and so $u + v \in \operatorname{null} T$. Finally, suppose $\lambda \in \mathbb{F}$. Then

$$T(\lambda v) = \lambda T(v) = \lambda \cdot 0 = 0,$$

and so $\lambda v \in \operatorname{null} T$. Thus $\operatorname{null} T$ is a subspace of V. □

First, recall the definition of injectivity:

Definition 3.2.2: Injectivity

A function $T : V \to W$ is **injective** if $T(v) = T(w)$ implies $v = w$; in other words, if $v \neq w$, then $T(v) \neq T(w)$.

It turns out that a trivial null space is necessary and sufficient for injectivity!

Proposition 3.2.2: Trivial Null Space Equals Injective

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

Proof. First, suppose T is injective. We know $0 \in \operatorname{null} T$, so suppose $v \in \operatorname{null} T$. Then

$$T(v) = 0 = T(0).$$

Hence for any $v \in V$, by injectivity we have $v = 0$, and so $\operatorname{null} T = \{0\}$.
Now, suppose that $\operatorname{null} T = \{0\}$. Let $u, v \in V$ such that $T(u) = T(v)$. Then

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

and so $u - v \in \operatorname{null} T$; but since $\operatorname{null} T = \{0\}$, we have $u - v = 0$, and so $u = v$. Hence T is injective. □

§3.2.2 Ranges

Now, we look at the set of outputs of a function.

Definition 3.2.3: Range

For $T \in \mathcal{L}(V, W)$, the **range** of T, denoted $\operatorname{range} T$, is the subset of W consisting of vectors of the form $T(v)$ for some $v \in V$:

$$\operatorname{range} T = \{T(v) \mid v \in V\}.$$

The range of the zero map, for instance, is $\{0\}$. If D is the differentiation map between polynomials, because every polynomial $p' = q$ is equal to another polynomial, $\operatorname{range} D = \mathcal{P}(\mathbb{R})$.
Like the null space, the range is indeed a subspace of W.

Proposition 3.2.3: Range is Subspace of W

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

Proof. Since T is a linear map, we know that $T(0) = 0$. Thus $0 \in \operatorname{range} T$. If $w_1, w_2 \in \operatorname{range} T$, then there exist $v_1, v_2 \in V$ such that $T(v_1) = w_1$, $T(v_2) = w_2$. Thus

$$T(v_1 + v_2) = T(v_1) + T(v_2) = w_1 + w_2,$$

and so $w_1 + w_2 \in \operatorname{range} T$. For $\lambda \in \mathbb{F}$, we have

$$T(\lambda v) = \lambda T(v) = \lambda w,$$

and so $\lambda v \in \operatorname{range} T$ as well. Hence $\operatorname{range} T$ is a subspace of W. □

Like the null space again, range correlates with surjectivity:

Definition 3.2.4: Surjectivity

A function $T : V \to W$ is **surjective** if $\operatorname{range} T = W$.

§3.2.3 Rank Nullity Theorem

Now, we get to one of the f̃undamentalt̃heorems of linear algebra.

Theorem 3.2.1: Rank-Nullity Theorem

If V is finite-dimensional and $T \in \mathcal{L}(V, W)$, then $\operatorname{range} T$ is finite-dimensional and

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

That is, the dimension of V is equal to the **nullity** (dimension of $\operatorname{null} T$) plus the **rank** (dimension of $\operatorname{range} T$).

Proof. Let $u_1, \ldots, u_m$ be a basis of $\operatorname{null} T$; thus $\dim \operatorname{null} T = m$. Since $u_1, \ldots, u_m$ is linearly independent in $\operatorname{null} T \subseteq V$, we can extend it to a basis

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

of V; thus $\dim V = m + n$. Now, we show $\dim \operatorname{range} T = n$.
Let $v \in V$. Since $u_1, \ldots, u_m, v_1, \ldots, v_n$ is a basis for V, we can write

$$v = a_1 u_1 + \ldots + a_m u_m + b_1 v_1 + \ldots + b_n v_n.$$

Applying T to both sides, we get

$$\begin{aligned} T(v) &= a_1 T(u_1) + \ldots + a_m T(u_m) + b_1 T(v_1) + \ldots + b_n T(v_n) \\ T(v) &= b_1 T(v_1) + \ldots + b_n T(v_n), \end{aligned}$$

since $T(u_i) = 0$ (due to null space properties). Hence $T(v_1) + \ldots + T(v_n)$ spans $T(v)$ (since $v \in V$ was arbitrary), and so $\operatorname{range} T$ is finite-dimensional.
To prove linear independence, let

$$\begin{aligned} b_1 T(v_1) + \ldots + b_n T(v_n) &= 0 \\ T(b_1 v_1 + \ldots + b_n v_n) &= 0 \\ b_1 v_1 + \ldots + b_n v_n &\in \operatorname{null} T. \end{aligned}$$

Thus we can write

$$\begin{aligned} b_1 v_1 + \ldots + b_n v_n &= a_1 u_1 + \ldots + a_m u_m \\ b_1 v_1 + \ldots + b_n v_n - a_1 u_1 - \ldots - a_m u_m &= 0. \end{aligned}$$

But $v_1, \ldots, v_n, u_1, \ldots, u_m$ is a basis for V, and so is linearly independent. Thus $T(v_1), \ldots, T(v_n)$ is a basis for $\operatorname{range} T$, as desired; and so $\dim V = m + n = \dim \operatorname{null} T + \dim \operatorname{range} T$. □

Using this, we can easily deduce information about the injectivity and surjectivity of linear maps. Intuitively, we get that no map to a "smaller" vector space is injective, and no map to a "larger" vector space is surjective (where size is determined by dimension).

Proposition 3.2.4: Map to Smaller Dimensional Space is not Injective

Suppose V, W are finite-dimensional vector spaces such that $\dim V > \dim W$. Then no linear map from V to W is injective.

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

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

Thus $\operatorname{null} T$ contains vectors other than 0, and so T is not injective. □

Proposition 3.2.5: Map to Larger Dimensional Space is not Surjective

Suppose V, W are finite-dimensional vector spaces such that $\dim V < \dim W$. Then no linear map from V to W is surjective.

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

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

Hence $\dim \operatorname{range} T < \dim W$, and so $\operatorname{range} T \neq W$. Therefore T is not surjective. □

This has important implications on solutions to systems of linear equations.

Definition 3.2.5: Homogeneous Linear Systems

A system of linear equations is **homogeneous** if all constants to the left of

$$\begin{aligned} \sum_{k=1}^{n} A_{1,k} x_k &= c_1 \\ &\vdots \\ \sum_{k=1}^{n} A_{m,k} x)k &= c_m \end{aligned}$$

are 0; that is, $c_i = 0$. A system is **inhomogeneous** if not all c_i are zero.

Example 21. ***When does a homogeneous system of linear equations have non-zero solutions?***
Consider the homogeneous system of linear equations

$$\begin{aligned} A_{1,1}x_1 + \ldots + A_{1,n}x_n &= 0 \\ &\vdots \\ A_{m,1}x_1 + \ldots + A_{m,n}x_n &= 0 \end{aligned}$$

where $A_{j,k} \in \mathbb{F}$. Obviously, $x_i = 0$ is a solution; the question is whether other solutions exist.
To convert this system into a linear map, define $T : \mathbb{F}^n \to \mathbb{F}^m$ by

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

That is, for every variable x_i, we map it to a column vector

$$x_i \longmapsto \begin{pmatrix} A_{1,i} \\ A_{2,i} \\ \vdots \\ A_{m,i} \end{pmatrix} x_i;$$

that is, every x_i contributes a little to each vector in $\mathbb{F}^m$:

$$x_i \longmapsto (A_{1,i}x_1, A_{2,i}x_i, \ldots, A_{m,i}x_i)\,.$$

In matrix form, each "row" (a.k.a. vector in $\mathbb{F}^m$) represents an equation, or vector in $\mathbb{F}^m$. $T(x_1, \ldots, x_n) = 0$ is the same as the homogeneous system of linear equations above. We wish to know if $\operatorname{null} T > \{0\}$. In other words, ***we can rephrase our question about non-zero solutions as: when is T not injective?***

Proposition 3.2.6: Homogeneous System of Linear Equations

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

Proof. Use the notation and result from the example above. Thus T is a linear map from $F^n \to F^m$, and we have a homogeneous system of m linear equations with n variables $x_1, \ldots, x_n$. From before, we see that T is not injective, and thus has non-zero solutions, if $n > m$, or there are more variables than equations. □

Now, looking at inhomogeneous systems of linear equations, we are also curious whether solutions exist.

Example 22. *Consider the inhomogeneous system of linear equations*

$$\begin{aligned} A_{1,1}x_1 + \ldots + A_{1,n}x_n &= c_1 \\ &\vdots \\ A_{m,1}x_1 + \ldots + A_{m,n}x_n &= c_m \end{aligned}$$

Now, the question is whether there is some choice of $c1, \ldots, c_m \in \mathbb{F}$ such that no solution exists.
Define $T : \mathbb{F}^n \to \mathbb{F}^m$ the same way:

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

The equation $T(x_1, \ldots, x_n) = (c_1, \ldots, c_m)$ is the same as the previous system of equations. Thus, we want to know if $\operatorname{range} T \neq \mathbb{F}^m$; that is, are there solutions for any choice of constants $c_1, \ldots, c_m$? Rephrasing, ***when is T not surjective?***

Proposition 3.2.7: Inhomogeneous System of Linear Equations

An inhomogeneous system of linear equations with more equations than variables has no solution for **some** (not all!) choice of the constant terms.

Proof. Use the notation and result from above; thus $T : F^n \to F^m$ describes a system of m equations and n variables. From before, T is not surjective if $n < m$. □

§3.3 Matrices

We know now that if $v_1, \ldots, v_n$ is a basis of V and $T : V \to W$ is a linear map, then $T(v_1), \ldots, T(v_n)$ determine the values of T on any vector $v \in V$. Matrices allow us to efficiently encode the values of $T(v_i)$ in terms of a basis of W.

Definition 3.3.1: Matrix of a Linear Map, $\mathcal{M}(T)$

Let $T : V \to W$ be a linear map, $v_1, \ldots, v_n$ be a basis for V, and $w_1, \ldots, w_m$ be a basis for W. The **matrix of** T, denoted $A = \mathcal{M}(T)$, with respect to these bases is the $m \times n$ matrix with entries $A_{j,k}$ defined by

$$T(v_k) = A_{1,k}w_1 + \ldots + A_{m,k}w_m.$$

Remark 4. *Note that since $w_1, \ldots, w_m$ is a basis, $A_{j,k}$ are determined uniquely; that is, there's always a unique way to write each $T(v_k)$ as a linear combination of $w_1, \ldots, w_m$. Thus, $\mathcal{M}(T)$ is determined uniquely by T.*

The indexing is quite weird; one way to remember how $\mathcal{M}(T)$ is constructed from T is by writing across the top of the matrix the basis vectors $v_1, \ldots, v_n$ (the domain), and the left the basis vectors $w_1, \ldots, w_m$ for which T maps:

$$\mathcal{M}(T) = \begin{array}{c} \\ w_1 \\ \vdots \\ w_m \end{array} \begin{array}{c} \begin{array}{cccc} v_1 & \ldots & v_k & \ldots & v_n \end{array} \\ \begin{pmatrix} & & A_{1,k} & & \\ & & \vdots & & \\ & & A_{m,k} & & \end{pmatrix} \end{array}.$$

Axler, pg. 71

Only the k-th column is included; basically, $T(v_k)$ can be computed by multiplying every component of v_k by every row of the k-th column. That is, the k-th column is "where v_k goes." The point is, once bases of V and W are agreed upon, the matrix of T, $\mathcal{M}(T)$, encodes T without losing information!

Example 23. *Suppose $T : \mathbb{R}^2 \to \mathbb{R}^3$ is a linear map given uniquely by*

$$T(1, 0) = (1, 2, 7) \text{ and } T(0, 1) = (3, 5, 9).$$

Then the matrix $\mathcal{M}(T)$ is given by

$$\begin{pmatrix} 1 & 3 \\ 2 & 6 \\ 7 & 9 \end{pmatrix}.$$

§3.3.1 $\mathcal{L}(V, W)$ as a Vector Space

Recall that $\mathcal{L}(V, W)$ denotes the set of linear maps from V to W. It turns out that $\mathcal{L}(V, W)$ can be given the structure of a vector space! That is, we can define vector addition and scalar multiplication in such a way to satisfy the properties of a vector space:

Definition 3.3.2: Addition and Scalar Multiplication in $\mathcal{L}(V, W)$

Let $S, T \in \mathcal{L}(V, W)$, $v \in V$, and $\lambda \in \mathbb{F}$.

- We define $S + T$ to be
$$(S + T)(v) = S(v) + T(v).$$
- We define λT to be
$$(\lambda T)(v) = \lambda T(v).$$

One can easily check that $S + T$ and λT are linear maps, and that $\mathcal{L}(V, W)$ forms a vector space (the reader is spared the menial work).

The additive identity in $\mathcal{L}(V, W)$ is given by the zero map $0 : V \to W$, $0(v) = 0$.
Like linear maps, a similar process can be applied for matrices. Let $\mathbb{F}^{m,n}$ denote the set of $m \times n$ matrices over $\mathbb{F}$.

Definition 3.3.3: Addition and Scalar Multiplication in $\mathbb{F}^{m,n}$

We define addition as component-wise addition; that is, for any $A = (a_{i,j}), B = (b_{i,j})$, $A + B = (a_{i,j} + b_{i,j})$. Scalar multiplication is similarly defined: for $A = (a_{i,j}), \lambda A = (\lambda a_{i,j})$.

With these operations, one can easily check that $\mathbb{F}^{m,n}$ is a vector space. We then look at the dimension of $\mathbb{F}^{m,n}$.

Proposition 3.3.1: $\dim \mathbb{F}^{m,n} = mn$

Suppose m, n are positive integers. With addition and scalar multiplication defined as above, $\mathbb{F}^{m,n}$ is a vector space with dimension mn.

Proof. Note that the additive identity is the $m \times n$ matrix whose entries all equal 0. Additionally, it is easy to see that the list of $m \times n$ matrices that have 0 in all entries except for a 1 in one entry is a basis of $\mathbb{F}^{m,n}$. Since there are mn such matrices, the dimension of $\mathbb{F}^{m,n}$ is mn. □

Now, we connect $\mathcal{L}(V, W)$ to $\mathbb{F}^{m,n}$; intuitively, these two structures should agree.

