# Vector Spaces

Paul L. Bailey

Department of Mathematics, Southern Arkansas University
*E-mail address*: plbailey@saumag.edu
*Date*: April 4, 2007

# Contents

# Preface

*Linear algebra* is the study of vector spaces and linear transformations. This document constitutes a brief course in linear algebra over the real numbers, with an initial emphasis on vectors in cartesian space. It is traditional to motivate the study of linear algebra by examining methods of solution to systems of linear equation; in this approach, matrices become arrays of coefficients of linear equations, and the definition of multiplication of matrices first appears as an arbitrary (and therefore difficult) concept which magically produces results.

We attempt instead to initially focus on the geometry of synthetic space and the concept of vectors as equivalence classes of arrows. These vectors can be added and stretched geometrically, which imposes an algebraic system on sets of vectors. Introduction of a coordinate system then transfers the geometric study of vectors in synthetic space to the algebraic study of points in cartesian space, producing the motivating examples of *vector spaces*.

The natural functions to examine between such spaces are those which preserve the geometrically inspired algebraic relations, which are exactly those which fix the origin and send lines to lines; such functions are the known as *linear transformations*, and can be described by matrices. In this context, multiplication of matrices corresponds to composition of linear transformations.

We understand that the precision of the language of set theory cannot productively be avoided in the study of mathematical objects; consequently, we start by reviewing this language, and use it throughout the text.

CHAPTER 1

# Sets and Functions

ABSTRACT. It is difficult to grasp advanced mathematics without fluent control over the concepts of set and function. This chapter rapidly lists some of what you should know. It is hoped that much of this is review; feel free to bring any questions you may have to the attention of the class.

## 1. Sets

A *set* is a collection of objects. The objects in a set are called *elements* of that set. Sometimes elements are referred to as *members* or *points*. If an element is in a set, we say that the element is *contained* in the set.

If two symbols $a$ and $b$ represent the same element, we write $a = b$. If the symbols $a$ and $b$ represent different elements, we write $a \neq b$. If an element $a$ is contained in a set $A$, this relation is written $a \in A$. If $a$ is not in $A$, this fact is denoted $a \notin A$. We assume that the statements $a \in A$ and $a = b$ are always either true or false, although we may not know which.

Two sets are considered equal when they contain the same elements:

$$A = B \Leftrightarrow [x \in A \Leftrightarrow x \in B].$$

The sets we will primarily be using are the standard sets of numbers, and those derived from them. These sets have standard names:

| | |
|---|---|
| Natural Numbers: | $\mathbb{N} = \{0, 1, 2, 3, \dots\}$ |
| Integers: | $\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$ |
| Rational Numbers: | $\mathbb{Q} = \{\frac{p}{q} \mid p, q \in \mathbb{Z}, q \neq 0\}$ |
| Real Numbers: | $\mathbb{R} = \{$ Infinite decimal expansions $\}$ |
| Complex Numbers: | $\mathbb{C} = \{a + ib \mid a, b \in \mathbb{R} \text{ and } i^2 = -1\}$ |

*Remark* 1.1. Some authors do not include zero in the set of natural numbers.

## 2. Subsets

Let $A$ and $B$ be sets. We say that $B$ is a *subset* of $A$ and write $B \subset A$ if $x \in B \Rightarrow x \in A$.

For our purposes, we consider $\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}$.

It is clear that $A = B$ if and only if $A \subset B$ and $B \subset A$.

A set with no elements is called an *empty set*. Since two sets are equal if and only if they contain the same elements, there is only one empty set, and it is denoted $\varnothing$. The empty set is a subset of any other set.

If $X$ is any set and $p(x)$ is a proposition whose truth or falsehood depends on each element $x \in X$, we may construct a new set consisting of all of the elements of $X$ for which the proposition is true; this set is denoted:

$$\{x \in X \mid p(x)\}.$$

An *interval* is a type of subset of the real numbers; it is the set of all real numbers between two points, called *endpoints*; we consider $\pm\infty$ to be valid endpoints. The distance between these endpoints is the *length* of the interval. This distance may be finite or infinite. Those intervals whose endpoints are contained in the set are called *closed*; those whose endpoints are not contained in the set are called *open*. Notation for intervals is standard:

$$\begin{aligned}
[a, b] &= \{x \in \mathbb{R} \mid a \le x \le b\} && \text{(finite closed)}\\
(a, b) &= \{x \in \mathbb{R} \mid a < x < b\} && \text{(finite open)}\\
[a, b) &= \{x \in \mathbb{R} \mid a \le x < b\} &&\\
(a, b] &= \{x \in \mathbb{R} \mid a < x \le b\} &&\\
(-\infty, b] &= \{x \in \mathbb{R} \mid x \le b\} && \text{(infinite closed)}\\
(-\infty, b) &= \{x \in \mathbb{R} \mid x < b\} && \text{(infinite open)}\\
[a, \infty) &= \{x \in \mathbb{R} \mid a \le x\} && \text{(infinite closed)}\\
(a, \infty) &= \{x \in \mathbb{R} \mid a < x\} && \text{(infinite open)}
\end{aligned}$$

## 3. Set Operations

Let $X$ be a set and let $A, B \subset X$.

The *intersection* of $A$ and $B$ is denoted by $A \cap B$ and is defined to be the set containing all of the elements of $X$ that are in both $A$ and $B$:

$$A \cap B = \{x \in X \mid x \in A \text{ and } x \in B\}.$$

The *union* of $A$ and $B$ is denoted by $A \cup B$ and is defined to be the set containing all of the elements of $X$ that are in either $A$ or $B$:

$$A \cup B = \{x \in X \mid x \in A \text{ or } x \in B\}.$$

We note here that there is no concept of "multiplicity" of an element in a set; that is, if $x$ is in both $A$ and $B$, then $x$ occurs only once in $A \cup B$.

The *complement* of $A$ with respect to $B$ is denoted $A \smallsetminus B$ and is defined to be the set containing all of the elements of $A$ which are not in $B$:

$$A \smallsetminus B = \{x \in X \mid x \in A \text{ and } x \notin B\}.$$

**Example 1.1.** Let $A = \{1, 3, 5, 7, 9\}$, $B = \{1, 2, 3, 4, 5\}$. Then $A \cap B = \{1, 3, 5\}$, $A \cup B = \{1, 2, 3, 4, 5, 7, 9\}$, $A \smallsetminus B = \{7, 9\}$, and $B \smallsetminus A = \{2, 4\}$. □

**Example 1.2.** Let $C = [1, 5] \cup (10, 16)$ and let $\mathbb{N}$ be the set of counting numbers. How many elements are in $C \cap \mathbb{N}$?

*Solution.* The set $C \cap \mathbb{N}$ is the set of natural numbers between 1 and 5 inclusive and between 10 and 16 exclusive. Thus $C \cap \mathbb{N} = \{1, 2, 3, 4, 5, 11, 12, 13, 14, 15\}$. Therefore $C \cap \mathbb{N}$ has 10 elements. □

A picture corresponds to each of these set operations; these pictures are called *Venn diagrams*. Use Venn diagrams to convince yourself of the following properties.

**Proposition 1.3.** *Let $X$ be a set and let $A, B, C \subset X$. Then*

- **(a)** $A = A \cup A = A \cap A$;
- **(b)** $\varnothing \cap A = \varnothing$;
- **(c)** $\varnothing \cup A = A$;
- **(d)** $A \subset B \Leftrightarrow A \cap B = A$;
- **(e)** $A \subset B \Leftrightarrow A \cup B = B$;
- **(f)** $A \cap B = B \cap A$;
- **(g)** $A \cup B = B \cup A$;
- **(h)** $(A \cap B) \cap C = A \cap (B \cap C)$;
- **(i)** $(A \cup B) \cup C = A \cup (B \cup C)$;
- **(j)** $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$;
- **(k)** $(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$;
- **(l)** $A \smallsetminus (B \cup C) = (A \smallsetminus B) \cap (A \smallsetminus C)$;
- **(m)** $A \smallsetminus (B \cap C) = (A \smallsetminus B) \cup (A \smallsetminus C)$;
- **(n)** $A \subset B \Rightarrow A \cup (B \smallsetminus A) = B$;
- **(o)** $A \subset B \Rightarrow A \cap (B \smallsetminus A) = \varnothing$;
- **(p)** $A \smallsetminus (B \smallsetminus C) = (A \smallsetminus B) \cup (A \cap B \cap C)$;
- **(q)** $(A \smallsetminus B) \smallsetminus C = A \smallsetminus (B \cup C)$;
- **(r)** $(A \smallsetminus B) \cup (B \smallsetminus A) = (A \cup B) \smallsetminus (A \cap B)$.

## 4. Product of Sets

An *ordered pair* is two elements in a specific order; if $a$ and $b$ are elements, the pair containing them with $a$ first and $b$ second is denoted $(a, b)$. Of course the notation conflicts with our notation for an open interval of real numbers, but this cannot be helped, since it is standard.

Ordered pairs obey the "defining property":

$$(a, b) = (c, d) \Leftrightarrow a = c \text{ and } b = d.$$

If $(a, b)$ is an ordered pair, then $a$ is called the *first coordinate* and $b$ is called the *second coordinate*.

Let $A$ and $B$ be sets. The *product* of $A$ and $B$ is denoted $A \times B$ and is defined to be the set of ordered pairs whose first coordinate is in $A$ and whose second coordinate is in $B$:

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

**Proposition 1.4.** *Let $X$ be a set and let $A, B, C \subset X$. Then*

**(a)** $(A \cup B) \times C = (A \times C) \cup (B \times C)$;
**(b)** $(A \cap B) \times C = (A \times C) \cap (B \times C)$;
**(c)** $A \times (B \cup C) = (A \times B) \cup (A \times C)$;
**(d)** $A \times (B \cap C) = (A \times B) \cap (A \times C)$;
**(e)** $(A \cap B) \times (C \cap D) = (A \times C) \cap (B \times D)$.

Similarly, we may speak of ordered triples $(a, b, c)$; the product of three sets $A$, $B$, and $C$ is

$$A \times B \times C = \{(a, b, c) \mid \ a \in A,\ b \in B, \text{ and } c \in C\ \}.$$

In general, we may speak of *ordered $n$ tuples* of the form $(a_1, \dots, a_n)$, where again the entry $a_i$ is known as the $i^{\text{th}}$ *coordinate* of the tuple. If $A_1, \dots, A_n$ are sets, their product is

$$\times_{i=1}^n A_i = \{(a_1, \dots, a_n) \mid a_i \in A_i\}.$$

The product of a set $A$ with itself $n$ times is denoted by $A^n$; thus

$$A^n = \{(a_1, \dots, a_n) \mid a_i \in A\}.$$

For example, the set of ordered triples of real numbers is denoted by $\mathbb{R}^3$.

**Example 1.5.** Let $A = [1, 3] \times [2, 4) \times (3, 5)$. How many elements are in the set $A \cap \mathbb{Z}^3$?

*Solution.* We have $B = [1, 3] \cap \mathbb{Z} = \{1, 2, 3\}$, $C = [2, 4) \cap \mathbb{Z} = \{2, 3\}$, and $D = (3, 5) \cap \mathbb{Z} = \{4\}$. Then

$$A \times \mathbb{Z}^3 = B \times C \times D = \{(1, 2, 5), (1, 3, 5), (2, 2, 5), (2, 3, 5), (3, 2, 5), (3, 3, 5)\},$$

a set with 6 elements. □

## 5. Functions

Let $A$ and $B$ be sets. A *function* from $A$ to $B$, denoted $f : A \to B$, is an assignment of every element in $A$ to a unique element in $B$; we say that $f$ *maps* $A$ *into* $B$, and that $f$ is a function *on* $A$. If $a \in A$, he element in $B$ to which $a$ is assigned is denoted $f(a)$; we say that *$a$ is mapped to $b$ by $f$*. We often think of $f$ as sending points in $A$ to locations in $B$. Functions obey the "defining property":

$$\text{for every } a \in A \text{ there exists a unique } b \in B \text{ such that } f(a) = b.$$

If $A$ is sufficiently small, we may explicitly describe the function by listing the elements of $A$ and where they go; for example, if $A = \{1, 2, 3\}$ and $B = \mathbb{R}$, a perfectly good function is described by $\{1 \mapsto 23.432, 2 \mapsto \pi, 3 \mapsto \sqrt{593}\}$.

However, if $A$ is large, the functions which are easiest to understand are those which are specified by some *rule* or *algorithm*. The common functions of single variable calculus are of this nature, for example, the polynomials in $x$, $\sin x$, $\log x$, etc.

Let $f : A \to B$ be a function. The *domain* of $f$ is $A$, and the *range* of $f$ is $B$.

If $C \subset A$, the *image* of $C$ is

$$f(C) = \{b \in B \mid f(c) = b \text{ for some } c \in C\}.$$

The image of a function is the image of its domain.

If $D \subset B$, the *preimage* of $D$ is

$$f^{-1}(D) = \{a \in A \mid f(a) \in D\}.$$

*Remark* 1.2. Some authors use the word range to mean what we have called the image of a function. In this case, they use the word codomain to mean what we have called the range.

We say that $f$ is *injective* (or *one to one*) if for every $a_1, a_2 \in A$ we have $f(a_1) = f(a_2) \Rightarrow a_1 = a_2$.

We say that $f$ is *surjective* (or *onto*) for every $b \in B$ there exists $a \in A$ such that $f(a) = b$. A function is surjective if and only if its range is equal to its image.

We say that $f$ is *bijective* if it is both injective and surjective. Such a function sets up a *correspondence* between the elements of $A$ and the elements of $B$.

**Example 1.6.** The function $f : \mathbb{Z} \to \mathbb{Z}$ given by $n \mapsto 2n$ is injective but not surjective. The function $g : \mathbb{Z} \times \mathbb{Z} \to \mathbb{Q}$ given by $(p, q) \mapsto \frac{p}{q}$ is surjective but not injective. □

**Example 1.7.** Let $f : \mathbb{R} \to \mathbb{R}$.

- **(a)** if $f(x) = x^3$, then $f$ is bijective.
- **(b)** if $f(x) = x^2$, then $f$ is neither surjective nor injective.
- **(c)** if $f(x) = x^3 - x$, the $f$ is surjective but not injective.
- **(d)** if $f(x) = \arctan x$, then $f$ is injective but not surjective.

## 6. Composition of Functions

Let $A$, $B$, and $C$ be sets and let $f : A \to B$ and $g : B \to C$. The *composition* of $f$ and $g$ is the function

$$g \circ f : A \to C$$

given by

$$g \circ f(a) = g(f(a)).$$

If $f$ and $g$ are injective, then $g \circ f$ is injective. If $f$ and $g$ are surjective, then $g \circ f$ is surjective.

The domain of $g \circ f$ is $A$ and the range is $C$. The image of $g \circ f$ is the image under $g$ of the image under $f$ of the domain of $f$.

**Example 1.8.** Let $A$ be the set of living things on earth, $B$ the set of species, and $C$ be the set of positive real numbers. Let $f : A \to B$ assign to each living thing its species, and let $g : B \to C$ assign to each species its average mass. Then $g \circ f$ guesses the mass of a living thing. $\square$

**Example 1.9.** Let $f : \mathbb{R} \to \mathbb{R}$ be given by $f(x) = x^2$ and let $g : \mathbb{R} \to \mathbb{R}$ be given by $g(x) = \sin x$. Then $g \circ f : \mathbb{R} \to \mathbb{R}$ is given by $g \circ f(x) = \sin x^2$ and $f \circ g : \mathbb{R} \to \mathbb{R}$ is given by $f \circ g(x) = \sin^2 x$. $\square$

**Example 1.10.** Let $f : \mathbb{R} \to \mathbb{R}^2$ be given by $f(t) : \langle 2\cos t, \sin t\rangle$. The image of $f$ is an ellipse in the plane. Let $s : \mathbb{R}^2 \to \mathbb{R}^3$ be given by $s(x, y) = (x, y, y^2 - x^2)$. The image of $s$ is a saddle surface.

Then the image of $s \circ f$ is a curve in $\mathbb{R}^3$ whose shape is roughly the boundary of a potato chip.

We may think of the ellipse as a road on a plane. Then think of $s$ as an earthquake which takes the plane and shifts it, warping its shape into a saddle. The road is carried along with the plane as it warps. The new position of the road is the image of the composition of the functions. $\square$

If $A$ is a set, define the *identity function* on $A$ to be the function $\mathrm{id}_A : A \to A$ given by $\mathrm{id}_A(a) = a$ for all $a \in A$. Identity functions are bijective, and have the property that if $f : A \to B$, then $f \circ \mathrm{id}_A = f$ and $\mathrm{id}_B \circ f = f$.

We say that $f$ is *invertible* if there exists a function $f^{-1} : B \to A$, called the *inverse* of $f$, such that $f \circ f^{-1} = \mathrm{id}_B$ and $f^{-1} \circ f = \mathrm{id}_A$.

**Proposition 1.11.** *A function is invertible if and only if it is bijective.*

If $f$ is injective, we define the *inverse* of $f$ to be a function $f^{-1} : f(A) \to A$ by $f^{-1}(y) = x$, where $f(x) = y$. Since an invertible function is bijective, it is injective, and this definition of inverse agrees with our previous one in this case.

If $f : A \to B$ is a function and $C \subset A$, we define a function $f \restriction_C : C \to B$, called the *restriction* of $f$ to $C$, by $f \restriction_C (c) = f(c)$. If $f$ is injective, then so is $f \restriction_C$.

The next two sections are a bit more advanced; you only need to skim them. The main idea that we will use from them is the notion of a set whose elements are functions.

## 7. Cardinality

The *cardinality* of a set is the number of elements in it. Two sets have the same cardinality if and only if there is a bijective function between them.

Let $\mathbb{N} = \{0, 1, 2, \dots\}$ be the set of natural numbers and for $n \in \mathbb{N}$ let $H_n = \{m \in \mathbb{N} \mid m < n\}$. A set $X$ is called *finite* if there exists a surjective function from $X$ to $H_n$ for some $n \in \mathbb{N}$. If there exists a bijective function $X \to H_n$, we say that the cardinality of $X$ is $n$, and write $|X| = N$.

A set $X$ is called *infinite* if there exists an injective function $\mathbb{N} \to X$.

**Proposition 1.12.** *A set is infinite if and only if it is not finite.*

**Proposition 1.13.** *Let $A$ be a finite set and let $f : A \to A$ be a function. Then $f$ is injective if and only if $f$ is surjective.*

**Proposition 1.14.** *Let $A$ and $B$ be finite sets. Then $|A \times B| = |A| \cdot |B|$.*

## 8. Collections

A *collection* is a set whose elements are themselves sets or functions.

Let $X$ be a set. The collection of all subsets of $X$ is called the *power set* of $X$ and is denoted $\mathcal{P}(X)$.

Let $\mathcal{C}$ be a collection of subsets of $X$; then $\mathcal{C} \subset \mathcal{P}(X)$. Define the *intersection* and *union* of the collection by

- $\cap\mathcal{C} = \{a \in A \mid a \in C \text{ for all } C \in \mathcal{C}\}$
- $\cup\mathcal{C} = \{a \in A \mid a \in C \text{ for some } C \in \mathcal{C}\}$

If $\mathcal{C}$ contains two subsets of $X$, this definition concurs with our previous definition for the union of two sets.

Let $A$ and $B$ sets. The collection of all functions from $A$ to $B$ is denoted $\mathcal{F}(A, B)$.

## 9. Summary

| Symbol | Meaning | Example |
|---|---|---|
| $\Rightarrow$ | implies | $p \Rightarrow q$ |
| $\Leftrightarrow$ | if and only if | $p \Leftrightarrow q$ |
| $\forall$ | for every | $\forall \epsilon > 0$ |
| $\exists$ | there exists | $\exists \delta > 0$ |
| $\vdash$ | such that | $\vdash p$ |

TABLE 1. Logical Connectives

| Set | Name | Definition |
|---|---|---|
| $\mathbb{N}$ | Natural Numbers | $\{1, 2, 3, \dots\}$ |
| $\mathbb{Z}$ | Integers | $\{\dots, -2, -1, 0, 1, 2, \dots\}$ |
| $\mathbb{Q}$ | Rational Numbers | $\{p/q \mid p, q \in \mathbb{Z}\}$ |
| $\mathbb{R}$ | Real Numbers | { Infinite decimal expansions } |
| $\mathbb{C}$ | Complex Numbers | $\{a + ib \mid a, b \in \mathbb{R}$ and $i^2 = -1\}$ |
| $\mathbb{R}^2$ | Euclidean Plane | $\{(a, b) \mid a, b \in \mathbb{R}\}$ |
| $\mathbb{R}^3$ | Euclidean Space | $\{(a, b, c) \mid a, b, c \in \mathbb{R}\}$ |

TABLE 2. Standard Sets

| Symbol | Meaning | Definition |
|---|---|---|
| $\in$ | is an element of | Example: $\pi \in \mathbb{R}$ |
| $\notin$ | is not an element of | Example: $\pi \notin \mathbb{Q}$ |
| $\subset$ | is a subset of | $A \subset B \Leftrightarrow (a \in A \Rightarrow a \in B)$ |
| $\cap$ | intersection | $A \cap B = \{x \mid x \in A$ and $x \in B\}$ |
| $\cup$ | union | $A \cup B = \{x \mid x \in A$ or $x \in B$ |
| $\smallsetminus$ | complement | $A \smallsetminus B = \{x \mid x \in A$ and $x \notin B\}$ |
| $\times$ | cartesian product | $A \times B = \{(a, b) \mid a \in A$ and $b \in B\}$ |

TABLE 3. Set Operations

Let $A$ and $B$ be sets. The notation $f : A \to B$ means $f$ *maps* $A$ *into* $B$; that is, $f$ is a function whose domain is $A$ and whose range is $B$.

## 10. Exercises

**Exercise 1.1.** Let $A$, $E$, $O$, $P$, and $S$ be the following subsets of the natural numbers:

- $A = \{n \in \mathbb{N} \mid n < 25\}$;
- $E = \{n \in A \mid n \text{ is even}\}$;
- $O = \{n \in A \mid n \text{ is odd}\}$;
- $P = \{n \in A \mid n \text{ is prime}\}$;
- $S = \{n \in A \mid n \text{ is a square}\}$;

Compute the following sets:

**(a)** $(E \cap P) \cup S$;
**(b)** $(E \cap S) \cup (P \smallsetminus O)$.
**(c)** $P \times S$;
**(d)** $(O \cap S) \times (E \cap S)$.

**Exercise 1.2.** Let $A$, $B$, and $C$ be the following subsets of $\mathbb{R}$:

- $A = [0, 100)$;
- $B = [\frac{1}{2}, \frac{505}{7}]$;
- $C = (-8, \pi]$.

Compute the number of points in the set $(A \times B \times C) \cap (\mathbb{Z} \times \mathbb{Z} \times \mathbb{Z})$.
(Hint: use Proposition 1.4(e) and Proposition 1.14.)

**Exercise 1.3.** Let $A$ and $B$ be subsets of a set $U$. The *symmetric difference* of $A$ and $B$, denoted $A \triangle B$, is the set of points in $U$ which are in either $A$ or $B$ but not in both.
**(a)** Draw a Venn diagram describing $A \triangle B$.
**(b)** Find two set expressions which could be used to define $A \triangle B$, and justify your answer.

**Exercise 1.4.** In each case, give an example of a function $f : \mathbb{R} \to \mathbb{R}$ such that:

**(a)** $f$ is neither injective nor surjective;
**(b)** $f$ is injective but not surjective;
**(c)** $f$ is surjective but not injective;
**(d)** $f$ is bijective.

**Exercise 1.5.** Let $f : \mathbb{R} \to \mathbb{R}$ be given by $f(x) = \sin \pi x$.
Let $A = \mathbb{Z}$ and $B = [\frac{1}{2}, 1]$.

**(a)** Find $f(A)$.
**(b)** Find $f^{-1}(B)$.

**Exercise 1.6.** Let $f : \mathbb{R} \to \mathbb{R}$ be given by $f(x) = x^3 - 5x^2 - 3x + 19$. Find $f^{-1}(4)$.

**Exercise 1.7.** Let $f : X \to Y$ be a function and let $A, B \subset X$.
**(a)** Show that $f(A \cup B) = f(A) \cup f(B)$.
**(b)** Show that $f(A \cap B) \subset f(A) \cap f(B)$.
**(c)** Give an example where $f(A \cap B) \neq f(A) \cap f(B)$.

**Exercise 1.8.** Let $f : X \to Y$ be a function and let $C, D \subset Y$.
**(a)** Show that $f^{-1}(C \cup D) = f^{-1}(C) \cup f(D)$.
**(b)** Show that $f^{-1}(C \cap D) = f^{-1}(C) \cap f(D)$.

**Exercise 1.9.** Let $A$, $B$, and $C$ be any sets. Determine which of the following statements is true, using Venn diagrams if necessary:

**(a)** $A \subset B \Rightarrow A \cap B = A$
**(b)** $A \subset B \Rightarrow B \setminus A = B$
**(c)** $A \setminus (B \cup C) = (A \setminus B) \cup (A \setminus C)$
**(d)** $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$

**Exercise 1.10.** For $a, b \in \mathbb{R}$, let $[a, b] = \{x \in \mathbb{R} \mid a \leq x \leq b\}$ be the closed interval between $a$ and $b$. How many elements are contained in the following sets?

**(a)** $([-2, 3] \cup [5, 9]) \cap \mathbb{Z}$
**(b)** $([\sqrt{2}, \pi] \cup (3^3, 2^5]) \cap \mathbb{Z}$
**(c)** $([1, 5] \times (3, 6)) \cap (\mathbb{Z} \times \mathbb{Z})$

CHAPTER 2

# Vectors

ABSTRACT. Our initial goal is to define "vector" and various vector operations both algebraically and geometrically, and to understand why these definitions are in agreement. Specifically, we will describe a correspondence

$$\{\text{classes of arrows in Euclidean space}\} \longleftrightarrow \{\text{points in cartesian space}\}$$

which preserves the vector operations.

## 1. Euclidean Space

Around 300 B.C. in ancient Greece, Euclid wrote *The Elements*, a collection of thirteen books which sets down the fundamental laws of *synthetic geometry*. This work starts with *points* and sets of points called *lines*, the notions of distance between points and angle between lines, and five postulates which regulate these ideas. The key postulate implies that two distinct points lie on exactly one line.

Geometric figures such as triangles and circles resided on an abstract notion of *plane*, which stretched indefinitely in two dimensions. The Greeks also analyzed solids such as regular tetrahedra, which resided in *space* which stretched indefinitely in three dimensions.

The ancient Greeks had very little algebra, so their mathematics was performed using pictures; no *coordinate system* which gave positions to points was used as an aid in their calculations. We shall refer to the uncoordinatized spaces of synthetic geometry as *Euclidean spaces*. Euclidean spaces are *flat* in the sense that if a Euclidean space contains two points, it contains the entire line which passes through these two points. Traditional Euclidean spaces come in four types: a point, a line, a plane, and space itself; these are Euclidean spaces of dimension zero, one, two, and three, respectively.

The notion of coordinate system arose in the *analytic geometry* of Fermat and Descartes after the European Renaissance (circa 1630). This technique connected the algebra which was flourishing at the time to the ancient Greek geometric notions. We refer to coordinatized lines, planes, and spaces as *cartesian spaces*; these are composed of *ordered $n$-tuples* of real numbers.

Just as coordinatizing Euclidean space yields a powerful technique in the understanding of geometric objects, so geometric intuition and the theorems of synthetic geometry aid in the analysis of sets of $n$-tuples of real numbers.

The concept of *vector* links the geometric world of Euclid to the more algebraic world of Descartes. Vectors may be defined and manipulated entirely in the geometric realm or entirely algebraically; ideally, we use the point of view that best serves our purpose. Typically, this is to understand (geometrically) or to compute (algebraically).

## 2. Cartesian Space

An *ordered $n$-tuple* of real numbers is an list $(x_1, \ldots, x_n)$, where $x_1, \ldots, x_n$ are real numbers, with the defining property that

$$(x_1, \ldots, x_n) = (y_1, \ldots, y_n) \Leftrightarrow x_1 = y_1, \ldots, x_n = y_n.$$

We define $n$-dimensional *cartesian space* to be the set $\mathbb{R}^n$ of ordered $n$-tuples of real numbers. The point $(0, \ldots, 0)$ is called the *origin*, and is labeled by $O$. The numbers $x_1, \ldots, x_n$ are called the *coordinates* of the point $(x_1, \ldots, x_n)$. The set of points of the form $(0, \ldots, 0, x_i, 0, \ldots, 0)$, where $x_i$ is in the $i^{\text{th}}$ slot, is known as the $i^{\text{th}}$ *coordinate axis*.

In $\mathbb{R}^2$, we often use the standard variables $x$ and $y$ instead of $x_1$ and $x_2$. In $\mathbb{R}^3$, we often use $x$, $y$, and $z$ instead of $x_1$, $x_2$, and $x_3$.

We wish to define the *distance* between two points in $\mathbb{R}^n$ in such a way that it will agree with our geometric intuition into the pictures produced by our graphs. Here we use the Pythagorean Theorem.

Let $P = (x_1, y_1)$ and $Q = (x_2, y_2)$. Then

$$d(P, Q) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}.$$

**Example 2.1.** The distance in $\mathbb{R}^2$ from $P = (-4, 3)$ to $Q = (2, 5)$ is

$$d(P, Q) = \sqrt{(2 - (-4))^2 + (5 - 3)^2} = \sqrt{36 + 4} = \sqrt{40} = 2\sqrt{10}.$$

□

Let $P = (x_1, y_1, z_1)$ and $Q = (x_2, y_2, z_2)$. Then

$$d(P, Q) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}.$$

**Example 2.2.** The distance $\mathbb{R}^3$ between $(2, 5, -1)$ and $(-4, 3, 8)$ is

$$d = \sqrt{(-4 - 2)^2 + (3 - 5)^2 + (8 - (-1))^2} = \sqrt{36 + 4 + 9} = \sqrt{49} = 7.$$

□

Let $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ be points in $\mathbb{R}^n$. The *distance* between $x$ and $y$ is defined by

$$d(x, y) = \sqrt{\sum_{i=1}^{n} (y_i - x_i)^2};$$

this formula, which is motivated by the Pythagorean Theorem, defines a function

$$d : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R},$$

called the *distance function*.

## 3. Graphing

For $n = 1$, 2, or 3, it is possible to draw a picture of a subset of $\mathbb{R}^n$. Such a picture is the *graph* of the set.

To graph a set of real numbers, draw a line and select a point to represent zero and a point to its right to represent one. Now mark off the other points accordingly; this process is *ruling*. Now plot real numbers accordingly.

We may graph ordered pairs and sets of order pairs by drawing perpendicular lines, called axes, which are ruled; each line represents a copy of the real numbers, and an ordered pair is plotted as the appropriate point. By convention, the horizontal axis is designated $x$ and represents the first coordinate, and the vertical axis is designated $y$ and represents the second coordinate. For example, the graph of the set $[0, 1] \times [1, 2]$ is a square which touches the $y$-axis and is lifted 1 unit above the $x$-axis. Note that the graph of a function $f$ is the graph of the set $\{(x, y) \in \mathbb{R}^2 \mid y = f(x)\}$.

We may also graph ordered triples of real numbers on a flat piece of paper, using perspective to give the illusion of depth. In this case, tradition demands that the first coordinate of an ordered triple is labeled $x$, the second $y$, and the third $z$; and that the positive $z$-axis points north, the positive $y$-axis points east, and the positive $x$-axis points southwest so that it appears to emanate from the page. Points and sets are plotted against this coordinate system in the natural way.

**Example 2.3.** Let $A = [1, 3]$, $B = [2, 4)$, $C = (3, 5)$, and $D = A \times B \times C$. Graph the set $D \cap \mathbb{Z}^3$.

*Solution.* We see that

$$\begin{aligned} D \times \mathbb{Z}^3 &= (A \times B \times C) \cap (\mathbb{Z} \times \mathbb{Z} \times \mathbb{Z}) \\ &= (A \cap \mathbb{Z}) \times (B \cap \mathbb{Z}) \times (C \cap \mathbb{Z}) \\ &= \{1, 2, 3\} \times \{2, 3\} \times \{4\} \\ &= \{(1, 2, 4), (1, 3, 4), (2, 2, 4), (2, 3, 4), (3, 2, 4), (3, 3, 4)\}. \end{aligned}$$

Plot these six points. □

**Example 2.4.** Draw the box with diagonal vertices $P(1, 1, 2)$ and $Q(4, -1, 4)$.

*Solution.* First we find the other six vertices. These are $(4, 1, 2)$, $(4, -1, 2)$, $(1, -1, 2)$, $(1, -1, 4)$, $(4, 1, 4)$, and $(1, 1, 4)$. Graph these and draw the edges which move parallel to a coordinate axis. □

## 4. Loci

We may consider subsets of $\mathbb{R}^n$ such that the coordinates of the points in the subset are related in some specified way. The common way of doing this is to consider *equations* with the coordinates as *variables*. The *solution set*, or *locus*, of an equation is the set of all points in $\mathbb{R}^n$ which, when their coordinates are plugged into the equation, cause the equality to be true.

Consider the solution set in $\mathbb{R}^3$ of the equation $z = 0$. This is the set of points of the form $(x, y, 0)$. This set is called the $xy$-plane, and is immediately identified with $\mathbb{R}^2$ in the natural way, via the correspondence $(x, y, 0) \leftrightarrow (x, y)$. Similarly, the solution sets of $x = 0$ and $y = 0$ are called the $yz$-plane and the $xz$-plane, respectively. Together, these sets are called *coordinate planes*.

**Example 2.5.** Find the locus in $\mathbb{R}^3$ of the equation $xyz = 0$.

*Solution.* If $xyz = 0$, either $x = 0$, $y = 0$, or $z = 0$. Thus the solution set is the union of the solution sets for these latter equations; that is, the locus of the equation $xyz = 0$ is the union of the coordinate planes. □

**Example 2.6.** Find an equation whose solution set in $\mathbb{R}^3$ is the union of the coordinate axes.

*Solution.* The $x$-axis is the set of points where $y = 0$ and $z = 0$. We can achieve the $x$-axis as the solution set of $y^2 + z^2 = 0$. Thus we can see that the solution set of

$$(x^2 + y^2)(x^2 + z^2)(y^2 + z^2) = 0$$

is the union of the coordinate axes. □

Now consider sets of points which simultaneously satisfy all of the equations in a collection of equations. Such sets are merely the intersection of the solution sets. For example, the solution set of $\{x = 0, y = 0\}$ is the $z$-axis.

If one of the variables is missing from an equation, its locus in $\mathbb{R}^3$ is a *curtain* (or *cylinder*), because the third variable can be anything.

**Example 2.7.** The locus in $\mathbb{R}^2$ of the equation $y = 2x + 1$ is a line, but in $\mathbb{R}^3$ it is a plane. The locus in $\mathbb{R}^3$ of the equation $z = \sin y$ is a rippled "plane"; any point of the form $(x, y, \sin y)$ is in the locus.

Let $P_0 = (x_0, y_0, z_0)$ be some fixed point in $\mathbb{R}^3$ and let $r \in \mathbb{R}$. Consider the equation $d(P, P_0) = r$, where $P = (x, y, z)$ is a variable point. The solution set of this equation is exactly the set of all points in $\mathbb{R}^3$ whose distance from $P_0$ is equal to $r$. This set is called the *sphere of radius $r$ centered at $P_0$*. Since distance is always positive, we may square both sides of the equation and obtain a new equation with the same solution set. Thus the equation of a sphere is

$$(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2 = r^2.$$

**Example 2.8.** Find the radius and center of the sphere given by $x^2 + y^2 + z^2 + 6x - 16 = 0$.

*Solution.* Complete the square. The locus of the above equation is the same as the locus of $x^2 + 6x + 9 + y^2 + z^2 = 16 + 9$, i.e., $(x + 3)^2 + y^2 + z^2 = 25$. Thus the center is $(-3, 0, 0)$ and the radius is 5. □

## 5. Coordinatization

In order to apply the techniques of analytic geometry to synthetic geometry or to a real-life problem, we must impose a coordinate system, a process we refer to as *coordinatization* of Euclidean space.

To coordinatize a line, we have only to select a single point as zero, and a direction for the positive numbers. The coordinate of a point is distance to zero, together with a negative sign if the point is on the negative side of zero; we call this *signed distance*.

To coordinatize a plane, select two lines which intersect at right angles to become the axes; the point of intersection becomes the origin. Select one ray from the origin as the positive $x$-axis; the positive $y$-axis is found by moving counterclockwise by 90 degrees. The coordinates of a point consist of the signed distance to the selected axes.

To coordinatize three dimensional space, we first select a point in Euclidean space and call it the origin. We then select three perpendicular lines that intersect at the origin as the axes. We must also select, on each axis, one of the two directions as the positive direction. By convention, this is done in such a way that the ordered system of axes constitute a right-handed orientation. We use the "right-hand rule": with your right hand, make a fist, let your thumb point up and your point your index finger out, parallel to your arm. Let your middle finger stick out perpendicular to your index finger. Then your axes should be oriented such that the index finger points in the positive $x$ direction, your middle finger points in the positive $y$ direction, and your thumb points in the positive $z$ direction.

Now the coordinates of a point are given by the signed distance of that point to the corresponding coordinate plane. No two points occupy the exact same location, so each point has its own unique coordinates.

Coordinatizing a Euclidean space gives us a cartesian space. These spaces have essentially the same properties. The reason for the distinction is to help us keep in mind that we may often select the coordinate system which best suits our needs in a particular problem.

## 6. Arrows

An *arrow* in Euclidean space is a directed line segment; it is a line segment with one end designated as its *tip* and the other as its *tail*.

Parenthetically, we note that we could be more precise here and define an arrow in a Euclidean space $E$ as an ordered pair $(P, Q) \in E \times E$, where $P$ is the tip and $Q$ is the tail. This is enough information to produce the line segment and differentiate the tip from the tail.

A nonzero arrow is determined by three attributes:

(1) *magnitude*, which is the distance between the tip and the tail;
(2) *direction*, which is the line on which it sits, and its orientation thereon;
(3) *position*, which is determined by its tail.

We do not exclude the possibility that the tip and the tail of an arrow are the same, in which case the arrow is just a point. Such an arrow is called a *zero arrow*; it has zero magnitude and no direction.

The *inverse arrow* of an arrow $\widehat{v}$ is the arrow $-\widehat{v}$, defined to be the same line segment with the tip and tail reversed.

Let $P$ and $Q$ be points in $\mathbb{R}^n$ and let $\widehat{PQ}$ denote the arrow whose tail is $P$ and whose tip is $Q$; this is the arrow *from $P$ to $Q$*. We may add two arrows if the tip of the first equals to the tail of the second. Thus

$$\widehat{PQ} + \widehat{QR} = \widehat{PR}.$$

The arrow $\widehat{PR}$ forms the third side of a triangle.

We would like to be able to add any two arrows, but the dependence on the positioning of the arrows in our definition prevents us. Thus we wish to consider only the magnitude and direction attributes of arrows, and ignore the position; this would allow us to "slide" arrows around in Euclidean space, and consider them to start at the tail or at the tip of some other arrow. This leads us to the concept of vector.

## 7. Vectors

We say that two arrows are *equivalent* if they have the same magnitude and direction, but not necessarily the same position. If $\widehat{v}$ is an arrow, define

$$\vec{v} = \{\widehat{w} \mid \widehat{w} \text{ is an arrow which is equivalent to } \widehat{v}\};$$

such a set is called an *equivalence class of arrows*, or a *vector*.

Let $\widehat{v}$ and $\widehat{w}$ be arrows. If $\widehat{w}$ is equivalent to $\widehat{v}$, we say that $\widehat{w}$ *represents* $\vec{v}$. Technically, the phrase "$\widehat{w}$ represents $\vec{v}$" means that $\widehat{w} \in \vec{v}$. Since any arrow is equivalent to itself, we see that in particular $\widehat{v}$ represents $\vec{v}$.

We can show that $\widehat{w} \in \vec{v}$ if and only if $\vec{w} = \vec{v}$. A vector is determined by two attributes:

(1) magnitude;
(2) direction.

Thus a vector is unpositioned direction and length.

All zero arrows are equivalent; thus there is a unique zero vector.

The *inverse vector* of a vector $\vec{v}$ is the vector $-\vec{v}$, defined to be the vector represented by any arrow $-\widehat{v}$, where $\widehat{v}$ represents $\vec{v}$.

If $P$ is the tail and $Q$ is the tip of an arrow, we write $\vec{PQ}$ for the vector represented by the arrow $\widehat{PQ}$.

For any vector $\vec{v}$ and any point $P \in \mathbb{R}^n$, there is a unique arrow $\widehat{w}$ such that $\widehat{w} \in \vec{v}$ and the tail of $\widehat{w}$ is equal to $P$. It is now possible to add the vectors $\vec{PQ}$ and $\vec{RS}$; let $\widehat{QT}$ be the unique arrow with the same magnitude and direction as $\widehat{RS}$, and define the geometric sum by $\vec{PQ} + \vec{RS} = \vec{PT}$. Note that $-\vec{PQ} = \vec{QP}$ and that $\vec{PQ} + \vec{QP}$ is the point $P$; thus adding the inverse vector produces the zero vector.

Now suppose that we have coordinatized affine space. Then each vector has exactly one representative which is an arrow whose tail is at the origin. Such an arrow is said to be in *standard position*. The tip of this arrow is a point in $\mathbb{R}^n$. Each vector corresponds to exactly one point in $\mathbb{R}^n$ in this way. If $P \in \mathbb{R}^n$, then $\vec{OP}$ is called the *position vector* of $P$.

This correspondence allows us to switch between the concepts of vectors in Euclidean space and points in cartesian space at will, blurring the distinction. We consider vectors and points in $\mathbb{R}^n$ as interchangeable; the viewpoint we adopt depends on the situation. Thus we may use the notation $\mathbb{R}^n$ to denote the set of all vectors in $n$-space.

We no longer put arrows over the vectors; we will specify in each case what set an element is coming from (in particular, whether it comes from $\mathbb{R}$ or from $\mathbb{R}^n$). We may denote the zero vector (the origin) either by $O$, or by 0 (when it cannot be confused with the zero scalar).

## 8. Vector Addition

Let $v = (v_1, v_2, \ldots, v_n)$ and $w = (w_1, w_2, \ldots, w_n)$ be vectors in $\mathbb{R}^n$. We define the *vector sum* of these vectors algebraically by adding the corresponding coordinates:

$$v + w = (v_1 + w_1, v_2 + w_2, \ldots, v_n + w_n).$$

**Proposition 2.9** (Primary Properties of Vector Addition)**.**
*Let $x, y, z \in \mathbb{R}^n$ and let $a, b \in \mathbb{R}$. Then*

- **(a)** $x + y = y + x$; *(Commutativity)*
- **(b)** $(x + y) + z = x + (y + z)$; *(Associativity)*
- **(c)** $x + 0 = x$; *(Existence of an Additive Identity)*
- **(d)** $x + (-x) = 0$; *(Existence of Additive Inverses)*

Geometrically, the vector sum $v+w$ corresponds to sliding an arrow representing $w$ over so that its tail is equal to the tip of $v$. That is, there is a unique arrow which represents the vector $w$ whose tail equals the tip of the vector $v$. We interpret $v+w$ geometrically to be the tip of this arrow. It is the endpoint of the diagonal of the parallelogram determined by $v$ and $w$.

