Linear Geometry

Paul L. Bailey

Department of Mathematics, BASIS Scottsdale
E-mail address: paul.bailey@basised.com
Date: February 13, 2019

(Version 2.50 Updated 02/13/2019)

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, emphasizing that the objects of study are spaces consisting of lines, and functions between them that map lines to lines. This is consistent with the usage of this material in vector calculus, and this text in intended to illuminate that subject.

Many authors begin the study of linear algebra by examining solution methods for systems of linear equations; in this approach, matrices become arrays of coefficients of linear equations, and the definition of multiplication of matrices appears as an unmotivated concept which somewhat mysteriously produces results.

The approach in this text is the emphasize the geometry at all times; solving systems of linear equations is a problem which naturally evolves from attempts to solve geometric problems involving lines, planes, and so forth. This is algebra in the service of geometry, and for this reason, we have chosen the title *Linear Geometry*.

We begin with the geometry of uncoordinatized euclidean 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 transfers the geometric study of vectors in euclidean space to the algebraic study of points in cartesian space.

We introduce *affine spaces* as subsets of $\mathbb{R}^n$ which are closed under lines, such as lines and planes. After developing explicit and implicit representations of affine spaces, we define *vector spaces* as subsets of $\mathbb{R}^n$ which are closed under algebraic vector operations. We see that vector spaces are exactly those affine spaces which pass through the origin, and that affine spaces are translations of vector spaces. This again links the geometry with the algebra.

The natural functions between affine spaces are those which send lines to lines, which we call *affine transformations*; the natural functions between vector spaces are those which preserve the vector operations, which we call *linear transformations*. We see that linear transformations are exactly those affine transformations on vector spaces which preserve the origin, and that affine transformations may be decomposed into a translation, followed by a linear transformation, followed by the inverse translation.

Working with such spaces and transformations produces systems of linear equations, which are initially solved using Gaussian elimination directly on the systems. The solutions thus obtained are themselves affine.

We then see how linear transformations can be described by matrices. In this context, multiplication of matrices corresponds to composition of linear transformations, and solving systems of linear equations using matrices is equivalent to finding the preimage of a point under a linear transformation. Such a preimage, in turn, is itself an affine space.

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.

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:	$\mathbb{N} = \{1, 2, 3, \dots\}$
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:	$\mathbb{R} = \{\text{infinite decimal expansions}\}$
Complex Numbers:	$\mathbb{C} = \{a + ib \mid a, b \in \mathbb{R} \text{ and } i^2 = -1\}$

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 $(a_1, \ldots, a_n)$, where again the entry a_i is known as the i^{th} *coordinate* of the tuple. If $A_1, \ldots, A_n$ are sets, their product is

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

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

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

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

$$A \times \mathbb{Z}^3 = B \times C \times D = \{(1, 2, 4), (1, 3, 4), (2, 2, 4), (2, 3, 4), (3, 2, 4), (3, 3, 4)\},$$

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":

for every $a \in A$ there exists a unique $b \in B$ 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 *codomain* 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 1.1. Some authors use the word range to mean what we have called the image of a function.

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. $\square$

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 $f\restriction_C$.

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 $\mathbb{N} = \{1, 2, 3, \dots\}$ be the set of natural numbers and for $n \in \mathbb{N}$ let

$$\mathbb{N}_n = \{m \in \mathbb{N} \mid m \leq n\} = \{1, 2, 3, \dots, n\}.$$

A set X is called *finite* if there exists a surjective function from X to $\mathbb{N}_n$ for some $n \in \mathbb{N}$. If there exists a bijective function $X \to \mathbb{N}_n$, 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

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

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 $i^2 = -1\}$
$\mathbb{R}^2$	Cartesian Plane	$\{(a, b) \mid a, b \in \mathbb{R}\}$
$\mathbb{R}^3$	Cartesian 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 $f^{-1}(C \cup D) = f^{-1}(C) \cup f^{-1}(D)$.
(b) Show that $f^{-1}(C \cap D) = f^{-1}(C) \cap f^{-1}(D)$.

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) $A \subset B \Rightarrow B \smallsetminus A = B$
(c) $A \smallsetminus (B \cup C) = (A \smallsetminus B) \cup (A \smallsetminus C)$
(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 in $\mathbb{R}^n$

Abstract. The primary goal of this chapter is to define the concept of "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 words and 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 $(x_1, \dots, x_n)$, where $x_1, \dots, x_n$ are real numbers, with the defining property that

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

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

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

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

Let $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_n)$ 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*.

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

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

□

Example 2.3. The distance in $\mathbb{R}^5$ between $(-2, -1, 0, 1, 2)$ and $(5, 4, 3, 2, 1)$ is

$$d = \sqrt{(5 - (-2))^2 + (4 - (-1))^2 + (3 - 0)^2 + (2 - 1)^2 + (1 - 2)^2} = \sqrt{96} = 4\sqrt{6}.$$

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 2.4. 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. $\square$

Example 2.5. 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. $\square$

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 *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. Two equations are *equivalent* if they have the same locus.

Example 2.6. The locus of $y = x$ in $\mathbb{R}^2$ is a diagonal line through the origin.

If we are considering and loci in $\mathbb{R}^n$ and one of the variables is missing from the equation, any value for that variable will satisfy the equation.

Example 2.7. The locus in $\mathbb{R}^3$ of the equation $y = x$ a vertical plane containing the line $y = x$ on the xy-plane, and the z-axis.

Example 2.8. 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.

Consider the locus in $\mathbb{R}^3$ of the equation $z = 0$. This is the set of points of the form $(x, y, 0)$. This set is 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 loci of $x = 0$ and $y = 0$ are the yz-plane and the xz-plane, respectively. Together, these sets are called *coordinate planes*.

Example 2.9. 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 locus is the union of the loci for these latter equations; that is, the locus of the equation $xyz = 0$ is the union of the coordinate planes. □

Consider the locus in $\mathbb{R}^3$, using variables x, y, and z, of the equation $y^2+z^2 = 0$. Since x can be anything, but y and z both must be zero, the locus is the set of points of the form $(x, 0, 0)$, which is the x-axis. Similarly, the loci of $x^2 + z^2 = 0$ and $x^2 + y^2 = 0$ are the y-axis and the z-axis, respectively. Together, these are called *coordinate axes*.

Example 2.10. Find the locus in $\mathbb{R}^3$ of the equation $(x^2+y^2)(x^2+z^2)(y^2+z^2) = 0$.

Solution. Since $ab = 0$ if and only if either $a = 0$ or $b = 0$, we obtain the union of loci of equations with one side equalling zero by multiplying. Thus we can see that the locus of $(x^2 + y^2)(x^2 + z^2)(y^2 + z^2) = 0$ is the union of the coordinate axes. □

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 locus 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 locus. Thus the equation of a sphere is

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

Example 2.11. 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. □

A *system of equations* in n-variables is a set of equations, each involving the same n-variables. The *locus* of the system is set of points in $\mathbb{R}^n$ which simultaneously satisfy all of the equations in the system. Such sets are merely the intersection of the loci of the individual equations. Two systems of equations are *equivalent* if they have the same locus.

Example 2.12. The locus of $\{x = 0, y = 0\}$ in $\mathbb{R}^3$ is the z-axis.

Typically, the equations are listed without the set braces.

Example 2.13. Find the locus in $\mathbb{R}^2$ of the system of equations

$$\begin{aligned} x + y &= 1; \\ x - y &= 3. \end{aligned}$$

Solution. If we add the equations, we obtain an equation which is also true for any point (x, y) in the locus. Here, the result is $2x = 4$, so $x = 2$. Substituting this into the first equation gives $2 + y = 1$, so $y = -1$. Both x and y are "bound" to specific values, so the locus is a set containing a point $\{(2, -1)\}$. □

Often, in order to understand the locus of a system of equations, we need to modify a system to obtain an equivalent system, which is more easily understood.

Example 2.14. Find the locus in $\mathbb{R}^3$ of the system of equations

$$\begin{aligned} x^2 + y^2 + z^2 &= 4; \\ x^2 + y^2 + (z-2)^2 &= 4. \end{aligned}$$

Solution. Subtracting these equations produces $-4z + 4 = 0$; dividing this by -4 and adding 4 to both sides gives $z = 1$. Substituting this result for the second equation, we obtain the equivalent system of equations

$$\begin{aligned} x^2 + y^2 + z^2 &= 4; \\ z &= 1. \end{aligned}$$

The locus of the first equation is a sphere of radius 2 centered at the origin. The locus of the second equation is a plane of height 1 parallel to and above the xy-plane. The intersection of a sphere and a plane is a circle on that plane. To find the radius of the circle, substitute $z = 1$ into the sphere to obtain

$$x^2 + y^2 = 3.$$

The locus of this equation in $\mathbb{R}^3$ is a cylinder surrounding the z-axis of radius $\sqrt{3}$. The circle is the intersection of this cylinder with the perpendicular plane $z = 1$; thus the radius of the circle is $\sqrt{3}$. □

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 its 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 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.* If the arrow starts at point P (P is the tail) and ends at Q (Q is the tip), denote this arrow by $\widehat{PQ}$. This is the arrow *from* P *to* Q.

To obtain more precision, we formally 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. The ordered pair is enough information to produce the line segment and differentiate the tip from the 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 thought of as a point. Such an arrow is called a *zero arrow.*

Each nonzero arrow determines a unique line, which is the line through the tip and the tail. There are two rays on the line whose endpoint is the tail of the arrow; the ray which contains the tip is known as the *orientation* of the arrow.

A nonzero arrow is determined by three attributes:

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

A zero arrow has zero magnitude and no direction; it is determined by its position.

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

We may add two arrows in a natural way if the tip of the first equals to the tail of the second; the sum is the defined to be the arrow which starts at the tail of the first and ends at the tip of the second, so that

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

The arrow $\widehat{PR}$ forms the third side of a triangle. Note that if we add the arrow $\widehat{PQ}$ to its inverse arrow $\widehat{QP}$, we obtain the zero arrow $\widehat{PP}$.

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. We need to be able to "slide" the arrows around via *parallel transport*; that is, we need to disregard the position of the arrow, and 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.

Say that two arrows are *equivalent* if they have the same direction and magnitude, but possibly different positions. Break the set of all arrows in a euclidean space into blocks, where the members of one block consist of all arrows which are equivalent to any other arrow in the block. We call such a block an *equivalence class* of arrows. Every arrow is in exactly one equivalence class. Any arrow in an equivalence class is called a *representative* of that class.

7. Vectors

A *vector* is an equivalence class of arrows. If $\widehat{v}$ is an arrow, define

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

this is the *vector represented by* $\widehat{v}$. If $\widehat{w}$ is equivalent to $\widehat{v}$, we say that $\widehat{w}$ *represents* $\vec{v}$. Naturally, since $\widehat{v}$ is equivalent to itself, $\widehat{v}$ represents $\vec{v}$. Technically, the phrase "$\widehat{w}$ represents $\vec{v}$" means that $\widehat{w} \in \vec{v}$, and in fact, $\widehat{w} \in \vec{v}$ if and only if $\vec{w} = \vec{v}$.

All zero arrows are equivalent; the *zero vector* is the equivalence class consisting of all of the zero arrows. 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 the arrow $-\widehat{v}$, where $\widehat{v}$ represents $\vec{v}$.

A nonzero vector is determined by two attributes:

(1) *direction*;
(2) *magnitude*.

Thus a vector is unpositioned direction and length.

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

For any vector $\vec{v}$ and any point P, 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 $\overrightarrow{PQ}$ and $\overrightarrow{RS}$; let $\widehat{QT}$ be the unique arrow with the same magnitude and direction as $\widehat{RS}$, and define the geometric sum by $\overrightarrow{PQ} + \overrightarrow{RS} = \overrightarrow{PT}$. Note that $-\overrightarrow{PQ} = \overrightarrow{QP}$ and that $\overrightarrow{PQ} + \overrightarrow{QP}$ is the point P; thus adding the inverse vector produces the zero vector.

Note that if $\overrightarrow{PQ} + \overrightarrow{QR} = \overrightarrow{PR}$, we may add the inverse of $\overrightarrow{PQ}$ to both sides of this equation to get $\overrightarrow{PR} - \overrightarrow{PQ} = \overrightarrow{QR}$. Let $\vec{v} = \overrightarrow{PQ}$, $\vec{w} = \overrightarrow{QR}$, and $\vec{x} = \overrightarrow{PR}$, this equation becomes $\vec{w} - \vec{v} = \vec{x}$. Thus the vector from the tip of $\vec{v}$ to the tip of $\vec{w}$ is $\vec{w} - \vec{v}$.

We now consider coordinatized cartesian space $\mathbb{R}^n$. Two arrows are equivalent in $\mathbb{R}^n$ if and only if the differences of the corresponding coordinates are equal. $A = (a_1, \ldots, a_n)$, $B = (b_1, \ldots, b_n)$, $C = (c_1, \ldots, c_n)$, and $D = (d_1, \ldots, d_n)$, then $\overrightarrow{AB}$ is equivalent to $\overrightarrow{CD}$ if and only if

$$b_i - a_i = d_i - c_i \quad \text{for all } i = 1, \ldots, n.$$

Equivalence classes of arrows in $\mathbb{R}^n$ are called vectors in $\mathbb{R}^n$.

Each vector in $\mathbb{R}^n$ 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 $\overrightarrow{OP}$ is called the *position vector* of P.