Proposition 3.3.2

Given vector spaces V, W over $\mathbb{F}$ with bases $v_1, \ldots, v_n$ and $w_1, \ldots, w_m$, for any $S, T \in \mathcal{L}(V, W)$, $\lambda \in \mathbb{F}$, we have

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

- $\mathcal{M}(\lambda T) = \lambda \mathcal{M}(T)$.

The proof should be relatively straightforward; let A, B represent $\mathcal{M}(S)$, $\mathcal{M}(T)$; then addition and scalar multiplication in $\mathcal{L}(V, W)$ should produce the same results as addition/multiplication in $\mathbb{F}^{m,n}$.
In fact, $\mathcal{M}(\cdot)$ is itself a map, from $\mathcal{L}(V, W)$ to $\mathbb{F}^{m,n}$. Indeed, this map is bijective!

Proposition 3.3.3

Let V, W be finite-dimensional vector spaces over $\mathbb{F}$, and choose bases $v_1, \ldots, v_n$, $w_1, \ldots, w_m$. Then

$$\mathcal{M}(\cdot) : \mathcal{L}(V, W) \to \mathbb{F}^{m,n}$$

is a bijective linear map.

Proof. The previous proposition shows that $\mathcal{M}(\cdot)$ is linear; to see bijective, for a given $A \in \mathbb{F}^{m,n}$, we need to show that there is a unique linear map $T : V \to W$ such that $\mathcal{M}(T) = A$.
Indeed, by definition of the matrix of a linear map, we wish to show that there exists a unique linear map such that, for each $i = 1, \ldots, n$, we have

$$T(v_i) = A_{1,i} w_1 + \ldots + A_{m,i} w_m.$$

But this is true because linear maps are freely and uniquely determined by their operations on a basis! □

Remark 5. *Recall that an invertible linear map or isomorphism is one that is bijective. Thus this proposition tells us that there is an isomorphism between $\mathcal{L}(V, W)$ and $\mathbb{F}^{m,n}$.*

§3.3.2 Composition of Linear Maps and Products of Matrices

Usually, vector multiplication doesn't make sense, but for some pairs of linear maps, a meaningful product exists.

Definition 3.3.4: 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

$$(ST)(u) = S \circ T(u) = S(T(u))$$

for $u \in U$.

In other words, ST is just the usual composition $S \circ T$ of two functions. One can easily verify that $ST \in \mathcal{L}(U, W)$ is a linear map from U to W.

Definition 3.3.5: Linear Operators

A linear map $T : V \to V$ from a vector space to itself is called a **linear operator**.

Proposition 3.3.4

The composition of linear maps is associative and distributive, an identity I exists. However, it is **not necessarily** commutative; that is, $ST \neq TS$. Moreover, $ST = 0$ does **not** imply that $S = 0$ or $T = 0$.

We now have products of linear maps; but what is the analogy for matrices?

Definition 3.3.6: Matrix Multiplication

Suppose $A \in \mathbb{F}^{m,n}$ is an $m \times n$ matrix, and $C \in \mathbb{F}^{n,p}$ is an $n \times p$ matrix. Then $AC \in \mathbb{F}^{m,p}$ is defined by the $m \times p$ matrix with $AC_{j,k}$-th entry given by:

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

In other words, the entry in row j, column k, of AC is computed by multiplying piecewise row j of A and row k of C.

Note that in order to multiply, the inner values (e.g. $m \times n$ and $n \times p$) must agree! That is, 3×2 and 2×4 would work, but 3×1 and 3×2 wouldn't.

Example 24. *Here we multiply a* 3×2 *and* 2×4 *matrix:*

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} \begin{pmatrix} 6 & 5 & 4 & 3 \\ 2 & 1 & 0 & -1 \end{pmatrix} = \begin{pmatrix} 10 & 7 & 4 & 1 \\ 26 & 19 & 12 & 5 \\ 42 & 31 & 20 & 9 \end{pmatrix}.$$

Proposition 3.3.5

Suppose $T : U \to V$, $S : V \to W$ are linear maps with fixed bases. Then

$$\mathcal{M}(ST) = \mathcal{M}(S)\mathcal{M}(T).$$

This proof is the result of tedious computation; see Axler p.74.

§3.4 Invertibility and Isomorphic Vector Spaces

§3.4.1 Invertible Linear Maps

We start by looking at invertible linear maps, and inverses in the context of linear maps.

Definition 3.4.1: Invertible, Inverse

A linear map $T \in \mathcal{L}(V, W)$ is called **invertible** if there exists a linear map $S \in \mathcal{L}(W, V)$ such that $ST = I_V$, and $TS = I_W$. $S \in \mathcal{L}(W, V)$ is called the **inverse** of

T.

Proposition 3.4.1: Inverse is Unique

An invertible linear map is unique.

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

$$S_1 = S_1 I = S_1(TS_2) = (S_1T)S_2 = IS_2 = S_2.$$

□

Thus, we can denote the inverse of a map $T \in \mathcal{L}(V, W)$ uniquely by $T^{-1} \in \mathcal{L}(W, V)$. It turns out that invertibility is equivalent to bijectivity:

Proposition 3.4.2

A linear map is invertible if and only if it is bijective.

Proof. We first sketch a proof, then proceed formally.
In order to be an inverse of $T \in \mathcal{L}(V, W)$, an inverse $T^{-1} \in \mathcal{L}(W, V)$ must "undo" T, and vice versa; that is, for any $v \in V$, $T^{-1}T(v) = v$, and for any $w \in W$, $TT^{-1}(w) = w$. Thus, it is clear that

- T must be injective: if it is not, then at least two elements in V map to the same element in w, so how is a T^{-1} supposed to "undo" this operation?
- T must be surjective: if it is not, then some element in $w \in W$ is not mapped to by T; that is, no such $v \in V$ exists such that $T(v) = w$. Then how is T supposed to "undo" w, if it never maps to w?

Now, we proceed formally. Suppose T is invertible; that is, some $T^{-1} \in \mathcal{L}(W, V)$ has the property $TT^{-1} = T^{-1}T = I$. Suppose $T(u) = T(v)$ for some $u, v \in V$. Then

$$u = T^{-1}T(u) = T^{-1}T(v) = v,$$

and so T is injective. Now, let w be any element in W. Then $w = T(T^{-1}(w))$, and so $w \in \operatorname{range} T$. Thus $\operatorname{range} T = W$, and so T is surjective, and thus a bijection.

Conversely, suppose $T \in \mathcal{L}(V, W)$ is a bijection. Then for every $w \in W$, there exists a unique $v \in V$ such that $T(v) = w$. Define $S \in \mathcal{L}(W, V)$ as the function that sends every w back to its original v; that is, since $w = T(v)$, we have $T(S(w)) = T(v) = w$. Clearly, $T \circ S$ is the identity map I_W. To show that $S \circ T = I_V$, consider any $v \in V$. Then

$$T(S \circ T(v)) = (T \circ S)T(v) = I_W T(v) = T(v).$$

Then $T(S \circ T(v)) = T(v)$; but since T is an injection, we have $S \circ T(v) = v$, and so $S \circ T = I_V$. Thus $S = T^{-1}$.
It remains to prove that S is a linear map. In order to satisfy linear map properties, we need to show that

$$S(a_1w_1 + a_2w_2) = a_1S(w_1) + a_2S(w_2).$$

Since T is injective, we really only need to show that

$$T(S(a_1w_1 + a_2w_2)) = T(a_1S(w_1) + a_2S(w_2)).$$

Indeed, we have

$$\begin{aligned} T(S(a_1w_1 + a_2w_2)) &= a_1w_1 + a_2w_2 \\ &= a_1T(S(w_1)) + a_2T(S(w_2)) \\ &= T(a_1S(w_1)) + T(a_2S(w_2)) \\ &= T(a_1S(w1) + a_2S(w_2)). \end{aligned}$$

Hence S is a linear map. □

§3.4.2 Isomorphic Vector Spaces

From this, we get the sense that if an invertible linear map exists between two vector spaces, then they are essentially the same; they differ only in the names of their elements.

Definition 3.4.2: Isomorphic, Isomorphism

An **isomorphism** is an invertible linear map. Two vector spaces V, W are **isomorphic** if there is an isomorphism between the two; we write $V \cong W$.

Isomorphic vector spaces are really the same name; one can picture any element $v \in V$ as being "relabeled" as $T(v) \in W$. Thus, a natural proposition follows.

Theorem 3.4.1: Dimension of Isomorphic Vector Spaces

Let V, W be finite-dimensional vector spaces. Then $V \cong W$ if and only if $\dim V = \dim W$.

Proof. First, suppose $V \cong W$. Then there exists an isomorphism $T \in \mathcal{L}(V, W)$; moreover, T is bijective. Thus $\operatorname{null} T = 0$, and $\operatorname{range} T = W$, and so $\dim \operatorname{null} T = 0$, $\dim \operatorname{range} T = \dim W$. Then

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

becomes $\dim V = \dim W$, as required.
Conversely, suppose $\dim V = \dim W$. Let $v_1, \ldots, v_n$ be a basis for V, and let $w_1, \ldots, w_n$ be a basis for W. Let $T \in \mathcal{L}(V, W)$ be defined as

$$T(c_1v_1 + \ldots + c_nv_n) = c_1w_1 + \ldots + c_nw_n.$$

Then T is a well-defined, unique linear map, since $v_1, \ldots, v_n$ is a basis. T is surjective, since $w_1, \ldots, w_n$ spans W; moreover, $\operatorname{null} T = \{0\}$, since $w_1, \ldots, w_n$ is linearly independent, so $c_1w_1 + \ldots + c_nw_n = 0$ iff $c_i = 0$ (alternatively, one can use the rank-nullity theorem and the fact that $\dim \operatorname{range} T = \dim W = \dim V$, so $\dim \operatorname{null} T$ is necessarily 0). Thus T is injective, and so T is an isomorphism. Hence $V \cong W$, as required. □

From this, we get that any finite-dimensional vector space V with dimension n is actually isomorphic to $\mathbb{F}^n$! But then, why don't we only study the vector spaces $\mathbb{F}^n$, if they're really the same as any other vector space with dimension n? Studying vector spaces abstractly is extremely useful; for example, the polynomial space of dimension 15 is quite useful in physics.
Our discussion about dimensions and isomorphism lead us to the following proposition:

Proposition 3.4.3

Suppose V, W are finite-dimensional vector spaces with $\dim V = \dim W$, and let $T \in \mathcal{L}(V, W)$ be a linear map. The following are equivalent:

1. T is invertible/bijective.
2. T is injective.
3. T is surjective.

Proof. Let $n = \dim V = \dim W$. Then $n = \dim V = \dim \operatorname{null} T + \dim \operatorname{range} T$. So $\dim \operatorname{null} T = 0 \iff \dim \operatorname{range} T = n \iff \dim \operatorname{null} T = 0$ AND $\dim \operatorname{range} T = n$. In other words, T is injective iff T is surjective iff T is bijective. □

Chapter 4

Eigenvalues, Eigenvectors, and Invariant Subspaces

In chapter 3, we studied linear maps from one vector space to another. We now turn our attention to maps from finite-dimensional vector spaces to themselves.

§4.1 Invariant Subspaces

We start by developing tools to help us understand the structure of operators. Recall that an **operator** is a linear map from a vector space to itself, and we denote the set of operators on V by $\mathcal{L}(V)$; in other words, $\mathcal{L}(V) = \mathcal{L}(V, V)$.
To better understand what an operator looks like, suppose $T \in \mathcal{L}(V)$. If we have a direct sum decomposition

$$V = U_1 \oplus \ldots \oplus U_m,$$

where each U_j is a proper subspace of V, then to understand the behavior of T, we need only understand its behavior on each individual subspace, $T|_{U_j}$. Dealing with each individual subspace is generally easier, since it's a smaller vector space.
However, we have a problem: $T|_{U_j}$ might not map U_j into itself; in other words, $T|_{U_j}$ may not be an operator on U_j. Thus, we must only consider decompositions of V where T maps each U_j into itself.

Definition 4.1.1: Invariatn Subspaces

Suppose $T \in \mathcal{L}(V)$. A subspace U of V is called **invariant** under T if $u \in U$ implies $T(u) \in U$. That is, T maps U into itself; or, $T|_U$ is an operator on U.

Some examples of operators include $\{0\}$, V, $\operatorname{null} T$, and $\operatorname{range} T$.
Do operators $T \in \mathcal{L}(V)$ have any invariant subspaces other than $\{0\}$ and V? Later, we will see that this is true **if** V is finite-dimensional and $\dim V > 1$ (for $\mathbb{F} = \mathbb{C}$) or $\dim V > 2$ (for $\mathbb{F} = \mathbb{R}$). Although the null space and range are invariant under T, they don't necessarily tell us anything new about the existence of non-trivial invariant subspaces; we can very well have $\operatorname{null} T = \{0\}$ and $\operatorname{range} T = V$ (i.e. T is invertible).

§4.1.1 Eigenvalues and Eigenvectors

We'll return later to invariant subspaces. We now turn to the simplest possible nontrivial invariant subspaces – those with dimension 1.
Take any $v \in V$ with $v \neq 0$ and let U equal the set of all scalar multiples of v:

$$U = \{\lambda v \mid \lambda \in \mathbb{F}\} = \operatorname{span}(v).$$

Then U is a 1-dimensional subspace of V (and every 1-dimensional subspace of V is of this form). If U is invariant under an operator $T \in \mathcal{L}(V)$, then $T(v) \in U$, and hence

there is a scalar $\lambda \in \mathbb{F}$ such that

$$T(v) = \lambda v.$$

Conversely, if $T(v) = \lambda v$ for some $\lambda \in \mathbb{F}$, then span (v) is a 1-dimensional subspace of V invariant under T.
Whenever we find such λ that have this property for an operator $T \in \mathcal{L}(V)$, we call them **eigenvalues** ("eigen" comes from the German word for "same", since they preserve the linear map structure).

Definition 4.1.2: Eigenvalues, Eigenvectors

Suppose $T \in \mathcal{L}(V)$. A number $\lambda \in \mathbb{F}$ is called an **eigenvalue** of T if there exists a non-zero $v \in V$ such that $T(v) = \lambda v$. Such a vector is called an **eigenvector** of T corresponding to eigenvalue λ.
In other words, a vector $v \in V$ is an **eigenvector** with **eigenvalue** λ if $v \neq 0$ and $Tv = \lambda v$.

Intuitively, a vector has a corresponding eigenvalue if applying T on the vector will only result in a "stretch" (in some direction) of the vector. We require that $v \neq 0$ because every scalar $\lambda \in \mathbb{F}$ satisfies $T(0) = \lambda \cdot 0$.