## 9. Scalar Multiplication

Let $v = (v_1, v_2, \ldots, v_n)$ and let $a$ be a real number; we often refer to real numbers as *scalars*. We define the *scalar multiplication* of $a$ times $v$ algebraically by multiplying each coordinate of $v$ by $a$:

$$a \cdot v = (av_1, av_2, \ldots, av_n).$$

The dot is usually omitted from the notation, so $a \cdot v$ is written as $av$.

**Proposition 2.10** (Primary Properties of Scalar Multiplication)**.**
*Let $x, y, z \in \mathbb{R}^n$ and let $a, b \in \mathbb{R}$. Then*

- **(a)** $1 \cdot x = x$; *(Scalar Identity)*
- **(b)** $(ab)x = a(bx)$; *(Scalar Associativity)*
- **(c)** $a(x + y) = ax + ay$; *(Distributivity of Scalar Mult over Vector Add)*
- **(d)** $(a + b)x = ax + bx$. *(Distributivity of Scalar Mult over Scalar Add)*

**Proposition 2.11** (Secondary Properties of Scalar Multiplication)**.**
*Let $x, y, z \in \mathbb{R}^n$ and let $a, b \in \mathbb{R}$. Let $O \in \mathbb{R}^n$ be the origin. Then*

- **(a)** $0 \cdot x = O$;
- **(b)** $a \cdot O = O$;
- **(c)** $-1 \cdot x = -x$;
- **(d)** $(-a)x = -(ax)$.

Geometrically, the scalar multiple $av$ is interpreted as the vector whose direction is that of $v$ but whose length is $|a||v|$. If $a < 0$, then the orientation of $av$ is opposite the orientation of $v$. Thus multiplying a vector by negative one reverses its orientation, and produces its negative.

The vector which proceeds from the tip of $v$ to the tip of $w$ is $w - v$. This is clear, since $v + (w - v) = w$.

We say that two nonzero vectors $v$ and $w$ are *parallel*, and write $v \| w$, if arrows representing $v$ and $w$ lie on parallel line segments. This happens exactly when $w = av$ for some nonzero $a \in \mathbb{R}$.

## 10. Norm of a Vector

The *norm* of a vector is the distance between the tip and the tail of a representing arrow. If the vector is in standard position in $\mathbb{R}^n$, its norm is the distance between the corresponding point and the origin. Thus if $x = (x_1, \ldots, x_n)$, the norm of $x$ is denoted $|x|$ and is given by

$$|x| = \sqrt{\sum_{i=1}^{n} x_i^2}.$$

Synonymous names for this quantity include *modulus*, *magnitude*, *absolute value*, or *length* of the vector.

**Example 2.12.** Let $v \in \mathbb{R}^3$ be given by $v = (2, 4, 8)$. Find $|v|$.

*Solution.* The norm is

$$|v| = \sqrt{2^2 + 4^2 + 8^2} = \sqrt{4 + 16 + 64} = \sqrt{84} = 2\sqrt{21}.$$

□

A *unit vector* is a vector whose norm is 1. In some sense, a unit vector represents pure direction (without length); if $u$ is a unit vector and $a$ is a scalar, then $au$ is a vector in the direction of $u$ with norm $a$.

Let $v$ be any nonzero vector. We obtain a unit vector in the direction of $v$ simply by dividing by the length of $v$. Thus the *unitization* of $v$ is

$$u = \frac{1}{|v|} v.$$

**Example 2.13.** Let $v \in \mathbb{R}^3$ be given by $v = (2, 4, 8)$. Find $|v|$. Find a unit vector in the same direction as $v$.

*Solution.* Since $|v| = 2\sqrt{21}$, the unitization of $v$ is

$$\frac{v}{|v|} = \left( \frac{1}{\sqrt{21}}, \frac{2}{\sqrt{21}}, \frac{4}{\sqrt{21}} \right).$$

□

## 11. Dot Product

Let $v = (v_1, v_2, \ldots, v_n)$ and $w = (w_1, w_2, \ldots, w_n)$ be vectors in $\mathbb{R}^n$. We define the *dot product* of $v$ and $w$ by the rule

$$v \cdot w = v_1 w_1 + v_2 w_2 + \cdots + v_n w_n.$$

There is no ambiguity caused by using a dot for scalar multiplication and vector dot product, because their definitions agree in the only case where there is overlap (namely, if $n = 1$). We usually drop the dot from the notation for scalar multiplication anyway (unless the vector is a known constant). Note that $v+w \in \mathbb{R}^n$ and $av \in \mathbb{R}^n$, but $v \cdot w \in \mathbb{R}$.

**Proposition 2.14** (Properties of Dot Product and Norm)**.**
*Let $x, y, z \in \mathbb{R}^n$ and $a \in \mathbb{R}$. Then*

- **(a)** $x \cdot x = |x|^2$;
- **(b)** $x \cdot y = y \cdot x$; *(Commutativity)*
- **(c)** $x \cdot (y + z) = (x \cdot y) + (x \cdot z)$; *(Distributivity over Vector Addition)*
- **(d)** $a(x \cdot y) = (ax) \cdot y = x \cdot (ay)$;
- **(e)** $x \cdot O = 0$;
- **(f)** $|ax| = |a||x|$.

*Remark.* Properties **(a)** through **(f)** are derived directly from the algebraic definitions. Properties **(c)** and **(d)** together are called *linearity of dot product.* □

The geometric interpretation of dot product is as useful as it is unanticipated from the definition. To understand it, we first need to understand the concept of projection.

Given a line $L$ in $\mathbb{R}^n$ and a point $P$ in $\mathbb{R}^n$ not on the line, there is a unique point $Q$ on the line which is closest to the point. The lines $L$ and $\overline{PQ}$ are perpendicular. The point $Q$ is the *projection* of $P$ onto $L$.

Let $v$ and $w$ be vectors in $\mathbb{R}^n$. There is a unique point on the line through $w$ which is the projection of the tip of $v$ onto this line. The vector whose tail is the origin and whose tip is this projected point is called the *vector projection* of $v$ onto $w$. The norm of this vector projection is the distance from the origin to this projected point and is called the *scalar projection* of $v$ onto $w$. Let $\text{proj}_w(v)$ denote the scalar projection of $v$ onto $w$.

Drop a perpendicular from the tip of $v$ onto the line through $w$ to obtain a right triangle. If $\theta$ is the angle between the vectors $v$ and $w$, we see that $\text{proj}_w(v) = |v| \cos\theta$.

To complete our geometric interpretation of dot product, we need a generalization of the Pythagorean theorem known as the *Law of Cosines.*

**Lemma 2.15** (Law of Cosines)**.** *For a triangle with angles $A, B, C$ and corresponding opposite sides of length $a, b, c$, we have*

$$c^2 = a^2 + b^2 - 2ab\cos(C).$$

*Proof.* We show this for the case where $B$ and $C$ are acute angles, the other cases being similar.

Drop a perpendicular from the angle $A$ to the opposite side. Call this distance $h$. Let $m$ be the distance from $C$ to the perpendicular. Then $a-m$ is the distance from $B$ to the perpendicular. Thus $(a-m)^2 + h^2 = c^2$ and $m^2 + h^2 = b^2$. Substituting $h^2 = b^2 - m^2$ into the first of these yields $a^2 - 2am + b^2 = c^2$. But $m = b\cos(C)$, proving the result. □

**Proposition 2.16.** *Let $v, w \in \mathbb{R}^n$ and let $\theta$ be the angle between $v$ and $w$. Then*

$$v \cdot w = |v||w|\cos\theta.$$

*Proof.* To prove this result, we use the Law of Cosines, a generalization of the Pythagorean Theorem. The Law of Cosines states that for any triangle whose sides have lengths $a$, $b$, and $c$ and whose angle opposite the side of length $c$ has angle $\theta$, then $c^2 = a^2 + b^2 - 2ab\cos\theta$.

To use this, consider the triangle whose vertices are the tips of $v$ and $w$. The vector from $v$ to $w$ is $w - v$, so the lengths of the sides of this triangle are $|v|$, $|w|$, and $|w - v|$. The Law of Cosines now gives us

$$|w - v|^2 = |v|^2 + |w|^2 - 2|v||w|\cos\theta.$$

Since the square of the modulus of a vector is its dot product with itself, we have

$$(w - v) \cdot (w - v) = v \cdot v + w \cdot w - 2|v||w|\cos\theta.$$

By distributativity of dot product over vector addition and other properties,

$$w \cdot w - 2v \cdot w + v \cdot v = v \cdot v + w \cdot w - 2|v||w|\cos\theta.$$

Cancelling like terms on both sides and then dividing by $-2$ yields

$$v \cdot w = |v||w|\cos\theta.$$

□

**Corollary 2.17.** *Let $v, w \in \mathbb{R}^n$ and let $\theta$ be the angle between $v$ and $w$. Then*

$$v \cdot w = |w|\mathrm{proj}_w(v).$$

*If $u$ is of unit length, then*

$$v \cdot u = \mathrm{proj}_u(v).$$

Geometrically, the dot product of $v$ and $w$ is the length of the projection $v$ onto $w$, times the length of $w$.

**Example 2.18.** Let $v = (5,2,1)$ and $w = (3,2,3)$. Find the scalar and vector projections of $v$ onto $w$, and find the angle between them.

*Solution.* We know that $v \cdot w = |w|\text{proj}_w(v)$. Thus

$$\text{proj}_w(v) = \frac{v \cdot w}{|w|} = \frac{15+4+3}{\sqrt{9+4+9}} = \frac{22}{\sqrt{22}} = \sqrt{22}.$$

Thus the scalar projection is the length of $w$, so vector projection is $w$ itself. This intuitively indicates that $v$ and $w$ form a right triangle, with the line segment between the origin and $v$ as the hypotenuse.

We also know that if $\theta$ is the angle between $v$ and $w$, then $\cos\theta = \frac{v \cdot w}{|v||w|} = \frac{\sqrt{22}}{\sqrt{29}}$, so the angle is approximately 29.4 degrees. □

We say that $v$ is *orthogonal* (or *perpendicular*) to $w$, and write $v \perp w$, if the angle $\theta$ between them is a right angle. This happens exactly when the cosine of this angle is zero: $\cos\theta = 0$. Also, by the definition of projection, this happens exactly when the vector projection of $v$ onto $w$ is the zero vector.

Dot product gives us a test for perpendicularity:

$$v \perp w \Leftrightarrow v \cdot w = 0.$$

Note that from this point of view, any vector is perpendicular to the zero vector.

**Example 2.19.** Let $v = (5,2,1)$ and $w = (3,2,3)$. Verify that these vectors form a right triangle.

*Solution.* From the previous example, we believe that the line segment between the points $v$ and $w$ is one of the legs. This leg is represented by the vector $x = w - v = (-2,0,2)$. Then $x \cdot w = -6 + 0 + 6 = 0$, so $x$ is orthogonal to $w$. □

We finish this section with a useful formula.

**Proposition 2.20.** *Let $v = (a,b)$ and $w = (c,d)$ be vectors in $\mathbb{R}^2$. The area of the parallelogram determined by $v$ and $w$ is $ad - bc$.*

*Proof.* The area of a parallelogram of height $h$ and base $s$ is $A = hs$. Consider $w$ to be the base; then $s = |w|$. Now the height is the scalar projection of $v$ onto a vector perpendicular to $w$. Let $x = (-d,c)$; then $w \cdot x = 0$, so $w \perp x$. Moreover, $|w| = |x|$. We have

$$A = hs = |w|\frac{x \cdot v}{|x|} = x \cdot v = ad - bc.$$

□

## 12. Cross Product

The dot product takes two vectors and produces a scalar. In three dimensions, there is a very useful operation that takes two vectors and produces a third vector. It is convenient to let $i = (1,0,0)$, $j = (0,1,0)$, and $k = (0,0,1)$. Then $(x,y,z) = xi + yj + zk$.

Let $v = (v_1, v_2, v_3)$ and $w = (w_1, w_2, w_3)$ be vectors in $\mathbb{R}^3$. The *cross product* of $v$ and $w$ is

$$v \times w = (v_2w_3 - v_3w_2, v_3w_1 - v_1w_3, v_1w_2 - v_2w_1).$$

This may be rewritten as

$$v \times w = (v_2w_3 - v_3w_2)i - (v_1w_3 - v_3w_1)j + (v_1w_2 - v_2w_1)k.$$

We remember this formula via a symbolic determinant. Recall that an $m \times n$ matrix is a rectangular array of real numbers with $m$ rows and $n$ columns. The determinant of a $2 \times 2$ matrix is

$$\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc.$$

The determinant of a $3 \times 3$ matrix is

$$\det \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix} = a_1 \det \begin{bmatrix} b_2 & b_3 \\ c_2 & c_3 \end{bmatrix} - a_2 \det \begin{bmatrix} b_1 & b_3 \\ c_1 & c_3 \end{bmatrix} + a_3 \det \begin{bmatrix} b_1 & b_2 \\ c_1 & c_2 \end{bmatrix}.$$

Thus

$$v \times w = \det \begin{bmatrix} i & j & k \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{bmatrix}.$$

**Proposition 2.21.**

**(a)** *$i \times j = k$;*
**(b)** *$j \times k = i$;*
**(c)** *$k \times i = j$.*

**Proposition 2.22** (Properties of Cross Product)**.** *Let $x, y, z \in \mathbb{R}^n$ and $d \in \mathbb{R}$. Then*

**(a)** *$x \times y = -(y \times x)$;*
**(b)** *$(dx) \times y = x \times (dy) = d(x \times y)$;*
**(c)** *$x \times (y + z) = (x \times y) + (x \times z)$;*
**(d)** *$(x + y) \times z = (x \times z) + (y \times z)$;*
**(e)** *$x \cdot (y \times z) = (x \times y) \cdot z$;*
**(f)** *$x \times (y \times z) = (x \cdot z)y - (x \cdot y)z$;*
**(g)** *$x \times 0 = 0$.*

*Proof.* Write each vector in terms of their components and use the algebraic definition of cross product. □

**Proposition 2.23.** *Let $v, w \in \mathbb{R}^3$. Then $(v \times w) \perp v$ and $(v \times w) \perp w$.*

*Proof.* To see that $(v \times w) \perp v$, we use the dot product.

$$\begin{aligned}(v \times w) \cdot v &= (v_2w_3 - v_3w_2)v_1 + (v_3w_1 - v_1w_3)v_2 + (v_1w_2 - v_2w_1)v_3 \\ &= v_2w_3v_1 - v_3w_2v_1 + v_3w_1v_2 - v_1w_3v_2 + v_1w_2v_3 - v_2w_1v_3 \\ &= 0.\end{aligned}$$

Similarly, $(v \times w) \cdot w = 0$ so $v \times w \perp w$. □

**Proposition 2.24.** *Let $v, w \in \mathbb{R}^3$, and let $\theta$ be the angle between $v$ and $w$. Then*

$$|v \times w| = |v||w| \sin \theta,$$

*which is the area of the parallelogram determined by $v$ and $w$.*

*Proof.* The area of the parallelogram determined by $v$ and $w$ is given by the formula area equals base times height. If we let $|v|$ be the base, then the height is simply $|w| \sin \theta$. Thus the area is $|v||w| \sin \theta$.

Now consider

$$\begin{aligned}|v \times w|^2 &= (v_2w_3 - v_3w_2)^2 + (v_3w_1 - v_1w_3)^2 + (v_1w_2 - v_2w_1)^2 \\ &= v_2^2w_3^2 - 2v_2v_3w_2w_3 + v_3^2w_2^2 \\ &+ v_3^2w_1^2 - 2v_1v_3w_1w_3 + v_1^2w_3^2 \\ &+ v_1^2w_2^2 - 2v_1v_2w_1w_2 + v_2^2w_1^2.\end{aligned}$$

Also,

$$\begin{aligned}(|v||w| \sin(\theta))^2 &= |v|^2|w|^2 \sin^2(\theta) \\ &= |v|^2|w|^2(1 - \cos^2(\theta)) \\ &= |v|^2|w|^2 - |v|^2|w|^2 \cos^2(\theta) \\ &= |v|^2|w|^2 - (v \cdot w)^2 \\ &= (v_1^2 + v_2^2 + v_3^2)(w_1^2 + w_2^2 + w_3^2) \\ &- (v_1w_1 + v_2w_2 + v_3w_3)^2 \\ &= v_2^2w_3^2 - 2v_2v_3w_2w_3 + v_3^2w_2^2 \\ &+ v_3^2w_1^2 - 2v_1v_3w_1w_3 + v_1^2w_3^2 \\ &+ v_1^2w_2^2 - 2v_1v_2w_1w_2 + v_2^2w_1^2.\end{aligned}$$

These last quantities are the same; taking square roots and noting that $\sqrt{\sin^2(\theta)} = \sin(\theta)$ since $\theta \in [0, \pi]$ yields the result. □

**Proposition 2.25.** *Let $v, w \in \mathbb{R}^3$. Then the triple $(v, w, v \times w)$ is oriented by the right-hand rule.*

*Remark.* The orientation of $v \times w$ is actually determined by the orientation given to the coordinate axes. The proof of this requires more advanced techniques than we currently have. The basic idea is the $v \times w$ changes continuously as the lengths of $v$ and $w$ and the angle between them change. Thus if can move $v$ to $i$ and $w$ to $j$ without getting a zero vector as the cross product, the orientation of $(v, w, v \times w)$ must be the same as that of $(i, j, i \times j)$, which is right handed. □

Geometrically, the cross product of $v, w \in \mathbb{R}^3$ is the unique vector $x \in \mathbb{R}^3$ which satisfies these three properties:

(1) $x \perp v$ and $x \perp w$ so that $x$ is perpendicular to the plane determined by $v$ and $w$;
(2) the length of $x$ is equal to the area of the parallelogram determined by $v$ and $w$;
(3) $x$ is oriented by the right hand rule.

**Proposition 2.26.** *Let $v, w \in \mathbb{R}^3$. Then $v \| w$ if and only if $v \times w = 0$.*

*Proof.* If $\theta \in [0, \pi]$, then $\sin\theta = 0$ if and only if $\theta = 0$ or $\theta = \pi$. □

**Example 2.27.** Find the area of the triangle with vertices $P(2, 4, 1)$, $Q(1, 2, 3)$, and $R(5, 0, 1)$.

*Solution.* Treat $P$ as a "translated origin". Let $v = Q - P = (-1, -2, 2)$ and $w = R - P = (3, -4, 0)$. The area of the triangle is half of the area of the parallelogram determined by $v$ and $w$, which we find via the cross product:

$$v \times w = (0 - 8)i - (0 - 6)j + (4 - (-6))k = (-8, 6, 10).$$

Thus the area of the triangle is half to the length of this vector:

$$\text{area} = \frac{1}{2}\sqrt{64 + 36 + 100} = 5\sqrt{2}.$$

□

**Example 2.28.** Let $v = (2, 5, 1)$ and $w = (3, 1, 2)$. Find a vector which is perpendicular to both $22v + 29w$ and $83v - 8w$.

*Solution.* These vectors are linear combinations of $v$ and $w$, and is therefore on the plane determined by $v$ and $w$. It suffices to find a vector which is perpendicular to this plane. We do this by crossing $v$ and $w$:

$$v \times w = (10 - 5)i - (4 - 3)j + (2 - 15)k = (5, -1, -13).$$

□

**Proposition 2.29.** *Let $x, y, z \in \mathbb{R}^3$. Then $x \cdot (y \times z)$ is a scalar quantity which is equal to the signed volume of the parallelepiped determined by the three vectors. The magnitude of this quantity is the volume and the sign detects whether the vectors have a right or left handed orientation in the order presented. We call $x \cdot (y \times z)$ the* scalar triple product.

*Proof.* The volume is equal to the base times the height. If $w = y \times z$, the height is simply the projection of $x$ onto this vector, $\text{proj}_w(x) = x \cdot w/|w|$. But the area of the base is $|w|$, so the base times the height is $x \cdot w$. □

The triple scalar product can the computed as a determinant.

$$x \cdot (y \times z) = \det \begin{bmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{bmatrix}.$$

**Example 2.30.** Do the points $O(0,0,0)$, $P(1,2,3)$, $Q(2,3,1)$, and $R(3,1,2)$ lie on the same plane?

*Solution.* We treat $P$, $Q$, and $R$ as vectors starting at the origin, and note that the four points lie on the same plane if and only if the volume of the parallelepiped determined by these vectors is zero. The triple scalar product is

$$P \cdot (Q \times R) = (6-1)1 - (4-3)2 + (2-9)3 = 5 - 2 - 21 = 18 \neq 0;$$

so no, they don't lie on the same plane. □

**Example 2.31.** Prove that the maximum volume of a parallelepiped with sides of length one is a cube.

*Solution.* First, draw a picture and give everything in the picture a name. Let $x, y, z \in \mathbb{R}^3$ have length one. Let $w = x \times y$. Let $\theta$ be the angle between $y$ and $z$. Let $\phi$ be the angle between $x$ and $w$. Note that $\theta, \phi \in [0, \pi]$.

The volume of a parallelepiped is base times height. The area of the base is the length of the cross product; since we have unit vectors, this is $\sin\theta$. The height is the projection of $x$ onto $w$; since $x$ has unit length, this is $\cos\phi$. Thus the volume is $\sin\theta\cos\phi$.

To maximize this product, maximize each of the factors; $\sin\theta$ is largest when $\theta = \pi/2$ and $\cos\phi$ is largest when $\phi = 0$. Thus the volume is maximized when $y \perp z$ and $x \| w$, which means that $x \perp y$ and $x \perp z$. This is a cube. □

## 13. Summary

- The set of points in Euclidean space, when labeled with coordinates, is called cartesian space. This is the set of all ordered $n$-tuples of real numbers, and is denoted $\mathbb{R}^n$. There is no geometric difference between Euclidean space and $\mathbb{R}^n$. The reason for the distinction is that there is more than one way to impose coordinates on Euclidean space; every rotation and translation of an axis system gives a different correspondence.
- Arrows have position, direction, and magnitude. Vectors have only direction and magnitude. Two arrows with the same magnitude and direction "represent" the same vector. We think of vectors as arrows which we can "slide around", to be placed at any convenient tail.
- Selection of a coordinate system creates a correspondence between vectors in Euclidean space and points in cartesian space, given by placing the tail of the vector at the origin and taking the corresponding point to be the tip. Every rotation of the axis system creates a different correspondence.
- The operations of vector addition, scalar multiplication, dot product, and cross product may be defined geometrically or algebraically, and these definitions respect the correspondence between vectors and points.
- The vector sum of two vectors crosses the diagonal of the parallelogram determined by the two vectors.
- The scalar product of a scalar times a vector is that vector stretched by a factor of the scalar.
- The dot product of two vectors is the length of the projection of one onto the other, adjusted by the length of the other.
- The cross product of two vectors is perpendicular to both of them, with length equal the area of the parallelogram determined by them, oriented by the right hand rule.
- Many formulas relating dot and cross products to projections, angles, and so forth can be derived from the above interpretations using pictures and simple geometric facts, and then computed with the algebraic definitions.
- The purpose of describing vectors in this way is to build up geometric intuition which will be helpful in solving problems using linear algebra.

## 14. Exercises

**Exercise 2.1.** Graph the box whose diagonal vertices are the points $(0,0,0)$ and $(1,4,2)$. Label each vertex of the box.

**Exercise 2.2.** Let $A=[0,1]$, $B=[1,2)$, and $C=(3,4]$. Graph the set
$$A\times A\times(B\cup C).$$

**Exercise 2.3.** Describe (and sketch if possible) the graph in $\mathbb{R}^3$ of the following equations, where $x,y,z$ are real variables:

- **(a)** $z=2$
- **(b)** $(x^2+y^2)z=0$
- **(c)** $x^2+y^2+z^2=0$
- **(d)** $x^2+y^2+z^2-4=0$
- **(e)** $x^2+y^2+z^2+4=0$
- **(f)** $x^2+y^2-z=0$

**Exercise 2.4.** Find the center and the radius of the sphere which is the solution set of the equation
$$x^2+y^2+z^2=4x+9y+36z.$$
Graph the sphere.

**Exercise 2.5.** Consider the line segment from $P_1(x_1,y_1,z_1)$ to $P_2(x_2,y_2,z_2)$. Convince yourself that its midpoint is
$$\Big(\frac{x_1+x_2}{2},\frac{y_1+y_2}{2},\frac{z_1+z_2}{2}\Big).$$

**Exercise 2.6.** Find an equation of a sphere if one of its diameters has endpoints $(2,1,4)$ and $(4,3,10)$.

**Exercise 2.7.** Draw the directed line segment $\vec{AB}$. Find and draw the equivalent the vector $v$ whose tail is at the origin.

- **(a)** $A(3,1)$, $B(3,3)$
- **(b)** $A(-3,5)$, $B(-2,0)$
- **(c)** $A(0,2,4)$, $B(5,2,-2)$

**Exercise 2.8.** Find the vector sum $v+w$ and illustrate geometrically.

- **(a)** $v=(0,1)$, $w=(1,0)$
- **(b)** $v=(2,4)$, $w=(5,1)$
- **(c)** $v=(-2,3)$, $w=(3,-2)$
- **(d)** $v=(1,0,1)$, $w=(0,1,0)$
- **(e)** $v=(1,2,3)$, $w=(-1,2,-3)$

**Exercise 2.9.** Find $|v|$, $v+w$, $v-w$, $2v$, and $3v-2w$.

- **(a)** $v=(1,2)$, $w=(3,4)$
- **(b)** $v=(-1,-2)$, $w=(2,1)$
- **(c)** $v=(3,2,-=1)$, $w=(0,6,7)$
- **(d)** $v=i-j$, $w=i+k$
- **(e)** $v=i+j+k$, $w=2i-3j-4k$

**Exercise 2.10.** Find a unit vector which has the same direction as $v$.

**(a)** $v = (3, 4)$
**(b)** $v = (5, -5)$
**(c)** $v = (1, 2, 3)$
**(d)** $v = i + j + k$

**Exercise 2.11.** Express $i$ and $j$ in terms of $v$ and $w$.

**(a)** $v = i + j$, $w = i - j$
**(b)** $v = 2i + 3j$, $w = i - j$

**Exercise 2.12.** Find $v \cdot w$.

**(a)** $v = (2, 4)$, $w = (-1, 4)$
**(b)** $v = (5, -1)$, $w = (7, 7)$
**(c)** $v = (1, 2, 3)$, $w = (3, 2, 1)$
**(d)** $v = (2, -4, 1)$, $w = (3, 3, 6)$

**Exercise 2.13.** Find the scalar and vector projections of $v$ onto $w$.

**(a)** $v = (2, 4)$, $w = (-1, 4)$
**(b)** $v = (5, -1)$, $w = (7, 7)$
**(c)** $v = (1, 2, 3)$, $w = (3, 2, 1)$
**(d)** $v = (2, -4, 1)$, $w = (3, 3, 6)$

**Exercise 2.14.** Find the values for $x \in \mathbb{R}$ such that $v \perp w$.

**(a)** $v = (3, x)$, $w = (-4, 3)$
**(b)** $v = (x, 12)$, $w = (x^3, 18)$
**(c)** $v = (t, 2t, 3t)$, $w = (5, x, -2)$

**Exercise 2.15.** Find the values for $x \in \mathbb{R}$ such that the angle between $v = (1, 1)$ and $w = (x, 1)$ is $60°$.

**Exercise 2.16.** Find the angle between the diagonal of a cube and one of its edges.

**Exercise 2.17.** Find $v \times w$.

**(a)** $v = (1, 0, 1)$, $w = (0, 1, 0)$
**(b)** $v = (1, 2, 3)$, $w = (1, 3, 5)$
**(c)** $v = (1, 1, 1)$, $w = (-1, 1, 1)$

**Exercise 2.18.** Let $v = (1, 2, 3)$ and $w = (3, 2, 1)$. Find the following.

**(a)** $|v|$
**(b)** $|w|$
**(c)** $x = v + w$
**(d)** $y = v - w$
**(e)** $|x|$
**(f)** $|y|$
**(g)** $v \cdot w$
**(h)** $x \cdot y$
**(i)** $v \times w$
**(j)** $x \times y$
**(k)** $|v \times w|$
**(l)** $|x \times y|$

**Exercise 2.19.** Let $x = (1, 2, 3)$, $y = (-2, 0, -3)$, and $z = (1, -2, 0)$.

**(a)** Draw each of these vectors emanating from the origin.
**(b)** Now draw $(x)$ emanating from the origin, $y$ with its tail at the tip of $x$, and $z$ with its tail at the tip of $y$.
**(c)** Find $x + y + z$. Does your result agree with your picture?

**Exercise 2.20.** (*Challenge*) The spheres

$$x^2 + y^2 + z^2 = 144 \quad \text{and} \quad (x-3)^2 + (y-4)^2 + (z-12)^2 = 25$$

intersect in a circle. Find the center of the circle.
(Hint: Let $O = (0, 0, 0)$, $P = (3, 4, 12)$, $Q$ be a point of intersection of the spheres, and $R$ be the center point; then $R$ is on the line $\overline{OP}$, and $\overline{OP} \perp \overline{QR}$.)

**Exercise 2.21.** Let $v$ and $w$ be vectors in $\mathbb{R}^2$. Give a geometric interpretation of and prove the following formulae:

**(a)** Cauchy Schwartz Inequality:

$$|v \cdot w| \leq |v||w|$$

**(b)** Triangle Inequality:

$$|v + w| \leq |v| + |w|$$

**(c)** Parallelogram Law:

$$|v + w|^2 + |v - w|^2 = 2|v|^2 + 2|w|^2$$

(Hint for (b) and (c): Use the Cauchy Schwartz Inequality, the distributivity of dot over sum, and the fact that $|v + w|^2 = (v + w) \cdot (v + w)$.)

**Exercise 2.22.** Find the volume of the parallelepiped determined by the vectors $x = (1, 2, 3)$, $y = (2, 3, 1)$, and $z = (-1, 0, t)$. Find $t$ such that these vectors are coplanar.

**Exercise 2.23.** Do the points $P(0, 1, 2)$, $Q(3, 7, 5)$, $R(-1, 0, 1)$, and $S(6, 2, 8)$ lie on the same plane? Can one change this answer by changing the $y$-coordinate of $Q$? What does this tell you?

**Exercise 2.24.** The following identities are true for $x, y, z \in \mathbb{R}^3$. Examine them for geometric content.

**(a)** $(x - y) \times (x + y) = 2(x \times y)$;
**(b)** $x \times (y \times z) = (x \cdot z)y - (x \cdot y)z$;
**(c)** $x \times (y \times z) + y \times (z \times x) + z \times (x \times y) = 0$.

(Hint: first consider the case of standard basis vectors; then consider the case of arbitrary unit vectors; then try to generalize to arbitrary vectors.)

**Exercise 2.25.** Let $v$ and $w$ be vectors in $\mathbb{R}^n$ and let $\theta$ be the angle between them. Recall that $v$ is *orthogonal* to $w$, written $v \perp w$, if $\theta = 90°$, which happens exactly when $v \cdot w = 0$,

Let $w, y \in \mathbb{R}^3$. Show that the vector

$$v = y - \frac{w \cdot y}{|w|^2} w$$

is orthogonal to $w$.

**Exercise 2.26.** Let $v$ and $w$ be vectors in $\mathbb{R}^n$ and let $\theta$ be the angle between them. Recall that the formula $v \cdot w = |v||w| \cos\theta$ implies that the scalar projection of $v$ onto $w$ is given by

$$\text{proj}_w(v) = \frac{v \cdot w}{|w|}.$$

- **(a)** Define a function $v : \mathbb{R} \to \mathbb{R}^2$ by $v(t) = (t, t^2)$. Graph the image of $v$.
- **(b)** Let $w = (1, 1) \in \mathbb{R}^2$. Graph the line through $w$.
- **(c)** Define a function $f : \mathbb{R} \to \mathbb{R}$ by $f(t) = \text{proj}_w(v(t))$. Find a formula for $f$ in terms of $t$.
- **(d)** Find $f([1, 2])$, the image of the closed interval $[1, 2]$ under the function $f$.
- **(e)** Interpret $\lim_{t\to\infty} f(t)$ geometrically.

**Exercise 2.27.** Let $f(t)$ be a real valued function given by

$$f(t) = |i \times (\cos t, \sin t, 0)|.$$

Find $f$ and interpret it geometrically, thinking of $t$ as time and noting that as $t$ changes, $(\cos t, \sin t, 0)$ sweeps out a unit circle in the $xy$-plane.

CHAPTER 3

# Linear Sets

ABSTRACT. Linear sets are subsets of Euclidean space which "look like" a lower dimensional Euclidean space. We use our results regarding vectors to investigate linear sets in cartesian space.

## 1. Linear Sets in $\mathbb{R}^n$

A *linear set* in $\mathbb{R}^n$ is a subset $L \subset \mathbb{R}^n$ with the property that if $P, Q \in L$ are distinct points in the linear set, then the entire line through $P$ and $Q$ is in $L$.

Let $P \in \mathbb{R}^n$ be a point and set $L = \{P\}$; we call a set containing one element a *singleton set*. The singleton set $L$ is vacuously a linear set, since there are not two distinct points in $L$. Blurring the distinction between $P$ and $\{P\}$, we say that points in $\mathbb{R}^n$ are linear sets.

If $L \subset \mathbb{R}^n$ is a line, then it is also a linear set. The linear subsets of $\mathbb{R}^2$ are points, lines, and the entire space $\mathbb{R}^2$. The linear subsets of $\mathbb{R}^3$ are points, lines, planes, and the entire space $\mathbb{R}^3$.

**Proposition 3.1.** *Let $L$ be a linear set in $\mathbb{R}^n$ and let $Q \in \mathbb{R}^n$. Then there exists a unique point $Q \in L$ such that $d(P, Q) \leq d(P, R)$ for every $R \in L$.*

*Proof.* In other words, there is a unique point $Q$ in $L$ which is closer to $P$ than any other point in $L$. To see this, suppose that there are two points $Q_1$ and $Q_2$ in $L$ such that $d(P, Q_1) = d(P, Q_2)$, and $d(P, Q_1) \leq d(P, R)$ for every $R \in L$. Then the entire line through $Q_1$ and $Q_2$ is in $L$; thus the midpoint between $Q_1$ and $Q_2$ is in $L$, and this midpoint is closer to $L$ then $Q_1$ and $Q_2$ by trigonometry. This shows that $Q_1 = Q_2$. □

The next proposition says that linear sets which pass through the origin are closed under vector addition and scalar multiplication.

**Proposition 3.2.** *Let $L \subset \mathbb{R}^n$ be a linear set such that $O \in L$, where $O$ is the origin. Then*

**(a)** *$v + w \in L$ for every $v, w \in L$;*
**(b)** *$tv \in L$ for every $v \in L$ and every $t \in \mathbb{R}$.*

*Proof.* Note that if $L = \{O\}$, the statement is true. Thus we assume that $L$ is contains more than just the origin.

Let $v \in L$ and $t \in \mathbb{R}$; we may assume the $v$ is nonzero. The entire line through $v$ and $O$ is in $L$; since $tv$ is on this line, it is in $L$, which shows **(b)**.

Let $v, w \in L$ and assume that they are nonzero. The line through the tips of $v$ and $w$ is contained in $L$; the midpoint of the line segment from the tip of $v$ to the tip of $w$ is $x = v + \frac{1}{2}(w - v)$. Thus $x \in L$. Therefore $2x \in L$; but $2x = 2v - (w - v) = v + w$. □

## 2. Lines in $\mathbb{R}^2$ via Vector Equations

A line in $\mathbb{R}^2$ is determined by a point on the line and the direction of the line. Let $P_0 = (x_0, y_0)$ be a point on the line, and let $v = (v_1, v_2)$ be a vector in the direction of the line; we call $v$ a *direction vector*. If we start at $P_0$ and move in the direction specified by $v$ for a period of time $t$ at a speed given by $|v|$, we arrive at the point

$$P = P_0 + tv.$$

If we let $t$ range throughout the real numbers, then the set of points satisfying this equation form a line. Thus we call $P = P_0 + tv$ the *vector equation* of the line; it is this form of the equation of a line most easily generalizes to higher dimensions.

If we label $P = (x, y)$, where we think of $x$ and $y$ as variables, the vector equation becomes

$$(x, y) = (x_0 + tv_1, y_0 + tv_2).$$

This produces two equations

$$x = x_0 + tv_1 \quad \text{and} \quad y = y_0 + tv_2.$$

These are called the *parametric equations* of the line. The variable $t$ is call the *parameter*.

If $v_1 v_2 \neq 0$, we may eliminate the parameter $t$ by solving the parametric equations for $t$ and setting the results equal to each other to obtain

$$\frac{x - x_0}{v_1} = \frac{y - y_0}{v_2};$$

this is called the *symmetric equation* of the line.

Solving the symmetric equation for $y$ produces the *functional equation*, or *slope-intercept form* of the equation of the line:

$$y = mx + b \quad \text{where} \quad m = \frac{v_2}{v_1} \text{ and } b = y_0 - mx_0.$$

On the other hand, given a line $y = mx + b$ in functional form, we immediately see that $P_0 = (0, b)$ is a point on it, and $(1, m)$ is its direction vector.

**Example 3.3.** Find the vector, parametric, and symmetric equations of the line in $\mathbb{R}^2$ through $Q(4, 1)$ and $R(-3, 3)$.

*Solution.* Let $v = R - Q = (-7, 2)$; this is a direction vector for the line. Then the vector equation is $P = Q + tv$, where $P = (x, y)$. In parametric form,

$$x = 4 - 7t \quad \text{and} \quad y = 1 + 2t.$$

Then symmetric equation is

$$\frac{x - 4}{-7} = \frac{y - 1}{2}.$$

□

## 3. Lines in $\mathbb{R}^2$ via Normal Equation

A line in $\mathbb{R}^2$ is also determined by a point on it and a direction which is perpendicular to the line. Again let $P_0 = (x_0, y_0)$ be a point on the line, and let $n = (n_1, n_2)$ be a vector which is perpendicular to the line; we call $n$ a *normal vector*. If $P = (x, y)$ is a general point on the line, then $P - P_0$ is a vector in the direction of the line, so

$$(P - P_0) \cdot n = 0;$$

this is called the *normal equation* of the line.

The normal equation may be rewritten as $(x - x_0, y - y_0) \cdot (n_1, n_2) = 0$, that is, $n_1(x - x_0) + n_2(y - y_0) = 0$, or finally $n_1 x + n_2 y = n_1 x_0 + n_2 y_0$. Setting $a = n_1$, $b = n_2$, and $c = n_1 x_0 + n_2 y_0$, we obtain the *general form* of the equation of a line

$$ax + by = c,$$

where $a, b, c \in \mathbb{R}$ are constants and $x$ and $y$ variables.

On the other hand, given a line $ax + by = c$ in general form, we immediately see that a vector normal to this line is $(a, b)$; for if $P_0 = (x_0, y_0)$ is a fixed point on the line, and if $P = (x, y)$ is an arbitrary point on the line, we have $ax + by = ax_0 + by_0$. Letting $n = (a, b)$, rewrite this as $P \cdot n = P_0 \cdot n$, which is equivalent to $(P - P_0) \cdot n = 0$. But $P - P_0$ is a direction vector for the line, so $n$ is perpendicular to it.

**Example 3.4.** Find the normal and general equations of the line in $\mathbb{R}^2$ through $Q(3, 1)$ and $R(-2, 3)$.

*Solution.* A direction vector for the line is $v = R - Q = (-5, 2)$. A vector perpendicular to this is $n = (2, 5)$, since $n \cdot v = -5(2) + 2(5) = 0$. The normal equation, then, is

$$(P - Q) \cdot n = 0,$$

where $P = (x, y)$. To get the general form, compute $(x - 3, y - 1) \cdot (2, 5) = 0$ implies $2x - 6 + 5y - 5 = 0$, so

$$2x + 5y = 11.$$

□

In $\mathbb{R}^2$ we have to distinct methods of using vectors to produce equations for lines; the vector equation and the normal equation. Do both of the techniques generalize to higher dimensions?

## 4. Lines in $\mathbb{R}^3$

A line in $\mathbb{R}^3$ is determined by a point on the line and the direction of the line. The direction may be specified by a *direction vector*.

Let $P_0 = (x_0, y_0, z_0)$ be a given point and let $v = (v_1, v_2, v_3)$ be a vector. If we start at $P_0$ and move in the direction $v$ for a period of time $t$ at a rate given by $|v|$, we arrive at the point

$$P = P_0 + tv.$$

If we let $t$ range throughout the real numbers, then the set of points satisfying this equation form a line. This is called the *vector equation* of the line.

If we label $P = (x, y, z)$, then

$$(x, y, z) = (x_0 + tv_1, y_0 + tv_2, z_0 + tv_3).$$

This gives us three equations

$$x = x_0 + tv_1, \qquad y = y_0 + tv_2, \qquad z = z_0 + tv_3.$$

These are called the *parametric equations* of the line. The variable $t$ is called the *parameter.*

**Example 3.5.** Find the vector and parametric equations of the line which passes through the points $Q(1, 3, 2)$ and $R(5, -2, 3)$.

*Solution.* Let $v$ be the vector from $Q$ to $R$. Thus $v = R - Q = (4, -5, 1)$. This is the direction of the line we seek. Letting $Q$ be a point on the line, we have that a point $P$ is on the line if $P = Q + tv = (1 + 4t, 3 - 5t, 2 + t)$. Thus the parametric equations of the line become $x = 1 + 4t$, $y = 3 - 5t$, and $z = 2 + t$. □

If $v_1$, $v_2$, and $v_3$ are nonzero, we may eliminate the parameter $t$ by simply solving the parametric equations for $t$ and setting all the results equal to each other. This yields

$$\frac{x - x_0}{v_1} = \frac{y - y_0}{v_2} = \frac{z - z_0}{v_3}.$$

These are called the *symmetric* equations of the line.

In this form, the symmetric equations point out that the locus of $P = P_0 + tv$ is somewhat independent of $t$; we could replace $t$ by $2t$ or $t^3$ and achieve the same line. Also the symmetric equations yield the following relationships:

$$\frac{y - y_0}{x - x_0} = \frac{v_1}{v_2}; \qquad \frac{z - z_0}{x - x_0} = \frac{v_1}{v_3}; \qquad \frac{z - z_0}{y - y_0} = \frac{v_2}{v_3}.$$

These equations may be recognized as the equations of the *projected lines*; that is, the line in $\mathbb{R}^3$ may be projected each of the coordinate planes, producing a line there whose equation is retrieved from the symmetric equations in this way.

**Example 3.6.** Find the slope-intercept form of the equation of the line which is the projection of the line $(1 + 4t, 3 - 5t, 2 + t)$ onto the $xy$-plane.

*Solution.* We merely eliminate the third coordinate. Thus the vector equation of the line is $(1+4t, 3-5t)$. Eliminating $t$ yields $\frac{x-1}{4} = \frac{y-3}{-5}$. Thus $y - 3 = -\frac{5}{4}(x - 1)$, so $y = -\frac{5}{4}x + \frac{7}{4}$. □

Given two lines in $\mathbb{R}^3$, exactly one of the following holds:

- They intersect in exactly one point.
- Their direction vectors are parallel, so we call them *parallel lines*.
- They do not lie on the same plane, in which case we call them *skew lines*.