If $P = (x_1, \ldots, x_n) \in \mathbb{R}^n$ is any point, the position vector of P will be denoted $\langle x_1, \ldots, x_n \rangle$. It makes sense to define the difference of points to be the vector from the first to the second; thus if $A = (a_1, \ldots, a_n)$, $B = (b_1, \ldots, b_n)$, we define

$$Q - P = \overrightarrow{QP} = \langle b_1 - a_1, \ldots, b_n - a_n \rangle.$$

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

8. Vector Addition

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

$$\vec{v} + \vec{w} = \langle v_1 + w_1, v_2 + w_2, \ldots, v_n + w_n\rangle.$$

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

Example 2.15. Let $\vec{v} = \langle 2, 3\rangle$ and $\vec{w} = \langle -1, 4\rangle$ be vectors in $\mathbb{R}^2$. Then

$$\vec{v} + \vec{w} = \langle 2 + (-1), 3 + 4\rangle = \langle 1, 7\rangle.$$

The zero vector in $\mathbb{R}^n$ corresponds to the origin, and is denoted $\vec{0}$, or $\vec{0}_n$ if we wish to emphasize that we are referring to the origin in $\mathbb{R}^n$. If $\vec{v} = \langle x_1, \ldots, x_n\rangle$, the inverse vector of $\vec{v}$ is

$$-\vec{v} = \langle -x_1, \ldots, x_n\rangle.$$

Example 2.16. We compute the sum of the vectors whose tips form an equilateral triangle inscribed in the unit circle of $\mathbb{R}^2$. Let $\vec{v}_1, \vec{v}_2, \vec{v}_3$ be vectors in the plane given by $\vec{v}_j = \langle \cos(2\pi j/3), \sin(2\pi j/3)\rangle$. Then $\vec{v}_1 = \langle -\sqrt{3}/2, 1/2\rangle$, $\vec{v}_2 = \langle -\sqrt{3}/2, -1/2\rangle$, and $\vec{v}_3 = \langle 1, 0\rangle$, so

$$(\vec{v}_1 + \vec{v}_2) + \vec{v}_3 = \langle -1, 0\rangle + \langle 1, 0\rangle = \vec{0}.$$

To indicate that the vector $\vec{v}$ is n-dimensional, we write $\vec{v} \in \mathbb{R}^n$.

Proposition 2.17 (Primary Properties of Vector Addition)**.**
Let $\vec{x}, \vec{y}, \vec{z} \in \mathbb{R}^n$. Then

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

Remark. These properties are derived directly from the definition. □

Subtraction of vectors means adding the inverse, which is given by

$$\vec{w} - \vec{v} = \vec{w} + (-\vec{v}).$$

The vector from the tip of $\vec{v}$ to the tip of $\vec{w}$ is $\vec{w} - \vec{v}$, which is clear, because $\vec{v} + (\vec{w} - \vec{v}) = \vec{w}$.

Example 2.18. Let $\vec{v} = \langle 1, -3, 0, 5\rangle$ and $\vec{w} = \langle 1, 2, 3, 4\rangle$ be vectors in $\mathbb{R}^4$. The inverse vector of $\vec{v}$ is $-\vec{v} = \langle -1, 3, 0, -5\rangle$, and the vector from the tip of $\vec{v}$ to the tip of $\vec{w}$ is

$$\vec{w} - \vec{v} = \vec{v} + (-\vec{w}) = \langle 1 - 1, 2 - (-3), 3 - 0, 4 - 5\rangle = \langle 0, 5, 3, -1\rangle.$$

9. Scalar Multiplication

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

$$a \cdot \vec{v} = \langle a\vec{v}_1, a\vec{v}_2, \ldots, a\vec{v}_n \rangle.$$

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

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

Example 2.19. Let $\vec{v} \in \mathbb{R}^3$ be given by $\vec{v} = \langle 3, -2, 8 \rangle$ and $a = 5$. Then the vector in the same direction as $\vec{v}$ but 5 times as long is

$$a\vec{v} = 5\langle 3, -2, 8 \rangle = \langle 15, -10, 8 \rangle.$$

Example 2.20. Let $\vec{v} \in \mathbb{R}^3$ be given by $\vec{v} = \langle 49, -14, 35 \rangle$. Then $\vec{v} = 7\langle 8, -2, 5 \rangle$.

Proposition 2.21 (Primary Properties of Scalar Multiplication)**.**
Let $\vec{x}, \vec{y}, \vec{z} \in \mathbb{R}^n$ and let $a, b \in \mathbb{R}$. Then

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

Remark. These properties are derived directly from the definition. □

Proposition 2.22 (Secondary Properties of Scalar Multiplication)**.**
Let $\vec{x}, \vec{y}, \vec{z} \in \mathbb{R}^n$ and let $a, b \in \mathbb{R}$. Then

- **(a)** $0 \cdot \vec{x} = \vec{0}$;
- **(b)** $a \cdot \vec{0} = \vec{0}$;
- **(c)** $-1 \cdot \vec{x} = -\vec{x}$;
- **(d)** $(-a)\vec{x} = -(a\vec{x})$.

Remark. These properties may be derived from the primary properties. □

We say that two nonzero vectors $\vec{v}$ and $\vec{w}$ are *parallel*, and write $\vec{v} \| \vec{w}$, if arrows representing $\vec{v}$ and $\vec{w}$ lie on parallel line segments. This happens exactly when $\vec{w}$ is a scalar multiple of $\vec{v}$:

$$\vec{v} \| \vec{w} \quad \Leftrightarrow \quad \vec{w} = a\vec{v} \text{ for some } a \in \mathbb{R}.$$

Example 2.23. Let $P = (4, -1, -1)$, $Q = (7, 2, 5)$, $R = (3, 4, 5)$, and $S = (5, 6, 9)$. Show that $\vec{PQ} \| \vec{RS}$.

Solution. We have $\vec{PQ} = Q - P = \langle 3, 3, 6 \rangle$, and $\vec{RS} = S - R = \langle 2, 2, 4 \rangle$. Let $\vec{v} = \langle 1, 1, 2 \rangle$. Then $\vec{PQ} = 3\vec{v}$ and $\vec{RS} = 2\vec{v}$, so $\vec{PQ} = \frac{3}{2}\vec{RS}$. Thus $\vec{PQ} \| \vec{RS}$. □

10. Linear Combinations

A *linear combination* of the vectors $\vec{v}_1, \ldots, \vec{v}_m \in \mathbb{R}^n$ is an expression of the form

$$a_1\vec{v}_1 + \cdots + a_m\vec{v}_m,$$

where $a_1, \ldots, a_m \in \mathbb{R}$. This expression produces another vector in $\mathbb{R}^n$.

Consider the case of two vectors $\vec{v}, \vec{w} \in \mathbb{R}^2$. A linear combination of $\vec{v}$ and $\vec{w}$ is an expression of the form $a\vec{v} + b\vec{w}$, where $a, b \in \mathbb{R}$.

Example 2.24. Let $\vec{v} = \langle 1, 1\rangle$ and $\vec{w} = \langle 1, -1\rangle$. Then $\vec{v}+\vec{w}$ is a linear combination $a\vec{v} + b\vec{w}$ of $\vec{v}$ and $\vec{w}$, with $a = b = 1$, and $\vec{v} + \vec{w} = \langle 2, 0\rangle$.

Let $\vec{i} = \langle 1, 0\rangle$ and $\vec{j} = \langle 0, 1\rangle$ in $\mathbb{R}^2$. Then every vector in $\mathbb{R}^2$ can be expressed as a linear combination of $\vec{i}$ and $\vec{j}$, thusly:

$$\langle x, y\rangle = \langle x, 0\rangle + \langle 0, y\rangle = x\langle 1, 0\rangle + y\langle 0, 1\rangle = x\vec{i} + y\vec{j}.$$

We say that $\vec{i}$ and $\vec{j}$ form a *basis* for $\mathbb{R}^2$, because every vector in $\mathbb{R}^2$ can be written as a linear combination of these two. This, however, is not the only basis.

Example 2.25. Let $\vec{v} = \langle 2, 1\rangle$, $\vec{w} = \langle -1, 2\rangle$, and $\vec{x} = \langle 5, 1\rangle$, Express $\vec{x}$ as a linear combination of $\vec{v}$ and $\vec{w}$.

Solution. We wish to find $a, b \in \mathbb{R}$ such that $\vec{x} = a\vec{v} + b\vec{w}$, that is, $\langle 5, 1\rangle = \langle 2a - b, a + 2b\rangle$. Now $\langle v_1, v_2\rangle = \langle w_1, w_2\rangle$ if and only if $v_1 = w_1$ and $v_2 = w_2$, so this gives us a system of equations in a and b:

$$\begin{aligned} 2a - b &= 5;\\ a + 2b &= 1. \end{aligned}$$

The first equation says $b = 2a - 5$, and substituting this into the second gives $a + (2a - 5) = 1$, so $3a = 6$, whence $a = 2$. Thus $b = 2(2) - 5 = -1$. This shows that $\vec{x} = 2\vec{v} - \vec{w}$. □

Next, consider a linear combination of two vectors in $\mathbb{R}^3$.

Example 2.26. Let $\vec{v} = \langle 2, -1, 3\rangle$, $\vec{w} = \langle -3, -1, 8\rangle$, $a = 2$, and $b = -3$. Then

$$a\vec{v} + b\vec{w} = 2\langle 2, -1, 3\rangle - 3\langle -3, -1, 8\rangle = \langle 4 + 9, -2 + 3, 6 - 24\rangle = \langle 13, 1, -18\rangle.$$

Let $\vec{i} = \langle 1, 0, 0\rangle$, $\vec{j} = \langle 0, 1, 0\rangle$, and $\vec{k} = \langle 0, 0, 1\rangle$ in $\mathbb{R}^3$. Then every vector in $\mathbb{R}^3$ can be expressed as a linear combination of these three vectors:

$$\langle x, y, z\rangle = x\vec{i} + y\vec{j} + z\vec{k}.$$

These are the *standard basis vectors* in $\mathbb{R}^3$.

11. Norm

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 $\vec{x} = \langle x_1, \dots, x_n \rangle$, the norm of $\vec{x}$ is denoted $\|\vec{x}\|$ and is given by

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

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

Example 2.27. Let $\vec{v} \in \mathbb{R}^3$ be given by $\vec{v} = \langle 2, 4, 8 \rangle$. Find $\|\vec{v}\|$.

Solution. The norm is

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

$\square$

Example 2.28. Let $\vec{v} \in \mathbb{R}^8$ be given by $\vec{v} = \langle 1, 3, 4, 2, 5, 4, 3, 1 \rangle$. Find $\|\vec{v}\|$.

Solution. The norm is

$$\|\vec{v}\| = \sqrt{1 + 9 + 16 + 4 + 25 + 16 + 9 + 1} = \sqrt{81} = 9.$$

$\square$

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

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

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

Example 2.29. Let $\vec{v} \in \mathbb{R}^3$ be given by $\vec{v} = \langle 2, 4, 8 \rangle$. Find $\|\vec{v}\|$. Find a unit vector in the same direction as v.

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

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

$\square$

12. Dot Product

Let $\vec{v} = \langle v_1, v_2, \ldots, v_n \rangle$ and $\vec{w} = \langle w_1, w_2, \ldots, w_n \rangle$ be vectors in $\mathbb{R}^n$. We define the *dot product* of $\vec{v}$ and $\vec{w}$ to be the real number $\vec{v} \cdot \vec{w}$ given by

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

Example 2.30. Let $\vec{v}, \vec{w} \in \mathbb{R}^3$ be given by $\vec{v} = \langle -2, -1, 4 \rangle$ and $\vec{w} = \langle -5, 12, 1 \rangle$. Then

$$\vec{v} \cdot \vec{w} = (-2)(-5) + (-1)(12) + (4)(1) = 10 - 12 + 4 = 2.$$

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 $\vec{v}+\vec{w} \in \mathbb{R}^n$ and $a\vec{v} \in \mathbb{R}^n$, but $\vec{v} \cdot \vec{w} \in \mathbb{R}$.

Proposition 2.31 (Properties of Dot Product and Norm)**.** *Let $\vec{x}, \vec{y}, \vec{z} \in \mathbb{R}^n$ and $a \in \mathbb{R}$. Then*

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

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

Drop a perpendicular from the tip of $\vec{v}$ onto the line through $\vec{w}$ to obtain a right triangle. If θ is the angle between the vectors $\vec{v}$ and $\vec{w}$, we see that

$$\mathrm{proj}_{\vec{w}}(\vec{v}) = \|\vec{v}\| \cos\theta.$$

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

Lemma 2.32 (Law of Cosines)**.** *Let A, B and C be the vertices of a triangle, whose corresponding opposite sides have lengths a, b, and c, respectively. Let θ be the angle at vertex C. Then*

$$c^2 = a^2 + b^2 - 2ab\cos\theta.$$

Proof. For simplicity, we assume that θ is the largest angle, so that the other two angles are acute. The other cases are similar.

Drop a perpendicular from vertex 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 $m = b\cos\theta$, proving the result. □

Proposition 2.33. *Let $\vec{v}, \vec{w} \in \mathbb{R}^n$ and let θ be the angle between $\vec{v}$ and $\vec{w}$. Then*

$$\vec{v}\cdot\vec{w} = \|\vec{v}\|\|\vec{w}\|\cos\theta.$$

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

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

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

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

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

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

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

$$\vec{v}\cdot\vec{w} = \|\vec{v}\|\|\vec{w}\|\cos\theta.$$

□

Corollary 2.34. *Let $\vec{v}, \vec{w} \in \mathbb{R}^n$ and let θ be the angle between $\vec{v}$ and $\vec{w}$. Then*

$$\mathrm{proj}_{\vec{w}}(\vec{v}) = \frac{\vec{v}\cdot\vec{w}}{\|\vec{w}\|}.$$

If $\vec{u}$ is of unit length, then

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

Proof. From trigonometry, $\mathrm{proj}_{\vec{w}}(\vec{v}) = \|\vec{v}\|\cos\theta$. But $\cos\theta = \frac{\vec{v}\cdot\vec{w}}{\|\vec{v}\|\|\vec{w}\|}$. The result follows. □

Geometrically, the dot product of $\vec{v}$ and $\vec{w}$ is the length of the projection $\vec{v}$ onto $\vec{w}$, divided by the length of $\vec{w}$.