Example 25. *Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be given by $T(x, y) = (2x, y)$. Does T have any eigenvalues? If so, what eigenvalues?*
***Yes**: there are two possible eigenvalues:*

- *$\lambda = 2$: for any eigenvector $(a, 0)$, we have $T(a, 0) = (2a, 0) = 2(a, 0)$.*
- *$\lambda = 1$: for any eigenvector $(0, b)$, we have $T(0, b) = (0, b) = 1(0, b)$.*

*Intuitively, T "stretches" a vector by 2 in the x-dimension. Thus, any vector on the x-axis will get doubled, and any vector **on** the y-axis will remain the same. However, any vector not on the axes will be distorted, not stretched.*

Proposition 4.1.1: Equivalent Conditions to be an Eigenvalue

Suppose V is finite-dimensional, $T \in \mathcal{L}(V)$, and $\lambda \in F$. Then the following are equivalent:

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

Proof. Conditions 1 and 2 are equivalent because the equation $T(v) = \lambda v$ is equivalent to the equation $(T - \lambda I)v = 0$; in other words, $T - \lambda I$ sends both 0 and v to zero. Hence $T - \lambda I$ is not injective (since it has a non-trivial kernel).
Conditions 2, 3, and 4 are equivalent by a previous theorem (stating that if dimensions are the same, injective iff surjective iff bijective/invertible). □

Because $T(v) = \lambda v$ if and only if $(T - \lambda I)v = 0$, a vector $v \in V$ with $v \neq 0$ is an eigenvector of T corresponding to λ if and only if $v \in \text{null}\,(T - \lambda I)$.
We see that, in the previous example, when trying to draw the eigenvectors, there isn't only one eigenvector associated with an eigenvalue; it's instead a space of eigenvectors that all fit into one form. We formalize this notion into an **eigenspace**.

Definition 4.1.3: Eigenspace

Let $T \in \mathcal{L}(V)$, and let $\lambda \in \mathbb{F}$ be an eigenvalue of T. The **eigenspace** of T corresponding to λ, denoted $E(\lambda, T)$, is

$$E(\lambda, T) = \{v \in V \mid T(v) = \lambda v\}.$$

Thus $E(\lambda, T)$ is the set of eigenvectors corresponding to λ, together with 0. This is a subspace of V: indeed, $E(\lambda, T) = \{v \in V \mid (T - \lambda I)(v) = 0\} = \text{null}\,(T - \lambda I)$; and null spaces are always subspaces.

Now, we show that eigenvectors corresponding to distinct eigenvalues are linearly independent.

Proposition 4.1.2: Linearly Independent Eigenvectors

Let $T \in \mathcal{L}(V)$. Suppose $\lambda_1, \ldots, \lambda_m$ are distinct eigenvalues of T and $v_1, \ldots, v_m$ are corresponding eigenvectors. Then $v_1, \ldots, v_m$ are linearly independent.

Proof. Suppose $v_1, \ldots, v_m$ are linearly dependent. Let v_j be the smallest j such that

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

(v_j exists by the Linear Dependence Lemma). Then

$$v_j = a_1 v_1 + \ldots + a_{j-1} v_{j-1};$$

applying T to both sides yields

$$\lambda_j v_j = a_1 \lambda_1 v_1 + \ldots + a_{j-1} \lambda_{j-1} v_{j-1}.$$

If we multiply λ_j to the first equation, then subtract from this equation, we get

$$0 = a_1(\lambda_j - \lambda_1)v_1 + \ldots + a_{j-1}(\lambda_j - \lambda_{j-1})v_{j-1}.$$

However, $v_1, \ldots, v_j$ are linearly independent (by construction); thus all the coefficients are 0; that is, $a_i(\lambda_j - \lambda_i) = 0$. Since λ_j is different from all $\lambda_{i \neq j}$, a_i must be 0. However, using this in the top equation implies that

$$v_j = 0v_1 + \ldots + 0v_{j-1} = 0,$$

a contradiction (since all eigenvectors must be non-zero). Hence $v_1, \ldots, v_j$ is linearly independent. □

Using this, we can identify the max number of possible eigenvalues.

Corollary 4.1.1: Number of Eigenvalues

Suppose V is finite-dimensional. Then each operator $T \in \mathcal{L}(V)$ has at most $\dim V$ distinct eigenvalues.

Proof. Let $T \in \mathcal{L}(V)$. Suppose $\lambda_1, \ldots, \lambda_m$ are distinct eigenvalues of T, with corresponding eigenvectors $v_1, \ldots, v_m$. Then the theorem above implies that the list $v_1, \ldots, v_m$ is linearly independent; thus $m \leq \dim V$ (since length of any linearly independent list $\leq$ length of any spanning list), as required. □

§4.2 Eigenvectors and Upper-Triangular Matrices

§4.2.1 Polynomials Applied to Operators

One main reason that a richer theory exists for operators, which map a vector space to itself, is that operators can be raised to powers. We start by defining this notion, and the key concept of applying a polynomial to an operator.
If $T \in \mathcal{L}(V)$, then TT makes sense and is also in $\mathcal{L}(V)$. Usually, we write T^2 instead of TT.

Definition 4.2.1: T^m

Suppose $T \in \mathcal{L}(V)$ and m is a positive integer.

- T^m is defined by
$$T^m = \underbrace{T \cdots T}_{m \text{ times}}.$$
- $T^0 = I_V$, the identity operator on V.
- If T is invertible with T^{-1} inverse, then T^{-m} is defined by
$$T^{-m} = (T^{-1})^m.$$

One can easily verify that if T is an operator, then

$$T^m T^n = T^{m+n} \text{ and } (T^m)^n = T^{mn}.$$

Definition 4.2.2: $p(T)$

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

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

for $z \in \mathbb{F}$. Then $p(T)$ is the operator defined by

$$p(T) = a_0 I + a_1 T + \ldots + a_m T^m.$$

This is a new use of the polynomial symbol p, since we're applying it to operators, not just variables in $\mathbb{F}$.

Example 26. *Consider $D : \mathcal{P}(\mathbb{R}) \to \mathcal{P}(\mathbb{R})$, the differentiation linear operator. What is $p(D)$?*
We have $p(T) = I + T^2$; thus, for a function $f \in \mathcal{P}(\mathbb{R})$, we have $p(T)(f) = (I + T^2)(f) = I(f) + T^2(f) = f + f''$.

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

Definition 4.2.3: 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 $z \in \mathbb{F}$.

Any two polynomials of an operator commute:

Proposition 4.2.1: Multiplicative properties

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

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

The proof is left as an exercise for the reader; both follow basically directly from properties of polynomials (one simply replaces z with T).

§4.2.2 Existence of Eigenvalues

Now, we come to one of the central results about operators on complex vector spaces.

Theorem 4.2.1: Operators on Complex Vector Spaces Have an Eigenvalue

Every operator on a finite-dimensional, non-zero, complex vector space has an eigenvalue.

Proof. Suppose V is a complex vector space with dimension $n > 0$, and $T \in \mathcal{L}(V)$. Choose a non-zero $v \in V$. Then

$$v, T(v), \ldots, T^n(v)$$

is not linearly independent, because V has dimension n and we have $n+1$ vectors. Thus there exist complex numbers $a_0, \ldots, a_n$, not all 0, such that

$$0 = a_0 v + a_1 T(v) + \ldots + a_n T^n(v).$$

Note that $a_1, \ldots, a_n$ cannot all be 0, since otherwise it would imply $a_0 v = 0$, or $a_0 = 0$, a contradiction of linear dependence (Basically, at least two–definitely more since otherwise a vector is a scalar multiple of another–of a_i must be non-zero).
Make the a_i the coefficients of a polynomial, which by the Fundamental Theorem of Algebra (every non-constant polynomial has a unique complex factorization) has a factorization

$$a_0 + a_1 z + \ldots + a_n z^n = c(z - \lambda_1) \cdots (z - \lambda_m),$$

where c is a non-zero complex number, each $\lambda_j \in \mathbb{C}$, and the equation holds for all $z \in \mathbb{C}$ (here m is not necessarily equal to n, because a_n may equal 0).
We then have

$$\begin{aligned} 0 &= a_0 v + a_1 T(v) + \ldots + a_n T^n(v) \\ &= (a_0 I + a_1 T + \ldots + a_n T^n) v \\ &= c(T - \lambda_1 I) \cdots (T - \lambda_m I) v = p(T)(v). \end{aligned}$$

Thus $T - \lambda_j I$ is not injective for at least one j, since at least one of the $T - \lambda_j I$'s must be 0 (both c, v are non-zero). In other words, T has an eigenvalue (recall that $T - \lambda_j I = 0$ implies $T(v) = \lambda_j v$, so λ_j is an eigenvalue). □

§4.2.3 Upper Triangular Matrices

When we looked at matrices of maps from one vector space to another, the matrix depended on the choice of basis of each of the two vector spaces. With operators, however, since it's a map from a vector space to itself, we only need one basis.

Definition 4.2.4: Matrix of an operator, $\mathcal{M}(T)$

Suppose $T \in \mathcal{L}(V)$ and $v_1, \ldots, v_n$ is a basis of V. The **matrix of** T with respect to this basis is the $n \times n$ matrix

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

whose entries $A_{j,k}$ is defined by

$$T(v_k) = A_{1,k} v_1 + \ldots + A_{n,k} v_n.$$

In other words, each column in the basis represents the result of applying the operator on that basis element.

Note that matrices of operators are square arrays, rather than the more general rectangular arrays. This means that squaring matrices is possible, since their dimensions always agree! (When no basis is specified, assume the standard basis.)
A central goal of linear algebra is to show that given an operator $T \in \mathcal{L}(V)$, there exists a basis of V with respect to which T has a reasonably simple matrix. This allows for much easier computations, since sparse matrices result are much easier to operate with. To make this more precise, we might try to choose a basis for V such that $\mathcal{M}(T)$ has many 0's. If V is a finite-dimensional complex vector space, since an eigenvalue/eigenvector pair exists, we can already form a matrix with 0's everywhere along the first column, except the first entry. We can let v an eigenvector be a part of the basis, since any non-zero vector can be extended to a basis.

Soon, we will see that we can choose a basis of V such that the matrix $\mathcal{M}(T)$ has even more 0's.

Definition 4.2.5: Diagonal of a Matrix

The **diagonal** of a square matrix consists of the entries along the diagonal.

Definition 4.2.6: Upper-Triangular Matrix

A matrix is **upper-triangular** if all the entries below the diagonal equal 0.

Typically, we represent an upper-triangular matrix in the form

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

The 0 in the matrix above indicates that all entries below the diagonal are 0.

§4.2.4 TODO: Finish Upper Triangular Matrices

Important Propositions/Theorems:

Proposition 4.2.2: Conditions for Upper-Triangular Matrices

Suppose $T \in \mathcal{L}(V)$ and $v_1, \ldots, v_n$ is a basis for V. Then the following are equivalent:

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

Theorem 4.2.2: Over $\mathbb{C}$, Every Operator has an Upper-Triangular Matrix

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

Proposition 4.2.3: Determination of Invertibility from Upper-Triangular Matrix

Suppose $T \in \mathcal{L}(V)$ has an upper-triangular matrix with respect to some basis of V. Then T is invertible if and only if all the entries on the diagonal of that upper triangular matrix are non-zero.

Proposition 4.2.4: Determination of Eigenvalues from 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.

§4.3 Eigenspaces and Diagonal Matrices

Definition 4.3.1: Diagonal Matrix

A **diagonal matrix** is a square matrix that is 0 everywhere except possibly along the diagonal.

Obviously, every diagonal matrix is upper-triangular; in general, a diagonal matrix has many more 0s than an upper-triangular matrix.
If an operator has a diagonal matrix with respect to some basis, then the entries along the diagonal are precisely the eigenvalues of the operator. We can then formulate a basis out of solely eigenvectors.

Definition 4.3.2: Eigenspace

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

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

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

For $T \in \mathcal{L}(V)$ and $\lambda \in \mathbb{F}$, the eigenspace $E(\lambda, T)$ is a subspace of V, since the null space of any linear map $T \in \mathcal{L}(V)$ is a subspace of V; and definitions imply that λ is an eigenvalue if and only if $E(\lambda, T) \neq \{0\}$. Moreover, if λ is an eigenvalue of an operator $T \in \mathcal{L}(V)$, then $T|_{E(\lambda,T)}$ is just multiplication by λ!

Proposition 4.3.1: Sum of Eigenspaces is Direct Sum

Suppose V is finite-dimensional and $T \in \mathcal{L}(V)$. Suppose also that $\lambda_1, \ldots, \lambda_m$ are distinct eigenvalues of T. Then

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

is a direct sum. Furthermore,

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

Proof. Suppose $u_1 + \ldots + u_m = 0$, where $u_j \in E(\lambda_j, T)$. Since eigenvectors corresponding to distinct eigenvalues are linearly independent, this implies that every $u_j = 0$. Thus $E(\lambda_1, T) + \ldots + E(\lambda_m, T)$ is a direct sum, as desired.
Direct sums is left as an exercise; it should follow quickly from a previous exercise on direct sums and dimensions. □

Definition 4.3.3: Diagonalizable

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

Proposition 4.3.2: Conditions Equivalent to Diagonalizability

Suppose V is finite-dimensional and $T \in \mathcal{L}(V)$. Let $\lambda_1, \ldots, \lambda_m$ denote the distinct eigenvalues of T. Then the following are equivalent:

- T is diagonalizable
- V has a basis consisting of eigenvectors of T
- There exist 1-dimensional subspaces $U_1, \ldots, U_n$ of V, each invariant under T, such that
$$V = U_1 \oplus \ldots \oplus U_n.$$
- $V = E(\lambda_1, T) \oplus \ldots \oplus E(\lambda_m, T)$.

Proof. An operator $T \in \mathcal{L}(V)$ has a diagonal matrix

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

with respect to a basis $v_1, \ldots, v_n \in V$ if and only if $T(v_j) = \lambda_j v_j$ for each v_j. Thus (1) and (2) are equivalent.
We skip the proof for (3) (I'm lazy).
Suppose V has a basis consisting of eigenvectors of T. Then every vector in V is a linear combination of eigenvectors of T, which implies that

$$V = E(\lambda_1, T) + \ldots + E(\lambda_m, T).$$

(This is true since each $E(\lambda_j, T)$ represents all scalar multiples of the eigenvector v; recall that this eigenspace is the result of scaling a vector along a line). Thus (4) holds. □

Unfortunately, not every operator is diagonalizable, even under complex vector spaces.