That two distinct intersecting lines on the same plane intersect in exactly one point is a result of Euclid's controversial fifth postulate. The only other claim being made here is the following intuitively clear proposition.

**Proposition 3.7.** *Two distinct lines have parallel direction vectors if and only if they lie on the same plane but do not intersect.*

**Example 3.8.** Determine whether or not the lines

$$(2+t, 3+2t, 4+3t) \quad \text{and} \quad (-3+2t, 3-t, -1+t)$$

are parallel, intersecting, or skew.

*Solution.* The direction vectors of the lines are $(1,2,3)$ and $(2,-1,1)$, which are not parallel; thus the lines are not parallel. To see if they intersect, let us call the parameter of the second line $s$ instead of $t$. Thus the second line becomes $(-3+2s, 3-s, -1+s)$.

The question becomes whether or not there are real numbers $s$ and $t$ such that $2+t=-3+2s$, $3+2t=3-s$, and $4+3t=-1+s$. We assume that there is such an $s$ and $t$ and try to find them. Adding the last two equations gives $7+5t=2$ so $5t=-5$ and $t=-1$. If $t=-1$, then the last equation gives $4-3=-1+s$ so $s=2$. Now plug $t=-1$ and $s=2$ into our lines and see that they give the same point, $(1,1,1)$. Thus the lines intersect there. $\square$

## 5. Planes in $\mathbb{R}^3$

A plane in $\mathbb{R}^3$ is determined by a point on the plane and a perpendicular direction. A vector which is perpendicular to a plane is called a *normal vector.*

Let $P_0 = (x_0, y_0, z_0)$ be a given point and let $n = (n_1, n_2, n_3)$ be a vector. Suppose $P = (x, y, z)$ is a point on the plane which passes through $P_0$ and is perpendicular to $n$. Then the arrow from $P_0$ to $P$ is on the plane, and the vector $P - P_0$ is perpendicular to the normal vector $n$. Thus

$$(P - P_0) \cdot n = 0.$$

The set of all $P$ which satisfy this equation constitute the plane; this is called the *normal equation* of the plane.

Writing this in coordinates gives $(x - x_0, y - y_0, z - z_0) \cdot (n_1, n_2, n_3) = 0$ so

$$n_1(x - x_0) + n_2(y - y_0) + n_3(z - z_0) = 0,$$

which can also be written

$$n_1 x + n_2 y + n_3 z = n_1 x_0 + n_2 y_0 + n_3 z_0.$$

On the other hand, the locus of an equation

$$ax + by + cz = d$$

is a plane with normal vector $(a, b, c)$. This is called the *general form* of the equation of a plane.

**Example 3.9.** Find the general form of the equation of the plane which passes through the point $P(2, 4, 1)$ with normal vector equal to the position vector.

*Solution.* We have that $n = (2, 4, 1)$. The equation of the plane, then, is $2(x-2)+4(y-4)+(z-1)=0$, which simplifies to $2x + 4y + z = 21$. □

If the plane is presented in the general form $ax + by + cz = d$ and $a, b, c, d$ are positive, the plane is particularly easy to graph. Simply find the axis intercepts by setting two of the variables to zero.

$$x - \text{intercept} = \frac{d}{a} \qquad y - \text{intercept} = \frac{d}{b} \qquad z - \text{intercept} = \frac{d}{c}$$

Plot these points and connect the dots to obtain a nice picture of the plane.

We realize that three points in $\mathbb{R}^3$ determine a plane; we now give three examples of finding the general form to the equation of such a plane.

**Example 3.10.** Find the equation of the plane which passes through the points $P(3,0,0)$, $Q(0,2,0)$, and $R(0,0,5)$, using intercept points.

*Solution.* The plane is of the form $ax+by+cz=d$. Let $d = 3\cdot 2\cdot 5 = 30$. We know that $3 = \frac{d}{a} = \frac{30}{a}$, so $a = 10$. Similarly, $b = 15$ and $c = 6$. Thus our plane is $10x + 15y + 6z = 30$. □

**Example 3.11.** Find the equation of the plane which passes through the points $P(0,1,4)$, $Q(1,0,3)$, and $R(-2,6,0)$, using dot product.

*Solution.* Let $v = Q - P = (1,-1,-1)$ and $w = R - P = (-2,5,-4)$. There is an entire plane's worth of vectors which are perpendicular to $v$, and a different plane's worth of vectors which are perpendicular to $w$; their intersection is a line which is perpendicular to both. Let $n = (n_1, n_2, n_3)$ be a direction vector for this line. Then $n$ is a normal vector for the plane we seek. Note that any nonzero vector along this line is a normal vector, so we anticipate some choice in our eventual solution.

Now $n$ is perpendicular to both $v$ and $w$, so

$$v \cdot n = 0 \quad \text{and } w \cdot n = 0.$$

Multiplying this out and thinking of the $n_i$'s as variables, this gives two equations in three variables:

$$\begin{aligned} n_1 - n_2 - n_3 &= 0 \\ -2n_1 + 5n_2 - 4n_3 &= 0 \end{aligned}$$

Multiply the first equation by 2, add the resulting equations, and simplify to see that $n_2 = 2n_3$. Plug this into the first equation and simplify to get $n_1 = 3n_3$.

Thus any vector of the form $n = (3n_3, 2n_3, n_3)$ is a normal vector. Set $n_3 = 1$ to get $n = (3,2,1)$. The equation of the plane is

$$n_1 x + n_2 y + n_3 z = n_1 x_0 + n_2 y_0 + n_3 z_0,$$

where $(x_0, y_0, z_0)$ is a point on the plane. Using $Q$ as this point, we have $(x_0, y_0, z_0) = (1,0,3)$, so

$$3x + 2y + z = 6.$$

□

**Example 3.12.** Find the equation of the plane which passes through the points $P(2,1,3)$, $Q(1,5,3)$, and $R(3,2,5)$, using cross product.

*Solution.* The vectors $v = Q - P = (-1,4,0)$ and $w = R - P = (1,1,2)$ lie on the plane. Thus their cross product is perpendicular to it, so we may use this as a normal vector.

$$n = v \times w = (6,2,-5)$$

Then use $P$ as a point on the plane, which gives the equation $6(x-2) + 2(y-1) - 5(z-3) = 0$, which simplifies to $6x + 2y - 5z = 6$. □

Given two planes in $\mathbb{R}^3$, exactly one of the following holds:

- Their normal vectors are parallel, so they are called *parallel planes*.
- They intersect in a line.

The *angle* between two planes is the angle between their normal vectors.

**Example 3.13.** Let $v = (1, 2, 2)$ and $w = (2, 0, 1)$. Let $Y$ be the plane spanned by $v$ and $j$ and let $Z$ be the plane spanned by $w$ and $k$. Find the line which is $Y \cap Z$ and the angle between $Y$ and $Z$.

*Solution.* Outline:

(1) Find the normal vectors using cross product;
(2) Cross the normals to find the direction vector of the line;
(3) Find a point on the line to produce the equation of the line;
(4) Dot the normals to find the angle between them.

□

**Example 3.14.** Let $T$ be the plane given by $5x + 3y + z = 4$ and let $P = (6, 2, 7)$. Find the distance from $P$ to $T$.

*Solution Method 1.* Find the line through $P$ in the direction of the normal vector of the plane. This line intersects the plane at a point $Q$. Then find the distance between $P$ and $Q$. □

*Solution Method 2.* Find any point $Q$ on the plane. Let $v = P - Q$. Find the unit normal $n$ to the plane. Project $v$ onto $n$.

We find $Q$ by plugging in arbitrary $x$ and $y$ and solving for $z$. It is easiest to use $x = 0$ and $y = 0$, which gives that $Q = (0, 0, 4)$ is on the plane.

Now find the unit normal vector of the plane. A normal vector is $(5, 3, 1)$, so the unit normal is $n = \frac{(5,3,1)}{\sqrt{35}}$.

Project the vector $v = P - Q = (6, 2, 3)$ onto the unit normal. This will give the distance.

$$\mathrm{proj}_n(v) = n \cdot v = \frac{30 + 6 + 3}{\sqrt{35}} = \frac{39}{\sqrt{35}}.$$

□

## 6. Lines in $\mathbb{R}^n$

A line in $\mathbb{R}^n$ is determined by a point $Q$ on the line and a direction vector $v$; the points on the line are those we encounter if we proceed from $Q$ in the direction of $v$. Each such point is of the form $Q + tv$, where we think of the real number $t$ as being the time spent travelling in that direction. Thus the line is the set of points $P$ of the form

$$P = tv + Q.$$

Note that the distance between $P$ and $Q$ is equal to $|t||v|$; we may think of $|v|$ as the velocity with which we proceed away from the point $Q$.

The equation $P = tv + Q$ is a *parametric equation*; here we have a parameter $t$ which is allowed to range throughout the entire set of real numbers. The line itself is not the locus of this equation; it is the set

$$L = \{P \in \mathbb{R}^n \mid P = tv + Q \text{ for some } t \in \mathbb{R}\}.$$

## 7. Hyperplanes in $\mathbb{R}^n$

The construction of a line in $\mathbb{R}^2$ and of a plane in $\mathbb{R}^3$ through the use of a normal vector is easily generalized to any dimension.

Define a *hyperplane* in $\mathbb{R}^n$ to be the set of all points perpendicular to a given vector $a$ and passing through a given point $P_0$. If $H \subset \mathbb{R}^n$ is such a hyperplane, then

$$H = \{P \in \mathbb{R}^n \mid (P - P_0) \cdot a = 0\}.$$

A hyperplane in $\mathbb{R}$ is a point; a hyperplane in $\mathbb{R}^2$ is a line, and a hyperplane in $\mathbb{R}^3$ is a plane in the standard sense. In general, a hyperplane in $\mathbb{R}^n$ is geometrically identical to a copy of $\mathbb{R}^{n-1}$ embedded in $\mathbb{R}^n$.

The *normal equation* of a hyperplane is

$$(P - P_0) \cdot a = 0,$$

where $P$ is a variable ranging through $\mathbb{R}^n$, $P_0$ is a specific point on the hyperplane, and $a$ is a vector which is perpendicular to the hyperplane.

The *general equation* of a hyperplane is

$$a_1 x_1 + \cdots + a_n x_n = b,$$

where $(a_1, \ldots, a_n)$ is the normal vector, $x_1, \ldots, x_n$ are coordinate variables, and $b \in \mathbb{R}$. This can be derived from the normal equation, and vice versa.

## 8. Linear Combinations and Spans

Let $A = \{v_1, \ldots, v_r\} \subset \mathbb{R}^n$ be a finite set of vectors from $\mathbb{R}^n$. A *linear combination* of the vectors in $A$ is an element of $\mathbb{R}^n$ of the form

$$a_1v_1 + \cdots + a_rv_r,$$

where $a_1, \ldots, a_r \in \mathbb{R}$. We may also call this a *linear combination from $A$*. We do not place any restrictions in our definitions regarding the relative size of $r$ and $n$; however, this relative size will play a role in what we will be able to conclude.

The *span* of $A$ is the subset $\text{span}(A) \subset \mathbb{R}^n$ defined by

$$\text{span}(A) = \{w \in \mathbb{R}^n \mid w \text{ is a linear combination from } A\}.$$

Let $X \subset \mathbb{R}^n$ be an arbitrary subset, not necessarily finite. Then define the span of $X$ to be the union of all spans of finite subsets of $X$:

$$\text{span}(X) = \{a_1v_1 + \cdots + a_rv_r \mid a_i \in \mathbb{R} \text{ and } v_i \in X \text{ for } i = 1, \ldots, r\}.$$

The span of $X$ is the set of all finite linear combinations of vectors in $X$; we do not have a definition for a linear combination of an infinite number of vectors (one could try to use limits here to get a definition in some cases).

**Proposition 3.15.** *Let $A = \{v_1, \ldots, v_r\} \subset \mathbb{R}^n$. Then*

**(a)** $A \subset \text{span}(A)$*;*
**(b)** $B \subset A \Rightarrow \text{span}(B) \subset \text{span}(A)$*;*
**(c)** $X \subset \text{span}(A) \Rightarrow \text{span}(X) \subset \text{span}(A)$*.*

*Proof.* Since the vector $v_i$ is a linear combination of the vectors in $A$ simply by taking $a_i = 1$ and $a_j = 0$ for $i \neq j$, we get **(a)**.

In light of this, **(b)** follows from **(c)**, so we prove **(c)**. Suppose that $X \subset \text{span}(A)$. Let $B = \{w_1, \ldots, w_s\} \subset X$ be a finite subset. It suffices to show that $\text{span}(B) \subset \text{span}(A)$. Pick an arbitrary vector $w \in \text{span}(B)$; it suffices to show that $w \in \text{span}(A)$.

Now $w = \sum_{j=1}^{s} b_jw_j$ for some real numbers $b_j$. Also, each vector $w_j$ is a linear combination of the $v_i$, that is, $w_j = \sum_{i=1}^{r} a_{ij}v_i$ for some real numbers $a_{ij}$. Thus

$$w = \sum_{j=1}^{s} b_jw_j = \sum_{j=1}^{s} b_j\left(\sum_{i=1}^{r} a_{ij}v_i\right) = \sum_{j=1}^{s}\sum_{i=1}^{r} b_ja_{ij}v_i = \sum_{i=1}^{r}\left(\sum_{j=1}^{s} a_{ij}b_j\right)v_i$$

We have expressed $w$ as a linear combination of the $v_i$s, thus $w \in \text{span}(A)$. □

**Proposition 3.16.** *Let $A = \{v_1, \ldots, v_r\} \subset \mathbb{R}^n$. Let $x, y \in \text{span}(A)$ and let $L$ be the line through $x$ and $y$. Then $L \subset \text{span}(A)$.*

*Exercise Hint.* Pick an arbitrary point on the line. It suffices to show that this point is in $\text{span}(A)$. First show that the point is in $\text{span}\{x, y\}$. □

*Remark* 3.1. The two propositions above remain true if $A$ is replaced by an infinite subset of $\mathbb{R}^n$.

## 9. Subspaces

A subset $W \subset \mathbb{R}^n$ is called a *subspace* of $\mathbb{R}^n$ if

**(S0)** $W$ is nonempty;
**(S1)** $x, y \in W \Rightarrow x + y \in W$;
**(S2)** $a \in \mathbb{R}, x \in X \Rightarrow ax \in W$.

If $W$ is a subspace of $\mathbb{R}^n$, this fact is denoted by $W \leq \mathbb{R}^n$.

Property **(S1)** says that $W$ is *closed under vector addition*, and property **(S2)** says that $W$ is *closed under scalar multiplication*. In the presence of these properties, property **(S0)** is equivalent to the assertion that the origin is an element of $W$. For if $0 \in W$, then $W$ is certainly nonempty; on the other hand, suppose that $W$ is nonempty and let $w \in W$. Then $-1w = -w \in W$ by property **(S2)**, so $0 = w + (-w) \in W$ by property **(S1)**.

**Example 3.17.** The set $\{0\}$, which contains only the origin, is a subspace, called the *trivial* subspace. Also, $\mathbb{R}^n$ is a subspace of itself.

**Example 3.18.** Let $v, w \in \mathbb{R}^3$ and let $W = \text{span}(v, w) = \{av + bw \mid a, b \in \mathbb{R}\}$. Then $W$ is a subspace of $\mathbb{R}^3$. To see this, first note that $0 = 0v + 0w \in W$, so **(S0)** is satisfied. Next select arbitrary vectors $a_1v + b_1w$ and $a_2v + b_2w$ from $V$ and note that their sum is $(a_1 + a_2)v + (b_1 + b_2)w$, which is also in $W$; thus **(S1)** is satisfied. Moreover, if $av + bw \in W$ and $c \in \mathbb{R}$, we have $cav + cbw \in W$; thus **(S2)** is satisfied.

The subspace $W$ is a plane through the origin in $\mathbb{R}^3$.

**Proposition 3.19.** *Let* $A \subset \mathbb{R}^n$. *Then* $\text{span}(A) \leq \mathbb{R}^n$.

*Reason.* Sums and scalar products of linear combinations from $A$ are linear combinations from $A$. □

**Proposition 3.20.** *Let* $X \subset \mathbb{R}^n$. *Then* $X \leq \mathbb{R}^n$ *if and only if* $\text{span}(X) = X$.

*Reason.* Suppose $X$ is a subspace of $\mathbb{R}^n$. We wish to show that $\text{span}(X) = X$. Since we already know that $X \subset \text{span}(X)$, it suffices to show that $\text{span}(X) \subset X$. Let $w \in \text{span}(X)$. It suffices to show that $w \in X$. Now $w$ is a finite linear combination of vectors from $X$. Since $X$ is a subspace, it is closed under addition and scalar multiplication, so all sums and scalar multiples of vectors in $X$ are also in $X$. Thus linear combinations of vectors from $X$ are also in $X$; thus $w \in X$.

Suppose that $\text{span}(X) = X$. Let $x, y \in X$ and $a \in \mathbb{R}$. Then $x + y$ is a linear combination of vectors from $X$, so $x + y \in \text{span}(X) = X$. Also $ax$ is a linear combination of vectors from $X$, so $ax \in \text{span}(X) = X$. Thus $X$ is closed under vector addition and scalar multiplication, i.e., $X$ is a subspace of $\mathbb{R}^n$. □

## 10. Bases

Let $W$ be a subspace of $\mathbb{R}^n$. A *basis* for $W$ is a subset $B \subset W$ such that

**(B1)** $\operatorname{span}(B) = W$;
**(B2)** $C \subsetneq B \Rightarrow \operatorname{span}(C) \subsetneq \operatorname{span}(B)$.

Together, these properties state that $B$ is a *minimal* spanning set. Later, we will show that every subspace has a basis, and that all bases have the same number of elements; we will call this number the *dimension* of the subspace.

**Example 3.21.** Let $v = (1, 1)$ and $w = (1, -1)$. Then $\{v, w\}$ is a basis for $\mathbb{R}^2$. Indeed, let $p = (x, y) \in \mathbb{R}^2$ be an arbitrary point; we wish to write $p$ as a linear combination of $v$ and $w$. This means that we wish to find real numbers $a, b \in \mathbb{R}$ such that $p = av + bw$, or $(x, y) = (a, a) + (b, -b)$. This leads to a pair of equations $x = a + b$ and $y = a - b$. Manipulate these to get $a = \frac{1}{2}(x + y)$ and $b = \frac{1}{2}(x - y)$. We have found $a$ and $b$ in terms of the coordinates of the point $p$, which shows that $p \in \operatorname{span}(v, w)$.

Neither $v$ nor $w$ span $\mathbb{R}^2$ by themselves, so $\{v, w\}$ is a minimal spanning set, so it is a basis.

**Proposition 3.22.** *Let $B \subset \mathbb{R}^n$. Then $B$ is a basis for $\mathbb{R}^n$ if and only if every vector in $\mathbb{R}^n$ can be written as a linear combination from $B$ in a unique way.*

*Remark.* We wish to show that the "minimality" property can be exchanged for the "uniqueness" property. We will show this later; think about why it is true. □

The $i^{\text{th}}$ *standard basis vector* for $\mathbb{R}^n$ is denoted $e_i$ and is defined to be the vector with 1 in the $i^{\text{th}}$ coordinate and zero in every other coordinate. The set of all such vectors is called the *standard basis* for $\mathbb{R}^n$.

For example, the standard basis for $\mathbb{R}^4$ is

$$\{e_1, e_2, e_3, e_4\} = \{(1, 0, 0, 0), (0, 1, 0, 0), (0, 0, 1, 0), (0, 0, 0, 1)\}.$$

This is indeed a basis; for example, we can write

$$(1, -3, \pi, \sqrt{2}) = e_1 - 3e_2 + \pi e_3 + \sqrt{2}e_4.$$

## 11. Linear Transformations

A *linear transformation* from $\mathbb{R}^n$ to $\mathbb{R}^m$ is a function

$$T : \mathbb{R}^n \to \mathbb{R}^m$$

which satisfies

**(L1)** $T(v + w) = T(v) + T(w)$ for all $v, w \in \mathbb{R}^n$;
**(L2)** $T(av) = aT(v)$ for all $v \in \mathbb{R}^n$ and $a \in \mathbb{R}$.

**Example 3.23.** Let $a, b \in \mathbb{R}$ be arbitrary constants. The function $T : \mathbb{R}^2 \to \mathbb{R}^1$ given by $T(x, y) = ax + by$ is linear. To see this, let $v = (x_1, y_1), (x_2, y_2) \in \mathbb{R}^2$. Then $v + w = (x_1 + x_2, y_1 + y_2)$, so

$$\begin{aligned} T(v + w) &= T((x_1 + x_2, y_1 + y_2) \\ &= a(x_1 + x_2) + b(y_1 + y_2) \\ &= (ax_1 + by_1) + (ax_2 + by_2) \\ &= T(v) + T(w). \end{aligned}$$

Now let $v = (x, y) \in \mathbb{R}^2$ and $c \in \mathbb{R}$; then

$$T(cv) = T(cx, cy) = acx + acy = c(ax + by) = cT(v).$$

Thus $T$ is linear.

**Example 3.24.** The function $P_i : \mathbb{R}^n \to \mathbb{R}$ given by $T(x_1, \dots, x_n) = x_i$ is linear; this is called *projection* onto the $i^{\text{th}}$ coordinate.

**Example 3.25.** Fix an arbitrary vector $w \in \mathbb{R}^n$. Then the function $T : \mathbb{R}^n \to \mathbb{R}$ given by $T(v) = v \cdot w$ is linear.

**Proposition 3.26.** *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then*

**(a)** $T(0) = 0$*;*
**(b)** $T(\text{span}(A)) = \text{span}(T(A))$*, where $A \subset \mathbb{R}^n$.*

*Proof.* Let $O_n$ denote the origin in $\mathbb{R}^n$ and let $O_m$ denote the origin in $\mathbb{R}^m$, to distinguish them from the 0 scalar. Then $T(O_n) = T(0 \cdot O_n) = 0 \cdot T(O_n) = O_m$, since 0 times anything in $\mathbb{R}^m$ is $O_m$.

Let $A \subset \mathbb{R}^n$; for simplicity assume that $A = \{v_1, \dots, v_r\}$ is a finite set. Then

$$\begin{aligned} T(\text{span}(A)) &= T\left(\left\{ \sum_{i=1}^{r} a_i v_i \,\middle|\, a_i \in \mathbb{R} \right\}\right) && \text{by definition of span} \\ &= \left\{ T\left( \sum_{i=1}^{r} a_i v_i \right) \,\middle|\, a_i \in \mathbb{R} \right\} && \text{by definition of image} \\ &= \left\{ \sum_{i=1}^{r} a_i T(v_i) \,\middle|\, a_i \in \mathbb{R} \right\} && \text{since } T \text{ is linear} \\ &= \text{span}(\{T(v_1), \dots, T(v_r)\}) && \text{by definition of span} \\ &= \text{span}(T(A)) && \text{by definition of image} \end{aligned}$$

□

**Proposition 3.27.** *A linear transformation is completely determined by its effect on the standard basis.*

*Proof.* This means that if we know the effect of a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ on the standard basis, then we know its effect on all of $\mathbb{R}^n$. This follows from the fact that if $v \in \mathbb{R}^n$, then $v = (a_1, \ldots, a_n)$ for some real numbers $a_i \in \mathbb{R}$. This is the same as saying that $v = \sum_{i=1}^n a_i e_i$; but since $T$ is linear, we have

$$\begin{aligned} T(v) &= T\left( \sum_{i=1}^n a_i e_i \right) \\ &= \sum_{i=1}^n T(a_i e_i) \\ &= \sum_{i=1}^n a_i T(e_i). \end{aligned}$$

□

*Remark* 3.2. The above argument shows that every vector in the image of a linear transformation is a linear combination of the images of the basis vectors.

*Remark* 3.3. The above argument proceeds without change if we replace the standard basis by any spanning set.

**Proposition 3.28.** *Let $w_1, \ldots, w_n \in \mathbb{R}^m$. Then there exists a unique linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ such that $T(e_i) = w_i$ for $i = 1, \ldots, n$.*

*Proof.* Define $T$ by $T(v) = \sum_{i=1}^n a_i T(e_i)$, where $v = (a_1, \ldots, a_n)$. This is linear and sends $e_i$ to the vector $w_i$. It is unique by the previous proposition. □

*Remark* 3.4. The above argument proceeds without change if we replace the standard basis by any finite spanning set.

**Example 3.29.** Define a linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^3$ by $T(e_1) = (1, 2, 0)$, $T(e_2) = (0, 1, 2)$, and $T(e_3) = (2, 0, 1)$. Let $v = (1, 2, 3)$. What is $T(v)$?

*Solution.* Note that $v = e_1 + 2e_2 + 3e_3$. Thus

$$T(v) = T(e_1) + 2T(e_2) + 3T(e_3) = (1, 2, 0) + (0, 2, 4) + (6, 0, 3) = (1, 4, 7).$$

□

**Example 3.30.** Define $T : \mathbb{R}^2 \to \mathbb{R}^2$ by $T(x,y) = (x-y, x+y)$. Discuss the geometric effect of $T$ on $\mathbb{R}^2$.

*Solution.* Note that $T(e_1) = T(1,0) = (1,1)$, and $T(e_2) = T(0,1) = (-1,1)$. On each of these vectors, $T$ has the effect of rotating by $\frac{\pi}{4}$ radians and dilating by a factor of $\sqrt{2}$. Thus, since $T$ is determined by its effect on a basis, we suspect that $T$ has this effect on the entire plane. Let's verify this.

Select an arbitrary vector $v = (x,y) \in \mathbb{R}^2$. Then

$$|T(v)| = |(x+y, x-y)| = \sqrt{(x+y)^2 + (x-y)^2} = \sqrt{2x^2 + 2y^2} = \sqrt{2}|v|.$$

Thus $T$ stretches $v$ by a factor of $\sqrt{2}$.

Now consider $v \cdot T(v) = (x,y) \cdot (x-y, x+y) = x^2 - xy + xy + y^2 = x^2 + y^2 = |v|^2$. If $\theta$ is the angle between $v$ and $T(v)$, we have

$$\cos(\theta) = \frac{v \cdot T(v)}{|v||T(v)|} = \frac{|v|^2}{\sqrt{2}|v|^2} = \frac{\sqrt{2}}{2}.$$

Thus $\theta = \frac{\pi}{4}$, and this is independent of which nonzero vector $v$ we choose. □

**Example 3.31.** Define $T : \mathbb{R}^2 \to \mathbb{R}^2$ by $T(x,y) = (x+y, x-y)$. Discuss the geometric effect of $T$ on $\mathbb{R}^2$.

*Solution.* We use ad hoc methods which we will later develop into a theory. By the same computation, this stretches every vector by a factor of $\sqrt{2}$. Since $T$ stretches every vector by the same amount, intuition tells us that any additional action of $T$ is either as a rotation about the origin, or as a reflection across a line through the origin.

But $T(1,0) = (1,1)$ and $T(0,1) = (-1,1)$; $e_1$ is rotated by $\frac{\pi}{4}$, but $e_2$ is rotated by $-\frac{3\pi}{4}$. Thus $T$ cannot be a dilating rotation. We look for an line across which $T$ is a dilating reflection. This line would need to bisect the angle between $e_1$ and $T(e_1)$, as well as the angle between $e_2$ and $T(e_2)$. The candidate is the line with angle $\frac{\pi}{8}$ radians. Let's verify this computationally.

Now if $T$ is a dilating reflection, then $T(v) = \sqrt{2}v$, where $v$ is the direction vector of the line of reflection. For $v = (x,y)$, this gives $\sqrt{2}x = x+y$ and $\sqrt{2}y = x-y$. Thus $\frac{x+y}{x} = \frac{x-y}{y}$, so $xy + y^2 = x^2 - xy$, whence $y^2 + 2xy - x^2 = 0$. Via the quadratic formula,

$$y = \frac{-2x \pm \sqrt{4x^2 + 4x^2}}{2} = -x \pm \sqrt{2}x.$$

Thus $T$ fixes (setwise) the lines $y = (\sqrt{2}-1)x$ and $y = -(\sqrt{2}+1)x$. The orientation of the line with positive slope is preserved, and the orientation of the line with negative slope is reversed. So the reflection occurs across the line $y = (\sqrt{2}-1)x$. □

## 12. Images and Preimages under Linear Transformations

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation and let $V$ be a subspace of $\mathbb{R}^n$. The *image* of $V$ under $T$ is denoted by $T(V)$ and is defined to be the set of all vectors in $\mathbb{R}^m$ which are "hit" by an element of $V$ under the transformation $T$:

$$T(V) = \{w \in \mathbb{R}^m \mid w = T(v) \text{ for some } v \in V\}.$$

Then $T(V)$ is actually a subspace of $\mathbb{R}^m$.

**Proposition 3.32.** *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation and let $V \leq \mathbb{R}^n$. Then $T(V) \leq \mathbb{R}^m$.*

*Proof.* In order to show that something is a subspace, we need to verify properties **(S0)**, **(S1)**, and **(S2)**.

**(S0)** Since $0 \in V$ and $T(0) = 0$, we see that $0 \in T(V)$.

**(S1)** Let $w_1, w_2 \in T(V)$. Then there exist vectors $v_1, v_2 \in V$ such that $w_1 = T(v_1)$ and $w_2 = T(v_2)$. We have $w_1 + w_2 = T(v_1) + T(v_2) = T(v_1 + v_2)$. Since $V$ is a subspace, $v_1 + v_2 \in V$; thus $w_1 + w_2 \in T(U)$.

**(S2)** Let $w \in T(V)$ and $a \in \mathbb{R}$. Then there exists $v \in V$ such that $T(v) = w$. We have $aw = aT(v) = T(av)$. Since $V$ is a subspace, $av \in U$; thus $aw \in T(V)$. ☐

**Example 3.33.** Let $V$ be the subspace of $\mathbb{R}^2$ spanned by the vector $v = (1, 1)$; that is, $V = \{(t, t) \mid t \in \mathbb{R}\}$ is a line through the origin of slope 1. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be given by $T(x, y) = (x - y, x + y)$; this is linear. Then $T(V)$ is the subspace of $\mathbb{R}^2$ spanned by $T(v) = (1 - 1, 1 + 1) = (0, 2)$; that is, $T(V)$ is the $y$-axis. Thus $T$ rotates $V$ by $\frac{\pi}{4}$ radians and expands it by a factor of $\sqrt{2}$. In fact, this is the effect of $T$ on the entire plane.

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation and let $W$ be a subspace of $\mathbb{R}^m$. The *preimage* of $W$ under $T$ is denoted by $T^{-1}(W)$ and is defined to be the set of all vectors in $\mathbb{R}^n$ which "hit" elements in $W$ under the transformation $T$:

$$T^{-1}(W) = \{v \in \mathbb{R}^n \mid T(v) = w \text{ for some } w \in W\}.$$

Then $T^{-1}(W)$ is actually a subspace of $\mathbb{R}^n$.

**Proposition 3.34.** *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation and let $W \leq \mathbb{R}^m$. Then $T^{-1}(W) \leq \mathbb{R}^n$.*

*Proof.* We verify properties **(S0)**, **(S1)**, and **(S2)**.

**(S0)** Since $0 \in W$ and $T(0) = 0$, we see that $0 \in T^{-1}(W)$.

**(S1)** Let $v_1, v_2 \in T^{-1}(W)$; then $T(v_1)$ and $T(v_2)$ are elements of $W$. Now $T(v_1 + v_2) = T(v_1) + T(v_2)$, and since $W$ is a subspace, this sum is also in $W$. Thus $v_1 + v_2 \in T^{-1}(W)$.

**(S2)** Let $v \in T^{-1}(W)$ and $a \in \mathbb{R}$. Then $T(av) = aT(v)$; since $T(v)$ is in $W$ and $W$ is a subspace, $aT(v) \in W$. Thus $av \in T^{-1}(W)$. ☐

**Example 3.35.** Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be the linear transformation given by $T(x, y) = (x - y, y - z, z - x)$. Let $W = \{0\} \subset \mathbb{R}^m$ be the trivial subspace of $\mathbb{R}^m$; here, 0 means the point $(0, 0, 0)$. The preimage is given by solving the equations

$$x - y = 0; \quad y - z = 0; \quad z - x = 0.$$

Any point of the form $(t, t, t)$, where $t \in \mathbb{R}$, is a solution. Thus $T^{-1}(W)$ is the line in $\mathbb{R}^3$ spanned by the vector $(1, 1, 1)$.

## 13. Kernels of Linear Transformations

The *kernel* of a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is the set of all vectors in the domain $\mathbb{R}^n$ which are sent to the origin in the range $\mathbb{R}^m$. We denote this set by $\ker(T)$:

$$\ker(T) = \{v \in \mathbb{R}^n \mid T(v) = 0\}.$$

**Proposition 3.36.** *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then $\ker(T) \leq \mathbb{R}^n$.*

*Proof.* We verify properties **(S0)**, **(S1)**, and **(S2)**.

**(S0)** We know that $T(0) = 0$; thus $0 \in \ker(T)$.

**(S1)** Let $v_1, v_2 \in \ker(T)$; this means that $T(v_1) = T(v_2) = 0$. Then $T(v_1+v_2) = T(v_1) + T(v_2) = 0 + 0 = 0$, so $v_1 + v_2 \in \ker(T)$.

**(S2)** Let $v \in \ker(T)$ and $a \in \mathbb{R}$. Then $T(av) = aT(v) = a \cdot 0 = 0$; thus $av \in \ker(T)$. □

*Alternate Proof.* Since $W = \{0\}$ is a subspace of $\mathbb{R}^m$ and $\ker(T)$ is the preimage of $W$, we know that $W$ is a subspace by a Proposition 3.34. □

**Example 3.37.** Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be given by $T(x, y, z) = (x, y, 0)$. This is projection onto the $xy$-plane, and is linear. The kernel is the $z$-axis.

**Proposition 3.38.** *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then $\ker(T) = \{0\}$ if and only if $T$ is injective.*

*Proof.* We must show both sides of the implication. Recall that $T$ is injective means that whenever $T(v_1) = T(v_2)$, we must have $v_1 = v_2$.

($\Rightarrow$) Suppose that $\ker(T) = \{0\}$. Let $v_1, v_2 \in \mathbb{R}^n$ such that $T(v_1) = T(v_2)$; we wish to show that $v_1 = v_2$. Then $T(v_1) - T(v_2) = 0$, so $T(v_1 - v_2) = 0$, so $v_1 - v_2 \in \ker(T)$. Since $\ker(T) = \{0\}$, we have $v_1 - v_2 = 0$, so $v_1 = v_2$. Therefore $T$ is injective.

($\Leftarrow$) Suppose that $T$ is injective. Let $v \in \ker(T)$; we wish to show that $v = 0$. But $T(v) = 0$ and $T(0) = 0$, and since $T$ is injective, we must have $v = 0$. □

If $W \leq \mathbb{R}^n$ is a subspace and $v \in \mathbb{R}^n$, the *translate* of $W$ by $v$ is the set

$$v + W = \{v + w \mid w \in W\}.$$

**Proposition 3.39.** *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Let $w \in \mathbb{R}^m$ be in the image of $T$ and let $v \in \mathbb{R}^n$ such that $T(v) = w$. Then*

$$T^{-1}(w) = v + \ker(T).$$

*Proof.* To show that two sets are equal, we show that each is contained in the other.

($\subset$) Let $x \in T^{-1}(w)$. Then $T(x) = w$, so $T(x) - w = 0$. Since $T(v) = w$, we have $T(x) - T(v) = T(x-v) = 0$. Thus $x - v \in \ker(T)$, so $x = v + (x-v) \in v + \ker(T)$.

($\supset$) Let $x \in v + \ker(T)$. Then $x = v + y$, where $y \in \ker(T)$. Thus $T(x) = T(v + y) = T(v) + T(y) = w + 0 = w$, so $x \in T^{-1}(w)$. □

## 14. Sums and Scalar Products of Linear Transformations

Let $S : \mathbb{R}^n \to \mathbb{R}^m$ and $T : \mathbb{R}^n \to \mathbb{R}^m$ be linear transformations identical domains and identical ranges. We define the sum of these linear transformations to be the function $S + T$ given by adding pointwise:

$$S + T : \mathbb{R}^n \to \mathbb{R}^m \quad \text{given by} \quad (S + T)(v) = S(v) + T(v).$$

**Proposition 3.40.** *Let $S : \mathbb{R}^n \to \mathbb{R}^m$ and $T : \mathbb{R}^n \to \mathbb{R}^m$ be linear transformations. Then $S + T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation.*

*Proof.* We verify properties **(L1)** and **(L2)**.

**(L1)** Let $v_1, v_2 \in \mathbb{R}^n$. Then

$$\begin{aligned}(S + T)(v_1 + v_2) &= S(v_1 + v_2) + T(v_1 + v_2)\\ &= S(v_1) + S(v_2) + T(v_1) + T(v_2)\\ &= S(v_1) + T(v_1) + S(v_2) + T(v_2)\\ &= (S + T)(v_1) + (S + T)(v_2).\end{aligned}$$

**(L2)** Let $v \in \mathbb{R}^p$ and $a \in \mathbb{R}$. Then

$$\begin{aligned}(S + T)(av) &= S(av) + T(av)\\ &= aS(v) + aT(v)\\ &= a(S(v) + T(v))\\ &= a(S + T)(v).\end{aligned}$$

□

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation and let $a \in \mathbb{R}$ be a scalar. We define the scalar product of $b$ and $T$ to be the function $bT$ given by multiplying pointwise:

$$bT : \mathbb{R}^n \to \mathbb{R}^m \quad \text{given by } (bT)(v) = bT(v).$$

**Proposition 3.41.** *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformations and let $a \in \mathbb{R}$. Then $bT : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation.*

*Proof.* We verify properties **(L1)** and **(L2)**.

**(L1)** Let $v_1, v_2 \in \mathbb{R}^n$. Then

$$bT(v_1 + v_2) = b(T(v_1) + T(v_2)) = bT(v_1) + aT(v_2).$$

**(L2)** Let $v \in \mathbb{R}^p$ and $a \in \mathbb{R}$. Then

$$bT(av) = baT(v) = abT(v) = a(bT(v)).$$

□

**Example 3.42.** Let $S(x, y) = (x - y, x + y)$ and $T(x, y) = (x + y, x - y)$. Then $(S + T)(x, y) = (2x, 2x)$. This has the effect of plummeting $(x, y)$ vertically onto the line $y = x$ and then stretching by a factor of 2.

## 15. Compositions of Linear Transformations

Let $S : \mathbb{R}^p \to \mathbb{R}^n$ and $T : \mathbb{R}^n \to \mathbb{R}^m$ be linear transformations. The *composition* of $S$ and $T$ is the function

$$T \circ S : \mathbb{R}^p \to \mathbb{R}^m \quad \text{given by} \quad (T \circ S)(v) = T(S(v)).$$

Then $T \circ S$ is actually a linear transformation.

**Proposition 3.43.** *Let $S : \mathbb{R}^p \to \mathbb{R}^n$ and $T : \mathbb{R}^n \to \mathbb{R}^m$ be linear transformations. Then $T \circ S : \mathbb{R}^p \to \mathbb{R}^m$ is a linear transformation.*

*Proof.* We verify properties **(L1)** and **(L2)**.

**(L1)** Let $v_1, v_2 \in \mathbb{R}^p$. Then

$$T(S(v_1 + v_2)) = T(S(v_1) + S(v_2)) = T(S(v_1)) + T(S(v_2)).$$

**(L2)** Let $v \in \mathbb{R}^p$ and $a \in \mathbb{R}$. Then

$$T(S(av)) = T(aS(v)) = aT(S(v)).$$

□

The *identity* transformation on $\mathbb{R}^n$ is the function $J_n = J : \mathbb{R}^n \to \mathbb{R}^n$ which sends every element to itself; that is, $J(v) = v$ for all $v \in \mathbb{R}^n$. This is clearly linear.

Actually, given any arbitrary set $A$, we can define the identity function on it.

Let $A$ be a set. The *identity function* on $A$ is the function

$$\mathrm{id}_A : A \to A \quad \text{given by } \mathrm{id}_A(a) = a.$$

Let $f : A \to B$ be a function. We say that $f$ is *invertible* if there exists a function $g : B \to A$ such that $g \circ f = \mathrm{id}_A$ and $f \circ g = \mathrm{id}_B$. The function $g$ is called the *inverse* of $f$, and is denoted by $f^{-1}$.

**Proposition 3.44.** *Let $f : A \to B$. Then $f$ is invertible if and only if $f$ is bijective.*

*Proof.* To show an if and only if statement, we show implication in both directions.

($\Rightarrow$) Suppose that $f$ is invertible. Then there exists a function $f^{-1} : B \to A$ such that $f^{-1}(f(a)) = a$ for every $a \in A$, and $f(f^{-1}(b)) = b$ for every $b \in B$.

We wish to show that $f$ is injective and surjective.

To show injectivity, we select arbitrary elements of $A$ which go to the same place under $f$ and show that they must have been the same element in the first place.

Let $a_1, a_2 \in A$ such that $f(a_1) = f(a_2)$. Then $f^{-1}(f(a_1)) = f^{-1}(f(a_2))$, so $a_1 = a_2$. Therefore $f$ is injective.

To show surjectivity, we select an arbitrary element of $B$ and find an element $a \in A$ such that $f(a) = b$.

Let $b \in B$. Let $a = f^{-1}(b)$. Then $f(a) = f(f^{-1}(b)) = b$. Therefore $f$ is surjective.

($\Leftarrow$) Suppose that $f$ is bijective. The for every $b \in B$ there exists a unique element $a \in A$ such that $f(a) = b$. Define $f^{-1} : B \to A$ by $f^{-1}(b) = a$. Clearly $f^{-1}$ is the inverse of $f$. □

A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is called invertible if it is invertible as a function. If $T$ is invertible, we have a function $S : \mathbb{R}^m \to \mathbb{R}^n$ such that $T \circ S = J_m$ and $S \circ T = J_n$. We will see that this implies that $m$ must equal $n$. For now, we content ourselves to be reassured that if $T$ is invertible, its inverse is also linear.

**Proposition 3.45.** *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a bijective linear transformation and let $S : \mathbb{R}^m \to \mathbb{R}^n$ be its inverse. Then $S$ is a linear transformation.*

*Proof.* We verify properties **(L1)** and **(L2)**.

**(L1)** Let $w_1, w_2 \in \mathbb{R}^m$. Since $T$ is surjective, there exist $v_1, v_2 \in \mathbb{R}^n$ such that $T(v_1) = w_1$ and $T(v_2) = w_2$. Then $S(w_1) = v_1$ and $S(w_2) = v_2$.

Now $w_1 + w_2 = T(v_1) + T(v_2) = T(v_1 + v_2)$, so $S(w_1 + w_2) = v_1 + v_2 = S(w_1) + S(w_2)$.

**(L2)** Let $w \in \mathbb{R}^m$ and $a \in \mathbb{R}$. There exists $v \in \mathbb{R}^n$ such that $T(v) = w$. Then $S(w) = v$.