Example 2.35. Let $\vec{v} = \langle 5, 2, 1\rangle$ and $\vec{w} = \langle 3, 2, 3\rangle$. Find the scalar and vector projections of $\vec{v}$ onto $\vec{w}$, and find the angle between them.

Solution. We know that $\vec{v} \cdot \vec{w} = \|\vec{w}\| \mathrm{proj}_{\vec{w}}(\vec{v})$. Thus

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

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

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

We say that $\vec{v}$ is *orthogonal* (or *perpendicular*) to $\vec{w}$, and write $\vec{v} \perp \vec{w}$, if the angle θ 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 $\vec{v}$ onto $\vec{w}$ is the zero vector.

Dot product gives us a test for perpendicularity:

$$\vec{v} \perp \vec{w} \quad \Leftrightarrow \quad \vec{v} \cdot \vec{w} = 0.$$

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

Example 2.36. Let $\vec{v} = \langle 5, 2, 1\rangle$ and $\vec{w} = \langle 3, 2, 3\rangle$. Verify that these vectors form a right triangle.

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

We finish this section with a useful formula.

Proposition 2.37. *Let $\vec{v} = \langle a, b\rangle$ and $\vec{w} = \langle c, d\rangle$ be vectors in $\mathbb{R}^2$. The area of the parallelogram determined by $\vec{v}$ and $\vec{w}$ is $|ad - bc|$.*

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

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

□

Example 2.38. Find the area of the triangle in $\mathbb{R}^2$ with vertices $P = (1, 2)$, $Q = (5, 3)$, and $R = (-3, 7)$.

Solution. We turn this into a problem involving vectors by treating P as a "translated origin"; subtract by P to translate the corresponding vertex to the origin. Thus let $\vec{v} = Q - P = \langle 4, 1\rangle$ and $\vec{w} = R - P = \langle -8, 4\rangle$. The area of the triangle is half of the area of the parallelogram, so

$$\mathrm{area}(\triangle PQR) = \frac{1}{2}((4)(4) - (1)(-8)) = 12.$$

□

13. 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, in defining the cross product, to use the standard basis vectors $\vec{i} = \langle 1,0,0\rangle$, $\vec{j} = \langle 0,1,0\rangle$, and $\vec{k} = \langle 0,0,1\rangle$; then $\langle x,y,z\rangle = x\vec{i} + y\vec{j} + z\vec{k}$.

Let $\vec{v} = \langle v_1, v_2, v_3\rangle$ and $\vec{w} = \langle w_1, w_2, w_3\rangle$ be vectors in $\mathbb{R}^3$. We define the *cross product* of $\vec{v}$ and $\vec{w}$ to be the vector $\vec{v} \times \vec{w}$ given by

$$\vec{v} \times \vec{w} = \langle v_2w_3 - v_3w_2, v_3w_1 - v_1w_3, v_1w_2 - v_2w_1\rangle.$$

This may be rewritten as

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

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 2×2 matrix is

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

The determinant of a 3×3 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

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

Proposition 2.39.

(a) $\vec{i} \times \vec{j} = \vec{k}$;
(b) $\vec{j} \times \vec{k} = \vec{i}$;
(c) $\vec{k} \times \vec{i} = \vec{j}$.

Proposition 2.40 (Properties of Cross Product). *Let $\vec{x}, \vec{y}, \vec{z} \in \mathbb{R}^n$ and $a \in \mathbb{R}$. Then*

(a) $\vec{x} \times \vec{y} = -(\vec{y} \times \vec{x})$;
(b) $(a\vec{x}) \times \vec{y} = \vec{x} \times (a\vec{y}) = a(\vec{x} \times \vec{y})$;
(c) $\vec{x} \times (\vec{y} + \vec{z}) = (\vec{x} \times \vec{y}) + (\vec{x} \times \vec{z})$;
(d) $(\vec{x} + \vec{y}) \times \vec{z} = (\vec{x} \times \vec{z}) + (\vec{y} \times \vec{z})$;
(e) $\vec{x} \cdot (\vec{y} \times \vec{z}) = (\vec{x} \times \vec{y}) \cdot \vec{z}$;
(f) $\vec{x} \times (\vec{y} \times \vec{z}) = (\vec{x} \cdot \vec{z})\vec{y} - (\vec{x} \cdot \vec{y})\vec{z}$;
(g) $\vec{x} \times \vec{0} = \vec{0}$.

Remark. To prove any of these identities, write each vector in terms of their components and use the algebraic definition of cross product.

Property **(a)** says that cross product is *anticommutative*. Properties **(b)** through **(d)** are referred to as the *linearity* of cross product. □

Proposition 2.41. *Let $\vec{v}, \vec{w} \in \mathbb{R}^3$. Then $(\vec{v} \times \vec{w}) \perp \vec{v}$ and $(\vec{v} \times \vec{w}) \perp \vec{w}$.*

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

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

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

Proposition 2.42. *Let $\vec{v}, \vec{w} \in \mathbb{R}^3$, and let θ be the angle between $\vec{v}$ and $\vec{w}$. Then*

$$\|\vec{v} \times \vec{w}\| = \|\vec{v}\|\|\vec{w}\| \sin\theta,$$

which is the area of the parallelogram determined by $\vec{v}$ and $\vec{w}$.

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

Now consider

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

Also,

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

These last quantities are the same, and since $\theta \in [0, \pi]$, we have $\sin\theta \geq 0$. Thus we take square roots to yield the result. □

Proposition 2.43. *Let $\vec{v}, \vec{w} \in \mathbb{R}^3$. Then the triple $(\vec{v}, \vec{w}, \vec{v} \times \vec{w})$ is oriented by the right-hand rule.*

Remark. The orientation of $\vec{v} \times \vec{w}$ 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 $\vec{v} \times \vec{w}$ changes continuously as the lengths of $\vec{v}$ and $\vec{w}$ and the angle between them change. Thus if can deform $\vec{v}$ to $\vec{i}$ and $\vec{w}$ to $\vec{j}$ without getting a zero vector as the cross product, the orientation of $(\vec{v}, \vec{w}, \vec{v} \times \vec{w})$ must be the same as that of $(\vec{i}, \vec{j}, \vec{i} \times \vec{j})$, which is right handed. □

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

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

Proposition 2.44. *Let $\vec{v}, \vec{w} \in \mathbb{R}^3$. Then $\vec{v} \| \vec{w}$ if and only if $\vec{v} \times \vec{w} = \vec{0}$.*

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

Example 2.45. Find the area of the triangle in $\mathbb{R}^3$ with vertices $P = (2, 4, 1)$, $Q = (1, 2, 3)$, and $R = (5, 0, 1)$.

Solution. Let $\vec{v} = Q - P = \langle -1, -2, 2\rangle$ and $\vec{w} = R - P = \langle 3, -4, 0\rangle$. The area of the triangle is half of the area of the parallelogram determined by $\vec{v}$ and $\vec{w}$, which we find via the cross product:

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

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

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

$\square$

Example 2.46. Let $\vec{v} = \langle 2, 5, 1\rangle$ and $\vec{w} = \langle 3, 1, 2\rangle$. Find a vector which is perpendicular to both $22\vec{v} + 29\vec{w}$ and $83\vec{v} - 8\vec{w}$.

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

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

$\square$

Proposition 2.47. *Let $\vec{x}, \vec{y}, \vec{z} \in \mathbb{R}^3$. Then $\vec{x} \cdot (\vec{y} \times \vec{z})$ 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 $\vec{x} \cdot (\vec{y} \times \vec{z})$ the* scalar triple product.

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

The triple scalar product can the computed as a determinant.

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

Example 2.48. 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 2.49. Show that the maximum volume of a parallelepiped with sides of length one is one.

Solution. First, draw a picture and give everything in the picture a name. Let $\vec{x}, \vec{y}, \vec{z} \in \mathbb{R}^3$ have length one. Let $\vec{w} = \vec{x} \times \vec{y}$. Let θ be the angle between $\vec{x}$ and $\vec{y}$. Let ϕ be the angle between $\vec{z}$ and $\vec{w}$. 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 $\vec{z}$ onto $\vec{w}$; since $\vec{z}$ 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 $\vec{x} \perp \vec{y}$ and $\vec{z} \| \vec{w}$, which means that $\vec{x} \perp \vec{z}$ and $\vec{y} \perp \vec{z}$. This is a cube. Thus its volume is one. □

Example 2.50. Find the volume of the tetrahedron in $\mathbb{R}^3$ with vertices $A = (2,0,1)$, $B = (1,2,0)$, $C = (0,1,2)$, and $D = (2,2,2)$.

Solution. Consider a solid obtained from a base region in a plane and a point P in space, and taking the union of all line segment from the point to a point in the base. Suppose A is the area of the base and h is the distance from P to the plane. By a combination of the formula of a cone, Cavelieri's principle, and proportionality argument, the volume of the solid is

$$V = \frac{1}{3}Ah.$$

Let $\vec{v} = B - A = \langle -1, 2, -1\rangle$, $\vec{w} = C - A = \langle -2, 1, 1\rangle$, and $\vec{x} = D - A = \langle 0, 2, 1\rangle$. The base of the tetrahedron is half of the parallelogram determined by $\vec{v}$ and $\vec{w}$. The height is the projection of the vector $\vec{x}$ onto the perpendicular vector $\vec{v} \times \vec{w}$. Thus, the volume is one sixth of the triple scalar product:

$$V = \frac{1}{6}\det\begin{bmatrix} -1 & 2 & -1 \\ -2 & 1 & 1 \\ 0 & 2 & 1 \end{bmatrix} = \frac{1}{6}[-(1-2) - 2(-2-0) - (-4-0)] = \frac{3}{2}.$$

□

We have previously seen that the area of the parallelogram determined by vectors $\vec{v}$ and $\vec{w}$ in $\mathbb{R}^2$ is the determinant of the 2×2 matrix whose rows are $\vec{v}$ and $\vec{w}$. We have now seen that the volume of a parallelepiped determined by vectors $\vec{v}$, $\vec{w}$, and $\vec{x}$ in $\mathbb{R}^3$ is the determinant of the 3×3 matrix whose rows are $\vec{v}$, $\vec{w}$, and $\vec{x}$. Later, we will see that this is not a coincidence.

14. 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 cartesian space. The reason for the distinction is that there is more than one way to impose coordinates on euclidean space.
- 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 translation or 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.
- We think of the difference of points as a vector, and the sum of a point and a vector as another point. However, if we need to add points, we convert them into the vector they represent without specifically mentioning this; if we need to consider a vector as the tip of the representing arrow in standard position, we also proceed without further mention.
- The vector sum of two vectors traverses 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.

15. 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 locus 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 $\overrightarrow{AB}$. Find and draw the equivalent the vector $\vec{v}$ 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 $\vec{v} + \vec{w}$ and illustrate geometrically.

(a) $\vec{v} = \langle 0, 1\rangle$, $\vec{w} = \langle 1, 0\rangle$
(b) $\vec{v} = \langle 2, 4\rangle$, $\vec{w} = \langle 5, 1\rangle$
(c) $\vec{v} = \langle -2, 3\rangle$, $\vec{w} = \langle 3, -2\rangle$
(d) $\vec{v} = \langle 1, 0, 1\rangle$, $\vec{w} = \langle 0, 1, 0\rangle$
(e) $\vec{v} = \langle 1, 2, 3\rangle$, $\vec{w} = \langle -1, 2, -3\rangle$

Exercise 2.9. Find $\|\vec{v}\|$, $\vec{v} + \vec{w}$, $\vec{v} - \vec{w}$, $2\vec{v}$, and $3\vec{v} - 2\vec{w}$.

(a) $\vec{v} = \langle 1, 2\rangle$, $\vec{w} = \langle 3, 4\rangle$
(b) $\vec{v} = \langle -1, -2\rangle$, $\vec{w} = \langle 2, 1\rangle$
(c) $\vec{v} = \langle 3, 2, -1\rangle$, $\vec{w} = \langle 0, 6, 7\rangle$
(d) $\vec{v} = \vec{i} - \vec{j}$, $\vec{w} = \vec{i} + \vec{k}$
(e) $\vec{v} = \vec{i} + \vec{j} + \vec{k}$, $\vec{w} = 2\vec{i} - 3\vec{j} - 4\vec{k}$

Exercise 2.10. Find a unit vector which has the same direction as $\vec{v}$.

(a) $\vec{v} = \langle 3, 4 \rangle$
(b) $\vec{v} = \langle 5, -5 \rangle$
(c) $\vec{v} = \langle 1, 2, 3 \rangle$
(d) $\vec{v} = \vec{i} + \vec{j} + \vec{k}$

Exercise 2.11. Express $\vec{i}$ and $\vec{j}$ in terms of $\vec{v}$ and $\vec{w}$.

(a) $\vec{v} = \vec{i} + \vec{j}$, $\vec{w} = \vec{i} - \vec{j}$
(b) $\vec{v} = 2\vec{i} + 3\vec{j}$, $\vec{w} = \vec{i} - \vec{j}$

Exercise 2.12. Find $\vec{v} \cdot \vec{w}$.

(a) $\vec{v} = \langle 2, 4 \rangle$, $\vec{w} = \langle -1, 4 \rangle$
(b) $\vec{v} = \langle 5, -1 \rangle$, $\vec{w} = \langle 7, 7 \rangle$
(c) $\vec{v} = \langle 1, 2, 3 \rangle$, $\vec{w} = \langle 3, 2, 1 \rangle$
(d) $\vec{v} = \langle 2, -4, 1 \rangle$, $\vec{w} = \langle 3, 3, 6 \rangle$

Exercise 2.13. Find the scalar and vector projections of $\vec{v}$ onto $\vec{w}$.