Example 27. *Show that the opeartor $T \in \mathcal{L}(\mathbb{C}^2)$ defined by*

$$T(w, z) = (z, 0)$$

is not diagonalizable. Solution: One can verify that 0 is the only eigenvalue of T, and further, that $E(0, T) = \{(w, 0) \in \mathbb{C}^2 \mid w \in \mathbb{C}\}$; we see this since the only instance when this is true is when $T(w, 0) = (0, 0)$ (any other $T(w, z) = (z, 0)$ will alter its orientation). Thus one can easily see that since V has dimension 2, it cannot have a basis consisting of only eigenvectors of T. Checking that a diagonal matrix doesn't exist is easy; with the above proposition, one need only check that one of the conditions fails.

The next result lets you determine the diagonalizability of an operator given its eigenvalues.

Theorem 4.3.1: Enough Eigenvalues implies Diagonalizability

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

Proof. Suppose $T \in \mathcal{L}(V)$ has $\dim V$ distinct eigenvalues $\lambda_1, \ldots, \lambda_{\dim V}$. Let $v_j \in V$ be an eigenvector corresponding to each distinct eigenvalue λ_j. Since eigenvectors corresponding to distinct eigenvalues are linearly independent, the list

$$v_1, \ldots, v_{\dim V}$$

is linearly independent. Since every linearly independent list of $\dim V$ vectors in V forms a basis for V, we know that $v_1, \ldots, v_{\dim V}$ forms a basis of V. Thus T has a basis of eigenvectors, and therefore is diagonalizable. □

§4.4 Determinants

We first explore the motivation behind determinants. A first introduction of determinants is a naive, computation-focused approach. Given a matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

we define $\det(A) = ad - bc$.
After that, state-school linear algebra courses usually stop there; but determinants are quite important in linear algebra, and understanding their motivation and intuition is paramount.
Let $T : \mathbb{R}^n \to \mathbb{R}^n$. Geometrically, we can think of this linear map as "morphing/stretching" a unit cube into a parallelopiped, or other shape; we can picture $e_1 \mapsto T(e_1)$, $e_2 \mapsto T(e_2), \ldots$ as a transformation of the unit cube. **The crucial notion of the determinant** of T is that it should measure the *oriented/signed* **"volume of scaling" of the image of the unit cube** under T. Orientation is important; a negative determinant would signify a change in orientation.
Recall: the vectors $T(e_1), \ldots, T(e_n)$ written with respect to the standard basis are the **columns of** $\mathcal{M}(T)$. So, we want a way to associate a number, the **determinant**, to any $n \times n$ matrix A; equivalently, to any n vectors $v_1, \ldots, v_n \in \mathbb{F}^n$.
WISH LIST. Let $\mathbb{F}$ be a field. We wish for a function

$$D : \underbrace{\mathbb{F}^n \times \ldots \times \mathbb{F}^n}_{n \text{ times}} \longrightarrow \mathbb{F}, \text{ or } D : \mathbb{F}^{n,n} \longrightarrow \mathbb{F}$$

that associates to a list of vectors in $\mathbb{F}^n$ with a number. Moreover, we want:

1. D is **multilinear**, i.e. for each $k = 1, \ldots, n$, we have

   $$D(v_1, \ldots, av_k + a'v_k', \ldots, v_n) = aD(v_1, \ldots, v_k, \ldots, v_n) + a'D(v_1, \ldots, v_k', \ldots, v_n).$$

   In other words, D is linear with respect to each k. For example, given a three dimensional object, if we fix width and length, we want linearity to hold when we stretch height. Moreover, each scalar for a dimension would affect the overall scaling; for example, increasing width by 2, height by 3, and length by 4 would result in $2 \cdot 3 \cdot 4 = 24$ times the volume.

2. D is **alternating**, i.e.

$$D(v_1, \ldots, v_j, \ldots, v_k, \ldots, v_n) = 0 \text{ if } v_j = v_k \text{ for } j \neq k.$$

In other words, the determinant of a matrix should be 0 if two columns (or rows) are equal (this has important implications; this means the determinant is 0 if two mapped basis vectors, e.g. Te_i, Te_j, are linearly dependent!). For example, suppose we have a three-dimensional object, with width and length vectors. If the height vector is in the span of the other two vectors, then the shape would have no volume (since it never leaves the 2D plane).

3. D is normalized, i.e. if we take the determinant of the unit cube, we should get 1 (which aligns with our intuition):

$$D(e_1, ..., e_n) = 1.$$

(We want this to prevent the zero map.)

Remark 6. ***CAUTION***: *$D : \mathbb{F}^{n,n} \to \mathbb{F}$ is **NOT** going to be a linear map. Consider T a linear map that morphs a unit square into a parallelopiped. If we take $2T$, then the linear map will double each side of the parallelopiped (which is **FOUR** times larger than the original!). In general, we have $\det(\lambda T) = \lambda^n \det T$ (where n is the number of dimensions).*

Proposition 4.4.1

Suppose $D : \mathbb{F}^n \times \ldots \times \mathbb{F}^n \to \mathbb{F}$ is a map that satisfies (1) and (2) above. Then

$$D(v_1, \ldots, v_j, \ldots, v_k, \ldots, v_n) = -D(v_1, \ldots, v_k, \ldots, v_j, \ldots, v_n) \text{ with } j \neq k.$$

In general, this should hold for multilinear, alternating maps. In mundane terms, **swapping two rows/columns flips the sign of the determinant**.

Proof. Consider $0 = D(v_1, \ldots, v_j + v_k, \ldots, v_j + v_k, \ldots, v_n)$ (which we get by (2)). Then

$$\begin{aligned} 0 &= D(v_1, \ldots, v_j + v_k, \ldots, v_j + v_k, \ldots, v_n) \\ &= D(v_1, \ldots, v_j, \ldots, v_j, \ldots, v_n) + D(v_1, \ldots, v_j, \ldots, v_k, \ldots, v_n) \\ &+ D(v_1, \ldots, v_k, \ldots, v_j, \ldots, v_n) + D(v_1, \ldots, v_k, \ldots, v_k, \ldots, v_n). \end{aligned}$$

By the second property, the first and last terms become 0. Hence

$$D(v_1, \ldots, v_j, \ldots, v_k, \ldots, v_n) = -D(v_1, \ldots, v_k, \ldots, v_j, \ldots, v_n).$$

□

Theorem 4.4.1: Determinant

There exists a unique multilinear, alternating, normalized function

$$D : \mathbb{F}^n \times \ldots \times \mathbb{F}^n \longrightarrow \mathbb{F}.$$

We call this function the **determinant**.

We postpone the proof of the existence and uniqueness of the determinant for later, and assume it's true. Especially beyond 3×3 matrices, the formula for the determinant becomes super messy; moreover, it is one of those topics in mathematics where it's more important to use it based on its properties, rather than knowing its definition/proof.

Proposition 4.4.2: Determinant of Linearly Dependent Lists

Let $A \in \mathbb{F}^{n,n}$ be a square matrix. If the columns $v_1, \ldots, v_n \in \mathbb{F}^n$ are linearly dependent, then $\det A = D(v_1, \ldots, v_n) = 0$.

Proof. Say $v_k = a_1v_1 + \ldots + a_{k-1}v_{k-1} + a_{k+1}v_{k+1} + \ldots + a_nv_n$, where $a_i \in \mathbb{F}$. Calculate

$$\begin{aligned} D(v_1, \ldots, v_n) &= D(v_1, \ldots, a_1v_1 + \ldots + a_{k-1}v_{k-1} + a_{k+1}v_{k+1} + \ldots + a_nv_n, \ldots, v_n) \\ &= a_1D(v_1, \ldots, v_1, \ldots, v_n) + a_2D(v_1, \ldots, v_2, \ldots, v_n) + a_nD(v_1, \ldots, v_n, \ldots, v_n) \end{aligned}$$

by multilinearity of D. But all of these determinants share a value in common, so by the alternating property, we get

$$D(v_1, \ldots, v_n) = 0 + \ldots + 0 = 0.$$

□

We'll show later that the converse is true as well; that is, if $\det A = 0$, then the columns are dependent.
Other useful cases of the determinant:

- Given a diagonal matrix, the determinant of the diagonal matrix is simply the multiple of each value:

$$\det \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} = \lambda_1 \cdot \ldots \cdot \lambda_n.$$

- What about upper triangular matrices? Consider the matrix

$$\begin{vmatrix} 1 & 0 & 5 \\ 0 & 1 & 7 \\ 0 & 0 & 1 \end{vmatrix} = \begin{vmatrix} 1 & 0 & 5-5 \\ 0 & 1 & 7-0 \\ 0 & 0 & 1-0 \end{vmatrix},$$

  since $D(v_1, \ldots, v_n) = D(v_1, \ldots, v_k + a_1v_1, \ldots, v_n)$ (since by breaking up the right hand side using multilinearity, we get $D(v_1, \ldots, v_k, \ldots, v_n) + a_1D(v_1, \ldots, v_1, \ldots, v_n)$). In other words, **adding columns/rows to different columns/rows does not affect the determinant**.

  In general, the determinant of an upper triangular matrix is $\lambda_1 \cdot \ldots \cdot \lambda_n$ (just the values on the diagonal); it follows that $\lambda_i = 0$ implies that $\det A = 0$. Indeed, if $\lambda_i = 0$, then the first i columns $v_1, \ldots, v_i$ of A lie in the span $(e_1, ..., e_{i-1})$. So the columns of A are dependent.

  If all $\lambda_i \neq 0$, then we can subtract the first column (which only has a non-zero entry in the first row) from the rest of the rows, "clearing" the first row; recall that **this doesn't affect the overall value of the determinant**. We repeat for all the other rows, and eventually get a diagonal matrix, which results in $\lambda_1 \cdot \ldots \cdot \lambda_n$. Geometrically, we get the notion that even if we "shear" a unit cube; that is, we shift it in a certain direction, the overall volume of the shape **does not change**.

§4.4.1 Existence and Uniqueness of Determinants

Theorem 4.4.2: Existence and Uniqueness of Determinants

Let $\mathbb{F}$ be a field, and n an integer. There exists a unique function

$$D : \mathbb{F}^n \times \ldots \times \mathbb{F}^n \longrightarrow \mathbb{F}$$

that is

- **multilinear**: D is linear in each of the copies of $\mathbb{F}^n$
- **alternating**: $D(v_1, \ldots, v_j, \ldots, v_k, \ldots, v_n) = 0$ if $v_j = v_k$. (We call this alternating because swapping rows/columns thus inverts the value of the determinant)
- **normalized**: $D(e_1, \ldots, e_n) = 1$.

Proof. We start with $n = 2$. If such a D exists, given arbitrary $a, b, c, d \in \mathbb{F}$, we look at

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = D((a, c), (b, d)).$$

This then becomes

$$\begin{aligned} D(ae_1 + ce_2, be_1 + de_2) &= aD(e_1, be_1 + de_2) + cD(e_2, be_1 + de_2) \\ &= abD(e_1, e_1) + adD(e_1, e_2) + bcD(e_2, e_1) + cdD(e_2, e_2) \\ &= ad - bc. \end{aligned}$$

Thus, if such a D exists, it is unique in that it must send $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ to $ad - bc$.

We now show that $D((a, c), (b, d)) = ad - bc$ really is multilinear, alternating, and normalized.

- Normalized: this is pretty easy, $D((1, 0), (0, 1)) = 1 - 0 = 1$.
- Alternating: similarly, $D((a, c), (a, c)) = ac - ac = 0$.
- Multilinear: regarding b, d as constants, $ad-bc$ is linear in a, c. Similarly, regarding a, c as constants, $ad-bc$ is linear in b, d (when we say linear in a, c, imagine $3x-4y$; such a function is linear. We can just treat a, c as variables, and b, d as constants). Slightly more formally, given fixed $b, d \in \mathbb{F}$, we wish to show that the function

  $$D_{b,d} : \mathbb{F}^2 \longrightarrow \mathbb{F}, \ (a, c) \mapsto ad - bc$$

  is linear (in the usual sense).

At first glance, this seems circular. We assume existence to prove uniqueness, then use the discovered uniqueness formula to prove existence. However, there really is no circularity! We use uniqueness to figure out what D **would** have to be, **if** such a D existed. But there's no guarantee that such a D is multilinear, alternating, and normalized.

Now, we prove for $n = 3$. If D exists, then for $a, b, c, d, e, f, g, h, i \in \mathbb{F}$,

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = D(ae_1 + be_2 + ce_3, de_1 + ee_2 + fe_3, ge_1 + he_2 + ie_3).$$

Instead of applying multilinearity 27 times (since there are $3 \cdot 3 \cdot 3 = 27$ ways to choose three elements from each coordinate), we only pick up the terms that give us non-zero terms (since there will be many repeats, thus alternating will guarantee that the determinant is zero). Thus, selecting only the ones with non-zero values, we get

$$aeiD(e_1, e_2, e_3) + bfgD(e_3, e_1, e_2) + cdhD(e_2, e_3, e_1) + afhD(e_1, e_3, e_2) + bdiD(e_2, e_1, e_3) + cegD(e_3, e_2, e_1).$$

(We obtain the above non-zero terms by choosing different e_1, e_2, e_3 from each coordinate). But we notice that this becomes

$$a(ei - fh) - b(fg - di) + c(dh - eg) = a\begin{vmatrix} e & f \\ h & i \end{vmatrix} - b\begin{vmatrix} d & f \\ g & i \end{vmatrix} + c\begin{vmatrix} d & e \\ g & h \end{vmatrix}.$$

We thus proved uniqueness of determinants for $n = 3$. Existence is left as an exercise for the reader. □

We now look at a combinatorial definition.

Definition 4.4.1: Permutations

A **permutation** is a bijection $\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}$, which can be notated in cycle notation. The **product** of permutations σ and τ is

$$\sigma\tau = \sigma \circ \tau : \{1, \ldots, n\} \to \{1, \ldots, n\}.$$

Remark 7. *Permutations of $\{1, \ldots, n\}$ form a group, the **symmetric group** $\mathcal{S}_n$.*

Definition 4.4.2: Transpositions

A **transposition** is a permutation is a single swap. For any transposition τ, $\tau^2 = I$.

Proposition 4.4.3

Every permutation is a product of transpositions.

Proof. Physical proof, using solo cups :) □

Definition 4.4.3: Sign

The **sign** of a permutation σ, denoted $\operatorname{sgn}(\sigma)$, is

$$\operatorname{sgn}(\sigma) = \begin{cases} +1 & \text{if } \sigma \text{ is the product of an even number of transpositions} \\ -1 & \text{if } \sigma \text{ is the product of an odd number of transpositions} \end{cases}.$$

We now use this to prove the existence and uniqueness of determinants for all n.