Now $T(av) = aT(v) = aw$, so $S(aw) = av = aS(v)$. □

**Example 3.46.** Let $S(x, y) = (x - y, x + y)$ and $T(x, y) = (x + y, x - y)$. Then $(T \circ S)(x, y) = (2x, -2y)$. This has the effect of reflecting the plane across the $x$-axis and stretching by a factor of 2.

## 16. Exercises

**Exercise 3.1.** Find the vector, parametric, and normal equations of the following lines in $\mathbb{R}^2$

- **(a)** The $x$-axis
- **(b)** The line whose functional equation is $y = -\frac{3}{2}x + 4$
- **(c)** The line whose general equation is $3x - 4y = 24$
- **(d)** The line through $(2, 3)$ and $(-2, 4)$
- **(e)** The line through $(1, 2)$ parallel to the vector $2i - 4j$
- **(f)** The line through $(-1, 1)$ perpendicular to the vector $(8, 1)$

**Exercise 3.2.** Find the vector, parametric, and symmetric equations of the following lines in $\mathbb{R}^3$

- **(a)** The $x$-axis
- **(b)** The line through $(1, 2, -1)$ and $(-1, 0, 1)$
- **(c)** The line through $(1, 2, 3)$ parallel to the vector $3i - 2j + k$
- **(d)** The line through $(-1, 2, -3)$ perpendicular to the plane $x + 2y - 4z = 8$
- **(e)** The intersection of the planes $2x - 3y + z = 3$ and $x + 2y + z = -3$

**Exercise 3.3.** Find the general equation of the following planes in $\mathbb{R}^3$

- **(a)** The plane through $(1, 2, 1)$ normal to the vector $(4, -1, 2)$
- **(b)** The plane through $(1, -2, -1)$ parallel to the plane $2x + 3y - z = 0$
- **(c)** The plane through $(1, 1, -1)$, $(2, 0, 2)$, and $(0, -2, 1)$
- **(d)** The plane through $(-2, 0, 1)$ perpendicular to the line $(5 + t, 2 - 2t, 3 + t)$

**Exercise 3.4.** Find the general equation of the line which is the set of all points in $\mathbb{R}^2$ equidistant between $(5, 9)$ and $(-4, 3)$.

**Exercise 3.5.** Find the general equation of the plane consisting of all points that are equidistant from the two points $(1, 1, 0)$ and $(0, 1, 1)$.

**Exercise 3.6.** Find the general equation of the plane that is the set of all points in $\mathbb{R}^3$ equidistant to $(1, 3, 2)$ and $(2, 0, 1)$.

**Exercise 3.7.** Find the parametric equations of the line in $\mathbb{R}^3$ which is the intersection of the planes $3x + 6y - 2z = 0$ and $x + 2y - z = 4$.

**Exercise 3.8.** Find the distance from the point $(6, -2)$ to the line $4x + y = 12$ in $\mathbb{R}^2$.

**Exercise 3.9.** Find the distance from the point $(4, 1, -3)$ to the plane $2x + 3y - z = 2$ in $\mathbb{R}^3$.

**Exercise 3.10.** Find the distance from the point $(5, 2, 1)$ to the line $(4 + 2t, 1 - t, -3 + 3t)$ in $\mathbb{R}^3$.

**Exercise 3.11.** Find the area of the triangle in $\mathbb{R}^2$ with vertices $(1, 2)$, $(5, -2)$, and $(-3, 5)$.

**Exercise 3.12.** Find the area of the triangle in $\mathbb{R}^3$ with vertices $(1, 2, -1)$, $(4, 3, 2)$, and $(-2, -3, 4)$.

**Exercise 3.13.** Find the volume of the tetrahedron in $\mathbb{R}^3$ with vertices $(1, 0, 2)$, $(0, 4, 1)$, $(-2, 4, 0)$, and $(3, 3, 3)$.

**Exercise 3.14.** Determine the value for $t$ such that $(4, 0, -2)$, $(0, 1, -5)$, $(2, 3, 4)$, and $(5, t, -2)$ are on the same plane.

**Exercise 3.15.** Let $A$ be the plane given by $x + 2y + 3z = 6$ and $B$ be the plane given by $3x + 2y + z = 6$. Let $L = A \cap B$ be the line of intersection of $A$ and $B$. Let $P = (1, 1, 1)$ and note that $P \in L$. Find the equation of the plane which is perpendicular to $L$ and passes through the point $P$, expressed in the form $ax + by + cz = d$.

**Exercise 3.16.** Let $S$ be the solution set of the equation $x^2 + y^2 + z^2 = 16$ Let $P = (4, 0, 0)$. Find the equation of a plane which passes through $P$ and intersects $S$ is a circle of radius $r$ for the following $r$:

**(a)** $r = 4$;
**(b)** $r = 2$;
**(c)** $r = 3$;
**(d)** $r = 1$.

**Exercise 3.17.** In $\mathbb{R}^2$, the set of points equidistant to two points is a line, and in $\mathbb{R}^3$ it is a plane. In $\mathbb{R}^4$, it is a three-dimensional hyperplane. Find the equation of the hyperplane in $\mathbb{R}^4$ which is the set of points equidistant to the points $(1, 2, 3, 0)$ and $(2, 0, -1, 1)$.

**Exercise 3.18.** Let $v_1 = (3, -1)$ and $v_2 = (-2, 5)$. Show that $\text{span}\{v_1, v_2\} = \mathbb{R}^2$ by writing and arbitrary point $(x, y) \in \mathbb{R}^2$ as a linear combination of $v_1$ and $v_2$.

**Exercise 3.19.** Let $A = \{v_1, \ldots, v_r\} \subset \mathbb{R}^n$. Let $w_1, w_2 \in \text{span}(A)$.

**(a)** Let $t \in \mathbb{R}$. Show that $t(w_1 - w_2) + w_1 \in \text{span}(A)$.
**(b)** Conclude that the line through $w_1$ and $w_2$ is contained in $\text{span}(A)$.

**Exercise 3.20.** Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear transformation of the plane. Suppose that $T(e_2) = aT(e_1)$ for some $a \in \mathbb{R}$. Show that the image of $T$ is a line through the origin in $\mathbb{R}^2$.

CHAPTER 4

# Matrices

## 1. Motivation for Matrices

A linear transformation is completely described by its effect on the standard basis. Given a linear transformation, we wish to compress this information (its effect on the standard basis) into as tight a package as we can; we will call this package a matrix.

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Let $x \in \mathbb{R}^n$ be an arbitrary vector, and call $T(x)$ the *destination* of $x$ under $T$. To understand the transformation $T$, we wish to find a formula for $T(x)$. Since $T(x)$ is an element of $\mathbb{R}^m$, it is a linear combination of the standard basis vectors $e_1, \dots, e_m$ in $\mathbb{R}^m$; thus there exist $c_i \in \mathbb{R}$ as $i$ runs from 1 to $m$ such that $T(x) = \sum_{i=1}^m c_i e_i$; we seek a formula for the coefficients $c_i$.

Now $x$ is a linear combination of the standard basis vectors in $\mathbb{R}^n$, so there exist $b_j \in \mathbb{R}$ as $j$ runs from 1 to $n$ such that $x = \sum_{j=1}^n b_j e_j$. Since $T$ is linear, we see that $T(x) = \sum_{j=1}^n b_j T(e_j)$; thus if we know where $T$ sends the standard basis vectors, we entirely understand $T$.

For each standard basis vector $e_j \in \mathbb{R}^n$, $T(e_j) \in \mathbb{R}^m$ so $T(e_j)$ is a linear combination of the standard basis vectors in $\mathbb{R}^m$. Fixing $j$, we see that there are real numbers $a_{ij}$, as $i$ runs from 1 to $m$, such that $T(e_j) = \sum_{i=1}^m a_{ij} e_i$. For our arbitrary vector $v$ we can write

$$T(x) = \sum_{j=1}^n b_j T(e_j) = \sum_{j=1}^n b_j \left( \sum_{i=1}^m a_{ij} e_i \right) = \sum_{i=1}^m \left( \sum_{j=1}^n a_{ij} b_j \right) e_i.$$

The final expression in the above equation reveals that the components of $T(x)$ are given by

$$c_i = \sum_{i=1}^n a_{ij} b_j = (a_{i1}, \dots, a_{in}) \cdot (b_1, \dots b_n);$$

that is, $T(x)$ is the vector whose $i^{\text{th}}$ coordinate is obtained by collecting the $i^{\text{th}}$ coordinates of the destinations of the standard basis vectors into one vector, and dotting that vector with $x$.

Thus $T$ is completely described by the numbers $a_{ij}$, as $i$ runs from 1 to $m$ and $j$ runs from 1 to $n$. These numbers form a mathematical object known as a matrix. The formula for $c_i$ above motivates our definition of matrix multiplication.

## 2. Matrices

Let $m, n$ be positive integers. An $m \times n$ *matrix* with real entries is an array of real numbers with $m$ rows and $n$ columns. We put brackets around the numbers; thus if $A$ is an $m \times n$ matrix, we write

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \dots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix},$$

where $a_{ij}$ is the real number in the $i^{\text{th}}$ row and the $j^{\text{th}}$ column. This can become a lot of writing; we use an abbreviated notation

$$(\text{number})_{\text{slot}}.$$

In our case,

$$A = (a_{ij})_{ij}.$$

This notation means that $a_{ij}$ is in the $ij^{\text{th}}$ slot. You may ask, "why repeat the $ij$"? The reason is, the number in the $ij^{\text{th}}$ slot is not always indexed by $ij$. For example, if $A$ is a $2 \times 3$ matrix written as $A = (2)_{ij}$ and $B$ is a $3 \times 2$ matrix written as $B = (3j - i)_{ij}$, then

$$A = \begin{bmatrix} 2 & 2 & 2 \\ 2 & 2 & 2 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 2 & 5 \\ 1 & 4 \\ 0 & 3 \end{bmatrix}.$$

The *transpose* of an $m \times n$ matrix $A = (a_{ij})_{ij}$ is the $n \times m$ matrix $A^t$ whose rows are the columns of $A$ and whose columns are the rows of $A$:

$$A^t = (a_{ji})_{ij}.$$

Note that $(A^t)^t = A$. An $m \times n$ matrix is called *square* if $m = n$. A matrix $A$ is *symmetric* if $A^t = A$; note that only square matrices can be symmetric.

A *row vector* is an $1 \times n$ matrix, and a *column vector* is a $m \times 1$ matrix. Note that if $v$ is a column vector, then $v^t$ is a row vector. From now on, whenever we need to consider a vector from $\mathbb{R}^n$ as a matrix, we consider to be a column vector.

Let $A$ be an $m \times n$ matrix. Denote the $i^{\text{th}}$ row of $A$ by $A_{(i)}$ and the $j^{\text{th}}$ column of $A$ by $A^{(j)}$. Thus $A_{(i)}$ is a $1 \times n$ row vector and $A^{(j)}$ is an $m \times 1$ column vector.

Let $v_1, \dots, v_n \in \mathbb{R}^m$. We consider these to be column vectors. Let

$$A = [v_1 \mid \cdots \mid v_n]$$

denote the matrix whose $j^{\text{th}}$ column is $v_j$; thus $A^{(j)} = v_j$.

## 3. Matrix Addition and Scalar Multiplication

Let $A = (a_{ij})_{ij}$ and $B = (b_{ij})_{ij}$ be $m \times n$ matrices. We define the *matrix sum* $A + B$ by

$$A + B = (a_{ij} + b_{ij})_{ij}.$$

We can only add matrices of the same size. Note that if $A$ is square, then $A + O = O + A = A$, where $O$ is the zero matrix of the same size.

Let $A = (a_{ij})_{ij}$ be an $m \times n$ matrix and let $c \in \mathbb{R}$. We define the *scalar multiplication* $cA$ by

$$cA = (ca_{ij})_{ij}.$$

We define $-A$ to be the scalar product of $-1$ and $A$.

Note that the sum of column vectors is a column vector, and a scalar multiple of a column vector is a column vector. Indeed, for the case of column vectors, the definitions of matrix addition and scalar multiplication agree with the definitions we previously gave for vectors in $\mathbb{R}^n$.

The *zero matrix* of size $m \times n$, denoted by $Z_{m\times n}$ or simply by $Z$, is the $m \times n$ matrix for which every entry is equal to zero: $Z_{m\times n} = (0)_{ij}$.

**Properties of Matrix Addition and Scalar Multiplication**

Let $A$ and $B$ be $m \times n$ matrices and let $c \in \mathbb{R}$ be a scalar. Then

- **(a)** $A + B = B + A$;
- **(b)** $(A + B) + C = A + (B + C)$;
- **(c)** $A + Z = A$;
- **(d)** $A + (-A) = Z$;
- **(e)** $c(A + B) = cA + cB$.

*Remark.* These properties are proved directly from the definitions. □

## 4. Matrix Multiplication

Let $A = (a_{ij})_{ij}$ be an $m \times n$ matrix and let $B = (b_{jk})_{jk}$ be an $n \times p$ matrix. We define the *matrix product* of $A$ and $B$ to be the $m \times p$ matrix $AB$ given by

$$AB = (c_{ik})_{ik}, \quad \text{where} \quad c_{ik} = \sum_{j=1}^{n} a_{ij} b_{jk}.$$

Viewing the $i^{\text{th}}$ row of $A$ and the $j^{\text{th}}$ column of $B$ as vectors in $\mathbb{R}^n$, we see that

$$c_{ik} = A_{(i)} \cdot B^{(j)}.$$

We have no definition for the product of an $m \times n$ matrix with a $p \times q$ matrix unless $n = p$. If $v, w \in \mathbb{R}^n$ as considered as column vectors, then $v^t w = v \cdot w$.

The *identity matrix* of dimension $n$, denoted by $I_n$ or simply by $I$, is the $n \times n$ matrix whose entries are one along the diagonal and zero everywhere else: $I_n = (\delta_{ij})_{ij}$, where $\delta_{ij}$ is the "Kronecker delta" defined by

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j; \\ 0 & \text{otherwise }. \end{cases}$$

**Properties of Matrix Multiplication** Let $A$ and $C$ be $m \times n$ matrices, $B$ and $D$ be $n \times p$ matrices, and $E$ be a $p \times q$ matrix. Let $c \in \mathbb{R}$ be a scalar. Then

- **(a)** $A(BE) = (AB)E$;
- **(b)** $I_m A = A$;
- **(c)** $AI_n = A$;
- **(d)** $(A + C)B = AB + CB$;
- **(e)** $A(B + D) = AB + AD$;
- **(f)** $c(AB) = A(cB)$;
- **(g)** $(AB)^t = B^t A^t$
- **(h)** $(AB)_{(i)} = A_{(i)} B$;
- **(i)** $(AB)^{(k)} = AB^{(k)}$;
- **(j)** $(AB)_{(i)}^{(k)} = A_{(i)} B^{(k)}$.

*Remark.* These properties may be proved directly from the definitions, although in some cases this could lead to a lot of notation. Of paramount importance to us are properties **(e)** and **(f)**, and we will soon examine them more closely. □

Matrix multiplication is NOT commutative.

Let $x = (x_1, \ldots, x_n)$ be a vector in $\mathbb{R}^n$. We view $x$ as a column vector, that is, as an $n \times 1$ matrix. Thus if $A = (a_{ij})_{ij}$ is an $m \times n$ matrix, the product $Ax$ is defined to be an $m \times 1$ matrix:

$$Ax = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ldots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{bmatrix}.$$

Using the distributive property of scalar multiplication over matrix addition, we see that

$$Ax = x_1 A^{(1)} + \cdots + x_n A^{(n)}.$$

This $m \times 1$ column vector is a linear combination of the columns of $A$.

## 5. Matrices and Linear Transformations

We now consider the geometric interpretation of the product of a matrix and a column vector. First we prove that this operation is linear.

**Proposition 4.1.** *Let $A$ be an $m \times n$ matrix. Then $Ae_j = A^{(j)}$, where $e_j$ is the $j^{\text{th}}$ standard basis vector in $\mathbb{R}^n$.*

*Proof.* Since $e_j = (0, \dots, 1, \dots, 0)$, with 1 in the $j^{\text{th}}$ slot, we have

$$Ae_j = 0 \cdot A^{(1)} + \cdots + 1 \cdot A^{(j)} + \cdots + 0 \cdot A^{(n)} = A^{(j)}.$$

□

**Proposition 4.2.** *Let $A$ be an $m \times n$ matrix. Then*

**(a)** *$A(x + y) = Ax + Ay$ for all $x, y \in \mathbb{R}^n$;*
**(b)** *$A(ax) = a(Ax)$ for all $x \in \mathbb{R}^n$, $a \in \mathbb{R}$.*

*Proof.* Let $x, y \in \mathbb{R}^n$. Then $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_n)$ for some $x_i, y_i \in \mathbb{R}$. By definition of vector addition and matrix multiplication,

$$\begin{aligned} A(x+y) &= (x_1 + y_1)A^{(1)} + \cdots + (x_n + y_n)A^{(n)} \\ &= (x_1 A^{(1)} + \cdots + x_n A^{(n)}) + (y_1 A^{(1)} + \cdots + y_n A^{(n)}) \\ &= Ax + Ay. \end{aligned}$$

Now let $a \in \mathbb{R}$. Then

$$\begin{aligned} A(ax) &= ax_1 A^{(1)} + \cdots + ax_n A^{(n)} \\ &= a(x_1 A^{(1)} + \cdots + x_n A^{(n)}) \\ &= a(Ax). \end{aligned}$$

□

**Proposition 4.3.** *Let $A$ be an $m \times n$ matrix. Define a function*

$$T_A : \mathbb{R}^n \to \mathbb{R}^m \quad \text{by} \quad T_A(x) = Ax.$$

*Then $T$ is a linear transformation.*

*Proof.* This is immediate from the previous proposition. □

**Proposition 4.4.** *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Define a matrix*

$$A_T = [T(e_1) \mid \cdots \mid T(e_n)].$$

*Then*

**(a)** *$T(x) = A_T x$ for all $x \in \mathbb{R}^n$;*
**(b)** *$T_{A_T} = T$;*
**(c)** *$A_{T_A} = A$.*

*Proof.* A linear transformation is completely determined by its effect on the standard basis. The effect of $A_T$ on the standard basis is the same as that of $T$; but $A_T$ induces a linear transformation, so it must be the transformation $T$. □

Thus $m \times n$ matrices correspond to linear transformations from $\mathbb{R}^n$ to $\mathbb{R}^m$. The zero matrix corresponds to the zero transformation (that transformation which sends every element to the origin), and the identity matrix corresponds to the identity transformation (that transformation which sends every element to itself).

We emphasize that the columns of a matrix $A$ are the destinations of the standard basis vectors.

**Example 4.5.** Find the matrix $R_\theta$ of the linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ which rotates the plane by an angle of $\theta$ radians.

*Solution.* We only need to discover what $T$ does to the standard basis vectors. We see that $T(e_1) = (\cos\theta, \sin\theta)$ and $T(e_2) = (-\sin\theta, \cos\theta)$. Then

$$R_\theta = A_T = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

Select a point on the unit circle to test this: Then

$$\begin{aligned} R_\theta \begin{bmatrix} \cos\alpha \\ \sin\alpha \end{bmatrix} &= \begin{bmatrix} \cos\theta\cos\alpha - \sin\theta\sin\alpha \\ \sin\theta\cos\alpha + \cos\theta\sin\alpha \end{bmatrix} \\ &= \begin{bmatrix} \cos(\theta+\alpha) \\ \sin(\theta+\alpha) \end{bmatrix}; \end{aligned}$$

this is what we would expect. □

**Example 4.6.** Find the matrix $F_\theta$ of the linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ which reflects the plane across a line whose angle with the $x$-axis is $\theta$.

*Solution.* We see that

$$T(e_1) = (\cos 2\theta, \sin 2\theta)$$

and that

$$T(e_2) = -(\cos(2\theta + \frac{\pi}{2}), \sin(2\theta + \frac{\pi}{2}) = (\sin 2\theta, -\cos 2\theta).$$

Thus

$$F_\theta = \begin{bmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{bmatrix}.$$

□

## 6. Matrices and Compositions of Linear Transformations

We now consider the geometric interpretation of matrix multiplication. Recall that if $T : \mathbb{R}^n \to \mathbb{R}^m$ and $S : \mathbb{R}_p \to \mathbb{R}_n$ are linear transformations, then the composition $T \circ S : \mathbb{R}^p \to \mathbb{R}^m$ given by $T \circ S(x) = T(S(x))$ is a linear transformation.

**Proposition 4.7.** *Let $A$ be an $m \times n$ matrix and let $B$ be an $n \times p$ matrix. Then*

$$T_{AB} = T_A \circ T_B.$$

*Proof.* This means that the transformation associated to a product of matrices is the composition of the associated transformations. To show this, we only need to show that these transformations have the same effect on an arbitrary basis vector $e_k \in \mathbb{R}^p$.

The $k^{\text{th}}$ column of $AB$ is equal to $A$ times the $k^{\text{th}}$ column of $B$, and we have seen that multiplying a matrix by $e_k$ picks out the $k^{\text{th}}$ column, so

$$T_{AB}(e_k) = (AB)^{(k)} = AB^{(k)}.$$

On the other hand,

$$T_A \circ T_B(e_k) = T_A(B^{(k)}) = AB^{(k)}.$$

Thus these transformations have the same effect on $e_k$, and we conclude that they are the same transformation. □

**Proposition 4.8.** *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ and $S : \mathbb{R}_p \to \mathbb{R}_n$ be linear transformations. Then*

$$A_{T \circ S} = A_T A_S.$$

*Proof.* This means that the matrix associated to a composition of transformations is the product of the associated matrices. It suffices to show that the $k^{\text{th}}$ column of $A_{T \circ S}$ is the same as the $k^{\text{th}}$ column of $A_T A_S$. But

$$A_{T \circ S}^{(k)} = T \circ S(e_k) = T(S(e_k))$$

and

$$A_T A_S^{(k)} = A_T A_S e_k = A_T S(e_k) = T(S(e_k)).$$

□

**Example 4.9.** Let $S : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation which stretches the plane horizontally by a factor of 2, and let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation which rotates the plane by 90 degrees (all rotations are counterclockwise). Then

$$A = A_S = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad B = A_T = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

Note that intuitively we see that $T \circ S$ and $S \circ T$ have different effects on the plane. Indeed,

$$BA = \begin{bmatrix} 0 & -1 \\ 2 & 0 \end{bmatrix} \quad \text{but} \quad AB = \begin{bmatrix} 0 & -2 \\ 1 & 0 \end{bmatrix}.$$

We see that matrix multiplication is NOT commutative.

**Example 4.10.** Show that the composition of rotations is a rotation whose angle is the sum of the original angles.

*Solution.* We compute with matrices:

$$\begin{aligned} R_\alpha R_\beta &= \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} \begin{bmatrix} \cos\beta & -\sin\beta \\ \sin\beta & \cos\beta \end{bmatrix} \\ &= \begin{bmatrix} \cos\alpha\cos\beta - \sin\alpha\sin\beta & -\cos\alpha\sin\beta - \sin\alpha\cos\beta \\ \sin\alpha\cos\beta + \cos\alpha\sin\beta & \sin\alpha\sin\beta + \cos\alpha\cos\beta \end{bmatrix} \\ &= \begin{bmatrix} \cos(\alpha+\beta) & -\sin(\alpha+\beta) \\ \sin(\alpha+\beta) & \cos(\alpha+\beta) \end{bmatrix}. \end{aligned}$$

□

## 7. Matrices and Invertible Linear Transformations

Recall that the identity transformation $J_n : \mathbb{R}^n \to \mathbb{R}^n$ is the function that has no effect on $\mathbb{R}^n$; it is given by $J_n(v) = v$. Since the identity matrix $I_n$ has no effect on the standard basis (viewed as column vectors), we see that

$$A_{J_n} = I_n \quad \text{and} \quad T_{I_n} = J_n.$$

Recall that a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is *invertible* if it is bijective, in which case there is an inverse function $T^{-1} : \mathbb{R}^m \to \mathbb{R}^n$ such that $T^{-1} \circ T = J_n$.

Suppose that $T : \mathbb{R}^n \to \mathbb{R}^m$ is an invertible linear transformation. We will see that this implies $m = n$; for now, just assume $m = n$. Then the matrix corresponding to $T$ is an $n \times n$ matrix; that is, it is square.

Suppose that $T$ is invertible; then $T \circ T^{-1} = T^{-1} \circ T = \mathrm{id}_{R^n}$. Then $A_T A_{T^{-1}} = A_{T^{-1}} A_T = A_{\mathrm{id}} = I$.

A matrix $A$ is called *invertible* if there exists a matrix $B$ such that $AB = BA = I$. We see that two matrices are invertible if and only if the corresponding linear transformations are bijective. The matrix $B$ is called the *inverse* of $A$, and is denoted by $A^{-1}$. Note that $A^{-1}$ is invertible (with inverse $A$).

**Properties of Matrix Inverses**

Let $A$, $B$, $C$, and $D$ be square matrices of the same size.

**(a)** Inverses are unique.
**(b)** If $A$ and $B$ are invertible, then so is $AB$, with $(AB)^{-1} = B^{-1}A^{-1}$.
**(c)** If $AC = DA = I$, then $C = D$.
**(d)** If $AB = I$, then $BA = I$, so $A$ and $B$ are invertible.

Proof of **(c)** goes as follows. If $AC = I$ and $DA = I$, then $DAC = DI = D$ and $DAC = UC = C$. Thus $C = DAC = D$.

Proof of **(d)** is postponed for now.

Now if $T : \mathbb{R}^n \to \mathbb{R}^n$ is bijective, then for every $b \in \mathbb{R}^n$ there exists a unique $x \in \mathbb{R}^n$ such that $T(x) = b$; indeed, we have $x = T^{-1}(b)$. In matrix form, this says that the matrix equation $Ax = b$ has a unique solution, given by $x = A^{-1}b$.

We would like a method to find $A^{-1}$. The idea is to "dissolve" $A$ by multiplying both sides of the equation $AX = I$ with invertible matrices: $E_1AX = E_1I = E_1$, then $E_2E_1AX = E_2E_1$, et cetera, at each step getting closer to the identity (e.g. $E_2E_1A$ looks more like the identity than $E_1A$), until finally we obtain $E_n \cdots E_1AX = E_n \cdots E_1$, where $E_n \cdots E_1A = I$, so $X = E_n \cdots E_1$. Now $X$ is the product of invertible matrices, so it is invertible, and it is the inverse of $A$ since $AX = I$.

## 8. Elementary Row Operations and Elementary Invertible Matrices

The invertible matrices $E_i$ mentioned above are called "elementary"; they correspond to *elementary row operations*. A row operation is a way of modifying a row of a matrix to change it into a different matrix. Tradition demands that we list three *elementary row operations*:

| | | |
|---|---|---|
| $\mathsf{R}_i + c\mathsf{R}_j$ | Type **E** | Multiply $j^{\text{th}}$ row by $c$ and add to $i^{\text{th}}$ row |
| $c\mathsf{R}_i$ | Type **D** | Multiply $i^{\text{th}}$ row by $c$ |
| $\mathsf{R}_i \leftrightarrow \mathsf{R}_j$ | Type **P** | Swap the $i^{\text{th}}$ row and the $j^{\text{th}}$ row |

For each of these three row operations, there is an invertible matrix $E$ such that $EA$ is the result of the row operation applied to $A$. To find $E$, just perform the row operation on the identity matrix.

$E(i,j;c)$ is $I$ except $a_{ij} = c$; $E(i,j;c)^{-1} = E(i,j;-c)$.

$D(i;c)$ is $I$ except $a_{ii} = c$; $D(i;c)^{-1} = D(i;c^{-1})$.

$P(i,j)$ is $I$ except $a_{ii} = a_{jj} = 0$ and $a_{ij} = a_{ji} = 1$; $P(i,j)^{-1} = P(i,j)$.

We give an organized algorithm for applying row operations to attempt to find the inverse of a matrix.

Algorithm for Row Reduction of a Square Matrix to Find an Inverse

- Make all entries below the diagonal into zero, starting with the second entry in the first column, proceeding downward, then doing the third column, etc.
- Make all diagonal entries equal to one.
- Make all entries above the diagonal zero, starting with the lowest entry in the last column, working upward in that column, then starting on the next to last column, etc.

Step one is always possible; it may be necessary to swap some rows to do this. Use only type **E** and **P** row operations.

Step two is possible if all diagonal entries are nonzero via use of type **D** row operations; otherwise, the matrix is not invertible. To see this, let $Q$ be the matrix obtained after step one, and suppose that $Q^{(i)}_{(i)} = 0$ is the first zero diagonal entry. Then all entries in column $i$ below the diagonal are also zero, so $Q^{(i)}$ is a linear combination of the previous columns (to see this may take some effort, but it is true); say $Q^{(i)} = a_1 Q^{(1)} + \cdots + a_{i-1} Q^{(i-1)}$. Then $a_1 e_1 + \cdots + a_{i-1} Q^{(i-1)} - e_i$ is in the kernel of $T_A$, so $T_A$ is not injective, and $A$ is not invertible.

Step three is possible whenever step two is possible. Use only type **E** row operations.

Thus every invertible matrix is a product of elementary invertible matrices. To see this, let $A$ be invertible and suppose that $X$ is its inverse. Then $AX = I$. Following the above algorithm, we obtain elementary invertible matrices $E_1, \ldots, E_r$ such that

$$X = E_r \cdots E_1 AX = E_r \cdots E_1 I = E_r \cdots E_1.$$

## 9. Introduction to Linear Equations

Consider the system of linear equations

$$3x_1 - 4x_2 = 11;$$
$$x_1 + 2x_2 = 7.$$

Solving this system means finding $x_1$ and $x_2$ which make the equations true.

The loci of the equations $3x_1 - 4x_2 = 11$ and $x_1 + 2x_2 = 7$ are lines in $\mathbb{R}^2$ (we have replaced the standard $x$ and $y$ by $x_1$ and $x_2$ because we want to use the variables $x$ and $y$ to indicate vectors). So we interpret this problem as finding the intersection of two lines.

A second geometric interpretation of the problem comes from forming the matrix of coefficients and the column vectors

$$A = \begin{bmatrix} 3 & -4 \\ 1 & 2 \end{bmatrix}, \quad x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix},$$

and considering the matrix equation $Ax = b$. Since $A$ corresponds to a linear transformation, solving the system of equations is equivalent to finding the preimage of the point $b$ under this linear transformation.

To solve this system, we can multiply the second equation by 2 and add it to the first to get $5x_1 = 25$, so $x_1 = 5$; then plug this into the second equation to get $5 + 2x_2 = 7$, so $x_2 = 1$.

Generalizing this solution technique to many equations in many unknowns could lead to a lot of confusion and difficulty without a more organized approach. We now search for a failsafe algorithm for finding the solution.

## 10. Linear Equations

A *linear equation* in $n$ variables $x_1, \dots, x_n$ is an equation of the form

$$a_1x_1 + \cdots + a_nx_n = b_1,$$

where $a_1, \dots, a_n, b_1 \in \mathbb{R}$ are fixed constants.

Let $a = (a_1, \dots, a_n)$, $x = (x_1, \dots, x_n)$, and $q = (\frac{b_1}{a_1}, 0, \dots, 0)$. The above equation becomes $a \cdot x = a \cdot q$, or

$$(x - q) \cdot a = 0.$$

We recognize this as the equation of a hyperplane in $\mathbb{R}^n$ through the point $q$ with normal vector $a$.

Consider an arbitrary system of linear equations

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= b_1 \\ \vdots &= \vdots \\ a_{i1}x_1 + \cdots + a_{in}x_n &= b_i \\ \vdots &= \vdots \\ a_{1n}x_1 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

where $a_{ij}, b_i \in \mathbb{R}$ are constants and $x_i$ are indeterminates.

Our goal is to use invertible matrices to help us solve such systems of linear equation; that is, we wish to find all vectors $x \in \mathbb{R}^n$ such that when we plug their coordinates into the equations, all of the resulting equations are true.

One geometric interpretation of this problem is to find the intersection of the hyperplanes in $\mathbb{R}^n$ which are the loci of the given equations.

A second geometric interpretation comes from forming the matrix $A = (a_{ij})_{ij}$. Then setting $x = (x_1, \ldots, x_n)$ and $b = (b_1, \ldots, b_m)$, we see that the solution set of the matrix equation $Ax = b$ is exactly the solution set of the system of equations. This matrix equation, stated in terms of linear transformations, is $T_A(x) = b$; solving means finding $T_A^{-1}(b)$, the preimage of the point $b$ under the linear transformation $T_A$.

Our approach to the problem uses matrices; we seek column vectors $x$ such that $Ax = b$.

A *general solution* is the set of all such column vectors $x$.

A *particular solution* is a specific such column vector $x$.

The system is called *homogeneous* if $b_i = 0$ for $i = 1, \ldots, m$. In this case, solving the system of equations means finding the kernel of $T_A$. Otherwise, the system is *nonhomogeneous.*

We have seen that if $b$ is in the image of $T_A$, say $T_A(v) = b$ for some $v \in \mathbb{R}^n$, then $T^{-1}(b) = v + \ker(T_A)$. If $T_A$ is injective, then $\ker(T_A)$ consists of a single point (the origin), so $T^{-1}(b) = \{v\}$. Otherwise, $\ker(T_A)$ is a nontrivial subspace, so $v + \ker(T_A)$ at least one line, and possibly a plane or more. Thus there are three possibilities:

(1) there are no solutions ($b$ is not in the image of $T_A$);
(2) there is exactly one solution ($T_A$ is injective);
(3) there are infinitely many solutions ($T_A$ has a nontrivial kernel).

If we have infinitely many solutions, they are of the form

$$v_0 + c_1 v_1 + \cdots + c_k v_k,$$

where $v_0, \ldots, v_k$ are vectors which span the kernel of $T_A$, $c_1, \ldots, c_k$ are free scalars, $v_0$ is a particular solution to $Ax = b$, and $c_1 v_1 + \cdots + c_k v_k$ is the general solution to the homogeneous equation $Ax = 0$ (the kernel of $T_A$).

Suppose that there exists an invertible matrix $E$ such that the matrix $EA$ has a particularly nice form. Then $Ax = b \Rightarrow EAx = Eb$; since $E$ is invertible, we have $EAx = Eb \Rightarrow E^{-1}EAx = E^{-1}Eb \Rightarrow Ax = b$. Thus the solution set of $Ax = b$ is exactly the solution set of $EAx = Eb$, so is suffices to find the solution set of $EAx = Eb$.

The nice form we refer to here is known as reduced row echelon form.

## 11. Reduced Row Echelon Form

A matrix is said to be in *row echelon form* if

**i.** All zero rows lie below nonzero rows;
**ii.** The first nonzero entry in any row appears in a column to the right of the first nonzero entry in any preceding row.

The first nonzero entry in a row is called a *pivot*.

Given a matrix $A$, there is a sequence of row operations which brings $A$ into row echelon form. The final product is not unique.

A matrix is said to be in *reduced row echelon form* if

**i.** It is in row echelon form;
**ii.** All the pivots equal 1;
**iii.** All nonpivot entries in a column containing a pivot are equal to 0.

Given a matrix $A$, there is a sequence of row operations which brings $A$ into row echelon form. Although the sequence of row operations is not unique, the final product is unique.

*Gaussian elimination* is an algorithm for using elementary row operations to bring a matrix into reduced row echelon form. There are two stages: *forward elimination* brings the matrix into row echelon form, and *backward elimination* brings the row echelon matrix into reduced row echelon form.

**Forward elimination:**

(1) Start with the first column, and proceed through all columns in order.
(2) If the diagonal entry in the column is zero, permute with the first available lower row so that the diagonal entry is nonzero (use $P$).
(3) Eliminate all entries below this one in order (use $E$).

Note that forward elimination does not use $D$. Also note that this algorithm is so specific, the sequence of elementary matrices and the row echelon form obtained is unique.

**Backward elimination:**

(1) Make all pivots equal to one (use $D$).
(2) Starting from the right, working upward then leftward, make all entries above a pivot equal to zero (use $E$).

To solve a system of linear equations $Ax = b$, form the augmented matrix $[A \mid b]$ and work on $A$ and $b$ simultaneously. Perform forward elimination and backward elimination on $A$, and then read off the solution. We describe this last step momentarily.

Once the matrix is in reduced row echelon form, it is easy to read off the general solution. We give an example, then list the exact steps to take.

**Example 4.11.** Consider the matrix equation

$$\begin{bmatrix} 1 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$$

Computing the matrix product on the left gives

$$\begin{bmatrix} x_1 + 2x_2 \\ x_3 \\ x_4 + 5x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$$

The solution set of this equation is a subset of $\mathbb{R}^6$, so we actually seek six dimensional vectors. Insert the free variables into the equation in an appropriate fashion to arrive at

$$\begin{bmatrix} x_1 + 2x_2 \\ x_2 \\ x_3 \\ x_4 + 5x_5 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} 1 \\ x_2 \\ 2 \\ 3 \\ x_5 \\ 4 \end{bmatrix}$$

By the definition of vector addition, this is the same as

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} + x_2 \begin{bmatrix} 2 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} 0 \\ 0 \\ 0 \\ 5 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 2 \\ 3 \\ 0 \\ 4 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

Subtract the free columns from both sides and use the distributive property to obtain

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 2 \\ 3 \\ 0 \\ 4 \end{bmatrix} + x_2 \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} 0 \\ 0 \\ 0 \\ -5 \\ 1 \\ 0 \end{bmatrix}$$

## 12. Solution Method

Let $A = [a_{ij}]_{ij}$ be an $m \times n$ matrix, $x = [x_1, \dots, x_n]^t$ an $n$-dimensional variable column vector, and $b = [b_1, \dots, b_m]$ an $m$-dimensional constant column vector. Then the solution set of the matrix equation $Ax = b$ is the solution set of a system of linear equations.

Let $O$ be the product of all the elementary matrices whose corresponding row operations bring the matrix $A$ into row echelon form, that is, those used in forward elimination. Set $Q = OA$, where $Q$ is in row echelon form. Let $c = Ob$. Then the solution set of $Ax = b$ is equal to the solution set of $OAx = Ob$, i.e., $Qx = c$.

At this point, we can tell if there is no solution: this happens when the a row of the nonaugmented matrix contains only zeros, but the corresponding entry of the augmentation column is nonzero. We can also tell if the solution is unique: this happens when the number of nonzero rows equals the number of columns.

Let $U$ be the product of all the elementary matrices whose corresponding row operations bring the matrix $A$ into reduced row echelon form; that is, $R = UA$, where $R$ is in reduced row echelon form. Let $d = Ub$. Then the solution set of $Ax = b$ is equal to the solution set of $UAx = Ub$, i.e., $Rx = d$. We describe how to read off the general solution from the matrix equation $Rx = d$.

We say that $R^{(j)}$ is a *basic column* if $R^{(j)}$ (or $Q^{(j)}$) contains a pivot; otherwise $R^{(j)}$ is a *free column*.

We say that $x_j$ is a *basic variable* if $A^{(j)}$ contains a pivot; otherwise $x_j$ is a *free variable*.

The general solution will be of the form

$$v_0 + c_1 v_1 + \cdots + c_k v_k,$$

where $k$ is the number of free variables; we have $k = n - r$, where $r$ is the number of nonzero rows.

The vector $v_0$ is the particular solution obtained by setting the free variables equal to 0 and solving for the basic variables.

The vectors $v_i$ are found by replacing $d$ by the zero vector, setting the $i^{\text{th}}$ free variable equal to 1 and the other free variables equal to 0, and solving for the basic variables.

We can read off the general solution from the reduced matrix as follows:

(1) eliminate any zero rows at the bottom of the reduced matrix;
(2) insert a zero row at row $i$ for every free variable $x_i$;
(3) multiply each free column by $-1$;
(4) add $e_i$ to each free column for every free variable $x_i$;
(5) the particular solution is now the augmentation column;
(6) the homogeneous solution is now the span of the adjusted free columns.

## 13. Geometric Interpretation of Systems of Linear Equations

We have two geometric interpretations for solving a system of linear equations: as the intersection of the loci of the equations, and as the preimage of a linear transformation. How do these geometric interpretations correspond?

Reconsider the system of linear equations

$$3x_1 - 4x_2 = 11;$$
$$x_1 + 2x_2 = 7.$$

There are two ways of viewing this problem geometrically:

We want to find a point $(x_1, x_2)$ which satisfies both equations, that is, which lies on both lines. This is an AND condition, and AND corresponds to the set operation of intersection (just as OR corresponds to the set operation of union); so we intersect the lines (which are, after all, subsets of $\mathbb{R}^2$) and find that the only point of intersection is $(11, 7)$.

The second geometric interpretation comes from putting the coefficients on the left hand side of the system of equations into a matrix $A$, the indeterminates into a column vector $x$ and the values on the left hand side into a column vector $b$:

$$A = \begin{bmatrix} 3 & -4 \\ 1 & 2 \end{bmatrix}; \quad x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}; \quad \text{and } b = \begin{bmatrix} 11 \\ 7 \end{bmatrix}.$$

We see that solving the system of equations is equivalent to solving the matrix equation

$$Ax = b.$$

But $A$ corresponds to a linear transformation $T_A$; thus we seek the preimage of $b$ under the linear transformation $T_A$.

How do these two geometric interpretations coincide?

## 14. Component Functions

Let $f : \mathbb{R}^n \to \mathbb{R}^m$. For each $i \in \{1, \ldots, m\}$, define a function

$$f_i : \mathbb{R}^n \to \mathbb{R} \quad \text{by } f_i(v) = \operatorname{proj}_{e_i} f(v);$$

this is called the $i^{\text{th}}$ *component function* of $f$.

**Example 4.12.** Let $f : \mathbb{R} \to \mathbb{R}^2$ be given by $f(t) = (\cos t, \sin t)$. Then $f_1 = \cos$ and $f_2 = \sin$. Note that the image of the function $f$ is a circle in $\mathbb{R}^2$.

As this example shows, we may turn our definition around; that is, given $m$ functions $f_1, \ldots, f_m : \mathbb{R}^n \to \mathbb{R}$, we construct a function $f : \mathbb{R}^n \to \mathbb{R}^m$ by defining $f(v) = (f_1(v), \ldots, f_m(v))$.

Now let $f_1 : \mathbb{R}^2 \to \mathbb{R}$ be given by $f_1(x, y) = 3x - 4y$ and let $f_2 : \mathbb{R}^2 \to \mathbb{R}$ be given by $f_2(x, y) = x + 2y$. The line in $\mathbb{R}^2$ which is the locus the equation $3x_1 - 4x_2 = 11$ is the preimage of 11 under the function $f_1$; the second line is the preimage of 7 under $f_2$. A solution $(x_1, x_2)$ for the system of equations is an element of the set $f_1^{-1}(11) \cap f_2^{-1}(7)$.

Define $f : \mathbb{R}^2 \to \mathbb{R}^2$ by $f(x) = (f_1(x), f_2(x))$; that is, $f(x_1, x_2) = (3x_1 - 4x_2, x_1 + 2x_2)$. Then the solution to the system of linear equations we started out with is the preimage of the point $(11, 7)$ under this new function; that is, we wish to find $v$ such that $f(v) = (11, 7)$, which is the same as saying that we wish to discover the set $f^{-1}(11, 7) = f_1^{-1}(11) \cap f_2^{-1}(7)$.