(a) $\vec{v} = \langle 2, 4 \rangle$, $\vec{w} = \langle -1, 4 \rangle$
(b) $\vec{v} = \langle 5, -1 \rangle$, $\vec{w} = \langle 7, 7 \rangle$
(c) $\vec{v} = \langle 1, 2, 3 \rangle$, $\vec{w} = \langle 3, 2, 1 \rangle$
(d) $\vec{v} = \langle 2, -4, 1 \rangle$, $\vec{w} = \langle 3, 3, 6 \rangle$

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

(a) $\vec{v} = \langle 3, t \rangle$, $\vec{w} = \langle -4, 3 \rangle$
(b) $\vec{v} = \langle t, 12 \rangle$, $\vec{w} = \langle t^3, 18 \rangle$
(c) $\vec{v} = \langle t, 2t, 3t \rangle$, $\vec{w} = \langle 5, t, -2 \rangle$

Exercise 2.15. Find the values for $t \in \mathbb{R}$ such that the angle between $\vec{v} = \langle 1, 1 \rangle$ and $\vec{w} = \langle t, 1 \rangle$ is $60°$.

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

Exercise 2.17. Find $\vec{v} \times \vec{w}$.

(a) $\vec{v} = \langle 1, 0, 1 \rangle$, $\vec{w} = \langle 0, 1, 0 \rangle$
(b) $\vec{v} = \langle 1, 2, 3 \rangle$, $\vec{w} = \langle 1, 3, 5 \rangle$
(c) $\vec{v} = \langle 1, 1, 1 \rangle$, $\vec{w} = \langle -1, 1, 1 \rangle$

Exercise 2.18. Let $\vec{v} = \langle 1, 2, 3 \rangle$ and $\vec{w} = \langle 3, 2, 1 \rangle$. Find the following.

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

Exercise 2.19. Let $\vec{x} = \langle 1, 2, 3\rangle$, $\vec{y} = \langle -2, 0, -3\rangle$, and $\vec{z} = \langle 1, -2, 0\rangle$.

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

Exercise 2.20. Find the volume of the parallelepiped determined by the vectors $\vec{x} = \langle 1, 2, 3\rangle$, $\vec{y} = \langle 2, 3, 1\rangle$, and $\vec{z} = \langle -1, 0, t\rangle$. Find t such that these vectors are coplanar.

Exercise 2.21. The vectors $\vec{v} = \langle 1, 0, 1\rangle$ and $\vec{w} = \langle 0, 1, 1\rangle$ form a $60°$ angle. Find a third vector $\vec{x}$ such that the origin and the tips of $\vec{v}$, $\vec{w}$, and $\vec{x}$ are the vertices of a regular tetrahedron.

Exercise 2.22. 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 2.23. 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 2.24. Let $\vec{v}$ and $\vec{w}$ be vectors in $\mathbb{R}^2$. Give a geometric interpretation of and prove the following formulae:

- **(a)** Cauchy Schwartz Inequality:
$$\|\vec{v} \cdot \vec{w}\| \leq \|\vec{v}\|\|\vec{w}\|$$
- **(b)** Triangle Inequality:
$$\|\vec{v} + \vec{w}\| \leq \|\vec{v}\| + \|\vec{w}\|$$
- **(c)** Parallelogram Law:
$$\|\vec{v} + \vec{w}\|^2 + \|\vec{v} - \vec{w}\|^2 = 2\|\vec{v}\|^2 + 2\|\vec{w}\|^2$$

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

Exercise 2.25. The following identities are true for $\vec{x}, \vec{y}, \vec{z} \in \mathbb{R}^3$. Examine them for geometric content.

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

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

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

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

is orthogonal to $\vec{w}$.

Exercise 2.27. Let $\vec{v}$ and $\vec{w}$ be vectors in $\mathbb{R}^n$ and let θ be the angle between them. Recall that the formula $\vec{v} \cdot \vec{w} = \|\vec{v}\|\|\vec{w}\| \cos\theta$ implies that the scalar projection of $\vec{v}$ onto $\vec{w}$ is given by

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

(a) Define a function $\vec{v} : \mathbb{R} \to \mathbb{R}^2$ by $\vec{v}(t) = \langle t, t^2 \rangle$. Graph the image of $\vec{v}$.
(b) Let $\vec{w} = \langle 1, 1 \rangle \in \mathbb{R}^2$. Graph the line through $\vec{w}$.
(c) Define a function $f : \mathbb{R} \to \mathbb{R}$ by $f(t) = \dfrac{\text{proj}_{\vec{w}}(\vec{v}(t))}{t^2}$. 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 2.28. Let $f(t)$ be a real valued function given by

$$f(t) = \|\vec{i} \times \langle \cos t, \sin t, 0 \rangle\|.$$

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

CHAPTER 3

Affine Spaces in $\mathbb{R}^n$

Abstract. Affine spaces are subsets of euclidean space which look like lower dimensional euclidean spaces in the sense that they are closed under lines. We use our results regarding vectors to investigate affine spaces in cartesian space.

1. Affine Spaces in $\mathbb{R}^n$

An *affine space* in $\mathbb{R}^n$ is a nonempty subset $L \subset \mathbb{R}^n$ with the property that if $P, Q \in L$ are distinct points in L, then the entire line through P and Q is a subset of L. If L is an affine space in $\mathbb{R}^n$, we call $\mathbb{R}^n$ the *ambient space* of 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 an affine space, 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 affine spaces.

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

Recall that two sets are disjoint if their intersection is empty; otherwise, we say that they are nondisjoint. We will show that the intersection of nondisjoint affine spaces is always an affine space.

If we intersect two lines in $\mathbb{R}^2$, we typically obtain a point, although we may obtain a line (if the lines are the same), or the empty set (if the lines are distinct and parallel). If we intersect two planes in $\mathbb{R}^3$, we typically obtain a line, although we may obtain a plane (if the planes are the same), or the empty set (if the planes are distinct and parallel). If we intersect two lines in $\mathbb{R}^3$, we typically obtain the empty set (if the lines are skew or distinct and parallel), although we may obtain a line (if the lines are the same), or a point (if the lines are nonparallel and reside on the same plane).

It may initially be surprising that it is possible to intersect two planes in $\mathbb{R}^4$ and obtain a point, but that is exactly the case if we intersect certain coordinate planes. To see this, let x, y, z, w be the coordinate variables in $\mathbb{R}^4$. The xy-plane is $\{(x, y, 0, 0) \mid x, y \in \mathbb{R}\}$, the yz-plane is $\{(0, y, z, 0) \mid y, z \in \mathbb{R}\}$, and the zw-plane is $\{(0, 0, z, w) \mid z, w \in \mathbb{R}\}$. Thus the xy-plane and the yz-plane intersect in a line (the y-axis), but the xy-plane and the zw-plane intersect in a point (the origin). Ask yourself this: given two arbitrary planes, the intersection is empty, a point, a line, or a plane; which has the highest probability?

We now build techniques of representing an affine space either parametrically or as the locus of a linear equation. We start with lines in $\mathbb{R}^2$, planes in $\mathbb{R}^3$, lines in $\mathbb{R}^3$, and work our way into higher dimensions.

2. Lines in $\mathbb{R}^2$ via Parametric 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 given point on the line, and let $\vec{v} = \langle v_1, v_2\rangle$ be a vector in the direction of the line; we call $\vec{v}$ a *direction vector*. If we start at P_0 and move in the direction specified by $\vec{v}$ for a period of time t at a speed given by $\|\vec{v}\|$, we arrive at the point

$$P = P_0 + t\vec{v}.$$

If we let t range throughout the real numbers, then the set of points satisfying this equation form a line. Thus we call $P = P_0 + t\vec{v}$ the *parametric equation* of the line; it is this form of the equation of a line that most easily generalizes to higher dimensions. The variable t is called the *parameter*.

If we label $P = (x, y)$, where we think of x and y as variables, the parametric 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 known as the *coordinate parametric equations* of the line.

If $v_1v_2 \neq 0$, 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 $\langle 1, m\rangle$ is its direction vector.

Example 3.1. Find the parametric, coordinate, symmetric, and functional equations of the line in $\mathbb{R}^2$ through $Q = (4, 1)$ and $R = (-3, 3)$.

Solution. Let $\vec{v} = R - Q = \langle -7, 2\rangle$; this is a direction vector for the line. Then the parametric equation is $P = Q + t\vec{v}$, where $P = (x, y)$. Substituting in the values, this becomes

$$(x, y) = (4, 1) + t\langle -7, 2\rangle.$$

In coordinate form,

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

The symmetric equation is

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

The functional equation is

$$y = -\frac{2}{7}x + \frac{15}{14}.$$

□

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

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 specific point on the line, and let $\vec{n} = \langle n_1, n_2 \rangle$ be a vector which is perpendicular to the line; we call $\vec{n}$ a *normal vector*. If $P = (x, y)$ is a general point on the line, then the arrow from P_0 to P lies on the line. Thus $P - P_0$ is a vector in the direction of the line, and $\vec{n}$ is perpendicular to $P - P_0$. Therefore

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

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

The normal equation may be rewritten as $\langle x - x_0, y - y_0 \rangle \cdot \langle n_1, n_2 \rangle = 0$, that is, $n_1(x - x_0) + n_2(y - y_0) = 0$, or finally $n_1x + n_2y = n_1x_0 + n_2y_0$. Setting $a = n_1$, $b = n_2$, and $c = n_1x_0 + n_2y_0$, 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 $\langle a, b \rangle$; 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 $\vec{n} = \langle a, b \rangle$, rewrite this as $P \cdot n = P_0 \cdot \vec{n}$, which is equivalent to $(P - P_0) \cdot \vec{n} = 0$. But $P - P_0$ is a direction vector for the line, so $\vec{n}$ is perpendicular to it.

Example 3.2. 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 $\vec{v} = R - Q = \langle -5, 2 \rangle$. A vector perpendicular to this is $\vec{n} = \langle 2, 5 \rangle$, since $\vec{n} \cdot \vec{v} = -5(2) + 2(5) = 0$. The normal equation, then, is

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

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

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

□

In $\mathbb{R}^2$, it is easy to find a perpendicular vector. If $\vec{v} = \langle a, b \rangle$, set $\vec{w} = \langle b, -a \rangle$. Then $\vec{v} \cdot \vec{w} = ab - ba = 0$, so $\vec{v} \perp \vec{w}$.

Example 3.3. Find the general equation of the parametric line $(1 + 2t, 2 - 3t)$.

Solution. The direction vector of the line is $\vec{v} = \langle 2, -3 \rangle$, and a point on it is $Q = (1, 2)$. Let $\vec{n} = \langle 3, -2 \rangle$; then $\vec{v} \perp \vec{n}$; set $P = (x, y)$, so that the general equation of the line is $P \cdot \vec{n} = Q \cdot \vec{n}$; since $Q \cdot \vec{n} = 3(1) - 2(2) = -1$, the general equation is

$$3x - 2y = -1.$$

□

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

4. Lines in $\mathbb{R}^3$ via Parametric Equations

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 $\vec{v} = \langle v_1, v_2, v_3\rangle$ be a vector. If we start at P_0 and move in the direction $\vec{v}$ for a period of time t at a rate given by $\|\vec{v}\|$, we arrive at the point

$$P = P_0 + t\vec{v}.$$

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

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 *coordinate parametric equations* of the line.

Example 3.4. Find the parametric and coordinate equations of the line which passes through the points $Q = (1, 3, 2)$ and $R = (5, -2, 3)$.

Solution. Let $\vec{v}$ be the vector from Q to R. Thus $\vec{v} = R - Q = \langle 4, -5, 1\rangle$. This is the direction of the line we seek. Letting Q be the designated point on the line, we have that a point P is on the line if $P = Q + t\vec{v} = (1 + 4t, 3 - 5t, 2 + t)$. Thus the coordinate 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 coordinate 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 set determined by $P = P_0 + t\vec{v}$ 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:

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

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

Example 3.5. 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:

- Their direction vectors are parallel, so we call them *parallel lines*. Parallel lines intersect if and only if they are the same line.
- They intersect in exactly one point.
- 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 intuitively clear proposition that two distinct lines in space have parallel direction vectors if and only if they lie on the same plane but do not intersect.

Example 3.6. 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 $\langle 1, 2, 3\rangle$ and $\langle 2, -1, 1\rangle$, which are not parallel; thus the lines are not parallel.

We realize that the t in the first parametrization represents a different quantity than the t in the second parametrization. To see if the lines intersect, we cannot simply solve for t; 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 $\mathbb{R}^3$ via Normal Equations

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 specific point and let $\vec{n} = \langle n_1, n_2, n_3 \rangle$ be a vector. Suppose $P = (x, y, z)$ is a general point on the plane which passes through P_0 and is perpendicular to $\vec{n}$. Then the arrow from P_0 to P is on the plane, and the vector $P - P_0$ is perpendicular to the normal vector $\vec{n}$. Thus

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

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

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

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

which can also be written

$$n_1x + n_2y + n_3z = n_1x_0 + n_2y_0 + n_3z_0.$$

On the other hand, the locus of an equation

$$ax + by + cz = d$$

is a plane with normal vector $\vec{n} = \langle a, b, c \rangle$. This is called the *general equation* of the plane. Note that $d = P_0 \cdot \vec{n}$, where P_0 is any point on the plane.

Example 3.7. Find the general equation of the plane which passes through the point $Q = (2, 4, 1)$ with normal vector equal to the position vector.

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

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 note that:

- If the plane passes through the origin, $d = 0$.
- Otherwise, d is not unique.