Proof. If the $n \times n$ determinant exists, then it must be

$$\begin{vmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{n,1} & \cdots & a_{n,n} \end{vmatrix} = D(a_{1,1}e_1 + \ldots + a_{n,1}e_n, \ldots, a_{1,n}e_1 + \ldots + a_{n,n}e_n).$$

Using multilinearity, we then get

$$\sum_{\sigma \in \mathcal{S}_n} a_{\sigma(1),1} \ldots a_{\sigma(n),n} D(e_{\sigma(1)}, e_{\sigma(2)}, \ldots, e_{\sigma(n)}),$$

where $D(e_{\sigma(1)}, e_{\sigma(2)}, \ldots, e_{\sigma(n)}) = \operatorname{sgn}(\sigma)$. This proves uniqueness of the determinant; again, existence is left as an exercise. □

However, we have not shown yet whether the sign function is actually well-defined. We now prove well-definedness of the sign function. To do this, we must show that if σ is the product of k transpositions, and is also the product of l transpositions, then k, l have the same parity. It's enough to be able to deduce the parity of the number of transpositions in a product expression for σ from σ itself.

Definition 4.4.4: Descent

A **descent** in σ is a pair (i, j) for $1 \leq i \leq j \leq n$ such that $\sigma(i) > \sigma(j)$.

For example, if $\sigma = (15)$, the descents of σ are: $(1,2), (1,3), (1,4), (1,5), (2,5), (3,5), (4,5)$.

How does a transposition affect the number of descents in a permutation σ?

It turns out that adjacent transpositions change the number of descents by ± 1; convince yourself of this using solo cups! With arbitrary transpositions, it turns out that it requires an odd number of transpositions, and so changes the number of descents by an odd amount.
Therefore,

Proposition 4.4.4

If σ is a product of an odd number of transpositions, then the number of descents of σ is odd. If σ is a product of an even number of transpositions, then the number of descents of σ is even.

We are done! The sign is well-defined, and so is the determinant. But why? The proposition above implies that if the number of descents is odd, then it **must** be the product of an odd number of transpositions; likewise with even. Thus, a permutation cannot be represented as both an odd and even number of transpositions, and so the sign of a permutation is the same, regardless of transposition representation.

Remark 8. *Note that the sign of σ and σ^{-1} are the same, since* $\operatorname{sgn}(\sigma_1\sigma_2) = \operatorname{sgn}(\sigma_1)\operatorname{sgn}(\sigma_2)$, *and* $\operatorname{sgn}\sigma\sigma^{-1} = \operatorname{sgn}(I) = 1$.

§4.4.2 Properties of Determinants

Now that we have rigorously proven the existence and uniqueness of the determinant, we now inspect properties of the determinant.

Definition 4.4.5: Transpose

Let $A \in \mathbb{F}^{m,n}$ be a matrix. The transpose of A, denoted A^T, is the $n \times m$ matrix given by $(A^T)_{i,j} = A_{j,i}$ (Intuitively, the transpose flips the matrix along the diagonal).

For example, $\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}^T = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}$.

Remark 9. *If $A = \mathcal{M}(T)$ for $T : V \to W$, then $A^T = \mathcal{M}(T^*)$ for $T^* : W^* \to V^*$, where T^* denotes the **dual** linear map [Axler Ch3].*

Proposition 4.4.5: Det of Transpose is Same

Let $A \in \mathbb{F}^{n,n}$. Then $\det(A^T) = \det(A)$.

Proof. $\det(A^T) = \sum_{\sigma \in \mathcal{S}_n} a^T_{\sigma(1),1} \dots a^T_{\sigma(n),n} \cdot \operatorname{sgn}(\sigma)$; here $a^T_{i,j}$ is the i,jth entry of A^T. But then by the inverse of transpositions,

$$\begin{aligned} \det(A^T) &= \sum_{\sigma \in \mathcal{S}_n} a^T_{\sigma(1),1} \dots a^T_{\sigma(n),n} \cdot \operatorname{sgn}(\sigma) \\ &= \sum_{\sigma \in \mathcal{S}_n} a_{1,\sigma(1)} \dots a_{n,\sigma(n)} \operatorname{sgn}(\sigma) \\ &= \sum_{\sigma \in \mathcal{S}_n} a_{\sigma^{-1}(1),1} \dots a_{\sigma^{-1}(n),n} \operatorname{sgn}(\sigma). \end{aligned}$$

We get this last step by re-ordering (since it's a bijection, all of $1, \dots, n$ are mapped uniquely by σ). However, since the sign of σ^{-1} is the same as the sign of σ; moreover, (convince yourself that this is true!) taking the sum of all permutations is the same as taking the sum of all inverses of permutations (due to group properties, a unique inverse exists!), and so the two sums are identical. Thus the determinant ends up being the same:

$$\det(A) = \det(A^T).$$

□

Definition 4.4.6: Cofactors

Let A be an $n \times n$ matrix. Let $A_{j,k}$ denote the $(n-1) \times (n-1)$ matrix obtained from A by deleting row j, column k. The numbers $(-1)^{j+k} \det(A_{j,k}) = C_{j,k}$ are the **cofactors** of A.

We can use cofactor expansion to thus find the determinant of a matrix!

Theorem 4.4.3: Cofactor Expansion

For each $1 \leq j \leq n$,

$$\det(A) = a_{j,1}C_{j,1} + \ldots + a_{j,n}C_{j,n}.$$

Example 28. *For* $\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$, *the determinant is*

$$\det(A) = 1 \cdot \begin{vmatrix} 5 & 6 \\ 8 & 9 \end{vmatrix} - 2 \begin{vmatrix} 4 & 6 \\ 7 & 9 \end{vmatrix} + 3 \begin{vmatrix} 4 & 5 \\ 7 & 8 \end{vmatrix}.$$

Proof. It suffices to show that the function given by the cofactor expansion is multilinear, alternating, and normalized (since there's only one unique function, we only need to prove its properties!).
Normalized is trivial; with alternating, two terms cancel (since they will have opposite signs), and the other ones have repeated columns (and so the determinant is zero). Multilinear is left as an exercise (but each of the n terms is linear in the columns of A). □

With cofactors, we can now find inverses of square matrices. Let A be an $n \times n$ matrix. The inverse of A, if it exists, is a matrix B such that

$$AB = I_n = BA.$$

We've seen the notion of invertibility for linear operators before; this is the same concept. If we view $A = \mathcal{M}(T)$ as the matrix of some linear operator $T \in \mathcal{L}(V)$, then A is invertible if and only if T is invertible (Why? If $AB = I = BA$, then $TS = ST = I$ where $\mathcal{M}(S) = B$). Also note that a linear operator $T \in \mathcal{L}(V)$ is invertible if and only if its basis is sent to another basis; e.g. $Te_1, \ldots, Te_n$ is a linearly independent set. In other words,

T is invertible if and only if the columns of $A = \mathcal{M}(T)$ are linearly independent.

Thus, there's another way of checking invertibility (which is the method used by most introductory linear algebra courses):

Proposition 4.4.6: Invertibility and Determinants

Let A be an $n \times n$ matrix. Then A is invertible (i.e. columns are L.I.) if and only if $\det(A) \neq 0$.

Proof. We've already proved that if the columns are linearly dependent, then $\det A = 0$. We now wish to show that if $\det A = 0$, then the columns are linearly dependent.
Proof postponed... □

Proposition 4.4.7

Computing Inverses) Let A be an $n \times n$ matrix. Let C be its matrix of cofactors. If $\det A \neq 0$, then A is invertible with $A^{-1} = \frac{1}{\det A} C^T$.

We can see this for 2×2 matrices: if $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, then

$$C = \begin{pmatrix} d & -c \\ -b & a \end{pmatrix}, C^T = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Next, note that $AC^T = \begin{pmatrix} ad-bc & 0 \\ 0 & ad-bc \end{pmatrix}$; thus

$$\frac{1}{ad-bc} AC^T = \frac{1}{\det A} AC^T = I,$$

giving us the result as required.

Proof. Note that

$$(AC^T)_{j,j} = a_{j,1}C^T_{1,j} + \ldots + a_{j,n}C^T_{n,j} = a_{j,1}C_{j,1} + \ldots + a_{j,n}C_{j,n} = \det A,$$

by the cofactor expansion formula for the determinant. Thus every diagonal entry is $\det A$; we now wish to show that every $(AC^T)_{k,j}$, $j \neq k$, is 0:

$$(AC^T)_{k,j} = a_{k,1}C^T_{1,j} + \ldots + a_{k,n}C^T_{n,j} = a_{k,1}C_{j,1} + \ldots + a_{k,n}C_{j,n}.$$

Recall now that the $C_{j,k}$ entry is given by

$$C_{j,k} = (-1)^{j+k} \det A_{j,k}.$$

This isn't immediately obvious, but now consider a matrix from A with row j replaced with row k; then $(AC^T)_{k,j}$ is exactly the determinant of this new matrix. But this determinant evaluates to 0, because it has two equal rows! Thus for all $j \neq k$,

$$(AC^T)_{k,j} = 0.$$

This completes the proof, since we now see that

$$AC^T = \det A \cdot I.$$

□

We now look at another property of determinants:

Proposition 4.4.8

Let A, B be $n \times n$ matrices. Then

$$\det AB = \det A \cdot \det B.$$

Proof. Case 1: if $\det A = 0$, then A not invertible, implying AB is not invertible (since otherwise, it has inverse C, so $(AB)C = I$ and $A(BC) = I$, a contradiction). Thus $\det AB = 0$.
Case 2: if $\det A \neq 0$, let's show, for all $n \times n$ matrices B,

$$\frac{\det AB}{\det A}$$

is multilinear, alternating, and normalized as a function on B (recall that since we proved that there is only one unique function on $F^{n\times n}$, if such a function exists, then that function is $\det B$. Proving this thus proves that $\det B = \frac{\det AB}{\det A}$).

- Normalized: if $B = I$, then $\det AB = \det A$, so $\frac{\det A}{\det A} = 1$.
- Alternating: Say B has two equal columns (since $\det A \neq 0$, so A doesn't have repeating columns). Then AB has two equal columns (verify this!), so

$$\frac{\det AB}{\det A} = 0.$$

- Multilinearity in columns of B: omitted, but verify! Essentially, one can split B into B', B'', and thus we can see multilinearity.

Thus $\det AB = \det A \det B$. □

§4.4.3 Determinants and Eigenvalues

Finally, we investigate how determinants can be used to compute eigenvalues and eigenvectors. Let $T : V \to V$, V finite-dimensional. How to find eigenvalues of T?
Recall that a scalar $\lambda \in \mathbb{F}$ is an eigenvalue of T if and only if

$$T - \lambda I \text{ is not invertible,}$$

but this is true if and only if $\mathcal{M}(T - \lambda I)$ is not invertible, with respect to some basis of V.
Let $A = \mathcal{M}(T)$. Then $\lambda \in \mathbb{F}$ is an eigenvalue if

$$\begin{vmatrix} a_{1,1} - \lambda & & & a_{1,n} \\ & a_{2,2} - \lambda & & \\ & & \ddots & \\ a_{n,1} & & & a_{n,n} - \lambda \end{vmatrix} = 0.$$

Example 29. *Consider the reflection across $y = x$, $T : \mathbb{R}^2 \to \mathbb{R}^2$, $T(x, y) = (y, x)$.*

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

Thus the eigenvalues of T are exactly the roots of the polynomial

$$\det \begin{pmatrix} 0 - \lambda & 1 \\ 1 & 0 - \lambda \end{pmatrix} = \lambda^2 - 1,$$

and so the eigenvalues are $\lambda_1 = 1,\ \lambda_2 = -1$. *Using this, we can now solve for the eigenspaces of* $1, -1$ *respectively:*

$$\begin{aligned} E(1,T) &= \{(x,y) \mid T(x,y) = 1(x,y)\} \\ &= \{(x,y) \mid (y,x) = (x,y)\} \\ &= \{(x,y) \mid y = x\}, \end{aligned}$$

and

$$\begin{aligned} E(-1,T) &= \{(x,y) \mid T(x,y) = -(x,y)\} \\ &= \{(x,y) \mid (y,x) = -(x,y)\} \\ &= \{(x,y) \mid y = -x\}. \end{aligned}$$

We call the polynomial of $\det(T - \lambda I)$ the **characteristic polynomial**. Note that T is a root of its own characteristic polynomial! For instance, $p(z) = z^2 - 1$ gives $p(T) = T^2 - I = 0$. This is the **Cayley-Hamilton Theorem**:

$$p(T) = 0 \text{ where } p(\lambda) = \det(M(T) - \lambda I).$$

Chapter 5

Inner Product Spaces

With vector spaces, we generalized the linear structure (of addition and scalar multiplication) in $\mathbb{R}^2$ and $\mathbb{R}^3$; however, we ignored other important features, such as length and angle. We now explore these ideas through **inner products**.

§5.1 Inner Products and Norms

§5.1.1 Inner Products

To motivate inner products, let's recall vectors in $\mathbb{R}^2$. The length of a vector $\vec{v} \in \mathbb{R}^2$ is called the **norm** of $\vec{v}$, denoted $\|\vec{v}\|$; and we calculate the norm of a vector $\vec{v} = (x_1, x_2)$ as follows:

$$\|\vec{v}\| = \sqrt{x_1^2 + x_2^2}.$$

Similarly, if $\vec{v} = (x_1, x_2, x_3) \in \mathbb{R}^3$, then

$$\|\vec{v}\| = \sqrt{x_1^2 + x_2^2 + x_3^2}.$$

Even though we can't visualize vectors in higher dimensions, the generalization to $\mathbb{R}^n$ is clear: for a vector $\vec{v} = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$, the norm of $\vec{v}$ is

$$\|\vec{v}\| = \sqrt{x_1^2 + x_2^2 + \ldots + x_n^2}.$$

However, clearly the norm is not linear on $\mathbb{R}^n$. To achieve this linearity, we introduce the **dot product**.

Definition 5.1.1: Dot Product

For $x, y \in \mathbb{R}^n$, the **dot product** of x and y, denoted $x \cdot y$, is

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

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

Importantly, the dot product of two vectors of $\mathbb{R}^n$ is a number, **not** a vector. Clearly,

$$x \cdot x = \|x\|^2$$

for all $x \in \mathbb{R}^n$. The dot product has the following properties:

- $x \cdot x \geq 0$ for all $x \in \mathbb{R}^n$.
- $x \cdot x = 0$ if and only if $x = 0$.
- For $y \in \mathbb{R}^n$ fixed, the map from $\mathbb{R}^n$ to $\mathbb{R}$ that sends $x \in \mathbb{R}^n$ to $x \cdot y$ is linear.