By a previous proposition, we see that the function $f$ is a linear transformation; let us relabel it by $T$.

Solving the system of equations is equivalent to finding the preimage of the point $(11, 7)$ under the linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ given by $T(x_1, x_2) = (3x_1 - 4x_2, x_1 + 2x_2)$. What is the effect of $T$ on the standard basis, and what is the matrix associated to $T$?

We have $T(e_1) = T(1, 0) = (3, 1)$ and $T(e_2) = T(0, 1) = (-4, 2)$. Thus the matrix which corresponds to $T$ is

$$A = \begin{bmatrix} 3 & -4 \\ 1 & 2 \end{bmatrix},$$

and finding the preimage of $(11, 7)$ under $T$ is equivalent to solving the matrix equation

$$Ax = b, \quad \text{where } x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \text{ and } b = \begin{bmatrix} 11 \\ 7 \end{bmatrix}.$$

In general, we have $m$ equations in $n$ unknowns. We obtain an $m \times n$ matrix $A$ of coefficients, an $n \times 1$ column vector $b$ of indeterminates, and an $m \times 1$ column vector of values. Solving the system is equivalent to solving the matrix equation $Ax = b$. The associated transformation $T = T_A$ is obtained by creating $m$ linear functions $T_i : \mathbb{R}^n \to \mathbb{R}$ given by the left hand sides of our equations; these become the component functions of $T : \mathbb{R}^n \to \mathbb{R}^m$. The preimage of each $T_i$ at $b_i$ is a hyperplane in $\mathbb{R}^n$. The solution set is the intersection of the hyperplanes, which is the same as the preimage of the point $b$ under the linear transformation $T$.

We may also view this as follows. Let $f : A \to B$ be any function, and let $C, D \subset B$. Then $f^{-1}(C \cap D) = f^{-1}(C) \cap f^{-1}(D)$.

Let $H_i$ be the hyperplane in $\mathbb{R}^m$ (the range of $T$) given by

$$H_i = \{(y_1, \ldots, y_m) \in \mathbb{R}^m \mid y_i = b_i\}.$$

Then

$$\{b\} = \cap_{i=1}^m H_i.$$

Let $L_i$ be the hyperplane in $R^n$ (the domain of $T$) which is the locus of the equation

$$a_{i1}x_i + \cdots + a_{in}x_n = b_i.$$

Then if $X$ is the solution set to our system of linear equations, we have

$$X = \cap_{i=1}^m L_i.$$

But $L_i = T^{-1}(H_i)$, and

$$X = \cap_{i=1}^m T^{-1}(H_i) = T^{-1}(\cap_{i=1}^m H_i) = T^{-1}(b).$$

## 15. Geometric Interpretation of the Solution Process

We have the matrix equation $Ax = b$, where $A$ is an $m \times n$ matrix. We know that $A$ corresponds to a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$. The columns of $A$ are the destinations of the standard basis vectors in $\mathbb{R}^n$ under the transformation $T$. We ask if $b$ is a linear combination of these destinations, in which case there is a solution to the equation.

When row reducing the augmented matrix $[A \mid b]$, we are in theory multiplying both sides of the equation $Ax = b$ by elementary invertible $m \times m$ matrices. Each such multiplication corresponds to an invertible linear transformation of $\mathbb{R}^m$, which is the range space of the linear transformation $T$. What in fact we are doing is transmuting $\mathbb{R}^m$ so that the *labeling* of the destinations of the standard basis vectors is more to our liking; in the process, $b$ is also moved to a new location. That is, we are relabeling the points in $\mathbb{R}^m$ so that we can see more clearly the manner in which $b$ is a linear combination of the destinations of the standard basis vectors.

## 16. Review and Example

*Gaussian elimination* is an algorithm for using elementary row operations to bring a matrix into reduced row echelon form. There are two stages: *forward elimination* brings the matrix into row echelon form, and *backward elimination* brings the row echelon matrix into reduced row echelon form. The algorithm is so specific that each stage is completely determined.

To solve a system of linear equations $Ax = b$, form the augmented matrix $[A \mid b]$ and work $A$ and $b$ simultaneously. Perform forward elimination and backward elimination, and then read off the solution.

Consider the system of linear equations

$$\begin{aligned} x_1 + 2x_2 + 2x_3 &= -7 \\ 3x_1 + 6x_2 &= 9 \\ -2x_1 - 4x_2 - x_3 &= -1 \end{aligned}$$

Let $A$ be the matrix of coefficients, $b$ be the column vector of values, and $x$ be the column vector of variables. The matrix equation $Ax = b$ is

$$\begin{bmatrix} 1 & 2 & 2 \\ 3 & 6 & 0 \\ -2 & -4 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -7 \\ 9 \\ -1 \end{bmatrix}$$

In augmented form, we write this as

$$\begin{bmatrix} 1 & 2 & 2 & | & -7 \\ 3 & 6 & 0 & | & 9 \\ -2 & -4 & -1 & | & -1 \end{bmatrix}.$$

**Forward elimination:**

(1) Start with the first column, and proceed through all columns in order.
(2) If the diagonal entry in the column is zero, permute with the first available lower row so that the diagonal entry is nonzero (use $P : \mathsf{R}_i \leftrightarrow \mathsf{R}_j$).
(3) Eliminate all entries below this one in order (use $E : \mathsf{R}_i + c\mathsf{R}_j$).

In our example, the row operations $\mathsf{R}_2 - 3\mathsf{R}_1$ and $\mathsf{R}_3 + 2\mathsf{R}_1$ produce

$$\begin{bmatrix} 1 & 2 & 2 & | & -7 \\ 0 & 0 & -6 & | & 30 \\ 0 & 0 & 3 & | & -15 \end{bmatrix}.$$

This completes the requirements for column one. Now do column two via $\mathsf{R}_3 + \frac{1}{2}\mathsf{R}-2$ to obtain

$$\begin{bmatrix} 1 & 2 & 2 & | & -7 \\ 0 & 0 & -6 & | & 30 \\ 0 & 0 & 0 & | & 0 \end{bmatrix};$$

this is in row echelon form, so this is the end of forward elimination. The row of zeros on the bottom tells us that the general solution contains infinitely many solutions.

**Backward elimination:**

(1) Make all pivots equal to one (use $D : c\mathsf{R}_i$).
(2) Starting from the right, working upward then leftward, make all entries above a pivot equal to zero (use $E : \mathsf{R}_i + c\mathsf{R}_j$).

Apply row operation $-\frac{1}{6}\mathsf{R}_2$ to obtain

$$\begin{bmatrix} 1 & 2 & 2 & | & -7 \\ 0 & 0 & 1 & | & -5 \\ 0 & 0 & 0 & | & 0 \end{bmatrix}.$$

Now all of the pivots are equal to one. Finally, row operation $\mathsf{R}_1 - 2\mathsf{R}_2$ produces

$$\begin{bmatrix} 1 & 2 & 0 & | & 3 \\ 0 & 0 & 1 & | & -5 \\ 0 & 0 & 0 & | & 0 \end{bmatrix}.$$

This is reduced row echelon form.

**Solution readoff:**

(1) eliminate any zero rows at the bottom of the reduced matrix;
(2) insert a zero row at row $i$ for every free variable $x_i$;
(3) multiply each free column by $-1$;
(4) add $e_i$ to each free column for every free variable $x_i$;
(5) the particular solution is now the augmentation column;
(6) the homogeneous solution is now the span of the adjusted free columns.

The basic variables are $x_1$ and $x_3$ and the free variable is $x_2$.

Step (1):

$$\begin{bmatrix} 1 & 2 & 0 & | & 3 \\ 0 & 0 & 1 & | & -5 \end{bmatrix}.$$

Step (2):

$$\begin{bmatrix} 1 & 2 & 0 & | & 3 \\ 0 & 0 & 0 & | & 0 \\ 0 & 0 & 1 & | & -5 \end{bmatrix}.$$

Step (3):

$$\begin{bmatrix} 1 & -2 & 0 & | & 3 \\ 0 & 0 & 0 & | & 0 \\ 0 & 0 & 1 & | & -5 \end{bmatrix}.$$

Step (4):

$$\begin{bmatrix} 1 & -2 & 0 & | & 3 \\ 0 & 1 & 0 & | & 0 \\ 0 & 0 & 1 & | & -5 \end{bmatrix}.$$

Step (5): A particular solution is $(3, 0, -5)$

Step (6): The homogeneous solution is the subspace of $\mathbb{R}^3$ spanned by the vector $(-2, 1, 0)$.

The general solution is

$$x = \begin{bmatrix} 3 \\ 0 \\ -5 \end{bmatrix} + x_2 \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}.$$

**How this works**

Row operations on matrices correspond to operations performed on the system of linear equations which produce equivalent systems; that is, systems with the same solution set. Therefore the original system

$$\begin{aligned} x_1 + 2x_2 + 2x_3 &= -7 \\ 3x_1 + 6x_2 &= 9 \\ -2x_1 - 4x_2 - x_3 &= -1 \end{aligned}$$

has the same solution set as the system given by the reduced row echelon form, which is

$$x_1 - 2x_2 = 3x_3 = -50 = 0$$

Now solution readoff proceeds as follows.

Step (1): the last equation contains no information, so we eliminate it.

Step (2) and Step (4): Insert an equation $x_2 = x_2$; this is certainly true.

Step (3): Solve equation $i$ for variable $x_i$. We subtract $-2x_2$ from both sides of equation one; this corresponds to multiplying a free column by $-1$.

$$\begin{aligned} x_1 &= 3 + 2x_2 \\ x_2 &= 0 + 1x_2 \\ x_3 &= -5 + 0x_2 \end{aligned}$$

In matrix form, this is

$$x = \begin{bmatrix} 3 \\ 0 \\ -5 \end{bmatrix} + x_2 \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}.$$

## 17. Exercises

**Exercise 4.1.** Let $d, e \in \mathbb{R}$ and set

$$A = \begin{bmatrix} 2 & 1 & 2 \\ 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix}, \quad B = \begin{bmatrix} 2 & 2 \\ 1 & 2 \\ 0 & 4 \end{bmatrix}, \quad E = \begin{bmatrix} 1 & 0 & 0 \\ e & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & d \end{bmatrix}.$$

Compute
**(a)** $AB$;
**(b)** $DEB$;
**(c)** $AD$;
**(d)** $A^{-1}$.

**Exercise 4.2.** Let

$$A = \begin{bmatrix} 3 & -2 \\ -1 & 4 \end{bmatrix}.$$

- **(a)** Find the inverse of $A$.
- **(b)** Use **(a)** to solve the system of linear equations

$$\begin{aligned} 3x_1 - 2x_2 &= 3 \\ -2x_1 + 4x_2 &= 5 \end{aligned}$$

**Exercise 4.3.** Find the general solution to the system of linear equations

$$\begin{aligned} x_1 + x_2 + 2x_4 + x_5 &= 4 \\ 2x_1 + x_2 - x_3 + x_5 &= 5 \\ 4x_1 + 3x_2 - x_3 + 4x_4 + 4x_5 &= 13 \end{aligned}$$

**Exercise 4.4.** Consider the system of linear equations

$$\begin{aligned} 2x_1 + 2x_2 + x_3 &= 3 \\ 3x_3 &= -7 \\ 5x_2 &= 2 \end{aligned}$$

- **(a)** Form the matrix $A$ of coefficients, the column vector $b$ of values, and the column vector $x$ of variables.
- **(b)** Find the matrix $A^{-1}$.
- **(c)** Use **(b)** to solve the matrix equation $Ax = b$.

**Exercise 4.5.** Consider the system of linear equations

$$\begin{aligned} x_1 - 2x_2 + x_3 &= 2 \\ -2x_1 + 4x_2 - x_3 &= -5 \\ x_1 - 2x_2 + 2x_3 &= 1 \end{aligned}$$

Let $A$ be the matrix of coefficients, $b$ the column vector of values, and $x$ the column vector of variables. Use Gaussian elimination (forward elimination, backward elimination, and solution readoff) to find the general solution to this system.

**Exercise 4.6.** Consider the system of linear equations

$$\begin{aligned} x_1 - 2x_2 + x_3 &= 2 \\ -2x_1 + 4x_2 - x_3 &= -5 \\ x_1 - 2x_2 + 2x_3 &= 1 \end{aligned}$$

Let $A$ be the matrix of coefficients, $b$ the column vector of values, and $x$ the column vector of variables. Use Gaussian elimination (forward elimination, backward elimination, and solution readoff) to find the general solution to this system.

**Exercise 4.7.** Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be the linear transformation whose corresponding matrix is

$$A = \begin{bmatrix} 1 & -2 & 1 \\ -2 & 4 & -1 \\ 1 & -2 & 2 \end{bmatrix}.$$

Let $b = (2, -5, 1) \in \mathbb{R}^3$. Viewing this as a column vector, we have

$$b = \begin{bmatrix} 2 \\ -5 \\ 1 \end{bmatrix}.$$

- **(a)** Use forward elimination to put $A$ into row echelon form $Q$. Record all row operations. Perform these row operations on the identity matrix $I$ to obtain the matrix $O$ which is the product of the elementary invertible matrices which correspond to the row operations. Now $OA = Q$. Does $T$ have a kernel?
- **(b)** Use backward elimination to put $Q$ into reduced row echelon form $R$. Record all row operations. Perform these row operations on $O$ to obtain a matrix $U$. Now $UA = R$.
- **(c)** Use solution readoff to find the kernel of $T$.
- **(d)** Compute $d = Ub$ and use solution readoff of find the preimage under $T$ to the point $(2, -5, 1)$.
- **(e)** Find the general solution to the system of linear equations

$$\begin{aligned} x_1 - 2x_2 + x_3 &= 2 \\ -2x_1 + 4x_2 - x_3 &= -5 \\ x_1 - 2x_2 + 2x_3 &= 1 \end{aligned}$$

**Exercise 4.8.** Let $\mathbb{R}^3 = \{(x, y, z) \mid x, y, z \in \mathbb{R}\}$ and view $\mathbb{R}^2 = \{(x, y) \mid x, y \in \mathbb{R}\}$ as a subspace of $\mathbb{R}^3$.
Let $T : \mathbb{R}^3 \to \mathbb{R}^2$ be the linear transformation which first rotates $\mathbb{R}^3$ by 90 degrees around the $z$-axis, then rotates $\mathbb{R}^3$ by 60 degrees around the $x$-axis, and then projects $\mathbb{R}^3$ onto the $xy$-plane.
**(a)** Find the matrix $A_T$ corresponding to $T$ (its columns are the destinations of the standard basis vectors).
**(b)** Find $T(x, y, z)$ (plug the column vector $([x, y, z]^t$ into $A_T$).
**(c)** Find $\ker(T) = \{(x, y, z) \in \mathbb{R}^3 \mid T(x, y, z) = (0, 0)\}$.

**Exercise 4.9.** Each $3 \times 3$ elementary matrix $E$ corresponds to a linear transformation $T_E : \mathbb{R}^3 \to \mathbb{R}^3$.
Describe its geometric effect on 3-space.

**Exercise 4.10.** Let $A$ be an $m \times n$ matrix.

We say that $A$ is *right invertible* if there exists a $n \times m$ matrix $B$ such that $AB = I_m$.

We say that $A$ is *left invertible* if there exists a $n \times m$ matrix $B$ such that $BA = I_n$.

Let $T_A : \mathbb{R}^n \to \mathbb{R}^m$ be the linear transformation corresponding to $A$. Exactly two of the following statements are true:
**(a)** $A$ is right invertible if and only if $T_A$ is injective;
**(b)** $A$ is right invertible if and only if $T_A$ is surjective;
**(c)** $A$ is left invertible if and only if $T_A$ is injective;
**(d)** $A$ is left invertible if and only if $T_A$ is surjective.
Decide which two statements are true and explain why they are true.

CHAPTER 5

# Vector Spaces

## 1. Introduction to Abstraction

Mathematics may be described as the study of *patterns*; not only patterns that arise is the real universe, but of all *possible* patterns that may arise in *any* universe.

*Abstraction* is the process of identifying the key attributes in a system and extracting them; these key attributes may be called *defining properties*. Then one investigates the implications of these attributes, disembodied from the system which originally motivated their investigation. The investigation is rigorous, and all conclusions drawn are supported purely by logic, not by the physical world. This support is called *proof*. Together, abstraction and proof are the *mathematical method*; they allow us to isolate and study patterns.

This method has three main benefits.

(1) The proof aspect ensures that all claims made are true (given the assumptions); additionally, it is often the proof itself that illuminates the situation so that we understand it more fully.
(2) The abstraction allows one to see exactly how the identified key attributes lead to the behavior being displayed in the motivating system.
(3) All results apply to *any* system which obeys the defining properties.

For example, the set $\mathbb{R}$ of real numbers has many attributes which effect how we think of them, for example:

- Order
- Distance
- Algebra

Although these attributes are interrelated in the case of the real numbers, by looking at systems which *a priori* have only order, distance, or algebra, we obtain a better understanding of how these attributes effect our understanding of the real numbers.

In particular, we will see that in the case of vector spaces, the abstracted properties are so universal that they appear repeatedly in various forms, and we benefit from the abstraction we now make.

We use standard English and assume the laws of logic as determined by truth tables, a preliminary knowledge of set theory, and the basic algebraic properties of the sets $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, and $\mathbb{R}$. We make no other assumptions, and proceed to develop a complete theory from this starting point. In our examples and applications, however, we may assume previous knowledge, such as calculus.

## 2. Vector Spaces

**Definition 5.1.** A *vector space* over $\mathbb{R}$ is a set $V$ together with a pair of operations

$$+ : V \times V \to V \quad \text{and} \quad \cdot : \mathbb{R} \times V \to V,$$

called *vector addition* and *scalar multiplication* respectively, satisfying

**(V1)** $v + w = w + v$ for all $v, w \in V$;
**(V2)** $v + (w + x) = (v + w) + x$ for all $v, w, x \in V$;
**(V3)** there exists $0_V \in V$ such that $v + 0_V = v$ for all $v \in V$;
**(V4)** for every $v \in V$ there exists $w \in V$ such that $v + w = 0_V$;
**(V5)** $1 \cdot v = v$ for every $v \in V$;
**(V6)** $(ab)v = a(bv)$ for every $v \in V$ and $a, b \in \mathbb{R}$;
**(V7)** $a(v + w) = av + aw$ for every $v, w \in V$ and $a \in \mathbb{R}$;
**(V8)** $(a + b)v = av + bv$ for every $v \in V$ and $a \in \mathbb{R}$.

The elements of $V$ are called *vectors*.

*Remark* 5.1. The $\cdot$ is usually suppressed in this notation (as in **(V6)**, **(V7)**, and **(V8)**); scalar multiplication is instead denoted by juxtaposition.

*Remark* 5.2. Elements are added two at a time. However, because of property **(V2)**, parentheses are useless to distinguish the order of addition. That is,

$$v_1 + \cdots + v_n$$

makes sense without inserting parentheses to denote the order in which the elements are added, since any order gives the same result.

*Remark* 5.3. In the absence of parentheses, the operations of vector addition and scalar multiplication are written with $\cdot$ having higher precedence over $+$. For example, if $a, b \in \mathbb{R}$ and $v, w \in V$, $ax + by$ means $(ax) + (by)$.

## 3. Examples of Vector Spaces

**Example 5.2.** Let $V = \{0\}$. Then $V$ is called the *trivial* vector space.

**Example 5.3.** Let $\mathbb{R}$ be the set of real numbers, together with their standard addition and multiplication.
Then $\mathbb{R}$ is a vector space.

**Example 5.4.** Let $\mathbb{R}^n$ be the set of ordered $n$-tuples of real numbers, together with vector addition and scalar multiplication as defined previously.
Then $\mathbb{R}^n$ is a vector space.

**Example 5.5.** Let $V \subset \mathbb{R}^n$ be a subspace of $\mathbb{R}^n$ under our previous interpretation, together with vector addition and scalar multiplication from $\mathbb{R}^n$.
Then $V$ is a vector space.

**Example 5.6.** Let $\mathcal{M}_{m\times n}$ be the set of $m \times n$ matrices with real entries, together with matrix addition and scalar multiplication as defined previously.
Then $\mathcal{M}_{m\times n}$ is a vector space.

**Example 5.7.** Let $I \subset \mathbb{R}$ be an open interval. Let $\mathcal{F}(I) = \{f : I \to \mathbb{R}\}$ be the set of all functions from $I$ into $\mathbb{R}$. Note that we have specified such a function if we specify its value at every point in $I$. Define addition and scalar multiplication by

$$(f + g)(t) = f(t) + g(t) \quad \text{where } f, g \in \mathcal{F}(I), t \in I;$$
$$(af)(t) = a(f(t)) \quad \text{where } f \in \mathcal{F}(I), a \in \mathbb{R}, t \in I.$$

Then $\mathcal{F}(I)$, together with these operations, is a vector space.

**Example 5.8.** Let $\mathcal{P}$ denote the set of all polynomial functions with real coefficients, and for each $n \in \mathbb{N}$, let $\mathcal{P}_n$ denote the set of all polynomial functions of degree less than or equal to $n$ with real coefficients:

$$\mathcal{P} = \{f : \mathbb{R} \to \mathbb{R} \mid f(x) = a_0 + a_1x + \cdots + a_nx^n \text{ where } a_i \in \mathbb{R}; n \in \mathbb{N}\};$$
$$\mathcal{P}_n = \{f \in \mathcal{P} \mid \deg(f) \leq n\}.$$

Define addition and scalar multiplication on these sets as in the case of $\mathcal{F}(I)$.
Then $\mathcal{P}$ and $\mathcal{P}_n$, together with these operations, are vector spaces.

**Example 5.9.** Let $V$ and $W$ be vector spaces. Define vector addition and scalar multiplication on the cartesian product $V \times W$ by

$$(v_1, w_1) + (v_2, w_2) = (v_1 + v_2, w_1 + w_2);$$
$$a(v, w) = (av, aw).$$

Then $V \times W$, together with these operations, is a vector space. Indeed, this is exactly how $\mathbb{R}^n$ is constructed from $\mathbb{R}$.

## 4. Properties of Vector Addition and Scalar Multiplication

**Proposition 5.10.** *Let $V$ be a vector space. Suppose that there exists $0_1, 0_2 \in V$ such that $v + 0_1 = v$ and $v + 0_2 = v$ for every $v \in V$. Then $0_1 = 0_2$.*

*Proof.* We have $0_1 = 0_1 + 0_2 = 0_2 + 0_1 = 0_2$. □

*Remark* 5.4. This proposition says the *the additive identity is unique.*

**Proposition 5.11.** *Let $V$ be a vector space and let $v \in V$. Suppose that there exists $w_1$ and $w_2$ such that $v + w_1 = 0_V$ and $v + w_2 = 0_V$. Then $w_1 = w_2$.*

*Proof.* Since $v + w_1 = 0_V$, we have $w_2 + (v + w_1) = w_2 + 0_V$. By **(V2)** on the left and **(V3)** on the right, we have, $(w_2 + v) + w_1 = w_2$. By property **(V1)**, $(v+w_2)+w_1 = w_2$, and by assumption on $w_2$, this gives $0_V + w_1 = w_2$. By property **(V2)**, $w_1 + 0_V = w_2$, and finally by property **(V3)** we obtain $w_1 = w_2$. □

*Remark* 5.5. This proposition says that *additive inverses are unique.* We denote the unique additive inverse of $v$ by $-v$. We shorten $w + (-v)$ to $w - v$.

**Proposition 5.12. (Cancellation Law)**
*Let $V$ be a vector space and let $v, w, x \in V$. Then $v + x = w + x \Rightarrow v = w$.*

*Proof.* Add $-x$ to both sides. □

**Proposition 5.13.** *Let $V$ be a vector space. Let $v \in V$ and $a \in \mathbb{R}$. Then*

- **(a)** $0v = 0_V$;
- **(b)** $a0_V = 0_V$;
- **(c)** $av = 0_V \Rightarrow a = 0$ *or* $v = 0_V$;
- **(d)** $(-1)v = -v$;
- **(e)** $(-a)v = -(av)$.

*Proof.*

**(a)** We have $v + (0v) = (1v) + (0v) = (1 + 0)v = 1v = v$; thus $0v$ acts like the additive identity, so it must be the additive identity by uniqueness.

**(b)** If $a = 0$, the result follows from **(a)**, so assume $a \neq 0$. We have $v+(a0_V) = (aa^{-1})v + a0_V = a(a^{-1}v) + a0_V = a(a^{-1}v + 0_V) = a(a^{-1}v) = v$; thus $v + (a0_V)$ acts like the additive identity, so it must be the additive identity.

**(c)** Exercise.

**(d)** Since $1v = v$, this is a special case of **(d)**.

**(e)** We have $av + (-a)v = (a + (-a))v = 0v = 0_V$; thus $(-a)v$ acts like the additive inverse of $a$, so it must be the additive inverse of $v$ by uniqueness. □

*Remark* 5.6. We now drop the subscript from $0_V$ and just write 0. We distinguish this from $0 \in \mathbb{R}$ by context.

## 5. Subspaces

**Definition 5.14.** Let $V$ be a vector space.

A *subspace* of $V$ is a subset $W \subset V$ which satisfies:

**(S0)** $0 \in W$;
**(S1)** $x, y \in W \Rightarrow x + y \in W$;
**(S2)** $a \in \mathbb{R}, x \in W \Rightarrow ax \in W$.

If $W$ is a subspace of $V$, this is denoted by $W \leq V$.

*Remark* 5.7. If $W \leq V$, then $W$ is a vector space under the same operations of addition and scalar multiplication. If is clear that if $U \leq W$ and $W \leq V$, then $U \leq W$.

*Remark* 5.8. Consider the condition

**(SE)** $W$ is nonempty.

In the presence of **(S1)** and **(S2)**, we see that **(SE)** is equivalent to **(S0)**. Clearly, if $0 \in W$, then $W$ is nonempty. On the other hand, suppose that $W$ is nonempty. Then $W$ contains a vector, say $w \in W$. By **(S2)**, we see that $0w = 0 \in W$.

**Example 5.15.** Let $V$ be a vector space; then there exists a zero element $0 \in V$, and $\{0\} \leq V$; this is called the *trivial* subspace.

**Example 5.16.** Let $V = \mathbb{R}^3$ and let $W = \{(x_1, x_2, x_3) \in V \mid x_1 + x_2 + x_3 = 0\}$. Show that $W \leq V$.

*Solution.* To show that $W$ is a subspace of $V$, we verify the three properties of being a subspace.

**(S0)** We wish to show that $0_V \in W$. Since $0 + 0 + 0 = 0$, we see that $(0, 0, 0) = 0_V \in W$.

**(S1)** We wish to show that the sum of two elements in $W$ is also an element in $W$. Let $x, y \in W$. Then $x, y \in \mathbb{R}^3$, so $x = (x_1, x_2, x_3)$ and $y = (y_1, y_2, y_2)$ for some real numbers $x_1, x_2, x_3, y_1, y_2, y_3 \in \mathbb{R}$. By definition of $W$, $x_1 + x_2 + x_3 = 0$ and $y_1 + y_2 + y_3 = 0$. Adding these equations and rearranging via properties **(V1)** and **(V2)** of $V$, we see that $(x_1 + y_1) + (x_2 + y_2) + (x_3 + y_3) = 0$. Thus $x + y = (x_1 + y_1, x_2 + y_2, x_3 + y_3)$ satisfies the defining condition of $W$, and must be an element of $W$.

**(S2)** We wish to show that any scalar multiple of an element in $W$ is also an element in $W$. Let $x = (x_1, x_2, x_3) \in W$ and let $a \in \mathbb{R}$. Then $x_1 + x_2 + x_3 = 0$, so $a(x_1 + x_2 + x_3) = a0 = 0$; by distributing, we get $ax_1 + ax_2 + ax_3 = 0$. Thus $ax = (ax_1, ax_2, ax_3) \in W$. □

**Example 5.17.** Fix $n \in \mathbb{N}$. Then $\mathcal{P}_n \leq \mathcal{P}$.

**Example 5.18.** Fix $m, n \in \mathbb{N}$ with $m \leq n$. Then $\mathcal{P}_m \leq \mathcal{P}_n$.

**Example 5.19.** A function $f : I \to \mathbb{R}$ is called *smooth* if it is infinitely differentiable on the interval $I$; that is, if derivatives of all orders exist and are continuous. The following are subspaces of $\mathcal{F}(I)$:

- $\mathcal{C}(I) = \{f \in \mathcal{F}(I) \mid f \text{ is continuous}\}$ ;
- $\mathcal{D}(I) = \{f \in \mathcal{F}(I) \mid f \text{ is smooth}\}$ .

Note that $\mathcal{D}(I) \leq \mathcal{C}(I) \leq \mathcal{F}(I)$.

**Proposition 5.20.** *Let $V$ be a vector space and let $W_1, W_2 \leq V$. Then $W_1 \cap W_2 \leq V$.*

*Proof.* We verify the three properties of a subspace.

**(S0)** Since $0 \in W_1$ and $0 \in W_2$, we have $0 \in W_1 \cap W_2$.

**(S1)** Let $x, y \in W_1 \cap W_2$. Then $x, y \in W_1$ and $x, y \in W_2$, so $x + y \in W_1$ and $x + y \in W_2$, because both of these sets are subspaces. Thus $x + y \in W_1 \cap W_2$.

**(S2)** Let $x \in W_1 \cap W_2$ and let $a \in \mathbb{R}$. Then $x \in W_1$ and $x \in W_2$, and since these are subspaces, we see that $ax \in W_1$ and $ax \in W_2$. Thus $ax \in W_1 \cap W_2$.

Therefore $W_1 \cap W_2 \leq V$. □

*Remark* 5.9. This argument generalizes so that the intersection of *any number* (even infinitely many) of subspaces is a subspace.

**Definition 5.21.** Let $V$ be a vector space and let $X, Y \subset V$. Define the sum of these sets to be the subset of $V$ given by

$$X + Y = \{x + y \mid x \in X, y \in Y\}.$$

**Proposition 5.22.** *Let $V$ be a vector space and let $W_1, W_2 \leq V$. Then $W_1 + W_2 \leq V$.*

*Proof.* We verify the three properties of a subspace.

**(S0)** Since $0 \in W_1$ and $0 \in W_2$, we see that $0 = 0 + 0 \in W_1 + W_2$.

**(S1)** Let $w_1, w_1' \in W_1$ and $w_2, w_2' \in W_2$ so that $w_1 + w_2$ and $w_1' + w_2'$ are arbitrary members of $W_1 + W_2$. Then $(w_1 + w_2) + (w_1' + w_2') = (w_1 + w_1') + (w_2 + w_2') \in W_1 + W_2$, by properties **(V1)** and **(V2)** of $V$ and by property **(S1)** of $W_1$ and $W_2$.

**(S2)** Let $w_1 \in W_1$ and $w_2 \in W_2$ so that $w_1 + w_2$ is an arbitrary member of $W_1 + W_2$ Let $a \in \mathbb{R}$. Then $a(w_1 + w_2) = aw_1 + aw_2 \in W_1 + W_2$, by property **(V7)** of $V$ and property **(S2)** of $W_1$ and $W_2$. □

*Remark* 5.10. It follows that any finite sum of subspaces is a subspace.

## 6. Spans

**Definition 5.23.** Let $V$ be a vector space and let $X \subset V$.

A *linear combination* of elements from $X$ is an element of $V$ of the form

$$a_1 v_1 + \cdots + a_n v_n \quad \text{where } a_i \in \mathbb{R} \text{ and } v_i \in X.$$

The *span* of $X$ is denoted by $\operatorname{span}(X)$ and is defined by

$$\operatorname{span}(X) = \{v \in V \mid v \text{ is a linear combination from } X\}.$$

The span of the empty set is defined to be $\{0\}$.

**Proposition 5.24.** *Let $V$ be a vector space and let $X, Y \subset V$. Then*

**(a)** $0 \in \operatorname{span}(X)$*;*
**(b)** $X \subset \operatorname{span}(X) \subset V$*;*
**(c)** $\operatorname{span}(X) = \operatorname{span}(\operatorname{span}(X))$*;*
**(d)** $X \subset Y \Rightarrow \operatorname{span}(X) \subset \operatorname{span}(Y)$*;*
**(e)** $X \subset \operatorname{span}(Y) \Rightarrow \operatorname{span}(X) \subset \operatorname{span}(Y)$*;*
**(f)** $X \leq V \Leftrightarrow X = \operatorname{span}(X)$*.*

*Proof.*

**(a)** If $x \in X$, then $0_V = 0x$ is a linear combination from $X$.

**(b)** To show that one set is a subset of another, it suffices to select an arbitrary element of the one set and show that it is in the other.

Let $x \in X$; then $x = 1x$, so $x$ is a linear combination from $X$. Thus $x \in \operatorname{span}(X)$. Since $x$ was arbitrary, we see that $X \subset \operatorname{span}(X)$.

Since $V$ is closed under addition and scalar multiplication, every linear combination of vectors from $V$ is also in $V$. Thus the span of any subset of $V$ is contained in $V$, i.e., $\operatorname{span}(X) \subset V$.

**(c)** To show that two sets are equal, we show that each is contained in the other. By **(b)**, we know that $\operatorname{span}(X) \subset \operatorname{span}(\operatorname{span}(X))$, so we only need to show that $\operatorname{span}(\operatorname{span}(X)) \subset \operatorname{span}(X)$.

Let $y \in \operatorname{span}(\operatorname{span}(X))$. Then there exist vectors $y_1, \ldots, y_n \in \operatorname{span}(X)$ and real number $a_1, \ldots, a_n \in \mathbb{R}$ such that $y = \sum_{i=1}^n a_i y_i$. Each $y_i$ is in $\operatorname{span}(X)$, so there exist a finite number (say $m_i$) of vectors $x_{i1}, \ldots, x_{im_i} \in X$ and real numbers $b_{i1}, \ldots, b_{im_i}$ such that $y_i = \sum_{j=1}^{m_i} b_{ij} x_{ij}$. Then $y = \sum_{i=1}^n \sum_{j=1}^{m_i} a_i b_{ij} x_{ij}$ is a linear combination from $X$, so $y \in \operatorname{span}(X)$.

**(d)** Suppose that $X \subset Y$. Then every linear combination from $X$ is a linear combination from $Y$. Thus $\operatorname{span}(X) \subset \operatorname{span}(Y)$.

**(e)** Suppose that $X \subset \operatorname{span}(Y)$. Then by **(d)**, $\operatorname{span}(X) \subset \operatorname{span}(\operatorname{span}(Y))$. But by **(c)**, $\operatorname{span}(\operatorname{span}(Y)) = \operatorname{span}(Y)$, so $\operatorname{span}(X) \subset \operatorname{span}(Y)$.

**(f)** To show an if and only if statement, we show implication in both directions.

($\Rightarrow$) Suppose that $X \leq V$. We know that $X \subset \operatorname{span}(X)$; we must show that $\operatorname{span}(X) \subset X$.

Let $x \in \operatorname{span}(X)$. Then $x$ is a linear combination from $X$. This means that $x$ is a finite sum of scalar multiples of elements of $X$. Since $X$ is a subspace, it is closed under vector addition and scalar multiplication. Thus each scalar multiple is in $X$, and the sum of these elements of $X$ is also in $X$. Thus $x \in X$.

($\Leftarrow$) Suppose that $X = \operatorname{span}(X)$. We have already noted that $0 \in X$. Let $x_1, x_2 \in X$ and let $a \in \mathbb{R}$. Then $x_1 + x_2$ and $ax_1$ are linear combinations from $X$, so they are in $X$ since $X = \operatorname{span}(X)$. Since $X$ satisfies **(S0)**, **(S1)**, and **(S2)**, we see that $X \leq V$. $\square$

**Definition 5.25.** Let $V$ be a vector space and let $X \subset V$.
We say that $X$ *spans* $V$ if $V = \operatorname{span}(X)$.
We say that $V$ is *finitely generated* if it is spanned by a finite set of vectors.

**Proposition 5.26.** *Let $V$ be a vector space and let $X, Y \subset V$.*
*If $X$ spans $V$ and $X \subset Y$, then $Y$ spans $V$.*

*Proof.* Suppose that $X$ spans $V$. Then every element of $V$ is a linear combination of elements from $X$. But since $X \subset Y$, all such linear combinations are also linear combinations from $Y$. Thus $Y$ spans $V$. □

*Remark* 5.11. Let $V$ be a finitely generated vector space; this means that there exists a finite set of vectors, say $X$, such that $X$ spans $V$. Suppose that $Y$ is an infinite subset of $V$ which spans $V$. The sets $X$ and $Y$ may not have any elements in common. The next proposition tells us than in spite of this, some finite subset of subset of $Y$ spans $V$.

**Proposition 5.27.** *Let $V$ be a finitely generated vector space and let $Y \subset V$ such that $Y$ spans $V$. Then there exists a finite subset $Z$ of $Y$ such that $Z$ spans $V$.*

*Proof.* Since $V$ is finitely generated, there exists some finite set

$$X = \{x_1, \ldots, x_n\} \subset V$$

such that $\operatorname{span}(X) = V$. But $X \subset \operatorname{span}(Y)$, so each of the vector $x_i \in X$ may be written as a linear combination of a finite number (say $m_i$) of vectors $z_{i1}, \ldots, z_{im_j} \in Y$. Let $Z = \{z_{ij} \mid i = 1, \ldots, n;\ j = 1, \ldots, m_i\}$. Then $Z$ is a finite set, and $X \subset \operatorname{span}(Z)$, so $V = \operatorname{span}(X) \subset \operatorname{span}(Z)$. Since $\operatorname{span}(Z) \subset V$, we have $\operatorname{span}(Z) = V$. □

**Example 5.28.** Let $V = \mathbb{R}^3$ and let $X$ be any set of vectors with the property that not all of them lie on the same plane. Then $X$ spans $V$. Moreover, one can pick out a finite subset of $X$ which spans $V$; soon we will see that one can choose this set with exactly three elements.

**Example 5.29.** Let $V = \mathbb{R}^n$. Let $e_i$ denote the vector whose $i^{\text{th}}$ entry is equal to one and whose other entries are equal to zero.
Let $X = \{e_1, \ldots, e_n\}$. Then $X$ spans $V$.

**Example 5.30.** Let $V = \mathcal{M}_{m\times n}$. Let $M_{ij}$ denote the $m \times n$ matrix whose $ij^{\text{th}}$ entry is equal to one and whose other entries are equal to zero.
Let $X = \{M_{ij} \mid i = 1, \ldots, m;\ j = 1, \ldots, n\}$. Then $X$ spans $V$.

**Example 5.31.** Let $V = \mathcal{P}_n$ and let $X = \{1, x, x^2, \ldots, x^n\}$. Then $X$ spans $\mathcal{P}_n$.

*Remark* 5.12. One can show that $\operatorname{span}(X)$ is *exactly* the intersection of all subspaces of $V$ which contain $X$.

## 7. Linear Independence

**Definition 5.32.** Let $V$ be a vector space and let $X \subset V$.

We say that $X$ is *linearly independent*, or simply *independent*, if whenever $v_1, \dots, v_n \in V$ are distinct elements of $X$ and $a_1, \dots, a_n \in \mathbb{R}$,

$$\sum_{i=1}^{n} a_i v_i = 0 \Rightarrow a_i = 0 \text{ for } i = 1, \dots, n.$$

In this case we may also say that the vectors in $X$ are linearly independent.

We say that $X$ is *linearly dependent*, or simply *dependent*, if it is not independent. In this case we may also say that the vectors in $X$ are linearly dependent.

*Remark* 5.13. If $X \subset V$ is dependent, there exists a *nontrivial dependence relation*; by this we mean that there exist distinct elements $x_1, \dots, x_n \in X$ and real numbers $a_1, \dots, a_n$, at least one of which is nonzero, such that

$$a_1 x_1 + \cdots + a_n x_n = 0.$$

Clearly, any set containing 0 is dependent.

**Proposition 5.33.** *Let $V$ be a vector space and let $X, Y \subset V$. If $Y$ is independent and $X \subset Y$, then $X$ is independent.*

*Proof.* Any nontrivial dependence relation among the elements of $X$ would be a nontrivial dependence relation among the elements of $Y$. □

**Proposition 5.34.** *Let $V$ be a vector space and let $X \subset V$. Then $X$ is independent if and only if for every finite subset $B = \{x_1, \dots, x_n\} \subset X$ and every $x \in \operatorname{span}(B)$ there exists a unique ordered $n$-tuple $(a_1, \dots, a_n) \in \mathbb{R}^n$ such that*

$$x = a_1 x_1 + \cdots + a_n x_n.$$

*Proof.*

($\Rightarrow$) We prove the contrapositive; suppose that the second condition is false, and prove that $X$ is dependent. Let $B \subset X$ be a finite subset whose span contains an element $x \in \operatorname{span}(B)$ which may not be expressed in a unique way as a linear combination from $B$. Since $x \in \operatorname{span}(B)$, there is at least one way to write it as a linear combination from $B$; thus the only way the condition can be false is if this expression is not unique.

Suppose that there exist distinct $n$-tuples $(a_i)_i$ and $(b_i)_i$ such that

$$x = \sum_{i=1}^{n} a_i x_i = \sum_{i=1}^{n} b_i x_i.$$

Subtracting yields $\sum_{i=1}^{n} (b_i - a_i) x_i = 0$; but for at least one $i$, we have $a_i \neq b_i$, so this is a nontrivial dependence relation among the $x_i$'s; thus $B$ is not independent, so neither is $X$.

($\Leftarrow$) Again we prove the contrapositive. Thus we suppose that $X$ is not independent. Then there exists some vectors $x_1, \dots, x_n \in X$ and some real numbers $a_1, \dots, a_n$, not all zero, such that $\sum_{i=1}^{n} a_i x_i = 0$. Since $0 \in \operatorname{span}\{x_1, \dots, x_n\}$, we see that 0 can be written in more than one way as a linear combination of vectors from $X$. Thus the second condition is also false. □

**Example 5.35.** Let $V = \mathbb{R}^n$ and let $X = \{e_1, \dots, e_n\}$.
Then $X$ is independent.

**Example 5.36.** Let $V = \mathcal{M}_{m\times n}$ and let $X = \{M_{ij} \mid i = 1, \dots, m;\ j = 1, \dots, n\}$.
Then $X$ is independent.

**Example 5.37.** Let $V = \mathcal{P}_n$ and let $X = \{1, x, x^2, \dots, x^n\}$.
Then $X$ is independent.

**Proposition 5.38.** *Let $V$ be a vector space and let $v, w \in V$ with $v \neq 0$. Then $v$ and $w$ are linearly dependent if and only if $w = cv$ for some $c \in \mathbb{R}$.*

*Proof.* Suppose that $v$ and $w$ are linearly dependent. Then there exists $a, b \in \mathbb{R}$, not both zero, such that $av + bw = 0$. If $a = 0$, then $bw = 0$, so $w = 0$ (since $b \neq 0$), whence $w = 0v$. Otherwise $a \neq 0$, and $v = -\frac{b}{a}w$.