For example, $x + 3y + 5z = 7$ and $2x + 6y + 10z = 14$ are equations of the same 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 3.8. 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 3.9. 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 $\vec{v} = Q - P = \langle 1, -1, -1\rangle$ and $\vec{w} = R - P = \langle -2, 5, -4\rangle$. There is an entire plane's worth of vectors which are perpendicular to $\vec{v}$, and a different plane's worth of vectors which are perpendicular to $\vec{w}$; their intersection is a line which is perpendicular to both. Let $\vec{n} = \langle n_1, n_2, n_3\rangle$ be a direction vector for this line. Then $\vec{n}$ 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 $\vec{n}$ is perpendicular to both $\vec{v}$ and $\vec{w}$, so

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

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 $\vec{n} = \langle 3n_3, 2n_3, n_3\rangle$ is a normal vector. Set $n_3 = 1$ to get $\vec{n} = \langle 3, 2, 1\rangle$. The equation of the plane is

$$n_1x + n_2y + n_3z = n_1x_0 + n_2y_0 + n_3z_0,$$

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 3.10. 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 $\vec{v} = Q - P = \langle -1, 4, 0\rangle$ and $\vec{w} = R - P = \langle 1, 1, 2\rangle$ lie on the plane. Thus their cross product is perpendicular to it, so we may use this as a normal vector; set

$$\vec{n} = \vec{v} \times \vec{w} = \langle 6, 2, -5\rangle$$

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:

- They intersect in a line.
- They have empty intersection. When this occurs, the normal vectors are parallel, and they are called *parallel planes*.
- They are equal.

Example 3.11. Consider the planes in $\mathbb{R}^3$ with general equations $x + 2y - z = -4$ and $2x - y + 3z = 7$. Find the parametric equation of the line which is the intersection of these planes.

Solution 1. We find two points of intersection; the direction vector will be their difference.

Set $z = 0$, so we have $x + 2y = 4$ and $2x - y = 7$. Then $y = 2x - 7$, so $x + 2(2x - 7) = -4$, whence $5x = 10$, so $x = 2$. Thus $y = -3$, and $Q = (2, -3, 0)$ is a point of intersection.

Set $z = 1$, so we have $x + 2y = -3$ and $2x - y = 4$. Then $y = 2x - 4$, so $x + 2(2x - 4) = -3$, whence $x = 1$. Thus $y = -2$, so $R = (1, -2, 1)$ is another point of intersection.

Let $\vec{v} = Q - R = \langle 1, -1, -1\rangle$; this is a direction vector for the line of intersection, and R is a point on it, so the parametric equation of the line is

$$(x, y, z) = R + t\vec{v} = (1 + t, -2 - t, 1 - t).$$

□

Solution 2. The normal vectors are $\vec{n}_1 = \langle 1, 2, -1\rangle$ and $\vec{n}_2 = \langle 2, -1, 3\rangle$. The line of intersection is on both planes, so it is perpendicular to both of these vectors; we compute

$$\vec{w} = \vec{n}_1 \times \vec{n}_2 = \det \begin{bmatrix} \vec{i} & \vec{j} & \vec{k} \\ 1 & 2 & -1 \\ 2 & -1 & 3 \end{bmatrix} = \langle 6 - 1, -2 - 3, -1 - 4\rangle = \langle 5, -5, -5\rangle.$$

Now $\vec{w}$ is a direction vector for the line, and since $\vec{v} = \langle 1, -1, -1\rangle$ is parallel to $\vec{w}$, it is also a direction vector. As above, a point on the plane is $(1, -2, 1)$, so the parametric equation of the line is

$$(x, y, z) = (1 + t, -2 - t, 1 - t).$$

□

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

Example 3.12. Find the angle between the planes $x+2y-z = -4$ and $2x-y+3z = 7$ in $\mathbb{R}^3$.

Solution. The normal vectors are $\vec{n}_1 = \langle 1, 2, -1\rangle$ and $\vec{n}_2 = \langle 2, -1, 3\rangle$. Thus we have

$$\cos\theta = \frac{\vec{n}_1 \cdot \vec{n}_2}{\|\vec{n}_1\|\|\vec{n}_2\|} = \frac{2 - 2 - 3}{\sqrt{1 + 4 + 1}\sqrt{4 + 1 + 9}} = \frac{-3}{2\sqrt{21}}.$$

We take the absolute value to obtain the supplementary angle. Thus

$$\theta = \arccos \frac{3}{2\sqrt{21}} \approx 70.89^\circ.$$

□

Given two vectors in $\mathbb{R}^3$, there is a unique plane which passes through their tips and the origin. We call this the plane *spanned* by these vectors. Every linear combination of the two vectors lies on the plane.

Example 3.13. Find the general equation of the plane spanned by $\vec{v} = \langle 2, -1, 4\rangle$ and $\vec{w} = \langle -2, 4, 3\rangle$.

Solution. The vectors lie on the plane, so the normal vector is perpendicular to both of them; it is

$$\vec{n} = \vec{v} \times \vec{w} = \det \begin{bmatrix} \vec{i} & \vec{j} & \vec{k} \\ 2 & -1 & 4 \\ -2 & 4 & 3 \end{bmatrix} = \langle -3 - 16, -8 - 6, 8 + 2\rangle = \langle -19, -14, 10\rangle.$$

Since the plane passes through the origin, the constant on the right hand side of its general equation is 0; thus the general equation is

$$-19x - 14y + 10z = 0.$$

□

Example 3.14. Let $\vec{v} = \langle 1, 2, 2\rangle$ and $\vec{w} = \langle 2, 0, 1\rangle$. Let Y be the plane spanned by $\vec{v}$ and $\vec{j}$ and let Z be the plane spanned by $\vec{w}$ and $\vec{k}$. 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 3.15. 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 $\vec{v} = P - Q$. Find the unit normal $\vec{n}$ to the plane. Project $\vec{v}$ onto $\vec{n}$.

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 $\langle 5, 3, 1\rangle$, so the unit normal is $\vec{n} = \frac{\langle 5,3,1\rangle}{\sqrt{35}}$.

Project the vector $\vec{v} = P - Q = \langle 6, 2, 3\rangle$ onto the unit normal. This will give the distance.

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

□

6. Planes in $\mathbb{R}^3$ via Parametric Equations

A plane in $\mathbb{R}^3$ may also be determined by one specific point on it, and two vectors indicating directions emanating from that point. Each point on the plane may then be reached by starting at the specified point, proceeding along the direction of one vector the necessary distance, and then proceeding in the direction of the other vector the necessary distance.

If $P = (x, y, z)$ is an arbitrary point in space, $P_0 = (x_0, y_0, z_0)$ is a point on the plane, $\vec{v} = \langle v_1, v_2, v_3\rangle$ is a direction along the plane, and $\vec{w} = \langle w_1, w_2, w_3\rangle$ is a different direction along the plane, then there exist real number $r, s \in \mathbb{R}$ such that

$$P = P_0 + r\vec{v} + s\vec{w};$$

this is the *parametric equation* of the plane. Expanding this, it reads

$$(x, y, z) = (x_0, y_0, z_0) + r\langle v_1, v_2, v_3\rangle + s\langle w_1, w_2, w_3\rangle.$$

This produces three equations

$$x = x_0 + rv_1 + sw_1; \quad y = y_0 + rv_2 + sw_2; \quad z = z_0 + rv_3 + sw_3;$$

these are the *coordinate parametric equations* of the plane.

Example 3.16. Find a parametric equation for the plane in $\mathbb{R}^3$ passing through the points $A = (1, 2, 3)$, $B = (3, 2, 1)$, and $C = (2, 0, 2)$.

Solution. We have a point on the plane; we need two vectors in the direction of the plane. Let $\vec{v} = A - C = \langle -1, 2, 1\rangle$ and $\vec{w} = B - C = \langle 1, 2, -1\rangle$. Then the parametric equation is

$$P = C + r\vec{v} + s\vec{w} = (2, 0, 2) + r\langle -1, 2, 1\rangle + s\langle 1, 2, -1\rangle = (2 - r + s, 2r + 2s, 2 + r - s).$$

□

Example 3.17. Find the general equation of the plane in $\mathbb{R}^3$ passing through $P_0 = (7, -3, 2)$ with direction vectors $\vec{v} = \langle 2, 1, -1\rangle$ and $\vec{w} = \langle 3, 1, 1\rangle$.

Solution. We have a point on the plane; we need a normal vector. Since $\vec{v}$ and $\vec{w}$ lie in the direction of the plane, a normal vector is perpendicular to both of them, so we use cross product to obtain it. We have

$$\vec{n} = \vec{v} \times \vec{w} = \det \begin{bmatrix} \vec{i} & \vec{j} & \vec{k} \\ 2 & 1 & -1 \\ 3 & 1 & 1 \end{bmatrix} = \langle 1 + 1, -3 + 2, 2 - 3\rangle = \langle 2, -1, -1\rangle.$$

Now $\vec{n} \cdot P_0 = 2(7) - 3 - 2 = 9$, so the general equation of the plane is

$$2x - y - z = 9.$$

□

Example 3.18. Find a parametric equation of the plane in $\mathbb{R}^3$ with normal equation $x + 2y + 3z = 6$.

Solution. We need to find a point on the plane, and two vectors in the direction of the plane. Three points on the plane are $A = (6, 0, 0)$, $B = (0, 3, 0)$, and $C = (0, 0, 2)$. Let $\vec{v} = A - C = \langle 6, 0, -2\rangle$ and $\vec{w} = B - C = \langle 0, 3, -2\rangle$, so that the plane is

$$(x, y, z) = A + r\vec{v} + s\vec{w} = (6 + 6r, 3s, -2r - 2s).$$

□

7. Lines and Planes in $\mathbb{R}^n$ via Parametric Equations

A line in $\mathbb{R}^n$ is determined by a point P_0 on the line and a direction vector $\vec{v}$; the points on the line are those we encounter if we proceed from P_0 in the direction of v. Each such point is of the form $P_0 + t\vec{v}$, 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

$$P = P_0 + t\vec{v} \quad \text{where } t \in \mathbb{R},$$

as t ranges through the real numbers; this is a *parametric equation* for the line. Note that the distance between P and P_0 is equal to $|t|\|\vec{v}\|$; we may think of $\|\vec{v}\|$ as the velocity with which we proceed away from the point P_0.

The parameter t 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

$$L = \{P \in \mathbb{R}^n \mid P = P_0 + t\vec{v} \text{ for some } t \in \mathbb{R}\}.$$

We can describe a plane in $\mathbb{R}^n$ if we know a point P_0 on the plane and two vectors $\vec{v}$ and $\vec{w}$ in the directions of the plane. Then the plane is the set of points P of the form

$$P = P_0 + r\vec{v} + s\vec{w} \quad \text{where } r, s \in \mathbb{R};$$

this is a *parametric equation* for the plane. The plane itself is the set

$$Z = \{P \in \mathbb{R}^n \mid P = P_0 + r\vec{v} + s\vec{w} \text{ for some } r, s \in \mathbb{R}\}.$$

8. Hyperplanes in $\mathbb{R}^n$ via Normal Equations

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 $\vec{a}$ and passing through a given point P_0. If $H \subset \mathbb{R}^n$ is such a hyperplane, then

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

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

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

where $\vec{a} = \langle a_1, \ldots, a_n \rangle$ is the normal vector perpendicular to the hyperplane, $P = (x_1, \ldots, x_n)$ is a variable point on the plane, with $x_1, \ldots, x_n$ being coordinate variables, and P_0 is a specific point on the plane.

The *general equation* of this hyperplane is

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

where $b \in \mathbb{R}$ is given by $b = P_0 \cdot \vec{n}$.

There are two types of hyperplanes; those that pass through the origin and those that do not. We will see that hyperplanes which pass through the origin have the additional property that they are closed under vector addition and scalar multiplication. In this way, they are both geometrically and algebraically identical to a copy of $\mathbb{R}^{n-1}$.

9. Affine Space Theory

Before we proceed to higher dimensional affine spaces, we formalize some of what we have begun to suspect with respect to their general properties.

Given a point and a line in a plane, there is a unique point on the line which is closer to the given point that any other point on the line. The next proposition says that this holds in general for affine spaces.

Proposition 3.19. *Let L be an affine space in $\mathbb{R}^n$ and let $P \in \mathbb{R}^n$. Then there exists a unique point $Q \in L$ such that $d(P,Q) \leq d(P,R)$ for every $R \in L$.*

Proof. Beyond the scope of this class. □

The next proposition states that the nonempty intersection of affine spaces is again an affine space.

Proposition 3.20. *Let $L, M \subset \mathbb{R}^n$ be nondisjoint affine spaces. Then $L \cap M$ is an affine space.*

Proof. If $L \cap M$ is a singleton, it is an affine space. Otherwise, let $P, Q \in L \cap M$ be distinct. Then $P, Q \in L$ and $P, Q \in M$, and since L and M are affine spaces, the entire line through P and Q is contained in L and is also contained in M. Therefore this line is contained in the intersection. □

We have described specific cases regarding two general methods of describing an affine space. These are

- **(a)** as the image of a parametric function;
- **(b)** as the locus of a system of linear equations.

In case **(a)**, the equations come from given direction vectors which lie along the affine space; that is, they radiate from some point on it. The affine space is built up from the given direction vectors; the more vectors we give, the larger the space may be.

In case **(b)**, the linear equations come from given normal vectors which are perpendicular to the affine space. The affine space is the intersection of the hyperplanes with the given normal vectors; the more vector we give, the smaller the space may be.

We now describe each of these representation techniques in more detail, and discuss how to move from one to the other.

10. Images of Parametric Transformations

We recall some definitions regarding sets. Let A and B be any sets and let $f : A \to B$ be any function; A is the *domain* and B is the *codomain*. If $a \in A$, we call $f(a)$ the *image* of a. If $C \subset A$, the *image* of C is

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

this is the subset of the codomain B consisting of the images of every $c \in C$. The *image* of f is $f(A)$.

Since affine spaces are defined as subsets of $\mathbb{R}^n$ which are closed under lines, the natural functions to consider between them are those which send lines to lines; that is, the image of a line is a line. For any two points P and Q, let $\overline{PQ}$ denote the line through P and Q.

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

$$L : \mathbb{R}^n \to \mathbb{R}^m \quad \text{satisfying} \quad L(\overline{AB}) = \overline{L(A)L(B)};$$

that is, the image of the line through A and B is the line through the images of A and B. So, an affine transformation is one which sends lines to lines. We will eventually see that affine transformations are exactly what we now describe as parametric transformations.