- $x \cdot y = y \cdot x$ for all $x, y \in \mathbb{R}^n$.

Now, how might we generalize the dot product from $\mathbb{R}^n$ to a general vector space? An intuitive notion may simply be to abstract the properties listed above. However, while this works for real vector spaces, we want a definition that works for complex vector spaces as well. We introduce the notion of **inner products**, but first, a review of complex numbers.

Definition 5.1.2: Complex Conjugates

Suppose $z = a + bi \in \mathbb{C}$. The **complex conjugate of** z, denote $\overline{z}$, is defined as

$$\overline{z} = a - bi \in \mathbb{C}.$$

Thus $z = \overline{z}$ if and only if $z \in \mathbb{R}$, i.e. $z = a + 0i$ for some $a \in \mathbb{R}$.

Note that complex conjugates respect addition and multiplication, in that

$$\overline{z + w} = \overline{z} + \overline{w} \text{ and } \overline{zw} = \overline{z}\,\overline{w}$$

for $z, w \in \mathbb{C}$. In other words, complex conjugation is an isomorphism between complex numbers.

Definition 5.1.3: Norm of Complex Number

The **absolute value**, or **norm**, of $z = a + bi$ is

$$|z| = \sqrt{a^2 + b^2} = \sqrt{(a + bi)(a - bi)} = \sqrt{z\overline{z}}.$$

Absolute values respect multiplication, i.e. $|zw| = |z|\,|w|$, but **not** addition! See the section on triangle inequalities for absolute values and addition.

Definition 5.1.4: 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:

- **Positive Definiteness**: $\langle v, v\rangle \geq 0$ for all $v \in V$, and $\langle v, v\rangle = 0$ if and only if $v = 0$.
- **Conjugate Symmetry**: $\langle u, v\rangle = \overline{\langle v, u\rangle}$ for all $u, v \in V$ (note that every real number equals its complex conjugate).
- **Linearity**: $\langle u + v, w\rangle = \langle u, w\rangle + \langle v, w\rangle$ for all $uv, w \in V$, and $\langle \lambda u, v\rangle = \lambda\,\langle u, v\rangle$ for $\lambda \in \mathbb{F}$.

Example 30. *Here are some examples of inner products:*

- *The* ***Euclidean inner product*** *on* $\mathbb{F}^n$*, defined by*

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

If $\mathbb{F}$ is real, this is simply the dot product. If $c_1, \ldots, c_n$ are positive numbers, then the Euclidean inner product can be extended to

$$\langle (w_1, \ldots, w_n), (z_1, \ldots, z_n) \rangle = c_1 w_1 \overline{z_1} + \ldots + c_n w_n \overline{z_n}.$$

- *An inner product can be defined on the vector space of continuous real-valued functions on the interval $[-1, 1]$ by*

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

With polynomials, a similar inner product may be defined

$$\langle p, q \rangle = \int_{0}^{\infty} p(x) q(x) e^{-x} dx.$$

Inner products exist over vector spaces; we give these vector spaces a name.

Definition 5.1.5: Inner Product Spaces

An **inner product space** is a vector space V along with an inner product on V.

The most prominent example of an inner product space is $\mathbb{F}^n$ with the Euclidean inner product, as defined above. When $\mathbb{F}^n$ is referred to as an inner product space, assume the inner product is the Euclidean inner product unless explicitly told otherwise.
For notational convenience, too, we now denote V an inner product space over $\mathbb{F}$.

Proposition 5.1.1: Properties of Inner Products

- For each fixed $u \in V$, the function that takes v to $\langle v, u \rangle$ is a linear map from V to $\mathbb{F}$.
- $\langle 0, u \rangle = 0$ for every $u \in V$.
- $\langle u, 0 \rangle = 0$ for every $u \in V$.
- $\langle u, v + w \rangle = \langle u, v \rangle + \langle u, w \rangle$ for all $u, v, w \in V$. For $\lambda \in \mathbb{F}$, $\langle u, \lambda v \rangle = \overline{\lambda} \langle u, v \rangle$. In other words, an inner product is **conjugate-linear** in the second term.

Proof. (1) follows from the conditions of linearity in the first term for inner products. (2) follows from (1), since every linear map takes 0 to 0. (3) follows from (1) and the conjugate symmetry property of linear maps.
For (4), suppose $u, v, w \in V$. Then

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

For (5), suppose $u, v \in V,\ \lambda \in \mathbb{F}$. Then

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

□

§5.1.2 Norms

Our motivation for defining inner products came initially from the norms of vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$, which represented the "length" of a vector. Now, we see that each inner product determines a norm.

Definition 5.1.6: Norm

For $v \in V$, the **norm** of v, denoted $\|v\|$, is defined by

$$\|v\| = \sqrt{\langle v, v\rangle}.$$

Example norms include the standard norm on $\mathbb{F}^n$, where

$$\|(z_1, \ldots, z_n)\| = \sqrt{|z_1|^2 + \ldots + |z_n|^2}.$$

In the vector space of continuous real-valued functions on $[-1, 1]$ (with inner product defined above), we have

$$\|f\| = \int_{-1}^{1} (f(x))^2 dx.$$

Now, let's look at some basic properties of norms.

Proposition 5.1.2: Properties of Norms

Suppose $v \in V$.

1. $\|v\| = 0$ if and only if $v = 0$.
2. $\|\lambda v\| = |\lambda|\,\|v\|$ for all $\lambda \in \mathbb{F}$.

Proof. (1) follows directly from inner products, since $\langle v, v\rangle = 0$ if and only if $v = 0$. Suppose $\lambda \in \mathbb{F}$. Then

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

Taking square roots now gives the desired equality. □

The above proof also illustrates that in general, working with norms squared is usually easier than working directly with norms.
Now, we come to a crucial definition.

Definition 5.1.7: Orthoganl

Two vectors $u, v \in V$ are **orthogonal** if $\langle u, v \rangle = 0$.

Order does not matter here, because $\langle u, v \rangle = 0$ if and only if $\langle v, u \rangle = 0$ (by properties of conjugates). We also sometimes say u is orthogonal to v. One should check that if u, v are non-zero vectors in $\mathbb{R}^2$, then

$$\langle u, v \rangle = \|u\|\|v\| \cos\theta,$$

where θ is the angle between u and v. This aligns with our intuition; two vectors in $\mathbb{R}^2$ are orthogonal if the angle between them is $90°$. In other words, *orthogonal* is a fancier, abstracted notion of *perpendicular*.
Let's look at some basic properties of orthogonality.

Proposition 5.1.3: Orthogonality and 0

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

Proof. Recall from earlier that $\langle 0, u \rangle = 0$ for every $u \in V$. Additionally, if $v \in V$, and $\langle v, v \rangle = 0$, by positive definiteness we need $v = 0$. □

Orthogonality can also be used to prove a familiar theorem:

Theorem 5.1.1: Pythagorean Theorem

Suppose $u, v \in V$ are orthogonal. Then

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

Proof.

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

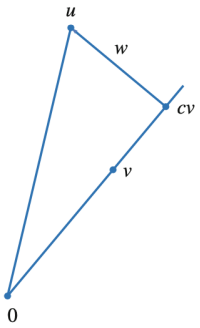
We get the second last step because orthogonality implies $\langle u, v \rangle = \langle v, u \rangle = 0$. □

Now, suppose $u, v \in V$ with $v \neq 0$. We would like to write u as a scalar multiple of v, plus a vector w orthogonal to v: In other words, we would like to *decompose* u into a vector v, plus some vector orthogonal to v.
To do this, let $c \in \mathbb{F}$ denote a scalar. Then

$$u = cv + (u - cv).$$

Thus we need to choose c so that v is orthogonal to $u - cv$. In other words, we need

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



From this, we want

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

So, we can decompose u into a scaled vector v, plus a vector $w = u - \frac{\langle u,v\rangle}{\|v\|^2}$ orthogonal to v:

$$u = \frac{\langle u, v\rangle}{\|v\|^2} v + \left(u - \frac{\langle u, v\rangle}{\|v\|^2} v\right).$$

In other words, we have proved the following result:

Proposition 5.1.4: Orthogonal Decomposition

Suppose $u, v \in V$ with $v \neq 0$. Set $c = \frac{\langle u,v\rangle}{\langle v,v\rangle} = \frac{\langle u,v\rangle}{\|v\|^2}$, and $w = u - cv$. Then

$$\langle w, v\rangle = 0 \text{ and } u = cv + w.$$

If we only look at cu, this is essentially the value of the vector v if **projected** onto the vector u; that is, if we look at the value of v only on u. This value has a special name:

Definition 5.1.8: Projection

Let $u, v \in V$ with $v \neq 0$. The **projection of v onto u**, denoted $\text{proj}_u(v)$, is defined by

$$\text{proj}_v(u) = \frac{\langle u, v\rangle}{\langle u, u\rangle} u.$$

We will see later that this has important implications for finding orthogonal bases.

Orthogonal decomposition is quite useful, especially in the proof of the following important equality:

Proposition 5.1.5: Cauchy-Schwarz Inequality

Suppose $u, v \in V$. Then

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

Equality holds if and only if one vector is a scalar multiple of another.

Proof. If $v = 0$, then both sides of the desired inequality equal 0, so suppose $v \neq 0$. Consider

$$u = \frac{\langle u, v\rangle}{\langle v, v\rangle} v + w$$

given above, where w is orthogonal to v. By the Pythagorean Theorem,

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

Hence

$$\|u\|^2 \|v\|^2 \geq |\langle u, v\rangle|^2,$$

and taking square roots then gives us the desired inequality.
From the above proof, Cauchy-Schwarz is an equality if and only if this statement is an equality:

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

Clearly, this is true only when $\langle w, w\rangle = 0$, or $w = 0$. But $w = 0$ if and only if u is a multiple of v (geometrically, the orthogonal vector between u and v is 0 only when one is a multiple of another). Thus equality holds if and only if one vector is a scalar multiple of another. □

The next result, called the **Triangle Inequality**, has the geometric interpretation that the length of each side of a triangle is less than the sum of the lengths of the other two sides. This also implies that the shortest path between two points is a line segment.

Proposition 5.1.6: Triangle Inequality

Suppose $u, v \in V$. Then

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

Again, equality holds if and only if one of u, v is a non-negative multiple of another.

Proof. We have

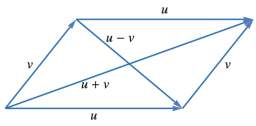
$$\begin{aligned} \|u + v\|^2 &= \langle u + v, u + v\rangle \\ &= \langle u, u\rangle + \langle u, v\rangle + \langle v, u\rangle + \langle v, v\rangle \\ &= \langle u, u\rangle + \langle v, v\rangle + \langle u, v\rangle + \overline{\langle u, v\rangle} \\ &\leq \|u\|^2 + \|v\|^2 + 2\,|\langle u, v\rangle|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

The second last inequality comes from the Cauchy-Schwarz Inequality, and taking square roots thus gives the desired inequality:

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

Equality holds if and only if $\langle u, v\rangle = \|u\|\|v\|$; but this only holds if one is a scalar multiple of another. □

This final result is called the parallelogram equality, because of its geometric interpretation: in every parallelogram, the sum of the squares of the lengths of the diagonals equals the sum of the squares of the lengths of the four sides.



The parallelogram equality.

Proposition 5.1.7: Parallelogram Equality

Suppose $u, v \in V$. Then

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

Proof.

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

as desired. □

§5.2 Orthonormal Bases

Definition 5.2.1: Orthonormal

A list of vectors is called **orthonormal** if each vector in the list has norm 1 and is orthogonal to all the other vectors in the list.
In other words, a list $e_1, \ldots, e_m$ of vectors in V is orthonormal if

$$\langle e_j, e_k\rangle = \begin{cases} 1 & \text{if } j = k, \\ 0 & \text{if } j \neq k \end{cases}$$

Example 31. *The standard basis in $\mathbb{F}^n$ is an orthonormal list. In addition,*

$$\left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right), \left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0\right), \left(\frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}, -\frac{2}{\sqrt{6}}\right)$$

is also an orthonormal list.

Why do we desire this result? It turns out that orthonormal lists are particularly easy to work with.

Proposition 5.2.1: Norm of Orthonormal Linear Combination

If $e_1, \ldots, e_m$ is an orthonormal list of vectors in V, then

$$\|a_1 e_1 + \ldots + a_m e_m\|^2 = |a_1|^2 + \ldots + |a_m|^2 .$$

Proof. Because each e_j has norm 1, this follows easily from repeated applications of the Pythagorean Theorem (which states that for orthogonal vectors $u, v \in V$, $\|u+v\|^2 = \|u\|^2 + \|v\|^2$; and the norm of a particular vector $\|a_i e_i\|^2 = |a_i|^2 \|e_i\|^2 = |a_i|^2$). □

This result has the following important corollary.

Corollary 5.2.1: 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 $a_1, \ldots, a_m \in \mathbb{F}$ are such that

$$a_1 e_1 + \ldots + a_m e_m = 0.$$

Then $|a_1|^2 + \ldots + |a_m|^2 = 0$; but this is true only when every $a_i = 0$. Thus $e_1, \ldots, e_m$ are linearly independent. □

Using this, we have the following definition.

Definition 5.2.2: Orthonormal Basis

An **orthonormal basis** of V is an orthonormal list of vectors in V that is also a basis of V.

For instance, the standard basis of $\mathbb{F}^n$ is also an orthonormal basis.
Using linear independence of orthonormal lists, we get the following result:

Proposition 5.2.2

Every orthonormal list of vectors in V with length $\dim V$ is an orthonormal basis of V.