On the other hand, if $w = cv$, then $cv - w = 0$ is a nontrivial dependence relation, so $v$ and $w$ are linearly dependent. □

**Example 5.39.** Let $V = \mathcal{D}(I)$, where $I \subset \mathbb{R}$ is an open interval. Let $f, g \in V$. If $f$ and $g$ are linearly dependent, then there exists a constant $c$ such that $f(t) = cg(t)$ for all $t \in I$. Assuming neither $f$ nor $g$ is the zero function, we see that $c \neq 0$. Then $f(t) = 0 \Leftrightarrow g(t) = 0$. Since differentiation is linear, we see that $f'(t) = 0 \Leftrightarrow g'(t) = 0$. This continues for all of the derivatives of $f$ and $g$.

Turning this around, one sees that if there exists $t \in I$ such that $f$ and $g$, or any of their derivatives, have the property that one is zero at $t$ and the other is not, then $f$ and $g$ are linearly independent.

**Example 5.40.** Let $V = \mathbb{R}^3$ and let

$$v_1 = (1, 2, -3), v_2 = (2, 0, 1), v_3 = (4, -4, 9) \in \mathbb{R}^3.$$

Show that the set $\{v_1, v_2, v_3\}$ is dependent.

*Solution.* If these vectors are indeed linearly dependent, then they all lie on the same plane in $\mathbb{R}^3$. We may choose any one of them, say $v_3$, are try to write it as a linear combination of the other two. That is, we want to find $x_1, x_2 \in \mathbb{R}$ such that $v_3 = x_1v_1 + x_2v_2$. Thinking of these vectors as column vectors, the equation we want to solve is

$$\begin{bmatrix} 4 \\ -4 \\ 9 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix} + x_2 \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}.$$

Setting

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 0 \\ -3 & 1 \end{bmatrix}, \quad x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad \text{and} \quad b = \begin{bmatrix} 4 \\ -4 \\ 9 \end{bmatrix},$$

we see that this is equivalent to solving the matrix equation $Ax = b$. We use Gaussian elimination (this is a relatively easy elimination; do it for practice) to obtain an alternate matrix equation

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} x = \begin{bmatrix} -2 \\ 3 \\ 0 \end{bmatrix}.$$

Thus $x_1 = -2$ and $x_2 = 3$. □

The last example does not completely give a test for linear independence.

**Example 5.41.** Let $V = \mathbb{R}^3$ and let

$$v_1 = (1, 2, -3), v_2 = (2, 4, -6), v_3 = (4, -4, 9) \in \mathbb{R}^3.$$

Show that the set $\{v_1, v_2, v_3\}$ is dependent.

*Attempt.* We try to write $v_3$ as a linear combination of $v_1$ and $v_2$. But it isn't! This is because $v_1$ and $v_2$ lie on the same line; so actually $v_3$ in independent from $v_1$ or from $v_2$. However, the set is dependent, because $2v_1 - v_2 + 0v_3 = 0$ is a nontrivial dependence relation. □

We now give a test for linear independence in $\mathbb{R}^m$. Let $X = \{v_1, \ldots, v_n\}$ be a subset of $\mathbb{R}^m$. Form the $m \times n$ matrix $A$ by putting the vectors in columns:

$$A = [v_1 \mid \cdots \mid v_n].$$

If $x = [x_1, \ldots, x_n]^t$ is a column vector in $\mathbb{R}^n$, we have seen that

$$\begin{aligned} Ax &= x_1 A^{(1)} + \cdots + x_n A^{(n)} \\ &= x_1 v_1 + \cdots + x_n v_n; \end{aligned}$$

that is, $Ax$ is a linear combination of the columns of $A$. Now $Ax = 0$ has a solution other than $x = (0, \ldots, 0)$ if and only if there is a nontrivial dependence relation among the $v_i$'s.

Form matrix $Q$ by performing forward elimination on $A$, so that $Q$ is in row echelon form; there is an invertible $m \times n$ matrix $O$ such that $Q = OA$. Since $O0 = 0$, we have

$$Ax = 0 \Leftrightarrow OAx = O0 \Leftrightarrow Qx = 0.$$

The solution to $Ax = 0$ is unique if and only if $Q$ has no free columns; otherwise, $Ax = 0$ has a nontrivial (i.e., nonzero) solution, which gives a nontrivial dependence relation among the columns of $A$.

**Example 5.42.** Let $V = \mathbb{R}^3$ and let

$$v_1 = (1, 2, -3), v_2 = (2, 0, 1), v_3 = (4, -4, 9) \in \mathbb{R}^3.$$

Show that the set $\{v_1, v_2, v_3\}$ is dependent.

*Solution.* Put the vectors in columns of a matrix $A$, so that

$$A = \begin{bmatrix} 1 & 2 & 4 \\ 2 & 0 & -4 \\ -3 & 1 & 9 \end{bmatrix}.$$

Perform forward elimination on $A$ is arrive at

$$Q = \begin{bmatrix} 1 & 2 & 4 \\ 0 & -4 & -12 \\ 0 & 0 & 0 \end{bmatrix}.$$

Since $Q$ has a free column, the vectors are not independent. □

## 8. Bases

**Definition 5.43.** Let $V$ be a vector space and let $X \subset V$.

We say that $X$ is a *basis* for $V$ if

**(B1)** $X$ spans $V$;
**(B2)** $X$ is independent.

**Definition 5.44.** Let $V$ be a vector space and let $X \subset V$.

We say that $X$ is a *spanning set* for $V$ if $X$ spans $V$.

We say that $X$ is a *minimal spanning set* for $V$ if

**(M1)** $X$ spans $V$;
**(M2)** $Y \subsetneq X \Rightarrow \text{span}(Y) \subsetneq V$.

**Proposition 5.45.** *Let $V$ be a vector space and let $X \subset V$. Then $X$ is a basis for $V$ if and only if $X$ is a minimal spanning set for $V$.*

*Proof.* To prove an if and only if statement, we prove the implication in both directions. Here it is clearly sufficient to show that **(B2)** is equivalent to **(M2)** in the presence of **(B1)**. Thus suppose that $X$ spans $V$. We prove the contrapositive in both directions.

($\Rightarrow$) Suppose that $X$ is a spanning set which is not minimal. Then there exists a smaller subset $Y \subsetneq X$ which spans. Let $x \in X \setminus Y$; then $x$ is a linear combination of vectors in $Y$, which demonstrates the presence of a nontrivial dependence relation in $X$. Thus $X$ is not independent.

($\Leftarrow$) Suppose that $X$ is a spanning set which is dependent. Then there exists a nontrivial dependence relation in $X$. This allows us to select some vector $x \in X$ and write it as a linear combination of the other vectors in $X$; let $Y = X \setminus \{x\}$. By Proposition 5.24 **(b)**, $Y \subset \text{span}(Y)$; also $x \in \text{span}(Y)$, so $X = Y \cup \{x\} \subset \text{span}(Y)$. Thus by Proposition 5.24 **(e)**,

$$V = \text{span}(X) \subset \text{span}(Y) \subset V,$$

which shows that $Y$ spans $V$. Thus $X$ is not a minimal spanning set. □

**Example 5.46.** Let $V = \mathbb{R}^3$ and let $W = \{(x, y, z) \in V \mid x + y + z = 0\}$. We have seen that $W \leq V$, so $W$ is a vector space. Actually, $W$ is a plane through the origin. Let $v_1 = (1, 0, -1)$ and $v_2 = (0, 1, -1)$ and let $X = \{v_1, v_2\}$. Then $X$ spans $W$: if $v = (x, y, z) \in W$, then $v = (x, y, -x - y)$, so $v = xv_1 + yv_2$. However, if we remove either vector from the set, the span of what remains is a line. Thus this set is a minimal spanning set, and so it is a basis.

**Proposition 5.47.** *Let $V$ be a vector space and let $X \subset V$. Then $X$ is a basis for $V$ if and only if every element of $V$ can be written as a linear combination from $X$ in a unique way.*

*Proof.* Exercise. □

**Proposition 5.48.** *Let $V$ be a vector space and let $X \subset V$ be independent. If $v \in V \smallsetminus \operatorname{span}(X)$, then $X \cup \{v\}$ is independent.*

*Proof.* Exercise. □

**Proposition 5.49.** *Let $V$ be a vector space and let $X \subset V$ be a spanning set. If $v \in V \smallsetminus X$, then $Y = X \cup \{v\}$ is dependent.*

*Proof.* If $v = 0$, there result is immediate. Otherwise, we may write $v$ as a linear combination of elements in $X$, so the expression of $v$ as a linear combination of elements in $Y$ is not unique; thus $Y$ is dependent. □

**Proposition 5.50.** *Let $V$ be a vector space and let $X = \{x_1, \dots, x_n\}$ be a dependent set. Then there exists $k \in \{1, \dots, n\}$ such that $x_k$ is a linear combination from $\{x_1, \dots, x_{k-1}\}$.*

*Proof.* Since $X$ is dependent, there is a nontrivial dependence relation

$$\sum_{i=1}^{n} a_i x_i = 0,$$

where not all $a_i$'s equal zero. Let $k$ be the largest integer between 1 and $n$ such that $a_k \neq 0$. Then

$$x_k = \frac{1}{a_k} \sum_{i=1}^{k-1} a_i x_i$$

is a linear combination of the preceding elements. □

*Remark* 5.14. Suppose $V$ and $X$ are as above, and note that $x_n$ is not necessary dependent on the preceding elements. For example, perhaps $V = \mathbb{R}^3$ and $x_1, \dots, x_{n-1}$ all lie on the same plane, but $x_n$ is perpendicular to it.

**Theorem 5.51.** *Let $V$ be a finitely generated vector space and let $X, Y \subset V$. If $X$ is independent and $Y$ spans, then $|X| \leq |Y|$.*

*Proof.* By Proposition 5.27, we may assume that $Y$ is finite, say $Y = \{y_1, \ldots, y_n\}$.

By way of contradiction (BWOC), suppose that $|X| > n$ and let

$$Z = \{z_1, \ldots, z_{n+1}\} \subset X;$$

then $Z$ is independent by Proposition 5.33. Label the elements of $Y$ and $Z$ so that all of those contained in $Y \cap Z$ are in the front:

$$Y = \{z_1, \ldots, z_i, y_{i+1}, \ldots, y_n\}.$$

By Proposition 5.49, the set

$$\{z_1, \ldots, z_{i+1}, y_{i+1}, y_{i+2}, \ldots, y_n\}$$

is dependent. By Proposition 5.50, one of these vectors is dependent on the preceding ones, and since the $z_i' s$ are linearly independent, there exists $k \in \{i+1, \ldots, n\}$ such that $y_k$ is a linear combination of $\{z_1, \ldots, z_{i+1}, y_{i+1}, \ldots, y_{k-1}\}$. Thus if we remove $y_k$ from the set, it will still span:

$$\operatorname{span}\{z_1, \ldots, z_{i+1}, y_{i+1}, \ldots, y_{k-1}, y_{k+1}, \ldots, y_n\} = V.$$

Continuing in this way, adding the next $z$ and removing a $y$, we see that after $n - i$ replacements we have

$$\operatorname{span}\{z_1, \ldots, z_n\} = V.$$

Thus the set $Z = \{z_1, \ldots, z_n\} \cup \{z_{n+1}\}$ is dependent by Proposition 5.49, producing a contradiction. □

*Remark* 5.15. There is an alternate proof of this proposition. Let $Z$ and $Y$ be as in the above proof. Since $Y$ spans, we have

$$z_j = \sum_{i=1}^{n} a_{ij} y_i \quad \text{for some } a_{ij} \in \mathbb{R}.$$

One may use Gaussian elimination to solve this system of linear equations to obtain a dependence relation among the $z$'s. However, for this to be used in a rigorous proof, one must first give a formal demonstration that Gaussian elimination works in general.

**Corollary 5.52. (Finite Dimension Theorem)**
*Let $V$ be a finitely generated vector space. Let $X$ and $Y$ be bases for $V$. Then $|X| = |Y|$.*

*Proof.* Exercise. □

**Example 5.53.** Let $V = \mathbb{R}^n$ and let $X = \{e_1, \ldots, e_n\}$.
Then $X$ is a basis for $V$.

**Example 5.54.** Let $V = \mathcal{M}_{m\times n}$ and let $X = \{M_{ij} \mid i = 1, \ldots, m;\ j = 1, \ldots, n\}$.
Then $X$ is a basis $V$.

**Example 5.55.** Let $V = \mathcal{P}_n$ and let $X = \{1, x, x^2, \ldots, x^n\}$.
Then $X$ is a basis $\mathcal{P}_n$.

**Proposition 5.56.** *Let $V$ be a finitely generated vector space and let $Y \subset V$ be a spanning set. Then there exists a subset $X \subset V$ with $X \subset Y$ such that $X$ is a basis for $V$.*

*Proof.* Since $Y$ is a spanning set, $Y$ contains a minimal spanning set, say $X$, which can be obtained simply by throwing out dependent vectors until none are left. Then $X$ is a basis by Proposition 5.45. ☐

*Remark* 5.16. In particular, every finitely generated vector space has a basis.

**Proposition 5.57.** *Let $V$ be a finitely generated vector space and let $X \subset V$ be independent. Then there exists a subset $Y \subset V$ with $X \subset Y$ such that $Y$ is a basis for $V$.*

*Proof.* If $X$ spans $V$, we are done.

Otherwise, there exists a vector $v$ which is in $V$ but not in $\operatorname{span}(X)$. Form the set $X \cup \{v\}$; this set is still independent by Proposition 5.48. Continue this process until the resulting set spans; this will happen in a finite number of steps since $V$ is finitely generated. ☐

**Definition 5.58.** Let $V$ be a finitely generated vector space and let $X \subset V$ be independent.

A *completion* of $X$ is a basis $Y$ for $V$ such that $X \subset Y$.

**Example 5.59.** Select any two vectors $v_1, v_2 \in \mathbb{R}^3$ that do not lie on the same line. Then the set $X = \{v_1, v_2\}$ is independent. Let $v_3$ be any vector in $\mathbb{R}^3$ which does not lie on the plane spanned by $X$. Then $X \cup \{v_3\}$ is a basis for $\mathbb{R}^3$.

## 9. Dimension

**Definition 5.60.** Let $V$ be a vector space.

The *dimension* of $V$ is the smallest cardinality of a spanning set for $V$, and is denoted by $\dim(V)$. The dimension of the trivial vector space is defined to be zero: $\dim\{0\} = 0$. If $\dim(V) \in \mathbb{N}$, we say that $V$ is *finite dimensional*.

*Remark* 5.17. If $V$ is finitely generated, then Corollary 5.52 tells us that the dimension of $V$ is the number of elements in *any* basis for $V$. We see that $V$ is finite dimensional if and only if $V$ is finitely generated.

**Example 5.61.** The dimension of $\mathbb{R}^n$ is $n$.

**Example 5.62.** The dimension of $\mathcal{M}_{m\times n}$ is $mn$.

**Example 5.63.** The dimension of $\mathcal{P}_n$ is $n+1$.

**Example 5.64.** The vector space $\mathcal{F}(I)$ is NOT finite dimensional.

**Proposition 5.65.** *Let $V$ be a vector space. Then $V$ is finite dimensional if and only if every independent subset is finite.*

*Proof.* If $V$ is finite dimensional, we already know that the cardinality of any independent set is less than or equal to the dimension of $V$.

Suppose $V$ is not finite dimensional; then $V$ is not finitely generated. Let $X \subset V$ be finite and independent. Then $X$ does not span $V$, so there exists a vector $v \in V \smallsetminus \operatorname{span}(X)$. The set $X \cup \{v\}$ is still independent by Proposition 5.50. We have taken an arbitrary independent set and produced a bigger one; continuing in this way we obtain an infinite independent set. □

**Proposition 5.66.** *Let $V$ be a finite dimensional vector space and let $W \leq V$. Then $W$ is finite dimensional, and* $\dim(W) \leq \dim(V)$.

*Proof.* Suppose that $W$ is not finite dimensional. Then $W$ has an independent subset of every cardinality. In particular, it has one whose cardinality is larger than the dimension of $V$, which contradicts Theorem 5.51. Thus $W$ is finite dimensional, so $W$ has a basis; this basis is a linearly independent subset of $V$, so its cardinality, which is the dimension of $W$, must be less than or equal to the dimension of $V$, again by Theorem 5.51. □

**Proposition 5.67.** *Let $V$ be a finite dimensional vector space and let $U, W \leq V$. Then*

$$\dim(U+W) = \dim(U) + \dim(W) - \dim(U \cap W).$$

*Proof.* Let $\dim(U) = p$, $\dim(W) = q$, and $\dim(U+W) = n$.

Let $X = \{x_1, \ldots, x_n\}$ be a basis for $U \cap W$. We complete this to a basis $Y = \{x_1, \ldots, x_n, u_1, \ldots, u_{p-n}\}$ for $U$ and $Z = \{x_1, \ldots, x_n, w_1, \ldots, w_{q-n}\}$ for $W$. We see that $B = \{x_1, \ldots, x_n, u_1, \ldots, u_p, w_1, \ldots, w_q\}$ spans $U + V$. But this is an independent set. To see this, let $a_1, \ldots, a_n, b_1, \ldots, b_p, c_1, \ldots, c_q \in \mathbb{R}$ such that

$$\sum_{i=1}^{n} a_i x_i + \sum_{j=1}^{p} b_j u_j + \sum_{k=1}^{q} c_k w_k = 0.$$

Then

$$\sum_{j=1}^{p} b_j u_j = -\sum_{i=1}^{n} a_i x_j - \sum_{k=1}^{q} c_k w_k.$$

The sum on the left is in $U$ and the sum on the right is in $W$, so the sum on the left is actually in $U \cap W$. Thus we have $d_1, \ldots, d_n$ such that

$$\sum_{j=1}^{p} b_j u_j = \sum_{i=1}^{n} d_i x_i.$$

Since $Y$ is a basis for $U$, we see that

$$b_1, \ldots, b_p = 0.$$

Similarly the $c_k$'s are all zero, whence the $a_i$'s are all zero. □

**Corollary 5.68.** *Let $V$ be a vector space and let $U \leq V$. Then $U = V$ if and only if $\dim(U) = \dim(V)$.*

*Proof.* Exercise. □

**Example 5.69.** Let $V = \mathbb{R}^3$.
Let $U = \operatorname{span}\{(1,2,0),(2,1,0)\}$, and $W = \operatorname{span}\{(1,0,2),(2,0,1)\}$. We see that $U$ is the $xy$-plane and $W$ is the $xz$-plane. The sum of $U$ and $W$ is all of $\mathbb{R}^3$. Their intersection is the $x$-axis. We see that

$$\dim(U+W) = 3 = 2 + 2 - 1 = \dim(U) + \dim(W) - \dim(U \cap W).$$

The proof above indicates that we can change our bases for $U$ and $W$:
$U = \operatorname{span}\{(1,0,0),(0,1,0)\}$ and $W = \operatorname{span}\{(1,0,0),(0,0,1)\}$, so that the union of these bases is a basis for $U + W$.

## 10. Exercises

**Exercise 5.1.** Let $V = \mathbb{R}^3$.
Set $v_1 = (1, -3, 2)$, $v_2 = (3, -7, -1)$, and $v_3 = (1, 1, 12)$.
Show that the set $\{v_1, v_2, v_3\}$ is linearly dependent.

**Exercise 5.2.** Let $V$ be a vector space. Let $v \in V$ and $a \in \mathbb{R}$.
Show that $av = 0_V \Rightarrow a = 0$ or $v = 0_V$.

**Exercise 5.3.** Let $V$ be a vector space and let $X \subset V$ be independent.
Show that if $v \in V \setminus \operatorname{span}(X)$, then $X \cup \{v\}$ is independent.

**Exercise 5.4.** Let $V$ be a finitely generated vector space.
Let $X$ and $Y$ be bases for $V$.
Show that $|X| = |Y|$.
(Hint: use Theorem 5.51.)

**Exercise 5.5.** Let $V$ be a vector space and let $U \leq V$. Show that $U = V$ if and only if $\dim(U) = \dim(V)$.

**Exercise 5.6.** Let $V = \mathbb{R}^3$ and let $a_1, a_2, a_3 \in \mathbb{R}$. Set

$$W(a_1, a_2, a_3) = \{(x_1, x_2, x_3) \in \mathbb{R}^3 \mid a_1x_1 + a_2x_2 + a_3x_3 = 0\}.$$

- **(a)** Show that $W(a_1, a_2, a_3) \leq V$.
- **(b)** Show that the general solution to the matrix equation

$$\begin{bmatrix} 1 & 5 & 4 \\ 2 & 1 & 0 \\ 0 & 3 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

  is a subspace of $V$.

CHAPTER 6

# Linear Transformations

## 1. Linear Transformations

**Definition 6.1.** Let $V$ and $W$ be a vector spaces.

A *linear transformation* from $V$ to $W$ is a function $T : V \to W$ such that

**(T1)** $T(x + y) = T(x) + T(y)$ for every $x, y \in V$;
**(T2)** $T(ax) = aT(x)$ for every $a \in \mathbb{R}$ and $x \in V$.

*Remark* 6.1. If we say, "let $T : V \to W$ be a linear transformation", but $V$ and $W$ have not yet been specified, it is implicit that $V$ and $W$ are arbitrary vector spaces.

**Proposition 6.2.** *Let $T : V \to W$ be a linear transformation. Then $T(0_V) = 0_W$.*

*Proof.* Since $0 \in \mathbb{R}$, by **(T2)** we have $T(0_V) = T(00_V) = 0T(0_V) = 0_W$. □

**Example 6.3.** Let $T : V \to W$ be given by $T(v) = 0_W$ for every $v \in V$. Then $T$ is a linear transformation, called the *zero* transformation.

**Example 6.4.** Let $T : V \to V$ be given by $T(v) = av$, where $a \in \mathbb{R}$ is a fixed real number. Then $T$ is a linear transformation, called *dilation by $a$*.

**Example 6.5.** Let $V$ be a vector space and let $X = \{x_1, \dots, x_n\}$ be a basis for $V$. Then every point $v \in V$ can be written in a unique way as a linear combination from $X$. Select a subset $Y = \{x_1, \dots, x_k\} \subset X$ and set $W = \text{span}(Y)$; note that $W$ is a vector space, and that $Y$ is a basis for $W$.

Define a function $T : V \to W$ by $T(v) = \sum_{i=1}^{k} a_i x_i$, where $v = \sum_{i=1}^{n} a_i x_i$. Then $T$ is a linear transformation, called *projection onto $W$*.

*Remark* 6.2. Linear transformations are so named because they take lines to lines (or to a point), planes to planes (or to lines or to a point), and so forth. We now show this.

## 2. Transformations and Bases

**Proposition 6.6.** *Let $T : V \to W$ be a linear transformation and let $X \subset V$. Then $T(\text{span}(X)) = \text{span}(T(X))$.*

*Proof.* To show that two sets are equal, we show that each is contained in the other.

Let $w \in T(\text{span}(X))$. Then $w = T(v)$ for some $v \in \text{span}(X)$. Since $v \in \text{span}(X)$, there exist vectors $x_1, \ldots, x_n \in X$ and real numbers $a_1, \ldots, a_n \in \mathbb{R}$ such that

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

Since $T$ is a linear transformation, it passes through summations and scalar multiplications, so

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

This latter expression is in the span of $T(X)$, so $w \in \text{span}(T(X))$.

Let $w \in \text{span}(T(X))$. Then there exist vectors $w_1, \ldots, w_m \in T(X)$ and real numbers $b_1, \ldots, b_m$ such that

$$w = \sum_{i=1}^{m} b_i w_i.$$

For each $i$, since $w_i \in \text{span}(X)$, the exists $x_i \in X$ such that $w_i = T(x_i)$. This gives

$$w = \sum_{i=1}^{m} b_i w_i = \sum_{i=1}^{m} b_i T(x_i) = \sum_{i=1}^{m} T(b_i x_i) = T(\sum_{i=1}^{m} b_i x_i).$$

Since $\sum_{i=1}^{m} b_i x_i \in \text{span}(X)$, we see that $w \in T(\text{span}(X))$. □

**Proposition 6.7.** *Let $V$ and $W$ be vector spaces. Let $X = \{v_1, \ldots, v_n\} \subset V$ be a basis for $V$. Let $Y = \{w_1, \ldots, w_n\} \subset W$. Then there exists a unique linear transformation $T : V \to W$ such that $T(v_i) = w_i$.*

*Proof.* For each $v \in V$, there exist unique real numbers $a_1, \ldots, a_n$ such that $v = \sum_{i=1}^{n} a_i v_i$. Define $T(v) = \sum_{i=1}^{n} a_i w_i$. It is clear that $T(v_i) = w_i$, and it is easy to verify that $T$ is linear. Uniqueness comes from the necessity of this definition, given that we require $T$ to be linear. □

**Corollary 6.8.** *Let $T : V \to W$ be a linear transformation. Then $T$ is completely determined by its effect on any basis for $V$.*

*Remark* 6.3. The above idea is a double edged sword. We completely know a transformation $T : V \to W$ if we know its effect on any basis for $V$. On the other hand, if we wish to construct a linear transformation, we only need to specify its effect on some basis.

## 3. Transformations and Subspaces

**Proposition 6.9.** *Let $T : V \to W$ be a linear transformation and let $U \leq V$. Then $T(U) \leq W$.*

*Proof.* We have $T(U) = T(\text{span}(U)) = \text{span}(T(U))$. Thus $T(U)$ is a subspace, since it equals its own span. □

**Proposition 6.10.** *Let $T : V \to W$ be a linear transformation and let $U \leq V$. If $U$ is finite dimensional, then $T(U)$ is finite dimensional, and*

$$\dim(T(U)) \leq \dim(U).$$

*Proof.* Suppose that $U$ is finite dimensional, and let $X \subset U$ be a basis for $U$. Then $T(U) = T(\text{span}(X)) = \text{span}(T(X))$, so $T(U)$ is spanned by the finite set $T(X)$. If $Y$ is a basis for $T(U)$, then $|Y| \leq |T(X)| \leq |X|$, that is, $\dim(T(U)) \leq \dim(U)$. □

**Example 6.11.** Let $T : \mathbb{R}^3 \to \mathbb{R}^2$ be given by $T(x, y, z) = (x, y)$.
Let $U$ be $yz$-plane; then $T(U)$ is the $y$-axis.

**Proposition 6.12.** *Let $T : V \to W$ be a linear transformation and let $U_1, U_2 \leq V$. Then $T(U_1 + U_2) = T(U_1) + T(U_2)$.*

*Proof.* We write this proof as a chain of logical equivalences.

$$\begin{aligned} w \in T(U_1 + U_2) &\Leftrightarrow w = T(u_1 + u_2) && \text{for some } u_1 \in U_1, u_2 \in U_2 \\ &\Leftrightarrow w = T(u_1) + T(u_2) && \text{because } T \text{ is linear} \\ &\Leftrightarrow w \in T(U_1) + T(U_2) && \text{by definition of image.} \end{aligned}$$

□

**Proposition 6.13.** *Let $T : V \to W$ be a linear transformation and let $U \leq W$. Then $T^{-1}(U) \leq V$.*

*Proof.* We verify the three properties of a subspace.

**(S0)** Since $T(0_V) = 0_W \in U$, we see that $0_V \in T^{-1}(U)$.

**(S1)** Let $v_1, v_2 \in T^{-1}(U)$. Then $T(v_1), T(v_2) \in U$. Thus $T(v_1) + T(v_2) \in U$ because $U$ is a subspace. But $T(v_1) + T(v_2) = T(v_1 + v_2)$ because $T$ is a linear transformation, which shows that $v_1 + v_2 \in T^{-1}(U)$.

**(S2)** Let $v \in T^{-1}(U)$ and $a \in \mathbb{R}$. Then $T(v) \in U$, so $aT(v) \in U$, whence $T(av) \in U$. Thus $av \in T^{-1}(U)$. □

**Example 6.14.** Let $T : \mathbb{R}^3 \to \mathbb{R}^2$ be given by $T(x, y, z) = (x, y)$. Let $U = \{0\}$. Then $T^{-1}(U)$ is the $z$-axis.

## 4. Kernels and Injectivity

**Definition 6.15.** Let $T : V \to W$ be a linear transformation.

The *kernel* of $T$ is the subset of $V$ denoted by $\ker(T)$ and defined as

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

*Remark* 6.4. Note that an alternate way of writing this is $\ker(T) = T^{-1}(0)$.

**Proposition 6.16.** *Let $T : V \to W$ be a linear transformation. Then $\ker(T) \leq V$.*

*Proof.* Since $\{0\} \leq W$, this follows from Proposition 6.13. □

**Proposition 6.17.** *Let $T : V \to W$ be a linear transformation. Then $T$ is injective if and only if $\ker(T) = \{0\}$.*

*Proof.*

($\Rightarrow$) Suppose that $T$ is injective. Let $v \in \ker(T)$. Then $T(v) = 0_W$; but $T(0_V) = 0_W$, so since $T$ is injective, $v = 0_V$. Thus $\ker(V) = \{0_V\}$.

($\Leftarrow$) Suppose that $\ker(T) = \{0_W\}$. Let $v_1, v_2 \in V$ such that $T(v_1) = T(v_2)$. Then $T(v_1) - T(v_2) = 0_W$, and since $T$ is linear, $T(v_1 - v_2) = 0_W$. Since $\ker(T) = \{0_V\}$, we have $v_1 - v_2 = 0_V$. Thus $v_1 = v_2$ Therefore $T$ is injective. □

**Proposition 6.18.** *Let $T : V \to W$ be a linear transformation. Then $T$ is injective if and only if for every independent subset $X \subset V$, $T(X)$ is independent.*

*Proof.* We prove the contrapositive in both directions.

($\Rightarrow$) Suppose that $X \subset V$ is independent but that $T(X)$ is dependent. Then there exists a nontrivial dependence relation

$$a_1T(x_1) + \cdots + a_nT(x_n) = 0,$$

where $x_i \in X$ and $a_i \in \mathbb{R}$, not all zero. Then $T(\sum_{i=1}^n a_i x_i) = 0$, so $\sum_{i=1}^n a_i x_i$ is a nontrivial member of $\ker(T)$. Thus $T$ is not injective.

($\Leftarrow$) Suppose that $T$ is not injective. Then its kernel is nontrivial, so there exists an nonzero vector $v \in V$ such that $T(v) = 0$. Since $v \neq 0$, the set $\{v\}$ is independent. But its image $T(v)$ is dependent. □

**Proposition 6.19.** *Let $T : V \to W$ be a linear transformation. Let $X$ be a basis for $V$. Then $T$ is injective if and only if $T(X)$ is a basis for $T(V)$.*

*Proof.* Suppose $X$ spans $V$. Then

$$T(V) = T(\mathrm{span}(X)) = \mathrm{span}(T(X)).$$

Now the result follows immediately from the preceding proposition. □

**Corollary 6.20.** *Let $T : V \to W$ be an injective linear transformation. Let $X$ be a basis for $V$. Then*

**(a)** *$T(X)$ is a basis for $T(V)$;*
**(b)** $\dim(V) = \dim(T(V))$.

## 5. Kernels and Cosets

**Definition 6.21.** Let $V$ be a vector space and let $W \leq V$.

A *coset* (or "translation") of $W$ is a subset of $V$ of the form

$$x + W = \{x + w \mid w \in W\},$$

where $x \in V$.

**Example 6.22.** Let $Z = \{(x, y, z) \in \mathbb{R}^3 \mid x = y = 0\}$. Then $Z$ is commonly known as the "$z$-axis". A coset of $Z$ is a set of the form $v + Z$, where $v \in \mathbb{R}^3$. In fact, we can always select $v$ to lie in the $xy$-plane; we see that $v + Z$ is a vertical line in $\mathbb{R}^3$, parallel to the $z$-axis, translated away by the vector $v$.

**Proposition 6.23.** *Let $V$ be a vector space and let $W \leq V$. Let $v_1, v_2 \in V$. Then*

- **(a)** $V = \cup_{v \in V}(v + W)$*;*
- **(b)** $(v_1 + W) \cap (v_2 + W) \neq \varnothing \Rightarrow (v_1 + W) = (v_2 + W)$.

*Proof.* Exercise. ☐

**Proposition 6.24.** *Let $V$ be a vector space and let $W \leq V$.*
*Then $v_1 + W = v_2 + W$ if and only if $v_2 - v_1 \in W$.*

*Proof.* Exercise. ☐

**Proposition 6.25.** *Let $V$ and $W$ be vector spaces. Let $T : V \to W$ be a linear transformation. Let $w \in T(V)$ and let $v \in T^{-1}(w)$. Then*

$$T^{-1}(w) = v + \ker(T);$$

*in words, the preimage of $w$ is a coset of the kernel.*

*Proof.* We show that each set is contained in the other.

Let $x \in T^{-1}(w)$. Then $T(x) = w$. Since $v \in T^{-1}(w)$, we have $T(v) = w$. Thus $T(x - v) = T(x) - T(v) = w - w = 0$, so $x - v \in \ker(T)$. Then $x = v + (x - v) \in v + \ker(T)$.

Let $x \in v + \ker(T)$. Then $x = v + k$ where $k \in \ker(T)$, so $T(x) = T(v + k) = T(v) + T(k) = w + 0 = w$, so $x \in T^{-1}(w)$. ☐

**Example 6.26.** A system of $m$ equations in $n$ variables gives a matrix equation

$$Ax = b.$$

The matrix $A$ corresponds to a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ given by $T(x) = Ax$. The solution set of the homogeneous equation $Ax = 0$ is the kernel of $T$. If $v$ is a particular solution to $Ax = b$, then the solution set is the homogeneous solution offset by the particular solution $v$.

## 6. Kernels and Direct Sums

**Definition 6.27.** Let $V$ be a vector space and let $U_1, U_2 \leq V$. We say that $V$ is a *direct sum* of $U_1$ and $U_2$, and write $V = U_1 \oplus U_2$, if

**(D1)** $V = U_1 + U_2$;
**(D2)** $U_1 \cap U_2 = \{0\}$.

**Proposition 6.28.** *Let $V$ be a vector space and let $X$ be a basis for $V$. Let $Y_1 \subset X$ and let $Y_2 = X \smallsetminus Y_1$. Let $U_1 = \operatorname{span}(Y_1)$ and let $U_2 = \operatorname{span}(Y_2)$. Then $V = U_1 \oplus U_2$.*

*Proof.* We verify the two properties of direct sum.

**(D1)** We always have $U_1 + U_2 \leq V$; we need to show that $V \subset U_1 + U_2$. If $v \in V$, then $V$ is a linear combination from $X$ because $X$ spans $V$. Since $X = Y_1 \cup Y_2$, $v$ can be written as a linear combination of some vectors from $Y_1$ plus a linear combination some vectors from $Y_2$. Such an element is in $U_1 + U_2$.

**(D2)** Let $v \in U_1 \cap U_2$. Then $v$ is a linear combination from $Y_1$ and also $v$ is a linear combination from $Y_2$. The difference of these is a linear combination from $X$ which equals zero; since $X$ is linearly independent, all of the coefficients must be zero. Thus $v = 0$. □

**Proposition 6.29.** *Let $V$ be a vector space. Let $U_1, U_2 \leq V$ such that $V = U_1 \oplus U_2$. Let $Y_1$ be a basis for $U_1$ and $Y_2$ be a basis for $U_2$. Then $Y_1 \cup Y_2$ is a basis for $V$.*

*Proof.* Exercise. □

**Corollary 6.30.** *Let $V$ be a finite dimensional vector space and let $U_1, U_2 \leq V$ such that $V = U_1 \oplus U_2$. Then $\dim(V) = \dim(U_1) + \dim(U_2)$.*

**Example 6.31.** Let $V = \mathbb{R}^3$. Let $U_1$ be a plane through the origin in $\mathbb{R}^3$ and let $U_2$ be a line through the origin in $\mathbb{R}^3$. Then $V = U_1 \oplus U_2$ if and only if the line $U_2$ does not lie on the plane $U_1$.

**Proposition 6.32.** *Let $T : V \to W$ be a linear transformation. Let $K = \ker(T)$. Then*

**(a)** *there exists $U \leq V$ such that $V = K \oplus U$;*
**(b)** *$T\restriction_U : U \to W$ is injective.*

*Proof.* Let $Y_1$ be a basis for $K$ and let $X$ be a completion of $Y_1$ to a basis for $X$. Let $Y_2 = X \smallsetminus Y_1$. Let $U = \operatorname{span}(Y_2)$. Then by Proposition 6.28, $V = K \oplus U$. This proves **(a)**.

Recall that $T\restriction_U : U \to W$ is the restriction of $T$ to the set $U$; that is, we only consider what $T$ does to elements of $U$. Let $u \in \ker(T\restriction_U)$. Then $T(u) = 0$, so $u \in K$. Thus $u \in K \cap U = \{0\}$, so $u = 0$. Thus the kernel of $T\restriction_U$ is trivial, so $T\restriction_U$ is injective by Proposition 6.17. □

## 7. Rank and Nullity

**Definition 6.33.** Let $V$ be a finite dimensional vector space and let $T : V \to W$ be a linear transformation. Let $\text{img}(T) = T(V)$ denote the image of $T$.

The *rank* of $T$ is the dimension of the image of $T$: $\text{rank} = \dim(\text{img}(T))$.

The *nullity* of $T$ is the dimension of the kernel of $T$: $\text{nullity} = \dim(\ker(T))$.

**Theorem 6.34. (Rank plus Nullity Theorem)**
*Let $V$ be a finite dimensional vector space and let $T : V \to W$ be a linear transformation. Then* $\dim(V) = \dim(\ker(T)) + \dim(\text{img}(T))$.

*Proof.* Let $K = \ker(T)$. By Proposition 6.32 **(a)**, there exists a subspace $U \leq V$ such that $V = K \oplus U$. Thus $\dim(V) = \dim(K) + \dim(U)$. By Proposition 6.32 **(b)**, the linear transformation $T \restriction_U : U \to W$ is injective, so $\dim(T(U)) = \dim(U)$. Thus

$$\dim(V) = \dim(K) + \dim(U) = \dim(\ker(T)) + \dim(\text{img}(T)).$$

□

**Corollary 6.35.** *Let $V$ and $W$ be a finite dimensional vector spaces of the same dimension. Let $T : V \to V$ be a linear transformation. Then $T$ is injective if and only if $T$ is surjective.*

*Proof.* Exercise. □

## 8. Composition of Linear Transformations

**Proposition 6.36.** *Let $S : U \to V$ and $T : V \to W$ be linear transformations. Then $T \circ S : U \to W$ is a linear transformation.*

*Proof.* We verify the two properties of a linear transformation.

**(T1)** Let $u_1, u_2 \in U$. Then

$$T(S(u_1 + u_1)) = T(S(u_1) + S(u_2)) = T(S(u_1)) + T(S(u_2)).$$

**(T2)** Let $u \in U$ and $a \in \mathbb{R}$. Then

$$T(S(au)) = T(aS(u)) = aT(S(u)).$$

□

**Example 6.37.** Let $S : \mathbb{R}^2 \to \mathbb{R}^2$ be dilation by $a$ and let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be dilation by $b$. Then $T \circ S : \mathbb{R}^2 \to \mathbb{R}^3$ is dilation by $ab$.

**Example 6.38.** Let $S : \mathbb{R}^2 \to \mathbb{R}^2$ be rotation by $\alpha$ degrees and let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be rotation by $\beta$ degrees. Then $T \circ S : \mathbb{R}^2 \to \mathbb{R}^3$ is rotation by $\alpha + \beta$ degrees.

**Definition 6.39.** Let $T : V \to W$ be a linear transformation.

We say that $T$ is *invertible* if there exists a linear transformation $S : W \to V$ such that $S \circ T = \mathrm{id}_V$ and $T \circ S = \mathrm{id}_W$. Such an $S$ is called the *inverse* of $T$; it is necessarily unique, and is denoted by $T^{-1}$.

**Proposition 6.40.** *Let $T : V \to W$ be a linear transformation. Then $T$ is invertible if and only if $T$ is bijective.*

*Proof.* Exercise. □

**Proposition 6.41.** *Let $T : U \to V$ be a linear transformation. Let $S : V \to W$ be an injective linear transformation. Then* $\ker(S \circ T) = \ker(T)$.

*Proof.* Exercise. □

*Remark* 6.5. Let $A$ be an $m \times n$ matrix and consider the matrix equation $Ax = 0$, where 0 is the zero $n \times 1$ column vector. The solution to this equation is the kernel of the corresponding linear transformation $T_A : \mathbb{R}^n \to \mathbb{R}^m$.

Let $B$ be $A$ in reduced row echelon form. Row reduction of $A$ corresponds to warping $m$-space by invertible transformations. Then $\ker(T_B) = \ker(T_A)$, because $B = UA$, where $U$ is a product of elementary invertible matrices and so it is invertible; then $T_U$ is injective. Therefore $\ker(T_B) = \ker(T_U \circ T_A) = \ker(T_A)$.

Moreover, the basic columns of $B$ are clearly linearly independent. Then the pullback of these basic columns via $U^{-1}$ gives linearly independent vectors in $\mathrm{img}(T) = T_A(\mathbb{R}^n)$, the image of $T_A$.

A basis for the kernel of $T_A$ is given by modifying the free columns of $B$ in the manner prescribed in solving $Ax = 0$.

A basis for the image of $T_A$ is given by the columns of $A$ corresponding to the basic columns of $B$.

**Example 6.42.** Let $e_1, \dots, e_4$ be the standard basis vectors for $\mathbb{R}^4$. Let

$$v_1 = (2, -4, 4), v_2 = (1, -1, 3), v_3 = (3, -7, 5), v_4 = (0, 2, 5) \in \mathbb{R}^3.$$

Let $T : \mathbb{R}^4 \to \mathbb{R}^3$ be the unique linear transformation given by $T(e_i) = v_i$. Find a basis for the image and the kernel of $T$.

*Solution.* Set

$$A = \begin{bmatrix} 2 & 1 & 3 & 0 \\ -4 & -1 & -7 & 2 \\ 4 & 3 & 5 & 5 \end{bmatrix}.$$

Row reduce $A$; the corresponding reduced row echelon matrix is

$$B = \begin{bmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

The basic variables are $x_1$, $x_2$, and $x_4$. The free variable is $x_3$. So the solution to $Ax = 0$ is

$$x_3 \begin{bmatrix} -2 \\ 1 \\ 1 \\ 0 \end{bmatrix};$$

thus $\{(-2, 1, 1, 0)\}$ is a basis for $\ker(T)$, and $\{(2, -4, 4), (1, -1, 3), (0, 2, 5)\}$ is a basis for $\mathrm{img}(T)$, the image of $T$. □

*Remark* 6.6. Let $Y = \{v_1, \dots, v_n\} \in \mathbb{R}^m$. We wish to determine whether or not the set $Y$ is independent. If $n > m$, we know they cannot be independent, so assume that $n \leq m$.

Form the matrix $A = [v_1 \mid \cdots \mid v_n]$. Corresponding to $A$ is a linear transformation $T_A : \mathbb{R}^n \to \mathbb{R}^m$. We know that $n = \dim(\mathbb{R}^n) = \dim(\ker(T_A)) + \dim(\mathrm{img}(T_A))$. Now $X$ is independent if and only if there span in $\mathbb{R}^m$ is a vector space of dimension $n$. This span is exactly $\mathrm{img}(T_A)$. Thus $X$ is independent if and only if $\dim(\mathrm{img}(T_A)) = n$. This is the case if and only if $\dim(\ker(T_A)) = 0$.