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

$$P : \mathbb{R}^n \to \mathbb{R}^m \quad \text{of the form} \quad P(t_1, \ldots, t_n) = P_0 + t_1\vec{v}_1 + \cdots + t_n\vec{v}_n.$$

where $t_1, \ldots, t_n$ are coordinate variables in the domain $\mathbb{R}^n$, $\vec{v}_1, \ldots, \vec{v}_n$ are fixed vectors in the codomain $\mathbb{R}^m$, and P_0 is a point in the codomain $\mathbb{R}^m$.

First, consider the case where $n = 1$ and $m = 3$. Then we have $P(t) = P_0 + t\vec{v}$, which we discussed as the parametric equation of a line. We now view this as a function; the variable t is on the copy of the real line which is the domain, and the variables x, y, and z are coordinate variables in the codomain $\mathbb{R}^3$. The function places the line into space. We may view this function as the path of a particle, where $P(t)$ is the particles location at time t. In this case, P_0 is its initial position, and $\|\vec{v}\|$ is its speed. From this point of view, the direction vector $\vec{v}$ may be called the *velocity vector*.

Next, consider the case where $n = 2$ and $m = 3$. Let r and s be the variables in the domain, and x, y, and z the variables in the codomain. The function P takes the plane with coordinates r and s and places it into space; $P(r, s)$ is where P sends the point (r, s).

The *dimension* of an affine space is smallest number of direction vectors which can be used to describe the space as the image of a parametric transformation. We need to be careful here, however, for some of the direction vectors may be a linear combination of other direction vectors.

For example, consider the function $P : \mathbb{R}^2 \to \mathbb{R}^3$ given by $P(r, s) = P_0 + r\vec{v} + s\vec{w}$ where $\vec{v} = \langle 6, -4, 2\rangle$, $\vec{w} = \langle -9, 6, -3\rangle$, and $P_0 = (1, 2, 3)$. Note that $\vec{w} = \frac{3}{2}\vec{v}$, so the direction vectors lie on the same line; thus the image is a line and not a plane.

In general, the dimension of the image is less than or equal to the dimension of the domain, and we have equality only if the direction vectors "point in different directions"; in the next chapter, we will formalize the meaning of this through the concept of linear independence.

11. Loci of Systems of Linear Equations

A *linear equation* in n variables $x_1, \ldots, x_n$ is an equation of the form

$$a_1 x_1 + \ldots a_n x_n = b.$$

Let P_0 be any point in the locus of this equation, say $P_0 = (b_1, \ldots, b_n)$. Then P_0 satisfies the equation, so $a_1 b_1 + \cdots + a_1 b_n = b$. If we let $P = (x_1, \ldots, x_n)$ denote an arbitrary point in the locus, and let $\vec{n} = \langle a_1, \ldots, a_n \rangle$, then the equation becomes $\vec{n} \cdot P = \vec{n} \cdot P_0$, so $(P - P_0) \cdot \vec{n} = 0$, which we recognize as the normal equation of a hyperplane in $\mathbb{R}^n$. Thus, *every* linear equation describes a hyperplane.

To construct a system of m equations in n variables, it is convenient to using double indexing of the coefficients. Thus we let a_{ij} denote the j^{th} coefficient in the i^{th} equation. In the way, we can write down a generic system of equations as

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1; \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2; \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_n. \end{aligned}$$

Each equation determines a hyperplane as its locus, and the locus of the system of equations is the intersection of the hyperplane.

In general, each equation cuts the dimension by one. However, again we must be careful; two of the equations may give the same hyperplane (this happens if one is a multiple of the other), or more subtly, one hyperplane may contain the intersection of some of the other hyperplanes.

For example, consider the system of equations

$$\begin{aligned} x + z &= 2; \\ x + y &= 3; \\ x - 2y + 3z &= 0. \end{aligned}$$

Subtracting the first from the second and third gives an equivalent system

$$\begin{aligned} x + z &= 2; \\ y - z &= 1; \\ -2y + 2z &= -2. \end{aligned}$$

The third equation divided by -2 equals the second, so it gives no additional information. If we move the z terms to the right hand side and insert the equation $z = 0 + z$, we obtain

$$\begin{aligned} x &= 2 - z; \\ y &= 1 + z; \\ z &= 0 + z. \end{aligned}$$

Writing these equations in vector form, we have $\langle x, y, z \rangle = \langle 2, 1, 0 \rangle + z\langle -1, 1, 1 \rangle$. Note that the variable z here is free, in the sense that it can take any real value. Letting $t = z$, we obtain a parametric form for the locus of the system of equations, and see that it is not a point but the line which is the image of the parametric function

$$\langle x, y, z \rangle = (2, 1, 0) + t\langle -1, 1, 1 \rangle.$$

12. Exercises

Exercise 3.1. Find the 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 $2\vec{\imath} - 4\vec{\jmath}$
(f) The line through $(-1, 1)$ perpendicular to the vector $(8, 1)$

Exercise 3.2. Find the 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 $3\vec{\imath} - 2\vec{\jmath} + \vec{k}$
(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 $(2,0,-1,1)$.

CHAPTER 4

Vector Spaces in $\mathbb{R}^n$

ABSTRACT. Vector spaces in $\mathbb{R}^n$ are subsets of cartesian space which are closed under vector addition and scalar multiplication. In this chapter, we see that vector spaces are affine spaces which pass through the origin, and that affine spaces are translations of vector spaces. The notions of span, linear independence, and basis are introduced. We also begin the study of linear transformations, which are functions between vectors spaces which preserve the vector operations.

1. Vector Spaces in $\mathbb{R}^n$

A *vector space* in $\mathbb{R}^n$ is a subset $V \subset \mathbb{R}^n$ satisfying

(S0) V is nonempty;
(S1) $\vec{v} + \vec{w} \in V$ for every $\vec{v}, \vec{w} \in V$;
(S2) $t\vec{v} \in V$ for every $\vec{v} \in V$ and $t \in \mathbb{R}$.

Property **(S1)** says that V is *closed under vector addition.* Property **(S2)** says that V 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 V. For if $\vec{0} \in \vec{v}$, then V is certainly nonempty; on the other hand, suppose that V is nonempty and let $v \in V$. Then $-1v = -v \in V$ by property **(S2)**, so $\vec{0} = \vec{v} + (-\vec{v}) \in V$ by property **(S1)**.

Example 4.1. Since $\mathbb{R}^n$ is nonempty and closed under vector addition and scalar multiplication, it is a vector space.

Example 4.2. Let $\vec{0} \in \mathbb{R}^n$ denote the origin, and let $V = \{\vec{0}\}$. Then V is a vector space, as $\vec{0} + \vec{0} = \vec{0} \in V$, and $t\vec{0} = \vec{0} \in V$ for all $t \in \mathbb{R}$.

Example 4.3. Let $\vec{v}, \vec{w} \in \mathbb{R}^3$ and let $V = \{a\vec{v} + b\vec{w} \mid a, b \in \mathbb{R}\}$. Then V is a vector space in $\mathbb{R}^3$. To see this, first note that $\vec{0} = 0\vec{v} + 0\vec{w} \in V$, so **(S0)** is satisfied. Next select arbitrary vectors $a_1\vec{v} + b_1\vec{w}$ and $a_2\vec{v} + b_2\vec{w}$ from V and note that their sum is $(a_1 + a_2)\vec{v} + (b_1 + b_2)\vec{w}$, which is also in V; thus **(S1)** is satisfied. Moreover, if $a\vec{v} + b\vec{w} \in V$ and $t \in \mathbb{R}$, we have $t(a\vec{v} + b\vec{w}) = (ta)\vec{v} + (tb)\vec{w} \in V$; thus **(S2)** is satisfied. In this example, the vector space V is a plane through the origin.

Let V be a vector space in $\mathbb{R}^n$. A *subspace* of V is a subset $W \subset V$ which is itself a vector space. Since $\mathbb{R}^n$ is itself a vector space, every vector space in $\mathbb{R}^n$ is a subspace of $\mathbb{R}^n$. The notation $W \leq V$ means that W is a subspace of V.

Example 4.4. Let $\vec{v}, \vec{w}, \vec{x} \in \mathbb{R}^3$ with $\vec{x} = \vec{v} + \vec{w}$, $V = \{a\vec{v} + b\vec{w} \mid a, b \in \mathbb{R}\}$, and $W = \{c\vec{x} \mid c \in \mathbb{R}\}$. Then $W, V \leq \mathbb{R}^3$, and $W \leq V$.

We see that lines and planes through the origin are also vector spaces; in fact, we have the following proposition.

Proposition 4.5. *A subset of $\mathbb{R}^n$ is a vector space if and only if it is an affine space which contains the origin.*

Proof. We prove both directions of implication.

($\Rightarrow$) Let V be a vector space; we wish to show that V is an affine space which contains the origin.

Then V is nonempty, so it contains some $\vec{v} \in V$. By **(S2)**, $-\vec{v} \in V$, so by **(S2)**, $\vec{v} + (-\vec{v}) = \vec{0} \in V$. Thus V contains the origin.

Let $\vec{v}, \vec{w} \in V$; we wish to show that the entire line through the tips of $\vec{v}$ and $\vec{w}$ is contained in V. The parametric equation of the line through the tips of $\vec{v}$ and $\vec{w}$ is $P = \vec{v} + t(\vec{w} - \vec{v})$. By **(S1)**, $\vec{w} - \vec{v} \in V$, so by **(S2)**, $t(\vec{w} - \vec{v}) \in V$. Adding $\vec{v}$ to this, **(S1)** says that $P \in V$.

($\Leftarrow$) Let A be an affine space which contains the origin; we wish to show that A is a vector space.

Let $\vec{v} \in A$ and $t \in \mathbb{R}$; if $\vec{v} = \vec{0}$, then $t\vec{v} = \vec{0} \in A$. Otherwise, the entire line through $\vec{v}$ and $\vec{0}$ is in A; since $t\vec{v}$ is on this line, it is in L, which shows **(S2)**.

Let $\vec{v}, \vec{w} \in A$ and assume that they are nonzero. The line through the tips of $\vec{v}$ and $\vec{w}$ is contained in A; the midpoint of the line segment from the tip of $\vec{v}$ to the tip of $\vec{w}$ is $\vec{x} = \vec{v} + \frac{1}{2}(\vec{w} - \vec{v})$. Thus $\vec{x} \in A$. Now $2\vec{x} \in A$ by the previous paragraph; but $2\vec{x} = 2\vec{v} - (\vec{w} - \vec{v}) = \vec{v} + \vec{w}$, so $\vec{v} + \vec{w} \in A$, which shows **(S1)**. $\square$

Let $A \subset \mathbb{R}^n$, and let $\vec{v} \in \mathbb{R}^n$ be a vector. The *translation* of A by $\vec{v}$ is

$$\vec{v} + A = \{\vec{v} + P \mid P \in A\}.$$

We may also write $A + \vec{v}$ to indicate $\{P + \vec{v} \mid P \in A\}$; this is the same set.

Proposition 4.6. *A translation of an affine space is an affine space.*

Proof. Let $\vec{v} + A$ be a translation of an affine space, where A is an affine space in $\mathbb{R}^n$ and $\vec{v} \in \mathbb{R}^n$. Let $P, Q \in \vec{v} + A$, and let R be a point on the line through P and Q. Then $R = P + t(Q - P)$ for some $t \in \mathbb{R}$.

Now $P - \vec{v}, Q - \vec{v} \in A$, and $(Q - P) = ((Q - \vec{v}) - (P - \vec{v}))$ is the vector between them; thus $R - \vec{v} = (P - \vec{v}) + t(Q - P)$ is on the line through $P - \vec{v}$ and $Q - \vec{v}$, so $R - \vec{v} \in A$. Thus $R \in \vec{v} + A$. $\square$

Translation produces a parallel linear set. For example, a translation of a line is a parallel line, and a translation of a plane is a parallel plane.

If P_0 is a point in $\mathbb{R}^n$, translation by P_0 means translation by the vector represented by the arrow in standard position with tip P_0. Thus, we may translate by points.

Proposition 4.7. *A subset of $\mathbb{R}^n$ is an affine space if and only if it is a translation of a vector space.*

Proof. We prove both directions of implication.

($\Rightarrow$) Let A be an affine space in $\mathbb{R}^n$. Then A is nonempty; let $\vec{x} \in A$, and let $\vec{w} = -\vec{x}$. Let $V = \vec{w} + A$; this is a translation of A, and $\vec{0} \in V$. Thus V is an affine space which contains the origin, and therefore is a vector space.

($\Leftarrow$) Let A be a translation of a vector space. Since a vector space is an affine space, A is a translation of an affine space, and is therefore itself an affine space. $\square$

2. Spans

Let $X \subset \mathbb{R}^n$ be a set of vectors from $\mathbb{R}^n$. A *linear combination from* X is an expression of the form

$$a_1\vec{v}_1 + \cdots + a_r\vec{v}_r,$$

where $\vec{v}_1, \ldots, \vec{v}_r \in X$ and $a_1, \ldots, a_r \in \mathbb{R}$. 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 X is the subset $\operatorname{span}(X) \subset \mathbb{R}^n$ defined by

$$\operatorname{span}(X) = \{\vec{v} \in \mathbb{R}^n \mid \vec{v} \text{ is a linear combination from } X\}.$$

If $\vec{v} \in X$, then $\vec{v}$ is a linear combination from X, with $r = 1$ and $a_1 = 1$. Thus, $X \subset \operatorname{span}(X)$. If $X \subset Y$, then $\operatorname{span}(X) \subset \operatorname{span}(Y)$, because any linear combination from X is a linear combination from Y. The next proposition may seem almost as clear, but as the proof shows, there is some work involved.

Proposition 4.8. *Let $Y \subset \mathbb{R}^n$. Then*

$$X \subset \operatorname{span}(Y) \Rightarrow \operatorname{span}(X) \subset \operatorname{span}(Y).$$

Proof. Suppose that $X \subset \operatorname{span}(Y)$. Pick an arbitrary vector $\vec{w} \in \operatorname{span}(X)$; it suffices to show that $\vec{w} \in \operatorname{span}(Y)$.