In general, given a basis $e_1, \ldots, e_m \in V$ and a vector $v \in V$, we know that there is some choice of scalars $a_1, \ldots, a_m \in \mathbb{F}$ such that

$$v = a_1 e_1 + \ldots + a_n e_n.$$

Computing the scalars $a_1, \ldots, a_m$ may be difficult for an arbitrary basis of V; however, this is quite easy for an orthonormal basis: simply take

$$a_j = \langle v, e_j \rangle !$$

Proposition 5.2.3

Suppose $e_1, \ldots, e_n$ is an orthonormal basis of V and $v \in V$. Then

$$v = \langle v, e_1 \rangle e_1 + \ldots + \langle v, e_n \rangle e_n,$$

and

$$\|v\|^2 = |\langle v, e_1\rangle|^2 + \ldots + |\langle v, e_n\rangle|^2 .$$

Proof. Because $e_1, \ldots, e_m$ is a basis, there exist scalars $a_1, \ldots, a_n \in \mathbb{F}$ such that

$$v = a_1 e_1 + \ldots + a_n e_n.$$

Because $e_1, \ldots, e_n$ is orthonormal, taking the inner product with e_j of both sides gives us

$$\langle v, e_j\rangle = a_1 \langle e_1, e_j\rangle + \ldots + a_n \langle e_n, e_j\rangle .$$

But $\langle e_i, e_j\rangle = 0$, unless $i = j$, in which case it equals 1; thus

$$\langle v, e_j\rangle = a_j.$$

The second equation follows immediately from above, and the previous proposition. □

Now that we see the usefulness of orthonormal bases, how do we find them? For example, does $\mathcal{P}_m(\mathbb{R})$ have an orthonormal basis? The following result will provide a procedural method to generating orthonormal bases from linearly independent bases.

Theorem 5.2.1: Gram-Schmidt Procedure

Suppose $v_1, \ldots, v_m$ is a linearly independent list of vectors in V. Let $e_1 = \frac{v_1}{\|v_1\|}$. For $j = 2, \ldots, m$, define e_j inductively by

$$e_j = \frac{v_j - \langle v_j, e_1\rangle e_1 - \ldots - \langle v_j, e_{j-1}\rangle e_{j-1}}{\|v_j - \langle v_j, e_1\rangle e_1 - \ldots - \langle v_j, e_{j-1}\rangle e_{j-1}\|}.$$

Then $e_1, \ldots, e_m$ is an orthonormal list of vectors in V such that

$$\text{span}\,(v_1, \ldots, v_j) = \text{span}\,(e_1, \ldots, e_j)$$

for $j = 1, \ldots, m$.

Proof. We show by induction on j that the desired conclusion holds. For $j = 1$, $\text{span}\,(v_1) = \text{span}\,(e_1)$, since v_1 is a positive multiple of e_1.
Suppose $1 < j < m$ and

$$\text{span}\,(v_1, \ldots, v_{j-1}) = \text{span}\,(e_1, \ldots, e_{j-1})$$

holds. Note that $v_j \notin \text{span}\,(v_1, \ldots, v_{j-1})$, since $v_1, \ldots, v_m$ is linearly independent; thus $v_j \notin \text{span}\,(e_1, \ldots, e_{j-1})$ either. Hence in the definition above, we are not dividing by zero, since v_j not in span means it is impossible for $v_j - \sum_{i=1}^{j-1} a_i e_i$ to be zero either. Dividing a vector by its norm produces a new vector with norm 1; thus $\|e_j\| = 1$.
Let $1 \leq k < j$. Then

$$\begin{aligned}\langle e_j, e_k\rangle &= \left\langle \frac{v_j - \langle v_j, e_1\rangle e_1 - \ldots - \langle v_j, e_{j-1}\rangle e_{j-1}}{\|v_j - \langle v_j, e_1\rangle e_1 - \ldots - \langle v_j, e_{j-1}\rangle e_{j-1}\|}, e_k \right\rangle \\ &= \frac{\langle v_j, e_k\rangle - \langle v_j, e_k\rangle}{\|v_j - \langle v_j, e_1\rangle e_1 - \ldots - \langle v_j, e_{j-1}\rangle e_{j-1}\|} \\ &= 0,\end{aligned}$$

where the second last step follows because $\langle e_j, e_k\rangle = 0$ for all $j \neq k$. Thus e_j is orthonormal to all previous vectors, and so $e_1, \ldots, e_j$ is an orthonormal list.

From the above definition of e_j, we see that $v_j \in \text{span}\,(e_1, \ldots, e_j)$. Since we've assumed that $\text{span}\,(e_1, \ldots, e_{j-1}) = \text{span}\,(v_1, \ldots, v_{j-1})$, we thus get

$$\text{span}\,(v_1, \ldots, v_j) \subset \text{span}\,(e_1, \ldots, e_j).$$

Both lists above are linearly independent (v's given, e's by orthonormality). Therefore both of these subspaces have dimension j, and hence they are equal. □

This definition of e_j seems extremely complicated, at first glance. However, closer inspection reveals somewhat of an intuitive process:

- First, we wish to convert a list of linearly independent vectors into a list of orthonormal vectors. Let's first forget about orthonormality, and look solely at orthogonality.
- We can start with any one vector; this will be the "base" vector, $v_1 = u_1$. In order to find the next *orthogonal* vector, we must find a vector orthogonal to u_1:
  - Recall that a vector $v_2 \in V$ can be decomposed into u_1, plus a vector orthogonal to it. To do this, we convert v_2 into $\frac{\langle u_1, v_2\rangle}{\langle u_1, u_1\rangle} u_1 + w$, where w is the vector orthogonal to u_1. **This is the vector we're interested in!**

  Thus, using orthogonal decomposition, we simply want the vector orthogonal to the base vector u_1, and so we get

$$u_2 = w = v_2 - \frac{\langle u_1, v_2\rangle}{\langle u_1, u_1\rangle} u_1 = v_2 - \text{proj}_{u_1}(v_2).$$

- Now, we wish to repeat this process with the rest of the vectors v_i. Like before, we wish to find a vector orthogonal to every other vector. We can repeat the processes above iteratively, by finding a vector (say v_i') orthogonal to one vector (say u_1), then using that to find another vector (say v_i'') orthogonal to both u_1 and another vector (say u_2), and so on until we reach a vector u_i orthogonal to all previous $u_{1 \le j < i}$. This gives us the desired (if complex) formula,

$$\begin{aligned} u_j &= v_j - \text{proj}_{u_1}(v_j) - \text{proj}_{u_2}(v_j) - \ldots - \text{proj}_{u_{j-1}}(v_j) \\ &= v_j - \frac{\langle u_1, v_j\rangle}{\langle u_1, u_1\rangle} u_1 - \ldots - \frac{\langle u_{j-1}, v_j\rangle}{\langle u_{j-1}, u_{j-1}\rangle} u_{j-1}. \end{aligned}$$

- This process gives us a list of *orthogonal* vectors; but if we desire an orthonormal list, we need the extra step of dividing by the norm of the result. Luckily, if each u_j is orthogonal, then every $\langle u_j, u_j\rangle = 1$, so we can focus only on the numerator. That's how we arrive at the final equation,

$$e_j = \frac{v_j - \langle v_j, e_1\rangle\, e_1 - \ldots - \langle v_j, e_{j-1}\rangle\, e_{j-1}}{\|v_j - \langle v_j, e_1\rangle\, e_1 - \ldots - \langle v_j, e_{j-1}\rangle\, e_{j-1}\|}.$$

Let's look at an example of the Gram-Schmidt process in action.

Example 32. *Find the orthonormal basis of $\mathcal{P}_2(\mathbb{R})$, where the inner product is given by*

$$\langle p, q\rangle = \int_{-1}^{1} p(x)q(x)dx.$$

Solution: We apply the Gram-Schmidt procedure to the basis $1, x, x^2$.
We have

$$e_1 = \frac{1}{\|1\|} = \frac{1}{\sqrt{\int_{-1}^{1} 1^2 dx}} = \frac{1}{\sqrt{2}}.$$

Now, the numerator for e_2 is

$$x - \langle x, e_1 \rangle e_1 = x - \left(\int_{-1}^{1} x \frac{1}{\sqrt{2}} dx \right) \frac{1}{\sqrt{2}} = x;$$

moreover,

$$\|x\|^2 = \int_{-1}^{1} x^2 dx = \frac{2}{3}.$$

Thus $\|x\| = \sqrt{\frac{2}{3}}$, and so $e_2 = \frac{x}{\sqrt{\frac{2}{3}}} = x\sqrt{\frac{3}{2}}$.
Finally, the numerator for e_3 is

$$\begin{aligned} & x^2 - \langle x^2, e_1 \rangle e_1 - \langle x^2, e_2 \rangle e_2 \\ &= x^2 - \left(\int_{-1}^{1} x^2 \sqrt{\frac{1}{2}} dx \right) - \left(\int_{-1}^{1} x^2 \sqrt{\frac{3}{2}} dx \right) x\sqrt{\frac{3}{2}} \\ &= x^2 - \frac{1}{3}. \end{aligned}$$

Moreover,

$$\|x^2 - \frac{1}{3}\|^2 = \int_{-1}^{1} (x^4 - \frac{2}{3}x^2 + \frac{1}{9}) dx = \frac{8}{45}.$$

Thus $\|x^2 - \frac{1}{3}\| = \sqrt{\frac{8}{45}}$, and so $e_3 = \sqrt{\frac{45}{8}}(x^2 - \frac{1}{3})$.
Thus

$$\frac{1}{\sqrt{2}},\ x\sqrt{\frac{3}{2}}, \sqrt{\frac{45}{8}}(x^2 - \frac{1}{3})$$

is an orthonormal basis of $\mathcal{P}_2(\mathbb{R})$.
But do these orthonormal bases always exist? It turns out that yes, they do.

Proposition 5.2.4: Existence of Orthonormal Basis

Every finite-dimensional inner product space has an orthonormal basis.

Proof. Suppose V is finite-dimensional. Choose a basis of V. Apply the Gram-Schmidt procedure to it, producing an orthonormal list with length $\dim V$. From before, we see that any orthonormal list of the right length ($\dim V$) is an orthonormal basis of V. □

Not only does an orthonormal basis always exist, we can also extend an orthonormal list of vectors to an orthonormal basis, using the Gram-Schmidt procedure.

Corollary 5.2.2: Orthonormal List Extends to Orthonormal Basis

Suppose V is finite-dimensional. Then every orthonormal list of vectors in V can be extended to an orthonormal basis of V.

Proof. Suppose $e_1, \ldots, e_m$ is an orthonormal list of vectors in V. Then $e_1, \ldots, e_m$ is linearly independent, and thus this list can be extended to a basis

$$e_1, \ldots, e_m, v_1, \ldots, v_n$$

of V. Now, using the Gram-Schmidt procedure we can produce an orthonormal list

$$e_1, \ldots, e_m, f_1, \ldots, f_n;$$

the first m vectors remain unchanged since they're already orthonormal. Then the list is an orthonormal basis. □

Now, let's explore some connections between orthonormal lists and upper-triangular matrices. We've shown before that every operator $T \in \mathcal{L}(V)$ has a basis that produces an upper-triangular matrix for $\mathcal{M}(T)$. Now, with inner product spaces, we're interested in whether there's an **orthonormal** basis that produces an upper-triangular matrix. With Gram-Schmidt, our wishes are fulfilled; as long as a basis admits an upper-triangular matrix, then we have what we need (this is always true for complex vector spaces, but only sometimes true for real vector spaces).

Proposition 5.2.5: Upper-Triangular w.r.t. Orthonormal Basis

Suppose $T \in \mathcal{L}(V)$. If T has an upper-triangular matrix with respect to some basis of V, then T has an upper-triangular matrix with respect to some orthonormal basis of V.

Proof. Suppose T has an upper-triangular matrix with respect to some basis $v_1, \ldots, v_n \in V$. Thus $\operatorname{span}(v_1, \ldots, v_j)$ is invariant under T for each $1 \leq j \leq n$.
Applying Gram-Schmidt to $v_1, \ldots, v_n$, we get an orthonormal basis $e_1, \ldots, e_n \in V$. Since

$$\operatorname{span}(e_1, \ldots, e_n) = \operatorname{span}(v_1, \ldots, v_n),$$

we conclude that $\operatorname{span}(e_1, \ldots, e_n)$ is invariant under T. Thus T has an upper-triangular matrix with respect to the orthonormal basis $e_1, \ldots, e_n$. □

This next result is an important application of the above result:

Theorem 5.2.2: Schur's Theorem

Suppose V is a finite-dimensional complex vector space and $T \in \mathcal{L}(V)$. Then T has an upper-triangular matrix with respect to some orthonormal basis of V.

Proof. Recall that T has an upper-triangular matrix with respect to some basis of V (since every operator in a complex vector space admits an upper-triangular matrix). Then, from above, T has an upper-triangular orthonormal basis. □

§5.2.1 Linear Functionals on Inner Product Spaces

Since linear maps into the scalar field $\mathbb{F}$ play an important role, we give them a special name.

Definition 5.2.3: 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 $\mathcal{L}(V, \mathbb{F})$.

Example 33. *The function $\varphi : \mathbb{F}^3 \to \mathbb{F}$ defined by*

$$\varphi(z_1, z_2, z_3) = 2z_1 - 5z_2 + z_3$$

is a linear functional on $\mathbb{F}^3$. We could write this linear functional in the form

$$\varphi(z) = \langle z, u \rangle$$

for every $z \in \mathbb{F}^3$, where $u = (2, -5, 1)$ (recall that unless explicitly stated otherwise, the default inner product for $\mathbb{F}^n$ is the Euclidean inner product; and every vector space is naturally an inner product space).

Example 34. *The function $\varphi : \mathcal{P}_2(\mathbb{R}) \to \mathbb{R}$ defined by*

$$\varphi(p) = \int_{-1}^{1} p(t) \left(\cos \pi(t)\right) dt$$

is a linear functional on $\mathcal{P}_2(\mathbb{R})$ (here the inner product is multiplication followed by integration on $[-1, 1]$). It's not obvious that there exists a $u \in \mathcal{P}_2(\mathbb{R})$ that satisfies

$$\varphi(p) = \langle p, u \rangle$$

for every $p \in \mathcal{P}_2(\mathbb{R})$, since we cannot take $u(t) = \cos(\pi t) \notin \mathcal{P}_2(\mathbb{R})$.

If $u \in V$, then the map that sends v to $\langle v, u \rangle$ is a linear functional on V. The next result shows that **every linear functional on V is of this form**! This result is indeed powerful; the above example demonstrates that there's not always an obvious candidate for u.

Theorem 5.2.3: Riesz Representation Theorem

Suppose V is finite-dimensional and φ is a linear functional on V. Then there is a unique vector $u \in V$ such that

$$\varphi(v) = \langle v, u \rangle$$

for every $v \in V$ (here, the inner product is the underlying inner product of the inner product space V).