Row reduce $A$ to obtain a matrix $B$; only forward elimination is necessary. Now $X$ is dependent if and only if $B$ has a free column, which is the case if and only if $B$ has a zero row (since $n \leq m$).

## 9. Isomorphisms

**Definition 6.43.** Let $V$ and $W$ be vector spaces.

An *isomorphism* from $V$ to $W$ is a bijective linear transformation $T : V \to W$. We say that $V$ is *isomorphic* to $W$, and write $V \cong W$, if there exists an isomorphism $T : V \to W$.

**Proposition 6.44.** *Let $V$ be a vector space.*
*Then* $\mathrm{id}_V : V \to V$ *is an isomorphism.*

*Proof.* Clear. □

**Proposition 6.45.** *Let $T : V \to W$ be an isomorphism.*
*Then $T^{-1} : W \to V$ is an isomorphism.*

*Proof.* Since $T$ is bijective, $T^{-1} : W \to V$ is a function. We verify the properties of a linear transformation.

**(T1)** Let $w_1, w_2 \in W$. Since $T$ is bijective, there exist unique elements $u_1, u_2 \in U$ such that $T(u_1) = w_1$ and $T(u_2) = w_2$. Now $T(u_2 + u_2) = T(u_1) + T(u_2) = w_1 + w_2$, so $T^{-1}(w_1 + w_2) = u_1 + u_2 = T^{-1}(w_1) + T^{-1}(w_2)$.

**(T2)** Let $w \in W$ and $a \in \mathbb{R}$. There exists a unique element $u \in U$ such that $T(u) = w$. Then $T(au) = aT(u) = aw$, so $T^{-1}(aw) = au = aT^{-1}(w)$. □

**Proposition 6.46.** *Let $S : U \to V$ and $T : V \to W$ be isomorphisms.*
*Then $T \circ S : U \to W$ is an isomorphism.*

*Proof.* We have seen that the composition of linear transformations is linear, and we always have that the composition of bijective functions is bijective. □

*Remark* 6.7. Let $U$, $V$, and $W$ be vector spaces. Then

**(a)** $V \cong V$;
**(b)** $V \cong W \Leftrightarrow W \cong V$;
**(c)** $U \cong V$ and $V \cong W \Rightarrow U \cong W$.

This says that isomorphism is an *equivalence relation*.

**Proposition 6.47.** *Let $T : V \to W$ be a linear transformation. Let $X$ be a basis for $V$. Then $T$ is an isomorphism if and only if $T(X)$ is a basis for $W$.*

*Proof.*

($\Rightarrow$) Suppose that $T$ is an isomorphism. Then $T$ is injective, so by Proposition 6.19, $T(X)$ is a basis for $T(V)$. But $T$ is also surjective, so $T(V) = W$, and the result follows.

($\Leftarrow$) Suppose that $T(X)$ is a basis for $W$. Then $T$ is clearly surjective, and by Proposition 6.19, $T$ is also injective. Thus $T$ is an isomorphism. □

*Remark* 6.8. In light of Proposition 6.7, we may construct an isomorphism between spaces by sending a basis to a basis.

**Definition 6.48.** Let $V$ be a finite dimensional vector space of dimension $n$.

An *ordered basis* for $V$ is an ordered $n$-tuple $(x_1, \ldots, x_n) \in V^n$ of linearly independent vectors from $V$.

*Remark* 6.9. Note that if $(x_1, \ldots, x_n)$ is an ordered basis, then $X = \{x_1, \ldots, x_n\}$ is a basis. With this understanding, we may say: "let $X$ be an ordered basis", by which we mean that $X$ is the basis which corresponds to an ordered basis.

**Theorem 6.49.** *Let $V$ be a finite dimensional vector space of dimension $n$. Let $X = \{x_1, \ldots, x_n\}$ be an ordered basis for $V$. Define a linear transformation*

$$\Gamma_X : V \to \mathbb{R}^n \quad \text{by} \quad \Gamma_X(x_i) = e_i.$$

*Then $\Gamma_X$ is an isomorphism.*

*Description.* We have already essentially proven this, so let us describe it in more detail.

Every element of $V$ may be written *in a unique way* as a linear combination of elements from $X$: if $v \in V$, then $v = \sum_{i=1} a_i x_i$ for some real numbers $a_1, \ldots, a_n$. Then

$$\Gamma_X(v) = \sum_{i=1}^n a_i \Gamma_X(x_i) = \sum_{i=1}^n \sum_{i=1}^n a_i e_i = (a_1, \ldots, a_n);$$

this is the linear transformation that sends the basis $X$ of $V$ to the standard basis for $\mathbb{R}^n$, whose existence, uniqueness, and linearity is guaranteed by Proposition 6.7. It is an isomorphism by Proposition 6.47. □

**Corollary 6.50.** *Let $V$ and $W$ be vector spaces of dimension $n$. Then $V \cong W$.*

*Proof.* Every finite dimensional vector space has a basis. Let $X$ be an ordered basis for $V$ and let $Y$ be an ordered basis for $W$. Since $\Gamma_Y : W \to \mathbb{R}^n$ is an isomorphism, it is invertible, and its inverse is also an isomorphism. Since the composition of isomorphisms is an isomorphism, we see that

$$\Gamma_Y^{-1} \circ \Gamma_X : V \to W$$

is an isomorphism, so $V \cong W$. □

*Remark* 6.10. Even though two vector spaces of the same dimension are isomorphic, there are many *ways* in which they are isomorphic. Indeed, each basis $X$ for $V$ gives a *different* isomorphism $\Gamma_X : V \to \mathbb{R}^n$. Controlling this is one of the challenges of linear algebra.

## 10. Computing Linear Transformations via Matrices

*Remark* 6.11. Let $V$ be a vector space of dimension $n$ and let $W$ be a vector space of dimension $m$. Let $T : V \to W$ be a linear transformation. If we know a basis for $V$ and for $W$, we can use matrices to compute information about $T$.

Let $X$ be a basis for $V$ and let $Y$ be a basis for $W$. Then $\Gamma_X : V \to \mathbb{R}^n$ is an isomorphism and $\Gamma_Y : W \to \mathbb{R}^m$ is an isomorphism. These isomorphisms pick off the coefficients of any vector in $V$ and $W$ and allow us to think of them as vectors in $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively. Actually, what we are doing is defining a transformation $S : \mathbb{R}^n \to \mathbb{R}^m$ given by $S = \Gamma_Y \circ T \circ \Gamma_X$. In this case,

$$T = \Gamma_Y^{-1} \circ S \circ \Gamma_X.$$

This can be written in diagram form:

$$\begin{array}{ccc} V & \xrightarrow{\;T\;} & W \\ {\scriptstyle \Gamma_X}\big\downarrow & & \big\downarrow{\scriptstyle \Gamma_Y} \\ \mathbb{R}^n & \xrightarrow[S]{} & \mathbb{R}^m \end{array}$$

This says that to compute $T(v)$, it suffices to push $v$ into $\mathbb{R}^n$ via $u = \Gamma_X(v)$, compute $S(u)$, then pull this result back to $W$ via $\Gamma_Y$.

But $S : \mathbb{R}^n \to \mathbb{R}^m$ corresponds to a matrix $A$, and we can compute $Au$ by matrix multiplication. This also allows us to compute kernels, images, and so forth via matrices.

**Example 6.51.** Let $v_1 = (1,0,0,0), v_2 = (1,0,1,0), v_3 = (1,0,0,1) \in \mathbb{R}^4$. Let $V$ be the subspace of $\mathbb{R}^4$ spanned by $\{v_1, v_2, v_3\}$; these form a basis for $V$. Let $W = \mathbb{R}^2$ Let $w_1 = (1,2), w_2 = (-1,0), w_3 = (3,2) \in W$. Let $T : V \to W$ be the unique linear transformation given $T(v_i) = w_i$. Find a basis for the kernel of $T$.

*Solution.* Let $e_1, e_2, e_3$ be the standard basis vectors for $\mathbb{R}^3$. Let $S : V \to \mathbb{R}^3$ be given by $T(v_i) = e_i$. Then $S$ is an isomorphism. Let $R : \mathbb{R}^3 \to \mathbb{R}^2$ be given by $T(e_i) = w_i$. The matrix for $R$ is

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

Row reduce $A$ to get

$$UA = \begin{bmatrix} 1 & 0 & 7 \\ 0 & 1 & 4 \end{bmatrix}.$$

The kernel of $R$ is spanned by the vector $(-7, -4, 1)$.

Now $T = S^{-1}RS$. Thus $ST = RS$. Then

$$\ker(T) = \ker(ST) = \ker(RS) = S^{-1}(\ker(R)).$$

Thus to find $\ker(T)$, pull the vector $(-7,-4,1)$ back through $S$ (find its preimage). This is $-7(1,1,0,0) - 4(1,0,1,0) + (1,0,0,1) = (-10,-7,-4,1)$. The kernel of $T$ is the span of this vector. $\square$

## 11. Vector Space of Linear Transformations

**Proposition 6.52.** *Let $V$ and $W$ be vector spaces and set*

$$\mathcal{L}(V,W) = \{T : V \to W \mid T \text{ is linear}\}.$$

*Let $S : V \to W$ and $T : V \to W$ be linear transformations. Let $a \in \mathbb{R}$. Define the* sum *$S + T$ and the* scalar product *$aT$ by their effect on any vector $v \in V$:*

- *$(S + T)(v) = S(v) + T(v)$ ;*
- *$(aT)(v) = aT(v)$ .*

*Then*

**(a)** *$S + T : V \to W$ and $aT : V \to W$ are linear transformations;*
**(b)** *$\mathcal{L}(V,W)$ is a vector space.*

*Reason.* The verification that $S + T$ and $aT$ are linear transformations is straightforward.

The proof that $\mathcal{L}(V,W)$ is a vector space comes down to the fact that all of the properties **(V1)** through **(V8)** of the vector space $W$ work pointwise on functions into $W$. □

*Remark* 6.12. The vector space $\mathcal{M}_{m\times n}$ of $m \times n$ matrices is isomorphic to $\mathbb{R}^{mn}$, as one expects. But also, we know that matrices correspond to linear transformations of cartesian spaces; we now describe this correspondence in terms of isomorphism.

**Proposition 6.53.** *Let $T_{ij} : \mathbb{R}^n \to \mathbb{R}^m$ be given by $T_{ij}(e_j) = e_i$ and $T_{ij}(e_k) = 0$ if $k \neq j$. Then $\{T_{ij} \mid i = 1,\dots,m;\ j = 1,\dots,n\}$ is a basis for $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$.*

*Reason.* One can show that this set is linearly independent and spans. □

**Proposition 6.54.** *Define a function*

$$\Omega_{m\times n} : \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m) \to \mathcal{M}_{m\times n} \quad \text{by} \quad \Omega_{m\times n}(T) = A_T,$$

*where $A_T = [T(e_1) \mid \cdots \mid T(e_n)]$ is the matrix corresponding to a transformation $T : \mathbb{R}^n \to \mathbb{R}^m$. Then $\Omega_{m\times n}$ is an isomorphism.*

*Reason.* The function $\Omega_{m\times n}$ sends the basis $\{T_{ij}\}$ for $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ to the basis $\{M_{ij}\}$ for $\mathcal{M}_{m\times n}$. □

**Proposition 6.55.** *Let $V$ and $W$ be finite dimensional vector spaces. Let $X = \{x_1,\dots,x_n\}$ be an ordered basis for $V$ and $Y = \{y_1,\dots,y_m\}$ be an ordered basis for $W$. Define a function*

$$\Omega_{Y,X} : \mathcal{L}(V,W) \to \mathcal{M}_{m\times n} \quad \text{by} \quad \Omega_{Y,X}(T) = A_S,$$

*where $S = \Gamma_Y \circ T\Gamma_X^{-1}$ and $A_S = [S(e_1) \mid \cdots \mid S(e_n)]$ is the matrix corresponding to $S$. Then $\Omega_{Y,X}$ is an isomorphism.*

*Reason.* In a manner similar to the case where $V = \mathbb{R}^n$ and $W = \mathbb{R}^m$, one can find a basis for $\mathcal{L}(V,W)$ that is sent by $\Omega_{Y,X}$ to the basis $\{M_{ij}\}$ for $\mathcal{M}_{m\times n}$. □

## 12. Linear Operators

**Definition 6.56.** Let $V$ be a vector space.
A *linear operator* on $V$ is a linear transformation $T : V \to V$.
Let $\mathcal{L}(V)$ denote the set of all linear operators on $V$.

Let $V$ be a vector space and let $S, T : V \to V$ be a linear operators. Then the composition $T \circ S : V \to V$ is a linear operator. Let us drop the $\circ$ from the notation and think of composition of linear operators as multiplication in the set $\mathcal{L}(V)$: thus $TS$ is the transformation $T \circ S$.

This multiplication distributes over addition of operators:

$$T(S + R) = TS + TR; \quad (T + S)R = TR + SR.$$

Thus $\mathcal{L}(V)$ is a set which comes equipped with two operations, addition of transformations and multiplication of transformations. The additive identity of this set is the zero transformation (which we denote by 0), and the multiplicative identity is the identity transformation $\mathrm{id}_V$, which we now denote by 1. Every transformation $T$ has an additive inverse $-T$. A transformation $T$ has a multiplicative inverse $T^{-1}$ if and only if $T$ has a trivial kernel.

Let $a \in \mathbb{R}$. Define $N_a : V \to V$ to be dilation by $a$: $N_a(v) = av$ for all $v \in V$. Then $N_a$ is a linear operator. Note that $N_a$ commutes with any other operator:

$$N_a T = T N_a.$$

Also note that $N_a T$ is exactly the transformation which we previously described by $aT$. When $N_a$ occurs on the left, we drop the $N$ from the notation, and simply write $aT$ instead of $N_a T$.

Let $T^2 = TT$, $T^3 = TTT$, and in general, let $T^n$ denote the composition of $T$ with itself $n$ times. This is $T$ applied to the space $V$ over and over. For example, if $T$ is rotation of $\mathbb{R}^2$ by an angle of 45 degrees, then $T^4$ is rotation by 180 degrees and $T^8$ is the identity transformation $I = \mathrm{id}_V$.

Let $T : V \to V$ be a linear operator. We see that any polynomial in $T$

$$L = T^n + a_{n-1}T^{n-1} + \cdots + a_1 T + a_0$$

is a linear operator. Its effect on $v \in V$ is given by distributing $v$ into the polynomial:

$$L(v) = T^n(v) + a_{n-1}T^{n-1}(v) + \cdots + a_1 T(v) + a_0.$$

## 13. Linear Algebra and Differential Equations

Consider the differential equation

$$y'' + by' + cy = g(t),$$

where $b, c \in \mathbb{R}$ and $g(t)$ is a smooth function on some open interval $I \subset \mathbb{R}$. To solve this differential equation means to find all smooth functions $y$ such that the function $y'' + by' + cy$ is equal to the function $g(t)$. We use linear algebra to analyze this situation.

Let $I \subset \mathbb{R}$ be an open interval and let $\mathcal{D}(I)$ be the set of smooth real valued functions defined by $I$; this is a vector space under addition and scalar multiplication of functions. Define $D : \mathcal{D}(I) \to \mathcal{D}(I)$ by $D(f) = f'$, the derivative of $f$. Then $D$ is a linear transformation. Any polynomial in $D$ is also a linear transformation, called a *differential operator*. Note that the kernel of $D$ is the set of all constant functions on $I$. This is a one dimensional subspace of $\mathcal{D}(I)$, spanned by the function $f(t) = 1$.

Let $b, c \in \mathbb{R}$ and let $g \in \mathcal{D}(I)$. Define a function

$$L : \mathcal{D}(I) \to \mathcal{D}(I) \quad \text{by} \quad L[y] = y'' + by' + cy.$$

Then $L$ is a differential operator:

$$L = D^2 + bD + c.$$

The general solution to the differential equation

$$y'' + by' + cy = g(t)$$

is of the form $y = y_h + y_p$, where $y_h$ is the general solution to the homogeneous differential equation $L[y] = 0$ and $y_p$ is a particular solution to the differential equation $L[y] = g(t)$. This comes from the fact that the solution to the homogeneous equation is the kernel of $L$, and the solution to the nonhomogeneous equation is a coset of this kernel.

One may attempt to solve the homogeneous differential equation $L[y] = 0$ by factoring the linear operator $L$:

$$L = (D - r_1)(D - r_2),$$

where $r_i = \frac{1}{2}(-b \pm \sqrt{b^2 - 4c})$ are the roots of the polynomial $L$. Now any solution to $(D - r_2)[y] = 0$ is also a solution to $L[y] = 0$, since $(D - r_1)[0] = 0$. Since $D - r_1$ and $D - r_2$ commute, the same can be said about solutions to $(D - r_1)[y]$. But this differential equation is very easy to solve:

$$(D - r)[y] = 0 \Leftrightarrow y' = ry \Leftrightarrow \log y = r + C \Leftrightarrow y = ke^{rt},$$

where $k = e^C$ is an arbitrary constant of integration. One can show that $\ker(L) = \operatorname{span}\{e^{r_1 t}, e^{r_2 t}\}$.

## 14. Exercises

**Exercise 6.1.** Let $V$ be a vector space.
Let $U_1, U_2 \leq V$ such that $V = U_1 \oplus U_2$. Let $Y_1$ be a basis for $U_1$ and $Y_2$ be a basis for $U_2$. Show that $Y_1 \cup Y_2$ is a basis for $V$.

**Exercise 6.2.** Let $V$ be a vector space and let $W \leq V$. Let $v_1, v_2 \in V$.

- **(a)** Show that $V = \cup_{v \in V}(v + W)$.
- **(b)** Show that $(v_1 + W) \cap (v_2 + W) \neq \varnothing \Rightarrow (v_1 + W) = (v_2 + W)$.

**Exercise 6.3.** Let $V$ be a vector space and let $W \leq V$.
Show that $v_1 + W = v_2 + W$ if and only if $v_2 - v_1 \in W$.

**Exercise 6.4.** Let $V$ be a finite dimensional vector space.
Let $U \leq V$ and let $T : V \to V$ be a linear transformation.

- **(a)** Show that $U = V$ if and only if $\dim(U) = \dim(V)$.
- **(b)** Show that $T$ is injective if and only if $T$ is surjective.

**Exercise 6.5.** Let $T : V \to W$ be a linear transformation.
Show that $T$ is invertible if and only if $T$ is bijective.

**Exercise 6.6.** Let $T : U \to V$ be a linear transformation.
Let $S : V \to W$ be an injective linear transformation.
Show that $\ker(S \circ T) = \ker(T)$.

**Exercise 6.7.** Let $T : V \to W$ be a linear transformation and let $U_1, U_2 \leq V$. In each case, prove or give a counterexample.

- **(a)** $T(U_1 \cap U_2) = T(U_1) \cap T(U_2)$;
- **(b)** $V = U_1 \oplus U_2 \Rightarrow T(V) = T(U_1) \oplus T(U_2)$.

**Exercise 6.8.** Let $T : V \to W$ be a linear transformation and let $U_1, U_2 \leq W$. In each case, prove or give a counterexample.

- **(a)** $T^{-1}(U_1 \cap U_2) = T^{-1}(U_1) \cap T^{-1}(U_2)$;
- **(b)** $W = U_1 \oplus U_2 \Rightarrow T^{-1}(W) = T^{-1}(U_1) \oplus T^{-1}(U_2)$.

**Exercise 6.9.** Let $\mathcal{P}_n$ denote the vector space of polynomial functions of degree less than or equal to $n$ with real coefficients:

$$\mathcal{P}_n = \{f(x) = a_0 + a_1 x + \cdots + a_n x^n \mid a_i \in \mathbb{R}\}.$$

Let $\Gamma : \mathcal{P}_4 \to \mathbb{R}^5$ be given by $\Gamma(x^{i-1}) = e_i$ for $i = 1, \ldots, 5$.
Let $D : \mathcal{P}_4 \to \mathcal{P}_4$ be given by $D(f) = \frac{df}{dx}$.
Let $T : \mathbb{R}^5 \to \mathbb{R}^5$ be given by $T = \Gamma \circ D \circ \Gamma^{-1}$.

- **(a)** Describe why $\Gamma$ is an isomorphism.
- **(b)** Find the matrix corresponding to the linear transformation $T$.
- **(c)** Find a basis for the image and the kernel of $T$.
- **(d)** Find a basis for the image and the kernel of $D$.

**Exercise 6.10.** Let $\mathcal{D}(\mathbb{R})$ denote the set of all smooth functions on $\mathbb{R}$.
Let $D : \mathcal{D}(\mathbb{R}) \to \mathcal{D}(\mathbb{R})$ be given by $D(f) = \frac{df}{dx}$.
Let $D^n : \mathcal{D}(\mathbb{R}) \to \mathcal{D}(\mathbb{R})$ denote $D$ composed with itself $n$ times.
Find $\ker(D^n)$; justify your answer.

CHAPTER 7

# Determinants and Eigenvectors

## 1. Transformations of a Vector Space

Let $V$ be a finite dimensional vector space of dimension $n$. Let $T : V \to V$ be a linear transformation. Transformations of this type (from a vector space into itself) are particularly interesting because they can be composed with themselves. Let $A$ be the corresponding matrix; then composing $T$ with itself corresponds to taking powers of $A$. Also, $T$ is an isomorphism if and only if $A$ is invertible. In this case, we think of $T$ as a warping of $n$-space.

Let $A$ be an $n \times n$ matrix given by $A = (a_{ij})_{ij}$.

We say that $A$ is *singular* if it is not invertible.

We say that $A$ is *scalar* if it is of the form $aI$, where $a \in \mathbb{R}$ and $I$ is the $n \times n$ identity matrix. This has the effect on $n$-space of dilating it by a factor of $a$ in every direction.

We say that $A$ is *diagonal* if all of its nondiagonal entries are zero, that is, if $a_{ij} = 0$ whenever $i \neq j$. This has the effect on $n$-space of expanding the $i^{\text{th}}$ axis by a factor of $a_{ii}$.

We say that $A$ is *upper triangular* if $a_{ij} = 0$ whenever $i > j$.

We say that $A$ is *lower triangular* if $a_{ij} = 0$ whenever $i < j$.

We say that $A$ is *triangular* if it is either upper triangular or lower triangular.

If $A$ is triangular and invertible, then $A$ can be reduced to a diagonal matrix by a sequence of row operations of type $\mathsf{R}_i + c\mathsf{R}_j$.

The process of Gaussian elimination shows that a matrix $A$ is invertible if and only if it is the product of elementary invertible matrices. Such a product is definitely invertible. On the other hand, if $A$ is invertible, we may find its inverse by row reducing the equation $AX = I$ to obtain $X = U$, where $U$ is the product of the matrices corresponding to the row operations we used. To examine this more closely, note that if $A$ is invertible, then for any $b \in \mathbb{R}^m$, there is a unique solution to the equation $Ax = b$, namely $x = A^{-1}b$, and this solution can be found by Gaussian elimination. In particular, if $x_i$ is the unique solution to $Ax = e_i$, then $A^{-1} = [x_1 \mid \cdots \mid x_n]$.

Thus if $A$ and $B$ are invertible matrices, we see that $AB$ is invertible if and only if both $A$ and $B$ are invertible.

## 2. Multilinear Functions

Let $V$ be a vector space and let $V^m$ denote the cartesian product of $V$ with itself $m$ times; this is the set of all ordered $m$-tuples of vectors from $V$.

A function $f : V^m \to \mathbb{R}$ is called *multilinear* if it is linear in each of its coordinates; that is, if

$$f(v_1, \dots, v_{i-1}, v_i + w_i, v_{i+1}, \dots, v_m) \\ = f(v_1, \dots, v_{i-1}, v_i, v_{i+1}, \dots, v_m) + f(v_1, \dots, v_{i-1}, w_i, v_{i+1}, \dots, v_m);$$

and

$$f(v_1, \dots, v_{i-1}, av_i, v_{i+1}, \dots, v_m) = af(v_1, \dots, v_{i-1}, v_i, v_{i+1}, \dots, v_m).$$

Let $f : V^n \to \mathbb{R}$ be multilinear and let $X$ be a basis for $V$. Then the value of $f$ is completely determined by the values of $f(x_{i_1}, \dots, x_{i_m})$, where the $x_i$'s range over all ordered choices of $m$ basis vectors.

A function $f : V^m \to \mathbb{R}$ is called *alternating* if exchanging positions changes the sign; that is, if

$$f(v_1, \dots, v_i, \dots, v_j, \dots, v_m) = -f(v_1, \dots, v_j, \dots, v_i, \dots, v_m).$$

Let $f : V^m \to \mathbb{R}$ be alternating. Suppose that two positions of an $n$-tuple are the same, say $v_i = v_j$. Then switching them gives the same value for $f$; but it must also give the negative value, since $f$ is alternating. Thus $f(v_1, \dots, v_m) = 0$ whenever two positions are the same.

**Example 7.1.** Let $V = \mathbb{R}^2$ and let $f : \mathbb{R}^2 \to \mathbb{R}$ be given by $f(v, w) = ad - bc$, where $v = (a, b)$ and $w = (c, d)$. Then $f$ is an alternating multilinear function.

Note that $f(e_1, e_2) = 1 \cdot 1 - 0 \cdot 0 = 1$.

Let $V$ be a finite dimensional vector space of dimension $n = m$ and let $f : V^n \to \mathbb{R}$ be an alternating multilinear function. Let $X = \{x_1, \dots, x_n\}$ be a basis for $V$. Then $f$ is completely determined by the value of $f(x_1, \dots, x_n)$. To see this, pick an arbitrary ordered $n$-tuple $(v_1, \dots, v_n)$. Write each of these as a linear combination of the vectors in $X$. Use multilinearity to break $f(v_1, \dots, v_n)$ into a sum of things of the form $f(x_{i_1}, \dots, x_{i_n})$. Use alternation to rearrange this into a sum of things of the form $\pm f(x_1, \dots, x_n)$.

A function $f : V^m \to \mathbb{R}$ is called *normalized* with respect to an ordered basis $\{x_1, \dots, x_n\}$ if $m = n$ and $f(x_1, \dots, x_n) = 1$.

**Proposition 7.2.** *Let $V$ be a vector space of dimension $n$ with ordered basis $X$. Then there exists a unique alternating multilinear function*

$$f : V^n \to \mathbb{R}$$

*which is normalized with respect to $X$.*

*Idea of Proof.* First one examines uniqueness. Suppose that $f$ and $g$ are alternating multilinear functions. By using multilinearity and alternation, one sees that the value of $f$ and $g$ on any order $n$-tuple $(v_1, \dots, v_n)$ of vectors is completely determined by their value on the ordered basis. This is a single real number. If the functions are normalized, then they must be the same.

Next one constructs a specific function which is multilinear, alternating, and normalized. We will do this momentarily. □

## 3. The General Determinant

Let $\mathcal{M}_{m\times n}$ be the set of all $m \times n$ matrices. If $m = n$, shorten this to $\mathcal{M}_n$.

A function $f : \mathcal{M}_n \to \mathbb{R}$ may be considered to be a multilinear function by considering its rows to be the coordinates of $V^n$, where $V = \mathbb{R}^n$.

**Proposition 7.3.** *There exists a unique alternating multilinear function*

$$\det : \mathcal{M}_n \to \mathbb{R},$$

*which is normalized with respect to the standard basis. This function is called the* determinant *function.*

We now describe how to construct such a function; the construction is inductive, which means that we construct the determinant of a $1\times 1$ matrix, and then construct the determinant of an $n \times n$ matrix in terms of determinants of $(n-1) \times (n-1)$ matrices.

Define the determinant of a $1 \times 1$ matrix to be the identity function (since a $1 \times 1$ function is merely a single real number).

Let $A = (a_{ij})_{ij}$ be an $n \times n$ matrix. Assume that the determinant of an $(n-1) \times (n-1)$ function has been defined.

Let $A_{ij}$ denote the matrix obtained from $A$ by deleting the $i^{\text{th}}$ row and the $j^{\text{th}}$ column. This matrix is called the $ij^{\text{th}}$ *minor* of $A$.

Let $a'_{ij} = \det(A_{ij})$. This number is called the $ij^{\text{th}}$ *cofactor* of $A$.

To compute the determinant of $A$, select any row or column of $A$. For each entry in the row of column, compute the cofactor of that entry. Then take the alternating sum of these cofactors. This process is called *expansion by minors.*

If we choose the $i^{\text{th}}$ row to expand along, the formula is

$$\det(A) = \sum_{j=1}^{n} (-1)^{j-1} a'_{ij}.$$

If we choose the $j^{\text{th}}$ column to expand along, the formula is

$$\det(A) = \sum_{i=1}^{n} (-1)^{i-1} a'_{ij}.$$

It is tedious and somewhat uninformative, but not terribly difficult, to use induction to show that this formula gives an alternating multilinear function which is normalized with respect to the standard basis. We move on.

## 4. Properties of the Determinant

Let $\det : \mathcal{M}_n \to \mathbb{R}$ be the unique function with the properties

**(a)** Multilinearity
**(b)** Alternation
**(c)** Normalization

From these properties, one can show

**(d)** If any row of $A$ is zero, then $\det(A) = 0$;
**(e)** If any two rows of $A$ are the same, then $\det(A) = 0$;
**(f)** If one row of $A$ is a scalar multiple of another, then $\det(A) = 0$;
**(g)** If $B$ is obtained from $A$ by a row operation of type $\mathsf{R}_i + c\mathsf{R}_j$, then $\det(B) = \det(A)$;
**(h)** If $A$ is diagonal, then $\det(A)$ is the product of the nonzero entries;
**(i)** If $A$ is triangular, then $\det(A)$ is the product of the diagonal entries.

Property **(d)** comes from multilinearity.
Property **(e)** comes from alternation, as we have already noted.
Property **(f)** is comes from multilinearity and **(e)**.
Property **(g)** results from multilinearity and **(e)**:

$$\begin{aligned}\det[x_1 \mid \cdots \mid x_i + cx_j \mid \cdots \mid x_j \mid \cdots \mid x_n] \\ = \det[x_1 \mid \cdots \mid x_i \mid \cdots \mid x_j \mid \ldots x_n] + c\det[x_1 \mid \cdots \mid x_j \mid \cdots \mid x_j \mid \ldots x_n] \\ = \det[x_1 \mid \cdots \mid x_i \mid \cdots \mid x_j \mid \ldots x_n] + 0.\end{aligned}$$

Property **(h)** comes from multilinearity and normalization.

Property **(i)** comes from **(g)** and **(h)** by noting that any triangular matrix can be obtained from a diagonal one by a sequence of row operations of the form $\mathsf{R}_i + c\mathsf{R}_j$.

We can now compute the determinants of the elementary invertible matrices.

- $\det(I) = 1$ by **(c)**;
- $\det(E(i, j; c)) = 1$ by **(i)**;
- $\det(D(i; c)) = c$ by **(h)**;
- $\det(P(i, j)) = -1$ by **(b)** and (c).

Since we know the effects of elementary invertible matrices on the rows of a matrix $A$, we can compute the following products.

If $E = E(i, j; c)$, then $\det(EA) = \det(A)$ by **(g)**.
If $E = D(i; c)$, then $\det(EA) = c\det(A)$ by **(a)**.
If $E = P(i, j)$, then $\det(EA) = -\det(A)$ by **(b)**.

In each of these cases, we have $\det(EA) = \det(E)\det(A)$. Thus if $U$ is a product of elementary invertible matrices, its determinant is the product of the determinants of the factors, and by sequential application of the above observation, we have $\det(UA) = \det(U)\det(A)$.

From this analysis of the determinant for elementary invertible matrices, we derive the following general properties.

**(j)** $A$ is invertible if and only if $\det(A) \neq 0$;
**(k)** $\det(AB) = \det(A)\det(B)$;

Let $R = OA$ be the result of forward elimination on $A$, where $O$ is the product of elementary invertible matrices. Then $\det(R) = \det(O)\det(A)$. Indeed, since forward elimination uses only $E$ and $P$ type matrices, $\det(O) = \pm 1$, where the sign is determined by the number of permutations used.

Since $A$ is square, $A$ is noninvertible if and only if $R$ has a zero row.

Suppose $A$ is noninvertible. Then $R$ has a zero row, so $\det(R) = 0$ by **(d)**, so $\det(A) = 0$. If $B$ is another matrix, then $AB$ is noninvertible, so $\det(AB) = 0 = \det(A)\det(B)$.

Suppose $A$ is invertible. Then its determinant is the product of elementary invertible matrices, so $\det(A) \neq 0$. If $B$ is another matrix, then $\det(AB) = \det(A)\det(B)$, as we previously noted. This proves **(j)** and **(k)**.

This also shows something more:

**(l)** $\det(A) = (-1)^p q$, where $p$ is the number of permutations used in forward elimination, and $q$ is the product along the diagonal of $R$;
**(m)** $\det(A^*) = \det(A)$.

We have $\det(R) = \det(O)\det(A)$. But $\det(R) = q$, and $\det(O) = (-1)^p$. This gives **(l)**.

If $A$ is invertible, then so is $A^*$:

$$(A^*(A^{-1})^*)^* = A^{-1}A = I = I^*,$$

so $(A^*)^{-1} = (A^{-1})^*$.

If $E$ is an elementary invertible matrix, then $\det(E) = \det(E^*)$. Suppose that $E$ and $F$ are matrices satisfying **(m)**, then

$$\det(EF) = \det(E)\det(F) = \det(E^*)\det(F^*) = \det(E^*F^*) = \det((EF)^*).$$

If $A$ is invertible, then $A$ is the product of elementary invertible matrices, and the result follows.

If $A$ is not invertible, then neither is $A^*$ thus $\det(A) = 0 = \det(A)^*$. This proves **(m)**.

## 5. Geometric Interpretation of Determinant

The *n-box* in $\mathbb{R}^m$ determined by the vectors $v_1, \dots, v_n \in \mathbb{R}^m$ is the set

$$\{t_1 v_1 + \cdots + t_n v_n \mid t_i \in [0,1]\}.$$

We define the *n-volume* of a box inductively by defining the 1-volume of a vector to be its length, and the $n$-volume of the box to be the height of the box times the $(n-1)$-volume of its base, where the height is the distance between $v_n$ and the span of $\{v_1, \dots, v_{n-1}\}$, and the base is the $(n-1)$-box determined by $v_1, \dots, v_{n-1}$. Let $\mathrm{vol}\{v_1, \dots, v_n\}$ denote this quantity.

If $m = n$, this definition of volume corresponds to the result we get by integrating the box via multiple integration.

The *orientation* of an ordered collection of vectors is determined by the $n$-dimension right hand rule. There are two distinct orientations (right and left handed); interchanging two vectors in an ordered collection switches the orientation.

The primary geometric interpretation of the determinant function is that $\det(A)$ is equal to the $n$-dimensional signed volume of the box determined by the columns of $A$, where $A$ is an $n \times n$ matrix. The sign is positive for right orientation and negative for left orientation.

This is the same thing as saying that $\det(A)$ is equal to the signed distortion of volume induced by the transformation $T_A : \mathbb{R}^n \to \mathbb{R}^n$. That is,

$$\mathrm{vol}(T_A(X)) = \pm \det(A) \mathrm{vol}(X),$$

where $X$ is any set of $n$ vectors in $\mathbb{R}^n$; the sign determines whether or not the transformation is orientation preserving or orientation reversing.

## 6. Linear Transformations as a Vector Space

We review some facts from the chapter "Linear Transformations".

If $S : V \to V$ and $T : V \to V$ are linear transformations, then $S + T : V \to V$ given by $(S+T)(v) = S(v) + T(v)$ is a linear transformation.

If $T : V \to V$ is a linear transformation and $a \in \mathbb{R}$, then $aT : V \to V$ given by $(aT)(v) = aT(v)$ is a linear transformation.

Thus the set of all linear transformation from $V$ to itself is a vector space, which we may denote by $\mathcal{L}(V)$.

If we fix a basis for $V$, we may write $S$ and $T$ as matrices. Then $A_{S+T} = A_S + A_T$ and $A_{aT} = aA_T$.

Since we may compose transformations from a vector space into itself, $\mathcal{L}(V)$ comes equipped with a multiplication. We can write $ST$ to mean $S \circ T$. This multiplication distributes over addition of linear transformations.

In particular, $T^n$ denotes $T$ composed with itself $n$ times. If we denote the transformation $a\mathrm{id}_V$ simply by $a$, we can form and factor polynomials such as

$$L = T^2 - 4T + 3 = (T-3)(T-1);$$

thus $L(v) = T(T(v)) - 4T(v) + 3v$.

## 7. Eigenvectors and Eigenvalues

Let $V$ be a finite dimensional vector space of dimension $n$ and let $T : V \to V$.

An *eigenvector* of $T$ is a nonzero vector $v \in V$ such that $T(v) = \lambda v$ for some $\lambda \in \mathbb{R}$. The number $\lambda$ is called an *eigenvalue* of $T$.

That is, a nonzero vector $v$ is an eigenvector of $T$ if and only if $T(v)$ is on the same line through the origin as $v$, so $T$ expands or contracts this line by a fixed factor; the eigenvalue associated to $v$ is this expansion factor.

Let $A$ be an $n \times n$ matrix. The *eigenvectors* and *eigenvalues* of $A$ are, by definition, the eigenvectors and eigenvalues of the corresponding linear transformation $T_A : \mathbb{R}^n \to \mathbb{R}^n$ given by $T_A(x) = Ax$.

**Proposition 7.4.** *Let $T : V \to V$ be a linear transformation Let $v \in V$ be an eigenvector with eigenvalue $\lambda$. Let $a \in \mathbb{R}$. Then $av$ is an eigenvector with eigenvalue $\lambda$.*

*Proof.* We have $T(av) = aT(v) = a\lambda v = \lambda(av)$. □

**Example 7.5.** Find the eigenvectors and eigenvalues of the linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ corresponding to the matrix

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

*Solution.* Since $T(e_1) = 2e_1$ and $T(e_2) = 3e_2$, we see that these are both eigenvectors with corresponding eigenvalues 2 and 3. Then all of the vectors on the $x$ and $y$ axis are also eigenvectors. However, if $v = ae_1 + be_2$, then $T(v) = 2ae_1 + 3be_2$ is a scalar multiple of $v$ if and only if either $a$ or $b$ is zero. Thus no other vectors are eigenvectors. □

**Example 7.6.** Find the eigenvectors and eigenvalues of the linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ corresponding to the matrix $A = \lambda I$.

*Solution.* Every nonzero vector in $\mathbb{R}^2$ is an eigenvector with eigenvalue $\lambda$. □

**Example 7.7.** Find the eigenvectors and eigenvalues of the linear transformation which rotates $\mathbb{R}^2$ by 90 degrees.

*Solution.* There are none. □

**Example 7.8.** Find the eigenvectors and eigenvalues of the linear transformation which reflects $\mathbb{R}^2$ across the $y$-axis.

*Solution.* Eigenvalue 1 corresponds to eigenvector $e_1$. Eigenvalue $-1$ corresponds to eigenvector $e_2$. □

## 8. Eigenspaces

Let $T : V \to V$ be a linear transformation with eigenvalue $\lambda$. The *eigenspace* of $\lambda$ is the set containing zero and of all eigenvectors of $T$ whose eigenvalue is $\lambda$:

$$\mathrm{eig}_\lambda(T) = \{v \in T \mid T(v) = \lambda v\}.$$

**Proposition 7.9.** *Let $T : V \to V$ be a linear transformation and let $a \in \mathbb{R}$. Then $\mathrm{eig}_a(T) \leq V$.*

*Proof.* Let $T - a : V \to V$ denote the linear transformation given by $(T - a)(v) = T(v) - av$. Then $v \in \mathrm{eig}_a(T)$ if and only if $(T - a)(v) = 0$. Thus $\mathrm{eig}_a(T) = \ker(T - a)$. The kernel of a linear transformation is always a subspace of the domain, so $\mathrm{eig}_a(T) \leq V$. □

The above proof points out that, in particular, $\mathrm{eig}_0(T) = \ker(T)$. We collect some facts regarding this.

**Proposition 7.10.** *Let $T : V \to V$ be a linear transformation. The following conditions are equivalent:*

- **i.** *$T$ is an isomorphism;*
- **ii.** *$T$ is bijective;*
- **iii.** *$T$ is surjective;*
- **iv.** *$T$ is injective;*
- **v.** $\ker(T) = \{0\}$*;*
- **vi.** $\mathrm{eig}_0(T) = \{0\}$*;*
- **vii.** $0$ *is not an eigenvalue of $T$.*

Let $T : V \to V$ be a linear transformation. The *total eigenspace* of $T$ is

$$\mathrm{eig}(T) = \mathrm{span}\{v \in V \mid \ v \text{ is an eigenvector of } T \ \}.$$

We now extend the concept of direct sum to more that one subspace.

Let $V$ be a vector space and let $U_1, \dots, U_n$ be subspaces. We say that $V$ is the *direct sum* of $U_1, \dots, U_n$, if

- **(D1)** $U_1 + \cdots + U_n = V$;
- **(D2)** $U_i \cap U_j = \{0\}$ whenever $i \neq j$.

In this case, we may write

$$V = \oplus_{i=1}^n U_i.$$

**Proposition 7.11.** *Let $T : V \to V$ be a linear transformation whose distinct eigenvalues are $\lambda_1, \dots, \lambda_n$. Then*

$$\mathrm{eig}(T) = \oplus_{i=1}^n \mathrm{eig}_{\lambda_i}(T).$$

*Proof.* It is clear from the definition that the vectors in $\mathrm{eig}_{\lambda_i}(T)$ span $\mathrm{eig}(T)$ as $\lambda_i$ ranges from $i = 1, \dots, n$. Also, if $v$ has eigenvalue $\lambda_i$, then it cannot also have a different eigenvalue $\lambda_j$. Thus the intersection of two of these eigenspaces is trivial. □

## 9. Finding Eigenvalues of a Matrix

Let $T : V \to V$ be a linear transformation; for simplicity, let us assume for the time being that $V = \mathbb{R}^n$. To find the eigenvectors and eigenvalues of $T$, we wish to solve the equation $T(v) = \lambda v$, where $\lambda$ is any real number. That is, we wish to solve

$$T(v) - \lambda v = 0.$$

Let us first try to find an appropriate $\lambda$.

If $A$ is the matrix corresponding to $T$, then this equation becomes

$$Av - \lambda I v = 0.$$

That is, we wish to find $\ker(A - \lambda I)$ whenever it is nontrivial. This kernel is nontrivial if and only if $\det(A - \lambda I) = 0$.

If we compute $\det(A - \lambda I)$, we obtain a polynomial in $\lambda$. The degree of this polynomial is exactly $\dim(V)$. Thus we define the *characteristic polynomial* of $A$ (or $T$) to be

$$\chi_A(\lambda) = \det(A - \lambda I).$$

We see that $\lambda$ is an eigenvalue if and only if $\chi_A(\lambda) = 0$, because this is exactly when $(A - \lambda I)$ has a nontrivial kernel.

Once one finds an eigenvalue $\lambda$, one can find the corresponding eigenvectors by solving $(A - \lambda I)x = 0$.

**Example 7.12.** Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be the linear transformation corresponding to the matrix

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

Find the eigenvectors and eigenvalues of $T$.