Since $\vec{w} \in \operatorname{span}(X)$, there exist vectors $\vec{w}_1, \ldots, \vec{w}_s \in X$ and real numbers $b_1, \ldots, b_r \in \mathbb{R}$ such that $\vec{w} = \sum_{j=1}^{s} b_j\vec{w}_j$.

Since $X \subset \operatorname{span}(Y)$, each vector $\vec{w}_j$ is a linear combination from Y. Each such linear combination may be written with finitely many vectors from Y; taking the union of sets of such vectors, there exists a finite set $Z = \{\vec{v}_1, \ldots, \vec{v}_r\} \subset Y$ such that each $\vec{w}_j$ is a linear combination from Z.

Thus, for each $j = 1, \ldots, s$, there exist real numbers $a_{1j}, \ldots, a_{rj} \in \mathbb{R}$ such that $\vec{w}_j = \sum_{i=1}^{r} a_{ij}\vec{v}_i$. Then, with liberal use of associativity of vector addition,

$$\vec{w} = \sum_{j=1}^{s} b_j\vec{w}_j = \sum_{j=1}^{s} b_j\left(\sum_{i=1}^{r} a_{ij}\vec{v}_i\right) = \sum_{j=1}^{s}\sum_{i=1}^{r} b_j a_{ij}\vec{v}_i = \sum_{i=1}^{r}\left(\sum_{j=1}^{s} a_{ij}b_j\right)\vec{v}_i$$

We have expressed $\vec{w}$ as a linear combination from Y. Therefore $\vec{w} \in \operatorname{span}(Y)$. □

Proposition 4.9. *Let $X \subset \mathbb{R}^n$ be nonempty. Then* $\operatorname{span}(X) \leq \mathbb{R}^n$.

Proof. Since X is nonempty, so is $\operatorname{span}(X)$. Moreover, sums and scalar products of linear combinations from X are linear combinations from X. Thus $\operatorname{span}(X)$ is closed under vector addition and scalar multiplication. □

Proposition 4.10. *Let $X \subset \mathbb{R}^n$. Then $X \leq \mathbb{R}^n$ if and only if* $\operatorname{span}(X) = X$.

Proof. Suppose X is a vector space in $\mathbb{R}^n$. We wish to show that $\operatorname{span}(X) = X$. Since we already know that $X \subset \operatorname{span}(X)$, it suffices to show that $\operatorname{span}(X) \subset X$. Let $\vec{w} \in \operatorname{span}(X)$. It suffices to show that $\vec{w} \in X$. Now $\vec{w}$ is a linear combination of vectors from X. Since X is a vector space, 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 $\vec{w} \in X$.

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

3. Linear Independence

Let V be a vector space in $\mathbb{R}^n$, and let $X \subset V$. We say that X *spans* V, or that X is a *spanning set* for V, if $\text{span}(X) = V$. Since $V = \text{span}(V)$, we know that V contains spanning sets. In fact, most of the vectors in V are a linear combination of only a few vectors in V.

A subset $X \subset V$ which spans V is minimal if no proper subset of X spans V; that is, if we remove any vector from X, it no longer spans V. Naturally, we would like to get the smallest possible spanning set, so we ask ourselves, under what conditions is a spanning set minimal?

Suppose that X is not minimal; then there is a vector $\vec{x} \in X$ such that if we remove $\vec{x}$, the span does not diminish; in that case, $\vec{x}$ is in the span of the other vectors in X. Thus $\vec{x}$ is a linear combination of the other vectors in X; $\vec{x} = \sum_{i=1}^{r-1} a_i\vec{v}_i$ for some $\vec{v}_i \in X$ and $a_i \in \mathbb{R}$. We can relabel $\vec{x} = \vec{v}_r$ and subtract it from both sides of this equation to obtain $\sum_{i=1}^{r} a_i\vec{v}_i = \vec{0}$.

Let $X \subset \mathbb{R}^n$; a *dependence relation* from X is an equation of the form

$$a_1\vec{v}_1 + \cdots + a_r\vec{v}_r = \vec{0},$$

where $\vec{v}_i \in X$ and $a_i \in \mathbb{R}$. This dependence relation is *trivial* if $a_i = 0$ for all i; otherwise, it is *nontrivial*. We see that we can write one vector in X as a linear combination of the other vectors if and only if X admits a nontrivial dependence relation.

We say that X is *linearly independent* if, for every $\vec{v}_1, \dots, \vec{v}_r \in X$ and $a_1, \dots, a_r \in \mathbb{R}$, we have

$$a_1\vec{v}_1 + \cdots + a_r\vec{v}_r = \vec{0} \quad \Rightarrow \quad a_1 = a_2 = \cdots = a_r = 0.$$

That is, X does not admit a nontrivial dependence relation. We see that X is a minimal spanning set for V if and only if X spans V and X is linearly independent.

Example 4.11. Let $\vec{v} = \langle 2, 3\rangle$ and $\vec{w} = \langle -4, 1\rangle$. Let $X = \{\vec{v}, \vec{w}\}$. Show that X is linearly independent.

Solution. Let $a\vec{v}+b\vec{w} = \vec{0}$ be a dependence relation; we wish to show that $a = b = 0$. We obtain a system of two equations in two variables

$$2a + 3b = 0;$$
$$-4a + b = 0.$$

Multiplying the first equation by 2 and adding, we get $7b = 0$, so $b = 0$, whence $a = 0$. $\square$

4. Bases

Let V be a vector space and let $B \subset V$. We say that B is a *basis* for V if

(B1) $\operatorname{span}(B) = V$;

(B2) B is linearly independent.

The plural of basis is *bases*.

Proposition 4.12. *Let V be a vector space, and let $B = \{\vec{v}_1, \ldots, \vec{v}_r\} \subset V$. Then B is a basis for V if and only if every vector in V can be written as a linear combination from B in a unique way.*

Proof. Suppose that B is a basis. Let $\vec{w} \in V$; since B is a basis, B spans V, so $\vec{w}$ can be written as a linear combination from B. Suppose we write $\vec{w}$ in two ways; then there are real numbers $a_1, \ldots, a_r, b_1, \ldots, b_r \in \mathbb{R}$ such that

$$\vec{w} = \sum_{i=1}^{r} a_i \vec{v}_i = \sum_{i=1}^{r} b_i \vec{v}_i.$$

Subtracting, we obtain

$$\sum_{i=1}^{r} (a_i - b_i)\vec{v}_i = \vec{0}.$$

Since B is independent, $a_i - b_i = 0$ for all i, so $a_i = b_i$ for all i; this shows that the expression of $\vec{w}$ as a linear combination is unique.

Suppose that every vector in V can be written as a linear combination from B in a unique way. Then B spans V. Suppose that

$$a_1\vec{v}_1 + \cdots + a_r\vec{v}_r = \vec{0}$$

is a dependence relation from B. This expresses $\vec{0}$ as a linear combination, and so it is unique. Since this equation is true for $a_1 = \cdots = a_r = 0$, this must be the case here. Thus B is linearly independent. $\square$

The i^{th} *standard basis vector* for $\mathbb{R}^n$ is denoted $\vec{e}_i$ and is defined to be the vector with 1 in the i^{th} coordinate and zero in every other coordinate; that is, $\vec{e}_1 = \langle 1, 0, \ldots, 0\rangle$, $\vec{e}_n = \langle 0, \ldots, 0, 1\rangle$, and

$$\vec{e}_i = \langle 0, \ldots, 0, 1, 0, \ldots, 0\rangle, \qquad \text{where the 1 is in the } i^{\text{th}} \text{ slot.}$$

The set $\{\vec{e}_i \in \mathbb{R}^n \mid i = 1, \ldots, n\}$ is called the *standard basis* for $\mathbb{R}^n$. Now if $\vec{v} = \langle a_1, \ldots, a_n\rangle = a_1\vec{e}_1 + \cdots + a_n\vec{e}_n$; the number a_i is call the i^{th} *component* of $\vec{v}$ with respect to the standard basis.

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

$$\{\vec{e}_1, \vec{e}_2, \vec{e}_3, \vec{e}_4\} = \{\langle 1,0,0,0\rangle, \langle 0,1,0,0\rangle, \langle 0,0,1,0\rangle, \langle 0,0,0,1\rangle\}.$$

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

$$\langle 1, -3, \pi, \sqrt{2}\rangle = \vec{e}_1 - 3\vec{e}_2 + \pi\vec{e}_3 + \sqrt{2}\vec{e}_4.$$

5. Dimension

We now show that any two bases have the same size, which will be the dimension of the vector space. First, we collect some observations in the form of lemmas.

Lemma 4.13. *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. □

Lemma 4.14. *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. □

Lemma 4.15. *Let V be a vector space and let $X \subset V$ be a spanning set. If $\vec{v} \in V \setminus X$, then $Y = X \cup \{\vec{v}\}$ is dependent.*

Proof. If $\vec{v} = \vec{0}$, then $1 \cdot \vec{v} = 0$ is a nontrivial dependence relation from Y, so Y is dependent; thus we may assume that $\vec{v} \neq \vec{0}$. Since X spans, we may write $\vec{v} = \sum_{i=1}^{m} a_i \vec{x}_i$ for some $a_i \in \mathbb{R}$ and $\vec{x}_i \in X$. Not all of the a_i's are zero, because $\vec{v} \neq 0$. Let $\vec{x}_{m+1} = \vec{v}$ and $a_{m+1} = -1$; then $\sum_{i=1}^{m+1} a_i \vec{x}_i = 0$ is a nontrivial dependence relation from Y. Thus Y is dependent. □

Lemma 4.16. *Let V be a vector space and let $X = \{\vec{v}_1, \dots, \vec{v}_n\}$ be a dependent set. Then there exists $k \in \{1, \dots, n\}$ such that $\vec{v}_k$ is a linear combination from $\{\vec{v}_1, \dots, \vec{v}_{k-1}\}$.*

Proof. Since X is dependent, there is a nontrivial dependence relation

$$\sum_{i=1}^{n} a_i \vec{v}_i = 0,$$

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

$$\vec{v}_k = \frac{1}{a_k} \sum_{i=1}^{k-1} a_i \vec{v}_i$$

is a linear combination of the preceding elements. □

Theorem 4.17. *Let V be a vector space in $\mathbb{R}^n$, and let $X, Y \subset V$ be finite subsets of V. If X is linearly independent and Y spans V, then*

$$|X| \leq |Y|.$$

Proof. Let $|Y| = n$ and $Y = \{\vec{y}_1, \ldots, \vec{y}_n\}$.

By way of contradiction,suppose that $|X| > n$ and let

$$Z = \{\vec{z}_1, \ldots, \vec{z}_{n+1}\} \subset X;$$

then Z is independent by Lemma 4.14. Label the elements of Y and Z so that all of those contained in $Y \cap Z$ are in the front, with $\vec{y}_i = \vec{z}_i$ for all $i \leq j$:

$$Y = \{\vec{z}_1, \ldots, \vec{z}_j, \vec{y}_{j+1}, \ldots, \vec{y}_n\}.$$

By Lemma 4.15, the set

$$\{\vec{z}_1, \ldots, \vec{z}_{j+1}, \vec{y}_{j+1}, \vec{y}_{k+2}, \ldots, \vec{y}_n\}$$

is dependent. By Lemma 4.16, one of these vectors is dependent on the preceding ones, and since the $\vec{z}_i's$ are linearly independent, there exists $k \in \{j+1, \ldots, n\}$ such that $\vec{y}_k$ is a linear combination of $\{\vec{z}_1, \ldots, \vec{z}_{j+1}, \vec{y}_{j+1}, \ldots, \vec{y}_{k-1}\}$. Thus if we remove $\vec{y}_k$ from the set, it will still span:

$$\text{span}\{\vec{z}_1, \ldots, \vec{z}_{i+1}, \vec{y}_{i+1}, \ldots, \vec{y}_{k-1}, \vec{y}_{k+1}, \ldots, \vec{y}_n\} = V.$$

Continuing in this way, adding the next z and removing a y, we see that after $n - j$ replacements we have

$$\text{span}\{\vec{z}_1, \ldots, \vec{z}_n\} = V.$$

Thus the set $Z = \{\vec{z}_1, \ldots, \vec{z}_n\} \cup \{\vec{z}_{n+1}\}$ is dependent by Lemma 4.15, producing a contradiction. □

Corollary 4.18. *Let V be a vector space in $\mathbb{R}^n$, and let $X, Y \subset V$ be finite bases of V. Then*

$$|X| = |Y|.$$

Proof. Since X and Y are basis, each spans and is independent. Since X is independent and Y spans, we have $|X| \leq |Y|$. But also, Y is independent and X spans, so $|Y| \leq |X|$. The result follows. □

If we know that V is spanned by a finite set, we can use Lemma 4.16 to throw out one superfluous vector at a time, until we arrive at a spanning set which is also independent. This is a basis for V. Thus, in this case, V has a basis.

Let V be a vector space in $\mathbb{R}^n$ The *dimension* of V is the cardinality of any basis for V, and is denoted $\dim(V)$.

Proposition 4.19. *Let $V, W \leq \mathbb{R}^n$. If $W \leq V$, then* $\dim(W) \leq \dim(V)$.

Proof. Let C be a basis for W and let B be a basis for V. Then $C \subset V$, and C is linearly independent, and B spans V. Thus $|C| \leq |B|$, so $\dim(W) \leq \dim(V)$. □

Let $V \leq \mathbb{R}^n$; we now show that any set of linearly independent vectors in V can be completed to a basis for V, and consequently, V has a basis unless V is trivial.