Proof. First, we show existence. Let $e_1, \ldots, e_n$ be an orthonormal basis for V. Then

$$\begin{aligned}\varphi(v) &= \varphi(\langle v, e_1\rangle e_1 + \ldots + \langle v, e_n\rangle e_n) \\ &= \langle v, e_1\rangle \varphi(e_1) + \ldots + \langle v, e_n\rangle \varphi(e_n) \\ &= \left\langle v, \overline{\varphi(e_1)}e_1 \right\rangle + \ldots + \left\langle v, \overline{\varphi(e_n)}e_n \right\rangle \\ &= \left\langle v, \overline{\varphi(e_1)}e_1 + \ldots + \overline{\varphi(e_n)}e_n \right\rangle\end{aligned}$$

for every $v \in V$ (recall that any v can be represented as $\langle v, e_1\rangle e_1 + \ldots + \langle v, e_n\rangle e_n$ if $e_1, \ldots, e_n$ is an orthonormal basis; and inner products are conjugate-linear with respect to their second terms). Thus setting

$$u = \overline{\varphi(e_1)}e_1 + \ldots + \overline{\varphi(e_n)}e_n,$$

we have $\varphi(v) = \langle v, u\rangle$, as desired.
Now, we show uniqueness. Suppose $u_1, u_2 \in V$ such that

$$\varphi(v) = \langle v, u_1\rangle = \langle v, u_2\rangle$$

for every $v \in V$. Then

$$0 = \langle v, u_1\rangle - \langle v, u_2\rangle = \langle v, u_1 - u_2\rangle$$

for every $v \in V$. Taking $v = u_1 - u_2$, properties of inner product spaces tells us that $v = u_1 - u_2 = 0$; thus $u1 = u_2$, and we get uniqueness. □

Not only does this prove the existence and uniqueness of $u \in V$ that satisfies

$$\varphi(v) = \langle v, u\rangle$$

for all $v \in V$, this also provides a formula for finding u:

$$u = \overline{\varphi(e_1)}e_1 + \ldots + \overline{\varphi(e_n)}e_n.$$

This seems to imply that u is dependent on the choice of orthonormal basis $e_1, \ldots, e_n$, as well as φ; but the Reisz Representation Theorem tells us that u is uniquely determined by φ; thus u is invariant of chosen orthonormal basis $e_1, \ldots, e_n$ of V.

Example 35. *Find $u \in \mathcal{P}_2(\mathbb{R})$ such that*

$$\int_{-1}^{1} p(t)(\cos(\pi t))dt = \int_{-1}^{1} p(t)u(t)dt$$

for all $p \in \mathcal{P}_2(\mathbb{R})$.

Solution: Let $\varphi(p) = \int_{-1}^{1} p(t)(\cos \pi t)dt$. Using the formula above and the orthonormal basis from the previous example, we have

$$\begin{aligned}u(x) = &\left(\int_{-1}^{1} \sqrt{\frac{1}{2}}(\cos(\pi t))dt\right)\sqrt{\frac{1}{2}} + \left(\int_{-1}^{1} \sqrt{\frac{3}{2}}t(\cos(\pi t))dt\right)\sqrt{3/2}x \\ &+ \left(\int_{-1}^{1} \sqrt{\frac{45}{8}}(t^2 - \frac{1}{3})(\cos(\pi t))dt\right)\sqrt{\frac{45}{8}}(x^2 - \frac{1}{3}).\end{aligned}$$

A little bit of calculus (:tf:) shows that

$$u(x) = -\frac{45}{2\pi^2}(x^2 - \frac{1}{3}).$$

§5.3 Orthogonal Complements and Minimization Problems

§5.3.1 Orthogonal Complements

Definition 5.3.1: Orthogonal Complement

If U is a subset of V, then the **orthogonal complement** of U, denoted $U^\perp$, is the set of all vectors in V that are orthogonal to every vector in U:

$$U^\perp = \{v \in V \mid \langle v, u\rangle = 0 \text{ for every } u \in U\}.$$

For example, if U is a line in $\mathbb{R}^3$, then $U^\perp$ is the plane containing the origin that is perpendicular to U. If U is instead a plane in $\mathbb{R}^3$, then $U^\perp$ is the line containing the origin that is perpendicular to U.

Proposition 5.3.1: Properties of Orthogonal Complements

1. If U is a subset of V, then $U^\perp$ is a subspace of V.
2. $\{0\}^\perp = V$.
3. $V^\perp = \{0\}$.
4. If U is a subset of V, then $U \cap U^\perp \subseteq \{0\}$.
5. If U and W are subsets of V and $U \subseteq W$, then $W^\perp \subseteq U^\perp$.

Proof.

1. Suppose U is a subset of V. Then $\langle 0, u\rangle = 0$ for every $u \in U$; thus $0 \in U^\perp$. Suppose $v, w \in U^\perp$. If $u \in U$, then
$$\langle v + w, u\rangle = \langle v, u\rangle + \langle w, u\rangle = 0 + 0 = 0.$$
Thus $v + w \in U^\perp$. Finally, suppose $\lambda \in \mathbb{F},\ v \in U^\perp$. If $u \in U$, then
$$\langle \lambda v, u\rangle = \lambda \langle v, u\rangle = \lambda \cdot 0 = 0.$$
Thus $\lambda v \in U^\perp$, and so $U^\perp$ is a subspace of V.
2. Clearly, any $v \in V$ satisfies $\langle v, 0\rangle = 0$. Thus every $v \in V^\perp$, so $V^\perp = V$.
3. Suppose $v \in V^\perp$. Then $\langle v, v\rangle = 0$, which forces $v = 0$. Thus $V^\perp = \{0\}$.
4. Suppose U is a subset of V and $v \in U \cap U^\perp$. Then $\langle v, v\rangle = 0$, which forces $v = 0$. Thus $U \cap U^\perp \subseteq \{0\}$.
5. Suppose U, W are subsets of V and $U \subseteq W$. Suppose $v \in W^\perp$. Since $u \in U \subseteq W$, $\langle v, u\rangle = 0$ for every $u \in U$. Thus $W^\perp \subseteq U^\perp$.

□

With direct sums, we can start to see why it is referred to as the orthogonal complement. Suppose U and W are subspaces of V such that $U \oplus W = V$ (in other words, every vector $v \in V$ can be written in exactly one way as a vector in U plus a vector in V). It turns out that every vector space can be decomposed into a subspace and its orthogonal complement.

Theorem 5.3.1: Direct Sum of Subspace and Orthogonal Complement

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

$$V = U \oplus U^{\perp}.$$

Proof. First, we will show that $V = U + U^{\perp}$. Suppose $v \in V$, and suppose $e_1, \ldots, e_m \in U$ is an orthonormal basis of U. Clearly,

$$v = \underbrace{\langle v, e_1 \rangle e_1 + \ldots + \langle v, e_m \rangle e_m}_{u} + \underbrace{v - \langle v, e_1 \rangle e_1 - \ldots - \langle v, e_m \rangle e_m}_{w}.$$

Let u and w be vectors as defined above. Clearly, $u \in U$. Since $e_1, \ldots, e_m$ is an orthonormal list, for each $j = 1, \ldots, m$ we have

$$\langle w, e_j \rangle = \langle v, e_j \rangle - \langle v, e_j \rangle = 0.$$

This is because taking the inner product of w with respect to e_j gives us $\langle v, e_j \rangle - \langle v, e_j \rangle \langle e_j, e_j \rangle = \langle v, e_j \rangle - \langle v, e_j \rangle$, since $\langle e_j, e_j \rangle = 1$ and $\langle e_j, e_k \rangle = 0$ for every $j \neq k$.
Thus w is orthogonal to every vector in $\operatorname{span}(e_1, \ldots, e_m)$. In other words, $w \in U^{\perp}$.
Thus we have written any vector $v \in V$ as $v = u + w$, where $u \in U$ and $w \in U^{\perp}$.
From the properties shown above, we know $U \cap U^{\perp} = \{0\}$. By Proposition 1.45 and above, we see that $V = U \oplus U^{\perp}$. □

Now, computing $\dim U^{\perp}$ is trivial; simply take

$$\dim U^{\perp} = \dim V - \dim U.$$

Another important consequence of the theorem is the following:

Proposition 5.3.2

Suppose U is a vinite-dimensional subspace of V. Then

$$U = \left(U^{\perp}\right)^{\perp}.$$

Proof. First, we show that

$$U \subseteq \left(U^{\perp}\right)^{\perp}.$$

Suppose $u \in U$. Then $\langle u, v \rangle = 0$ for every $v \in U^{\perp}$ (by definition). Since u is thus orthogonal to every vector in $U^{\perp}$, we have $u \in \left(U^{\perp}\right)^{\perp}$.
To show $\left(U^{\perp}\right)^{\perp} \subseteq U$, suppose $v \in \left(U^{\perp}\right)^{\perp}$. We can write $v = u + w$ (since the complement of a subset, here $U^{\perp}$, is a subspace of V), where $u \in U$ and $w \in U^{\perp}$. Then

$$w = v - u \in U^{\perp}.$$

Since $v \in \left(U^{\perp}\right)^{b}ot$ and $u \in \left(U^{\perp}\right)^{\perp}$ (from above, since $U \subseteq \left(U^{\perp}\right)^{\perp}$), we have $v - u \in \left(U^{\perp}\right)^{\perp}$. But this means $v - u \in U^{\perp} \cap \left(U^{\perp}\right)^{\perp}$, which implies that $v - u$ is orthogonal to itself; in other words, $\langle v - u, v - u \rangle = 0$, so $v - u = 0$, and so $v = u \in U$, so $v \in U$ as well. Thus every vector in $\left(U^{\perp}\right)^{\perp}$ is in U as well, so $\left(U^{\perp}\right)^{\perp} \subseteq U$, completing the proof. □

We now define an operator $\mathcal{P}_U$ for every finite-dimensional subspace U of V.

Definition 5.3.2: Orthogonal Projection

Suppose U is a finite-dimensional subspace of V. The **orthogonal projection of V onto U** is the operator $\mathcal{P}_U \in \mathcal{L}(V)$, defined as follows:

For $v \in V$, write $v = u + w$ where $u \in U$ and $w \in U^\perp$. Then $\mathcal{P}_U(v) = u$.

Since given a subpsace, every vector space uniquely decomposes into that subspace and its orthogonal complement, $\mathcal{P}_U$ is well-defined.

Example 36. *Suppose $x \in V$ with $x \neq 0$ and $U = \operatorname{span}(x)$. Show that*

$$\mathcal{P}_U(v) = \frac{\langle v, x \rangle}{\|x\|^2} x$$

for every $v \in V$.

Solution: Suppose $v \in V$. Then

$$v = \frac{\langle v, v \rangle}{\|x\|^2} x + \left(v - \frac{\langle v, x \rangle}{\|x\|^2} x \right),$$

where the first term on the right is in $\operatorname{span}(x)$ (and thus in U) and the second term on the right is orthogonal to x (see the proof of the direct sum theorem above for why), and thus in $U^\perp$. Thus $\mathcal{P}_U(v)$ equals the first term on the right, as desired.

Proposition 5.3.3: Properties of Orthogonal Projections

Suppose U is a finite-dimensional subspace of V and $v \in V$. Then

1. $\mathcal{P}_U \subseteq \mathcal{L}(V)$;
2. $\mathcal{P}_U u = u$ for every $u \in U$;
3. $\mathcal{P}_U w = 0$ for every $w \in U^\perp$;
4. $\operatorname{range} \mathcal{P}_U = U$;
5. $\operatorname{null} \mathcal{P}_U = U^\perp$;
6. $v - \mathcal{P}_U v \in U^\perp$;
7. $\mathcal{P}_U^2 = \mathcal{P}_U$;
8. $\|\mathcal{P}_U v\| \leq \|v\|$;
9. For every orthonormal basis $e_1, \ldots, e_m \in U$,

$$\mathcal{P}_U v = \langle v, e_1 \rangle e_1 + \ldots + \langle v, e_m \rangle e_m.$$

Proof. To show that $\mathcal{P}_U$ is a linear map on V, suppose $v_1, v_2 \in V$. Write them as

$$v_1 = u_1 + w_1, \; v_2 = u_2 + w_2,$$

where $u_1, u_2 \in U$ and $w_1, w_2 \in U^\perp$. Thus $\mathcal{P}_U v_1 = u_1$ and $\mathcal{P}_U v_2 = u_2$. Since $v_1 + v_2 = (u_1 + u_2) + (w_1 + w_2)$, clearly

$$\mathcal{P}_U(v_1 + v_2) = u_1 + u_2 = \mathcal{P}_U v_1 + \mathcal{P}_U v_2.$$

Similarly with $\lambda \in \mathbb{F}$, $\lambda v_1 = \lambda u_1 + \lambda w_1$, so

$$\mathcal{P}_U(\lambda v_1) = \lambda u_1 = \lambda \mathcal{P}_U v_1.$$

Thus $\mathcal{P}_U \in \mathcal{L}(V)$ is a valid linear map.

All but the last two follow trivially.

For the second last one, let $v = u + w$ with $u \in U,\ w \in U^\perp$. Then

$$\|\mathcal{P}_U v\|^2 = \|u\|^2 \leq \|u\|^2 + \|w\|^2 = \|v\|^2,$$

where the last equality comes from the Pythagorean Theorem.

The last property follows from the proof of the direct sum decomposition.

□

§5.3.2 Minimization Problems

This problem often arises: given a subspace U of V and a point $v \in V$, find a point $u \in U$ such that $\|v - u\|$ is as small as possible. It turns out that this minimization problem is actually solved by taking $u = \mathcal{P}_U v$.

Proposition 5.3.4: Minimizing Distance to Subspace

Suppose U is a finite-dimensional subspace of V, $v \in V$, and $u \in U$. Then

$$\|v - \mathcal{P}_U v\| \leq \|v - u\|.$$

In other words, the distance between v and $\mathcal{P}_U v$ is smaller than the distance between v and any vector u in U. Equality only holds if $u = \mathcal{P}_U v$.

Proof. We have

$$\begin{aligned} \|v - \mathcal{P}_U v\|^2 &\leq \|v - \mathcal{P}_U v\|^2 + \|\mathcal{P}_U v - u\|^2 \\ &= \|(v - \mathcal{P}_U v) + (\mathcal{P}_U v - u)\|^2 \\ &= \|v - u\|^2. \end{aligned}$$

The first inequality holds because $\|\mathcal{P}_U v - u\|^2 \geq 0$, and the secone equality holds due to the Pythagorean Theorem (since $v - \mathcal{P}_U v \in U^\perp$, and $\mathcal{P}_U v - u \in U$). Taking square roots gives the desired inequality.
Inequality becomes equality if and only if the above inequality is an equality, which happens if and only if $\mathcal{P}_U v - u = 0$, or $\mathcal{P}_U v = u$. □