*Solution.* First we find the eigenvalues. The characteristic polynomial is

$$\chi_A(\lambda) = \det(A - \lambda I) = (2 - \lambda)^2(1 + \lambda).$$

Thus the eigenvalues are 2 and $-1$.

Now we find the eigenvectors. We find $\ker(A - 2I) = \operatorname{span}\{(1, 0, 1)\}$ and $\ker(A + I) = \operatorname{span}\{(1, -3, 4)\}$. □

## 10. Matrices with Respect to a Basis

Let $V$ be a vector space of dimension $n$. Let $X = \{x_1, \ldots, x_n\}$ be a basis for $V$. If $v \in V$, then there exist unique real numbers $a_1, \ldots, a_n \in \mathbb{R}$ such that $v = \sum_{i=1}^n a_i x_i$.

Let $\Gamma_X : V \to \mathbb{R}^n$ be given by $\Gamma_X(v) = (a_1, \ldots, a_n)$, where $v = \sum_{i=1}^n a_i x_i$. Recalling that any transformation is completely determined by its value on a basis, we see that $\Gamma_X$ is the unique transformation from $V \to \mathbb{R}^n$ which sends $x_i$ to $e_i$. Since $\Gamma_X$ sends a basis to a basis, it is an isomorphism.

Let $T : V \to V$ be a linear transformation. The *matrix of $T$ with respect to the basis $X$* is the $n \times n$ matrix $B$ which corresponds to the linear transformation

$$\Gamma_X \circ T \circ \Gamma_X^{-1} : \mathbb{R}^n \to \mathbb{R}^n.$$

We view this via the commutative diagram

$$\begin{array}{ccc} V & \xrightarrow{T} & V \\ {\scriptstyle \Gamma_X}\downarrow & & \downarrow{\scriptstyle \Gamma_X} \\ \mathbb{R}^n & \xrightarrow[\Gamma_X \circ T \circ \Gamma_X^{-1}]{} & \mathbb{R}^n \end{array}$$

The columns of $B$ represent the destinations of the basis vectors in $X$ under the transformation $T$, written in terms of the basis $X$.

For example, if the $4^{\text{th}}$ column of $B$ is $(1, 0, 3, -2)$, then $T(x_4) = x_1 + 3x_3 - x_4$.

## 11. Matrices with Respect to a Basis in $\mathbb{R}^n$

Let $V = \mathbb{R}^n$ and let $X \subset \mathbb{R}^n$ be a set of $n$ linearly independent vectors in $\mathbb{R}^n$. Then $X$ is a basis for $\mathbb{R}^n$, but $X$ is not necessarily the standard basis.

Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation. Then $T$ has a corresponding matrix, say $A$.

Since $\Gamma_X^{-1} : \mathbb{R}^n \to \mathbb{R}^n$, it has a corresponding matrix, say $C$. It is easy to see what the matrix inverse of $C$ is; since $\Gamma_X^{-1}(e_i) = x_i$, then

$$C = [x_1 \mid \cdots \mid x_n].$$

Thus the matrix $B$ of $T$ with respect to the basis $X$ is

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

We may also write this as a commutative diagram

$$\begin{array}{ccc} \mathbb{R}^n & \xrightarrow{A} & \mathbb{R}^n \\ {\scriptstyle C}\big\uparrow & & \big\downarrow{\scriptstyle C^{-1}} \\ \mathbb{R}^n & \xrightarrow[C^{-1}AC]{} & \mathbb{R}^n \end{array}$$

Let $A$ and $B$ be $n \times n$ matrices. We say that $A$ and $B$ are *conjugate* (or *similar*) if there exists an invertible $n \times n$ matrix $C$ such that $B = C^{-1}AC$. Note that $A$ is invertible if and only if $B$ is invertible.

Suppose that $A$ and $B$ are conjugate matrices, and the $B = C^{-1}AC$. Can we express the action of $B$ on $\mathbb{R}^n$ in terms of the action of $A$? Since $C$ is invertible, the columns of $C$ are a basis for $\mathbb{R}^n$. Let $X = \{x_1, \ldots, x_n\}$ be this basis. Now $Ax_i$ may be written in terms of the basis $X$:

$$Ax_i = \sum_{j=1}^{n} b_{ij} x_j.$$

Then

$$C^{-1}Ax_i = \sum_{j=1}^{n} b_{ij} e_j.$$

On the other hand,

$$BC^{-1}x_i = Be_i.$$

Thus, since $BC^{-1} = C^{-1}A$, we have

$$Be_i = \sum_{j=1}^{n} b_{ij} e_j,$$

which shows that $B = (b_{ij})$.

In words, the columns of $B$ represent the destinations of the nonstandard basis vectors $x_i$ under the transformation $T_A$ (corresponding to $A$) when these destinations are *written in terms of the basis* $X$.

**Example 7.13.** Find the matrix of a linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ which reflects the plane across the line $y = 2x$.

*Solution.* If we find a nice basis, then this transformation is easy.

Let $x_1 = (1, 2)$. Then $x_1$ is on the line $y = 2x$, so $T(x_1) = x_1$. Let $x_2 = (-2, 1)$; then $x_2$ is perpendicular to $x_1$, since $x_1 \cdot x_2 = -2 + 2 = 0$. Thus $T(x_2) = -x_2$.

Thus the matrix of $T$ with respect to this basis is

$$B = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

Let

$$C = \begin{bmatrix} 1 & -2 \\ 2 & 1 \end{bmatrix}; \text{ then } C^{-1} = \begin{bmatrix} \frac{1}{5} & \frac{2}{5} \\ \frac{-2}{5} & \frac{1}{5} \end{bmatrix}.$$

Therefore

$$A = CBC^{-1} = \begin{bmatrix} \frac{-3}{5} & \frac{4}{5} \\ \frac{4}{5} & \frac{3}{5} \end{bmatrix}.$$

□

**Proposition 7.14.** *Let $T : V \to V$ be a linear transformation.*
*Let $v_1, \dots, v_n \in V$ be eigenvectors with distinct eigenvalues.*
*Then $\{v_1, \dots, v_n\}$ is independent.*

*Proof.* Let $d_i$ be the eigenvalue corresponding to $v_i$. Suppose that the set is not independent; then one of these vectors is in the span of the previous vectors. Let $k$ be the smallest integer such that this is true, so that

$$v_k = a_1 v_1 + \dots a_{k-1} v_{k-1},$$

where $\{v_1, \dots, v_{k-1}\}$ is independent. Multiplying this equation by $d_k$ gives

$$d_k v_k = \sum_{i=1}^{k-1} a_{k-1} v_{k-1},$$

but applying $A$ gives

$$d_k v_k = \sum_{i=1}^{k-1} a_i d_i v_i.$$

Subtracting these gives

$$0 = \sum_{i=1}^{k-1} (d_k - d_i) a_i v_i.$$

Since the $d_i$'s are distinct, this is a nontrivial dependence relation, contradicting the fact that $\{v_1, \dots, v_{k-1}\}$ is independent. □

**Corollary 7.15.** *Let $T : V \to V$ be a linear transformation, where* $\dim(V) = n$.
*Let $v_1, \dots, v_n \in V$ be eigenvectors with distinct eigenvalues.*
*Then*

**(a)** *$\{v_1, \dots, v_n\}$ is a basis for $V$;*
**(b)** $\operatorname{eig}(T) = V$;
**(c)** $V = \oplus_{i=1}^{n} \operatorname{eig}_{\lambda_i}(T)$.

## 12. Diagonalization

Let $A$ be an $n \times n$ matrix.

We say that $A$ is *diagonalizable* if there exists a diagonal matrix $D$ and an invertible matrix $C$ such that $D = C^{-1}AC$.

We say that $\mathbb{R}^n$ has a *basis of eigenvectors* of $A$ if their exist $n$ linearly independent eigenvectors of $A$. When this happens, they form a basis.

**Proposition 7.16.** *Let $A$ be an $n \times n$ matrix. Then $A$ is diagonalizable if and only if $\mathbb{R}^n$ has a basis of eigenvectors of $A$.*

*Proof.* Suppose that $A$ is diagonalizable, and let $D$ be diagonal and $C$ invertible such that $D = C^{-1}AC$. Then $D = (d_{ij})$, where $d_{ij} = 0$ unless $i = j$.

The columns of $C$ are a basis of eigenvectors of $A$. They are linearly independent because $C$ is invertible; to see that they are eigenvectors, let $x_i$ be the $i^{\text{th}}$ column of $C$. Then

$$Ax_i = CDC^{-1}x_i = CDe_i = C(d_ie_i) = d_iCe_i = d_ix_i.$$

Suppose that $A$ has a basis of eigenvectors $X = \{x_1, \dots, x_n\}$ with corresponding eigenvalues $d_1, \dots, d_n$. Form the square matrix $D$ with $d_i$'s along the diagonal and 0 elsewhere. Let $C = [x_1 \mid \cdots \mid x_n]$. Then $D$ is $A$ written with respect to the basis $X$, so $D = C^{-1}AC$. $\square$

Here is a criterion for diagonalizability.

**Proposition 7.17.** *Let $A$ be an $n \times n$ matrix with $n$ distinct eigenvalues. Then $A$ is diagonalizable.*

*Proof.* Each eigenvalue corresponds to a different eigenvector. These are linearly independent. $\square$

It is sometimes useful or necessary to consider linear transformations composed with themselves. If the transformation corresponds to a diagonalizable matrix, we are in luck.

**Proposition 7.18.** *Let $B = C^{-1}AC$. Then $B^n = C^{-1}A^nC$.*

**Proposition 7.19.** *Let $D = (d_{ij})$ be diagonal. Then $D^n = (d_{ij}^n)$.*

Thus if $A$ is diagonalizable and $D = C^{-1}AC$, then $A = CDC^{-1}$, so $A^n = CD^nC^{-1}$ is relatively easy to compute.

**Example 7.20.** Let

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

**(a)** Diagonalize $A$.
**(b)** Find $A^8$.

*Solution.* The characteristic polynomial of $A$ is

$$\begin{aligned} \chi_A(\lambda) &= (-2-\lambda)[(2-\lambda)^2] + 3(2-\lambda) \\ &= [(-1)(2+\lambda)(2-\lambda)+3](2-\lambda) \\ &= [\lambda^2 - 1](2-\lambda) \\ &= (\lambda+1)(\lambda-1)(2-\lambda). \end{aligned}$$

Thus the eigenvalues are 1, 2, and $-1$. Corresponding eigenvectors are $x_1 = (-1,0,3)$, $x_2 = (0,1,0)$, and $x_3 = (-1,0,1)$. Let $C = [x_1 \mid x_2 \mid x_3]$. Then

$$D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -1 \end{bmatrix}; \quad \text{where} \quad D = C^{-1}AC \quad \text{and} \quad C^{-1} = \frac{1}{2}\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ -3 & 0 & -1 \end{bmatrix}.$$

Thus $A^8 = CD^8C^{-1}$ is easy to compute. Try this.

In this *particular* example, simply squaring $A$ will reveal that something nice happens, which explains the result above (if you tried it). □

## 13. Finding Eigenvalues of a Linear Transformation

Let $V$ be an arbitrary finite dimensional vector space of dimension $n$. We turn to the question to finding eigenvalues of a linear transformation $T : V \to V$. By definition, the eigenvalues of $T$ should not depend on any particular basis we select for $V$.

Select an ordered basis $X = \{x_1, \ldots, x_n\}$ for $V$. If we know the value of $T$ on each of the basis vectors $x_i$, we can find the matrix $A$ of $X$ with respect to this basis; $A$ is the matrix corresponding to the transformation

$$\Gamma_X \circ T \circ \Gamma_X^{-1} : \mathbb{R}^n \to \mathbb{R}^n.$$

Then we can compute the characteristic polynomial $\det(A - \lambda I)$ and attempt to find its roots; these roots should be our eigenvalues.

The matrix $A$, however, depends on the basis $X$ we chose for $V$. The question arises as to whether or not we get the same result if we choose a different basis for $V$. To see that we do get the same result, we formulate two propositions.

**Proposition 7.21.** *Let $V$ be a finite dimensional vector space of dimension $n$. Let $T : V \to V$ be a linear transformation. Let $X$ and $Y$ be ordered bases for $V$. Let $A$ be the matrix of $T$ with respect to $X$. Let $B$ be the matrix of $T$ with respect to $Y$. Then there exists a matrix $C$ such that $B = C^{-1}AC$.*

*Proof.* By definition of the matrix of a transformation with respect to a basis, we know that $A$ is the matrix corresponding to the transformation $\Gamma_X \circ T \circ \Gamma_X^{-1}$ and the $B$ is the matrix corresponding to the transformation $\Gamma_Y \circ T \circ \Gamma_Y^{-1}$. Let $C$ be the matrix corresponding to the transformation $\Gamma_X \circ \Gamma_Y^{-1} : \mathbb{R}^n \to \mathbb{R}^n$. Note that $C^{-1}$ corresponds to $\Gamma_Y^{-1} \circ \Gamma_X$. Then

$$\Gamma_Y \circ T \circ \Gamma_Y^{-1} = \left(\Gamma_Y \circ \Gamma_X^{-1}\right) \circ \left(\Gamma_X \circ T \circ \Gamma_X^{-1}\right) \circ \left(\Gamma_X \circ \Gamma_Y^{-1}\right);$$

thus $B = C^{-1}AC$. □

This proposition states that matrices of the same transformation with respect to different bases are conjugate. Diagrams help explain this; the transformation diagram

$$\begin{array}{ccc}
\mathbb{R}^n & \xrightarrow{\Gamma_X \circ T \circ \Gamma_X^{-1}} & \mathbb{R}^n \\
\Gamma_X \big\uparrow & & \big\uparrow \Gamma_X \\
V & \xrightarrow{T} & V \\
\Gamma_Y \big\downarrow & & \big\downarrow \Gamma_Y \\
\mathbb{R}^n & \xrightarrow[\Gamma_Y \circ T \circ \Gamma_Y^{-1}]{} & > \mathbb{R}^n
\end{array}$$

is converted into the matrix diagram

$$\begin{array}{ccc}
\mathbb{R}^n & \xrightarrow{A} & \mathbb{R}^n \\
C \big\uparrow & & \big\downarrow C^{-1} \\
\mathbb{R}^n & \xrightarrow[B=C^{-1}AC]{} & \mathbb{R}^n
\end{array}$$

in a manner identical to a change of basis within $\mathbb{R}^n$. It is not hard to see what $C$ is; its columns are the destinations in $\mathbb{R}^n$ of the ordered basis $Y$ under the transformation $\Gamma_X$.

**Proposition 7.22.** *Let $V$ be a finite dimensional vector space of dimension $n$. Let $T : V \to V$ be a linear transformation. Let $X$ and $Y$ be ordered bases for $V$. Let $A$ be the matrix of $T$ with respect to $X$. Let $B$ be the matrix of $T$ with respect to $Y$. Then $\chi_A(\lambda) = \chi_B(\lambda)$.*

*Proof.* We compute

$$\begin{aligned}
\chi_B(\lambda) &= \det(B - \lambda I) \\
&= \det(C^{-1}AC - \lambda I) \\
&= \det(C^{-1}AC - \lambda C^{-1}IC) \\
&= \det(C^{-1}(A - \lambda I)C) \\
&= \det(C^{-1})\det(A - \lambda I)\det(C) \\
&= \det(A - \lambda I) \\
&= \chi_A(\lambda).
\end{aligned}$$

□

This says that we can think of the characteristic polynomial as an *invariant* of a transformation as opposed to an invariant of a matrix which changes as the basis changes. This also tells us that we can find the eigenvalues of a linear transformation by selecting any basis and computing the eigenvalues with respect to that basis.

Let $V$ be a finite dimensional vector space and let $T : V \to V$ be a linear transformation. The *characteristic polynomial* of $T$ is $\chi_T(\lambda) = \det(A - \lambda I)$, where $A$ is the matrix of $T$ with respect to *any* basis.

APPENDIX A

# Fundamental Subspaces of a Matrix

## 1. Transpose Transformations

**Definition A.1.** Let $A$ be an $m \times n$ matrix.

The *transpose* of $A$, which we denote by $A^*$, is the $n \times m$ matrix whose $j^{\text{th}}$ row is the $j^{\text{th}}$ column of $A$.

**Definition A.2.** Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation.

The *transpose* of $T$ is the linear transformation $T^* : \mathbb{R}^m \to \mathbb{R}^n$ given by $T^*(w) = A^*w$, where $A = [T(e_1) \mid \cdots \mid T(e_n)]$.

*Remark* A.1. Let $A$ be an $m \times n$ matrix. Recall that $A$ corresponds to a linear transformation $T_A : \mathbb{R}^n \to \mathbb{R}^m$ which is given by $T_A(v) = Av$. Then $A^*$ corresponds to a linear transformation $T_{A^*} : \mathbb{R}^m \to \mathbb{R}^n$ which is given by $T_{A^*}(w) = A^*(w)$. Thus $T_A^* = T_{A^*}$.

Let $T = T_A$ be the transformation corresponding to $A$. We know that the columns of $A$ are the destinations of the standard basis vectors of $\mathbb{R}^n$ under the transformation $T$. Thus the image of $T$ is spanned by these vectors. On the other hand, the columns of $T^*$ are the rows of $A$, so the image of $T^*$ is a subspace of $\mathbb{R}^n$ which is spanned by the rows of $A$. We now investigate the relationship between the image of $T^*$ and the kernel of $T$.

## 2. Column Spaces and Row Spaces

**Definition A.3.** Let $A$ be an $m \times n$ matrix.

The *image* of $A$ is the image of $T_A$.

The *kernel* of $A$ is the kernel of $T_A$.

The *column space* of $A$ is the subspace of $\mathbb{R}^m$ spanned by the columns of $A$, and is denoted by $\operatorname{col}(A)$.

The *row space* of $A$ is the subspace of $\mathbb{R}^n$ spanned by the rows of $A$, and is denoted by $\operatorname{row}(A)$.

The *null space* of $A$ is the set $\{x \in \mathbb{R}^n \mid Ax = 0\}$, and is denoted by $\ker(A)$.

The *rank* of $A$ is the dimension of the column space of $A$.

The *nullity* of $A$ is the dimension of the null space of $A$.

Let $A$ be an $m \times n$ matrix. The four fundamental subspaces associated to $A$ are $\operatorname{col}(A)$, $\operatorname{row}(A)$, $\ker(A)$, and $\ker(A^*)$.

**Proposition A.4.** *Let $A$ be an $m \times n$ matrix. Perform forward elimination on the matrix $A$ to achieve $B = OA$, where $O$ is invertible and $B$ is in row echelon form. Perform backward elimination on $B$ to achieve $C = UA$, where $U$ is invertible and $C$ is in reduced row echelon form. Then*

- **(a)** $\text{col}(A) = \text{row}(A^*)$;
- **(b)** $\text{col}(A) = \text{img}(T_A)$;
- **(c)** $\text{row}(A) = \text{img}(T_A^*)$;
- **(d)** *the rank of $A$ is equal to the number of basic columns of $B$ (or of $C$);*
- **(e)** *the nullity of $A$ is equal to the number of free columns of $B$ (or of $C$);*
- **(f)** $\ker(A) = \ker(B) = \ker(C)$;
- **(g)** $\text{row}(A) = \text{row}(B) = \text{row}(C)$;
- **(h)** $\dim(\text{col}(A)) = \dim(\text{row}(A))$;
- **(i)** *the nonzero rows of $B$ (or of $C$) form a basis for* $\text{row}(A)$;
- **(j)** *the last $m - r$ rows of $O$ (or of $U$) form a basis for* $\ker(A^*)$, *where $r$ is the rank of $A$.*

*Proof.*

**(a)** $\text{col}(A) = \text{row}(A^*)$
This follows from the definition of transpose.

**(b)** $\text{col}(A) = \text{img}(T_A)$
This follows fact that the image of $T_A$ is spanned by the destinations of the standard basis vectors; these destinations are the columns of $A$.

**(c)** $\text{row}(A) = \text{img}(T_A^*)$
This follows from **(a)** and **(b)**.

**(d)** *the rank of $A$ is equal to the number of basic columns of $B$ (or of $C$)*
The rank of $B$ is clearly equal to the number of basic columns of $B$. The rank of $A$ equals the rank of $B$ because $B = UA$, where $U$ is an invertible matrix. The transformation $T_U$ is an isomorphism, so a basis for the image of $T_A$ is sent by $T_U$ to a basis for the image of $T_B$.

**(e)** *the nullity of $A$ is equal to the number of free columns of $B$ (or of $C$)*
The nullity of $A$ is the number of free columns by the Rank Plus Nullity Theorem: $\dim(\ker(A)) = \dim(\ker(T_A)) = n - \dim(\text{img}(T_A))$; since $\dim(\text{img}(T_A))$ is the number of basic columns, $n - \dim(\text{img}(T_A))$ must be the number of free columns.

**(f)** $\ker(A) = \ker(B) = \ker(C)$
This is given by the fact that composing on the left with an injective transformation does not change the kernel of a transformation. Since $B = OA$, we have $\ker(A) = \ker(T_A) = \ker(T_O \circ T_A) = \ker(T_{OA}) = \ker(OA) = \ker(B)$. Similarly, $\ker(A) = \ker(C)$.

**(g)** $\mathrm{row}(A) = \mathrm{row}(B) = \mathrm{row}(C)$

If $E$ is an elementary invertible matrix and $D$ is any compatibly sized matrix, then the rows of $ED$ are a linear combination of the rows of $D$; one sees this by considering the effect of the corresponding elementary row operation on $D$. Thus $\mathrm{row}(ED) \subset \mathrm{row}(D)$. But $E^{-1}$ is also an elementary invertible matrix, so $\mathrm{row}(D) = \mathrm{row}(E^{-1}ED) \subset \mathrm{row}(ED)$, which shows that $\mathrm{row}(ED) = \mathrm{row}(D)$ and $E$ does not change the row space.

Since $B = OA$ and $O$ is a product of elementary invertible matrices, we see that $\mathrm{row}(B) = \mathrm{row}(OA) = \mathrm{row}(A)$. Similarly, $\mathrm{row}(C) = \mathrm{row}(A)$.

**(h)** $\dim(\mathrm{col}(A)) = \dim(\mathrm{row}(A))$

It is apparent from the definition of row echelon form that the nonzero rows of $B$ form a basis for the row space of $B$.

By **(d)**, $\dim(\mathrm{col}(A)) = \dim(\mathrm{col}(B))$. The dimension of $\mathrm{col}(B)$ is equal to the number of pivots in $B$ (or $C$), which is equal to the number of nonzero rows of $B$ (or $C$), which is equal to the dimension of $\mathrm{row}(B)$. Thus $\dim(\mathrm{col}(A)) = \dim(\mathrm{col}(B)) = \dim(\mathrm{row}(B)) = \dim(\mathrm{row}(A))$.

**(i)** *the nonzero rows of $B$ (or of $C$) form a basis for* $\mathrm{row}(A)$

The nonzero rows of $B$ (respectively $C$) form a basis for $\mathrm{row}(B)$ (respectively $\mathrm{row}(C)$). By **(g)**, $\mathrm{row}(A) = \mathrm{row}(B)$, and the result follows.

**(j)** *the last $m - r$ rows of $O$ (or of $U$) form a basis for* $\ker(A^*)$

We show this for $O$; the proof for $U$ is the identical.

Set $k = m - r$ and note that $\dim(\ker(A^*)) = k$. This follows from the Rank Plus Nullity Theorem and **(g)**: we have $r = \dim(\mathrm{col}(A)) = \dim(\mathrm{row}(A^*)) = \dim(\mathrm{col}(A^*))$. Thus $\dim(\ker(A^*)) = m - \dim(\mathrm{col}(A^*)) = m - r$.

Since $O$ is invertible, its rows are linearly independent. Indeed, $T_O$ is an isomorphism, so $\ker(O) = \{0\}$; thus $\dim(\mathrm{row}(O)) = \dim(\mathrm{col}(O)) = \dim(\mathbb{R}^m) - \dim(\ker(O)) = m$, since $\dim(\ker(O)) = 0$. Then $\mathrm{row}(O) = \mathbb{R}^m$, so the rows of $O$ are a basis for $\mathbb{R}^m$.

Thus the last $k$ rows of $O$ are linearly independent, so if these vectors are in $\ker(A^*)$, they are a basis for it. We only need to show that they are in $\ker(A^*)$.

Since $B = OA$, we have $B^* = A^*O^*$. The last $k$ rows of $B$ are zero, so the last $k$ columns of $B^*$ are zero. If $x^*$ is one of the last $k$ rows of $O$, then $x$ is one of the last $k$ columns of $O^*$, and $A^*x$ is one of the last $k$ columns of $B^*$; that is, it is zero. Thus $x$ is in the kernel of $A^*$. □

## 3. Perpendicular Decompositions

**Proposition A.5.** *Let $U \leq \mathbb{R}^n$. Set*

$$\perp(U) = \{v \in \mathbb{R}^n \mid u \cdot v = 0 \text{ for all } u \in U\}.$$

*Then*

**(a)** $\perp(U) \leq \mathbb{R}^n$;
**(b)** $U \cap \perp(U) = \{0\}$;
**(c)** $\perp(\perp(U)) = U$.

*Proof.* Exercise. □

**Proposition A.6.** *Let $A$ be an $m \times n$ matrix. Then*

**(a)** $\mathrm{row}(A) = \perp(\ker(A))$ *and* $\mathbb{R}^n = \mathrm{row}(A) \oplus \ker(A)$;
**(b)** $\mathrm{col}(A) = \perp(\ker(A^*))$ *and* $\mathbb{R}^m = \mathrm{col}(A) \oplus \ker(A^*)$.

*Proof.* In light of the fact that $\mathrm{col}(A) = \mathrm{row}(A^*)$, if we prove **(a)**, then **(b)** will follow simply by replacing $A$ with $A^*$. Thus we prove **(a)**.

The coordinates of $Ax$ are the dot products of the rows of $A$ with the vector $x$. If $x \in \ker(A)$, the $Ax = 0$ (the zero vector). Thus each of the coordinates of $Ax$ is equal to 0 (the zero scalar). This shows that each row of $A$ is perpendicular to any vector in the kernel of $A$. Then any vector in the span of these rows is also perpendicular, because dot product is linear.

On the other hand, if $x$ is not in the kernel, then it has a nonzero dot product with one of the rows, so it is not perpendicular to the row space. Therefore $\perp(\ker(A)) = \mathrm{row}(A)$.

Since the row space of $A$ is perpendicular to the kernel of $A$, we see that $\mathrm{row}(A) \cap \ker(A) = \{0\}$. Now combine the Subspace Dimension Formula, the fact that $\dim(\mathrm{row}(A)) = \dim(\mathrm{col}(A))$, and the Rank plus Nullity Theorem to obtain

$$\begin{aligned}\dim(\mathrm{row}(A) + \ker(A)) &= \dim(\mathrm{row}(A)) + \dim(\ker(A)) + \dim(\mathrm{row}(A) \cap \ker(A)) \\ &= \dim(\mathrm{col}(A)) + \dim(\ker(A)) + 0 \\ &= \dim(\mathbb{R}^n).\end{aligned}$$

Since $\mathrm{row}(A) + \ker(A)$ is a subspace of $\mathbb{R}^n$ with the same dimension, it must be all of $\mathbb{R}^n$. Therefore $\mathbb{R}^n = \mathrm{row}(A) \oplus \ker(A)$. □

**Corollary A.7.** *Let $U \leq \mathbb{R}^m$. Then $\mathbb{R}^m = U \oplus \perp(U)$.*

*Proof.* Let $\{u_1, \ldots, u_r\}$ be a basis for $U$. Form the matrix

$$A = [u_1 \mid \cdots \mid u_r \mid 0 \cdots \mid 0].$$

Then $\mathrm{col}(A) = U$, and $\perp(\mathrm{col}(A)) = \ker(A^*)$ with $\mathbb{R}^m = \mathrm{col}(A) \oplus \ker(A^*)$. □

**Example A.8.** Let $U$ be the subspace of $\mathbb{R}^m$ spanned by the vectors $\{v_1, \ldots, v_n\}$.

**(a)** Find a basis for $U$.
**(b)** Find a basis for $\perp(U)$.

*Method of Solution.* Form the $m \times n$ matrix $A = [v_1 \mid \cdots \mid v_n]$. Use forward elimination only to row reduce the augmented matrix $[A \mid I]$ to an augmented matrix $[B \mid O]$. A basis for $U$ is given by the columns of $A$ which correspond to the basic columns of $B$. Since $U = \mathrm{col}(A)$, a basis for $\perp(U)$ is given by the last $m - r$ rows of $O$, where $r = \dim(U)$. □

## 4. Exercises

**Exercise A.1.** Let $U \leq \mathbb{R}^n$. Show that

**(a)** $\perp(U) \leq \mathbb{R}^n$;
**(b)** $U \cap \perp(U) = \{0\}$;
**(c)** $\perp(\perp(U)) = U$.

**Exercise A.2.** Let

$$A = \begin{bmatrix} 2 & 0 & -1 & 4 & 1 \\ -2 & 0 & 2 & -2 & 0 \\ 0 & 0 & 1 & 2 & 2 \end{bmatrix}.$$

Let $T : \mathbb{R}^5 \to \mathbb{R}^3$ be the linear transformation given by $T(v) = Av$.

**(a)** Find a basis for $\mathrm{img}(T)$ and for $\ker(T)$.
**(b)** Find a basis for $\perp(\mathrm{img}(T))$ and for $\perp(\ker(T))$.

**Exercise A.3.** Let $U$ be the subspace of $\mathbb{R}^4$ spanned by the vectors

$$v_1 = (1, 0, -1, 1),\ v_2 = (2, 1, 1, 0),\ \text{and } v_3 = (0, -1, -3, 2).$$

**(a)** Find a basis for $U$.
**(b)** Find a basis for $\perp(U)$.
**(c)** Find a matrix $A$ such that $U = \ker(A)$.

**Exercise A.4.** Let

$$A = \begin{bmatrix} 1 & 2 & 0 & 2 & 1 \\ 0 & 1 & 1 & 1 & 0 \\ 2 & 1 & 1 & 0 & 1 \\ 0 & 4 & 0 & 5 & 1 \end{bmatrix}.$$

Find a basis for each of the four fundamental spaces associated to $A$.

APPENDIX B

# Matrix Techniques

Abstract. This appendix collects matrix techniques for solving problems in linear algebra. None of these techniques should be applied without an understanding of why they work.

## 1. Elementary Invertible Matrices

The identity matrix is denoted by $I$.

The elementary invertible matrices are

- $E(i,j;c)$ is $I$ except $a_{ij} = c$;
- $D(i;c)$ is $I$ except $a_{ii} = c$;
- $P(i,j)$ is $I$ except $a_{ii} = a_{jj} = 0$ and $a_{ij} = a_{ji} = 1$.

The inverses of the elementary invertible matrices are

- $E(i,j;c)^{-1} = E(i,j;-c)$;
- $D(i;c)^{-1} = D(i;c^{-1})$;
- $P(i,j)^{-1} = P(i,j)$.

Let $E$ be an elementary invertible matrix. Multiplying on the left of $A$ to form $EA$ has the indicated effect on the rows of $A$. Multiplying on the right of $A$ to form $AE$ has the analogous effect on the columns of $A$.

| | |
|---|---|
| $E(i,j;c)$ | Multiply the $j^{\text{th}}$ row by $c$ and add to the $i^{\text{th}}$ row |
| $D(i;c)$ | Multiply the $i^{\text{th}}$ row by $c$ |
| $P(i,j)$ | Swap the $i^{\text{th}}$ row and the $j^{\text{th}}$ row |

## 2. Gaussian Elimination

Let $A$ denote the original matrix.
Let $B = OA$ be the result of forward elimination, where $O$ is invertible.
Let $C = UA$ be the result of backward elimination, where $U$ is invertible.
Let $M$ be the modified augmented matrix obtained by solution readoff.
The basic columns of $B$ or $C$ are the columns containing the pivots.
The free columns of $B$ or $C$ are the other columns.
The basic columns of $A$ or $M$ correspond to the basic columns of $B$ or $C$.
The free columns of $A$ or $M$ correspond to the free columns of $B$ or $C$.
Let $r$ be the number of basic columns of $B$ or $C$.
Let $k$ be the number of free columns of $B$ or $C$.
The basic rows of $O$ or $U$ are the first $r$ rows.
The free rows of $O$ or $U$ are the last $m - r$ rows.

**Forward Elimination**

(1) Start with the first nonzero column.
(2) If the top entry in the column is zero, permute with a lower row so that the top entry is nonzero (use $P$).
(3) Eliminate all entries below this one (use $E$).
(4) Repeat this process, disregarding the current top row and all rows above it.

**Backward Elimination**

(1) Make all pivots equal to one (use $D$).
(2) Starting from the right, working upward then leftward, make all entries above a pivot equal to zero (use $E$).

**Solution Readoff**

(1) insert a zero row at row $i$ for every free variable $x_i$;
(2) multiply each free column by $-1$;
(3) add $e_i$ to each free column;
(4) the particular solution is now the augmentation column;
(5) the homogeneous solution is now the span of the free columns.

## 3. Finding a Basis for Fundamental Subspaces

The four fundamental subspaces associated to $A$ are the column space $\mathrm{col}(A)$, the row space $\mathrm{row}(A)$, the kernel $\ker(A)$, and the kernel of the transpose $\ker(A^*)$.

The primary techniques for finding a basis of these spaces are:

**(F1)** The basic columns of $A$ are a basis for $\mathrm{col}(A)$.
**(F2)** The nonzero rows of $B$ or $C$ are a basis for $\mathrm{row}(A)$.
**(F3)** The free columns of $M$ are a basis for $\ker(A)$.
**(F4)** The free rows $O$ or $U$ are a basis for $\ker(A^*)$.

These secondary techniques are implied by the primary techniques:

**(F5)** The basic columns of $A$ are a basis for $\mathrm{row}(A^*)$.
**(F6)** The nonzero rows of $B$ or $C$ are a basis for $\mathrm{col}(A^*)$.

To avoid backward elimination, row reduce $A^*$ instead of $A$ and apply techniques **(F2)** and **(F4)** instead of **(F1)** and **(F3)**.

## 4. Finding a Basis for a Span

Let $X = \{w_1, \dots, w_n\} \subset \mathbb{R}^m$ and let $W = \mathrm{span}(X)$.
Form the $m \times n$ matrix $A = [w_1 \mid \cdots \mid w_n]$.
Reduce $A$ and apply **(F1)**; a basis for $W$ is a basis for $\mathrm{col}(A)$.
Reduce $A^*$ and apply **(F2)**; a basis for $W$ is a basis for $\mathrm{row}(A^*)$.

## 5. Test for Linear Independence

Let $X = \{w_1, \dots, w_n\} \subset \mathbb{R}^m$.
If $n > m$, then $X$ is dependent.
Form the $m \times n$ matrix $A = [w_1 \mid \cdots \mid w_n]$.
Reduce $A$; if $n = r$, then $X$ is independent, otherwise it is not.

## 6. Test for Spanning

Let $X = \{w_1, \dots, w_n\} \subset \mathbb{R}^m$.
If $n < m$, then $X$ does not span $\mathbb{R}^m$.
Form the $m \times n$ matrix $A = [w_1 \mid \cdots \mid w_n]$.
Reduce $A$; if $m = r$, then $X$ spans $\mathbb{R}^m$; otherwise it does not.

## 7. Test for a Basis

Let $X = \{w_1, \dots, w_n\} \subset \mathbb{R}^m$.
If $n > m$, then $X$ is not a basis.
If $n < m$, then $X$ is not a basis.
If $n = m$, then $X$ is a basis if and only if $X$ spans.
If $n = m$, then $X$ is a basis if and only if $X$ is independent.

## 8. Finding the Inverse

If $A$ is not square, it cannot be invertible.

Reduce $A$ to $B$.

If $r < n$, then $A$ is not invertible.

Reduce $B$ to $C$; then $A^{-1} = U$.

## 9. Finding the Determinant I

If $A$ is not square, the determinant of $A$ is undefined.

Select any row or column and expand along it.

Along the $i^{\text{th}}$ row:

$$\det(A) = \sum_{j=1}^{n} (-1)^{j-1} a_{ij} \det(A_{ij}).$$

Along the $j^{\text{th}}$ column:

$$\det(A) = \sum_{i=1}^{n} (-1)^{i-1} a_{ij} \det(A_{ij}).$$

Here, $A_{ij}$ is the $ij^{\text{th}}$ minor matrix of $A$.

## 10. Finding the Determinant II

If $A$ is not square, the determinant of $A$ is undefined.

Reduce $A$ to $B$ via forward elimination using $E$ and $P$ but not $D$.

If $r < n$, then $\det(A) = 0$.

If $r = n$, then $B$ is upper triangular and $\det(B)$ is the product of the diagonal entries.

Thus $\det(A) = (-1)^p \det(B)$, where $p$ is the number of $P$ matrices used in forward elimination.

## 11. Finding Eigenvalues and Eigenvectors

Let $A$ be an $n \times n$ matrix.

The characteristic polynomial of $A$ is

$$\chi_A(\lambda) = \det(A - \lambda I);$$

this is a polynomial of degree $n$.

Then $a$ is an eigenvalue of $A$ if and only if $a$ is a root of $\chi_A(\lambda)$.

To find eigenvectors associated to $a$, find a basis for $\ker(A - aI)$.

## 12. Test for Diagonalizability

Let $A$ be an $n \times n$ matrix.

Then $A$ is diagonalizable if and only if $\mathbb{R}^n$ has a basis of eigenvectors of $A$.

To diagonalize $A$, find a basis of eigenvectors and construct the matrix $C$ which has these eigenvectors as columns.

Then $B = C^{-1}AC$ is diagonal.

APPENDIX C

# Gram-Schmidt Orthonormalization Process

ABSTRACT. This appendix describes the Gram-Schmidt process for taking a basis for a subspace of $\mathbb{R}^n$ and producing a orthonormal basis.

let $v, w \in \mathbb{R}^n$. We say that $v$ and $w$ are *orthogonal* if $v \cdot w = 0$. Recall that this occurs exactly when $w$ is perpendicular to $v$. We now generalize this to sets of vectors.

A subset $X \subset \mathbb{R}^n$ is *orthogonal* if $X$ is a set of nonzero vectors such that for every distinct $x, y \in X$, we have $x \cdot y = 0$. An orthogonal set of vectors $X$ is *orthonormal* if for every $x \in X$, we have $|x| = 1$.

**Proposition C.1.** *Let $X \subset \mathbb{R}^n$ be orthogonal. Then $X$ is independent.*

*Proof.* Let $\sum_{i=1}^m a_i x_i = 0$ be a dependence relation from $X$. Let $j$ be between 1 and $m$, and take the dot product of both sides of the dependence relation with $x_j$:

$$\left(\sum_{i=1}^m a_i x_i\right) \cdot x_j = 0 \cdot x_j.$$

Since dot product is linear, this gives

$$\sum_{i=1}^m a_i (x_i \cdot x_j) = 0.$$

Since $X$ is orthogonal, this becomes

$$a_i x_j \cdot x_j = a_i |x_j| = 0.$$

Since $x_j$ is nonzero, we conclude that $a_j = 0$. Since $j$ was arbitrary, $a_i = 0$ for all $i$. This shows independence. □

**Proposition C.2.** *Let $V \leq \mathbb{R}^n$ and set $X = \{x_1, \ldots, x_m\} \subset V$ be an orthonormal basis for $V$. Then every for $v \in V$, we have*

$$v = \sum_{i=1}^{m} (v \cdot x_i) x_i.$$

*Proof.* Let $v \in V$ and $X$ is a basis for $V$, then $v$ is a linear combination of the elements of $X$; that is,

$$v = \sum_{i=1}^{m} a_i x_i$$

for some $a_1, \ldots, a_m \in \mathbb{R}$. Let $x_j \in X$. Taking the dot product of both sides of this equation with $x_j$, we have

$$\begin{aligned} v \cdot x_j &= \left( \sum \right)_{i=1}^{m} a_i x_i \Big) \cdot x_j \\ &= a_j (x_j \cdot x_j) \quad (\text{ because } x_i \cdot x_j = 0 \text{ for } i \neq j) \\ &= a_j \quad (\text{because } |x_j| = 1) \end{aligned}$$

This is all we needed to show. □

**Proposition C.3. Gram-Schmidt Process**
*Let $V \leq \mathbb{R}^n$. Then $V$ has an orthonormal basis.*

*Proof.* Let $Y = \{y_1, \ldots, y_m\}$ be a basis for $V$. Set $x_1 = y_1$. The *vector projection* of $y_2$ onto $x_1$ is $\frac{x_1 \cdot y_2}{|x_1|^2} x_1$. The difference between this and $y_2$ is perpendicular to $x_1$; thus let $x_2 = y_2 - \frac{x_1 \cdot y_2}{|x_1|^2} x_1$. Continuing in this way, inductively define

$$x_k = y_k - \left( \sum_{i=1}^{k-1} \frac{x_i \cdot y_k}{|x_i|^2} x_i \right).$$

Let $X = \{x_1, \ldots, x_m\}$. We claim that $X$ is an orthogonal basis for $V$. It is easy to see that $\text{span}(X) = \text{span}(Y)$, and since they have the same cardinality, we must have that $X$ is a basis for $Y$.

To check that $X$ is an orthogonal set of vectors, we apply dot product. Let $x_k \in X$ and select $j < k$. By the principle of induction, we may assume that $x_i \cdot x_j = 0$ for $i < k$; we wish to show that this implies that $x_k \cdot x_j = 0$. Thus compute:

$$\begin{aligned} x_k \cdot x_j &= \left[ y_k - \left( \sum_{i=1}^{k-1} \frac{x_i \cdot y_k}{|x_i|^2} x_i \right) \right] \cdot x_j \\ &= y_k \cdot x_j - \left( \sum_{i=1}^{k-1} \frac{x_i \cdot y_k}{|x_i|^2} x_i \cdot x_j \right) \\ &= y_k \cdot x_j - \left( \frac{x_j \cdot y_k}{|x_j|} x_j \cdot x_j \right) \\ &= y_k \cdot x_j - x_j \cdot y_k \qquad = 0. \end{aligned}$$

Thus $x_k \perp x_j$.

To obtain an orthonormal basis from $X$, divide each element $x \in X$ by its length. □

# Bibliography

[EP88] Edwards, C.H., and Penney, David, *Elementary Linear Algebra*, Prentice-Hall, Inc. (1986)

[KH05] Kolman,Bernad and Hill,David, *Introductory Linear Algebra: An Applied First Course*, $8^{\text{th}}$ edition, Prentice Hall (2005)

[Ha58] Halmos, Paul R., *Finite-Dimensional Vector Spaces*, Undergraduate Texts in Mathematics, Springer-Verlag (1958,1974,1987)

[Ha60] Halmos, Paul R., *Naive Set Theory*, Undergraduate Texts in Mathematics, Springer-Verlag (1960,1974)

[MT76] Marsden, Jerrold, and Tromba, Anthony, *Vector Calculus*, W.H.Freeman and Company (1976)

[Sm78] Smith, Larry, *Linear Algebra*, Undergraduate Texts in Mathematics, Springer-Verlag (1978)

[St87] Stewart,James *Calculus*, $2^{\text{nd}}$ edition, Brooks/Cole Publishing Company (1987,1991)