Lemma 4.20. *Let $X \subset \mathbb{R}^n$ be linearly independent, and let $\vec{v} \in \mathbb{R}^n \smallsetminus \operatorname{span}(X)$. Then $X \cup \{\vec{v}\}$ is linearly independent.*

Proof. Since $\mathbb{R}^n$ is spanned by the n standard basis vectors, Theorem 4.17 implies that $|X| \leq n$, say $|X| = k \leq n$; let $\vec{x}_1, \dots, \vec{x}_k$ be the distinct vectors in X, and set $\vec{x}_{k+1} = \vec{v}$. Then

$$X \cup \{\vec{v}\} = \{\vec{x}_1, \dots, \vec{x}_k, \vec{x}_{k+1}\}.$$

If this set is linearly dependent, one of these vectors is a linear combination of the preceding vectors, by Lemma 4.16; but $\vec{x}_{k+1}$ is not a linear combination of the preceding vectors, since $\vec{v} = \vec{x}_{k+1} \notin \operatorname{span} X$, and since X is independent, $\vec{x}_i$ is a not linear combination of the preceding vectors for $i = 1, \dots, k$. So $X \cup \{\vec{v}\}$ cannot be linearly dependent. $\square$

Proposition 4.21. *Let $V \leq \mathbb{R}^n$ and let $X \subset V$ be linearly independent. Then there exists a set $B \subset V$ with $X \subset B$ such that B is a basis for V.*

Proof. If $\operatorname{span}(X) = V$, then X is a basis for V containing X. Otherwise, there exists $\vec{v} \in V \smallsetminus \operatorname{span}(X)$. Set $Y = X \cup \vec{v}$; then Y is independent. If $\operatorname{span}(Y) = V$, we have a basis containing X; otherwise, continue in this way until a basis for V is linearly independent spanning set is obtained. This will happen in no more than $n - |X|$ steps, since any set of $n + 1$ or more vectors in $\mathbb{R}^n$ is dependent. $\square$

If V is a vector space, $X \subset V$ is linearly independent, and B is a basis for V containing X, we call B a *completion* of X to a basis for V.

Proposition 4.22. *Let $V \leq \mathbb{R}^n$. Then V has a basis.*

Proof. We consider the empty set to be linearly independent, and the span of the empty set to be the $\{\vec{0}\}$. Thus if $V = \{\vec{0}\}$, its basis is the empty set.

Otherwise, the empty set can be completed to a basis for V. $\square$

6. 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

(T1) $T(\vec{v} + \vec{w}) = T(\vec{v}) + T(\vec{w})$ for all $\vec{v}, \vec{w} \in \mathbb{R}^n$;
(T2) $T(a\vec{v}) = aT(\vec{v})$ for all $\vec{v} \in \mathbb{R}^n$ and $a \in \mathbb{R}$.

Example 4.23. Let $a, b \in \mathbb{R}$ be arbitrary constants. The function $T : \mathbb{R}^2 \to \mathbb{R}^1$ given by $T\langle x, y\rangle = ax + by$ is linear, where $T\langle x, y\rangle$ means $T(\langle x, y\rangle)$. To see this, let $\vec{v} = \langle x_1, y_1\rangle, \vec{w} = \langle x_2, y_2\rangle \in \mathbb{R}^2$. Then $\vec{v} + \vec{w} = \langle x_1 + x_2, y_1 + y_2\rangle$, so

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

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

$$T(c\vec{v}) = T\langle cx, cy\rangle = acx + acy = c(ax + by) = cT(\vec{v}).$$

Thus T is linear.

Example 4.24. The function $P_i : \mathbb{R}^n \to \mathbb{R}$ given by $T\langle x_1, \ldots, x_n\rangle = x_i$ is linear; this is called *projection* onto the i^{th} coordinate.

Example 4.25. Fix an arbitrary vector $\vec{w} \in \mathbb{R}^n$. Then the function $T : \mathbb{R}^n \to \mathbb{R}$ given by $T(\vec{v}) = \vec{v} \cdot \vec{v}$ is linear.

Proposition 4.26. *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then*

(a) $T(\vec{0}_n) = \vec{0}_m$*;*
(b) $T(\operatorname{span}(X)) = \operatorname{span}(T(X))$*, where* $X \subset \mathbb{R}^n$*.*

Proof. To prove **(a)**, we use transformation property **(T2)** to see that

$$T(\vec{0}_n) = T(0 \cdot \vec{0}_n) = 0 \cdot T(\vec{0}_n) = \vec{0}_m,$$

because 0 scalar multiplied by any vector in $\mathbb{R}^m$ is $\vec{0}_m$.

To prove **(b)**, let $X \subset \mathbb{R}^n$; for simplicity assume that $X = \{\vec{v}_1, \ldots, \vec{v}_r\}$ is a finite set. Then

$$\begin{aligned} T(\operatorname{span}(X)) &= T\left(\left\{ \sum_{i=1}^{r} a_i\vec{v}_i \,\middle|\, a_i \in \mathbb{R}\right\}\right) && \text{by definition of span} \\ &= \left\{ T\left(\sum_{i=1}^{r} a_i\vec{v}_i\right) \,\middle|\, a_i \in \mathbb{R}\right\} && \text{by definition of image} \\ &= \left\{ \sum_{i=1}^{r} a_iT(\vec{v}_i) \,\middle|\, a_i \in \mathbb{R}\right\} && \text{since } T \text{ is linear} \\ &= \operatorname{span}(\{T(\vec{v}_1), \ldots, T(\vec{v}_r)\}) && \text{by definition of span} \\ &= \operatorname{span}(T(X)) && \text{by definition of image} \end{aligned}$$

$\square$

Proposition 4.27. *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 $\vec{v} \in \mathbb{R}^n$, then $\vec{v} = \langle a_1, \ldots, a_n \rangle$ for some real numbers $a_i \in \mathbb{R}$. This is the same as saying that $\vec{v} = \sum_{i=1}^n a_i \vec{e}_i$; but since T is linear, we have

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

Thus if we know $T(\vec{e}_i)$ for all i, then we completely understand T. □

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

Proposition 4.28. *Let $\vec{w}_1, \ldots, \vec{w}_n \in \mathbb{R}^m$. Then there exists a unique linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ such that $T(\vec{e}_i) = \vec{w}_i$ for $i = 1, \ldots, n$.*

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

So, in order to define a linear transformation, we simply need to indicate where the basis vectors are sent, and "extend linearly", which means defining the transformation T by the above formula.

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

Example 4.29. Define a linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^3$ by $T(\vec{e}_1) = \langle 1, 2, 0 \rangle$, $T(\vec{e}_2) = \langle 0, 1, 2 \rangle$, and $T(\vec{e}_3) = \langle 2, 0, 1 \rangle$. Let $\vec{v} = \langle 1, 2, 3 \rangle$. What is $T(\vec{v})$?

Solution. Note that $\vec{v} = \vec{e}_1 + 2\vec{e}_2 + 3\vec{e}_3$. Thus

$$T(\vec{v}) = T(\vec{e}_1) + 2T(\vec{e}_2) + 3T(\vec{e}_3) = \langle 1, 2, 0 \rangle + \langle 0, 2, 4 \rangle + \langle 6, 0, 3 \rangle = \langle 1, 4, 7 \rangle.$$

□

7. Standard Linear Transformations

Linear transformations fix the origin, so assume that the origin is fixed in the following types of geometric transformations. The notation $\vec{v} \mapsto \vec{w}$ is read "$\vec{v}$ maps to $\vec{w}$", and it means that the transformation sends $\vec{v}$ to $\vec{w}$.

Example 4.30. (Rotations)
In $\mathbb{R}^2$, a rotation is determined by a single angle θ. It can be defined by

$$\vec{e}_1 \mapsto \langle \cos\theta, \sin\theta\rangle; \quad \vec{e}_2 \mapsto \langle -\sin\theta, \cos\theta\rangle.$$

The full transformation is constructed by extending linearly.

In $\mathbb{R}^3$, a rotation is determined by a line through the origin, known as the *axis of rotation*, and an angle of rotation about this line. For example, to rotate by $120°$ about the line (t, t, t). Then the rotation is defined by

$$\vec{e}_1 \mapsto \vec{e}_2; \quad \vec{e}_2 \mapsto \vec{e}_3; \quad \vec{e}_3 \mapsto \vec{e}_1.$$

Example 4.31. (Reflections)
In $\mathbb{R}^2$, a reflection is determined by a line through the origin; the line is fixed by the reflection. For example, to reflect across the line $y = x$, define a transformation by

$$\vec{e}_1 \mapsto \vec{e}_2; \quad \vec{e}_2 \mapsto \vec{e}_1.$$

In $\mathbb{R}^3$, a reflection is determined by a plane through the origin; the plane is fixed by the reflection. For example, reflection through the plane $x = y$ is given by

$$\vec{e}_1 \mapsto \vec{e}_2; \quad \vec{e}_2 \mapsto \vec{e}_1; \quad \vec{e}_3 \mapsto \vec{e}_3.$$

Example 4.32. (Projections)
Let V be a subspace of $\mathbb{R}^n$. For each $\vec{v} \in \mathbb{R}^n$, there is a unique point on V which is closest to the tip of $\vec{v}$. Projection onto V is the transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ which maps $\vec{v}$ to the vector represented by this point.

For example, suppose V is a line through the origin with direction vector $\vec{w}$. Then the projection of $\vec{v}$ onto V is given by the vector projection of $\vec{v}$ onto $\vec{w}$. Thus we can define

$$T : \mathbb{R}^n \to \mathbb{R}^n \quad \text{given by } T(\vec{v}) = \frac{\vec{v}\cdot\vec{w}}{\vec{w}\cdot\vec{w}}\vec{w}.$$

It is more difficult to compute projections onto higher dimensional subspaces.

Example 4.33. (Dilations)
The transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ given by $T(\vec{v}) = a\vec{v}$ dilates $\mathbb{R}^n$ by a factor of a.

It is also possible to stretch $\mathbb{R}^n$ by different amounts in different directions. For example, $T : \mathbb{R}^2 \to \mathbb{R}^2$ given by $T\langle x, y\rangle = \langle 2x, 3y\rangle$ stretches by a factor of 2 horizontally and a factor of 3 vertically. This is not a dilation.

Example 4.34. (Sheering)
We can tilt $\mathbb{R}^n$ by fixing some standard basis directions and adding a multiple of the fixed basis vectors to other basis vectors; this is known as sheering.

For example, consider the transformation of $\mathbb{R}^2$ given by

$$\vec{e}_1 \to \vec{e}_1; \quad \vec{e}_2 \to \vec{e}_2 + 2\vec{e}_1.$$

This fixes the x-axis, but tilts the y-axis; the y-axis is stretched as it tilts, so that the height of a point above the x-axis does not change.

Example 4.35. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be the unique linear transformation define by $\vec{e}_1 \mapsto \vec{e}_2$, $\vec{e}_2 \mapsto \vec{e}_3$, and $\vec{e}_3 \mapsto \vec{e}_1$. We see that the line $\langle t, t, t\rangle$ and the plane $x + y + z = 0$ are fixed by this transformation, and in fact, T is a rotation by 120° about this fixed line.

Example 4.36. Define $T : \mathbb{R}^2 \to \mathbb{R}^2$ by $T\langle x, y\rangle = \langle x - y, x + y\rangle$. Discuss the geometric effect of T on $\mathbb{R}^2$.

Solution. Note that $T(\vec{e}_1) = T\langle 1, 0\rangle = \langle 1, 1\rangle$, and $T(\vec{e}_2) = T(0, 1) = \langle -1, 1\rangle$. 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 $\vec{v} = \langle x, y\rangle \in \mathbb{R}^2$. Then

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

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

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

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

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

Example 4.37. Define $T : \mathbb{R}^2 \to \mathbb{R}^2$ by $T\langle x, y\rangle = \langle x + y, x - y\rangle$. 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 $T\langle 1, 0\rangle = \langle 1, 1\rangle$ and $T\langle 0, 1\rangle = \langle -1, 1\rangle$; $\vec{e}_1$ is rotated by $\frac{\pi}{4}$, but $\vec{e}_2$ 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 $\vec{e}_1$ and $T(\vec{e}_1)$, as well as the angle between $\vec{e}_2$ and $T(\vec{e}_2)$. The candidate is the line with angle $\frac{\pi}{8}$ radians. Let's verify this computationally.

Now if T is a dilating reflection, then $T(\vec{v}) = \sqrt{2}\vec{v}$, where $\vec{v}$ is the direction vector of the line of reflection. For $\vec{v} = \langle x, y\rangle$, 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 $y = (\sqrt{2}-1)x$. □

8. 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 4.38. *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 $\vec{0} \in V$ and $T(\vec{0}) = \vec{0}$, we see that $\vec{0} \in T(V)$.

(S1) Let $\vec{w}_1, \vec{w}_2 \in T(V)$. Then there exist vectors $\vec{v}_1, \vec{v}_2 \in V$ such that $\vec{w}_1 = T(\vec{v}_1)$ and $\vec{w}_2 = T(\vec{v}_2)$. We have $\vec{w}_1 + \vec{w}_2 = T(\vec{v}_1) + T(\vec{v}_2) = T(\vec{v}_1 + \vec{v}_2)$. Since V is a subspace, $\vec{v}_1 + \vec{v}_2 \in V$; thus $\vec{w}_1 + \vec{w}_2 \in T(V)$.

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

Example 4.39. Let V be the subspace of $\mathbb{R}^2$ spanned by the vector $\vec{v} = \langle 1, 1\rangle$; that is, $V = \{\langle t, t\rangle \mid t \in \mathbb{R}\}$ is a line through the origin of slope 1. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be given by $T\langle x, y\rangle = \langle x - y, x + y\rangle$; this is linear. Then $T(V)$ is the subspace of $\mathbb{R}^2$ spanned by $T(\vec{v}) = \langle 1 - 1, 1 + 1\rangle = \langle 0, 2\rangle$; 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:

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

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

Proposition 4.40. *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 $\vec{0} \in W$ and $T(\vec{0}) = \vec{0}$, we see that $\vec{0} \in T^{-1}(W)$.

(S1) Let $\vec{v}_1, \vec{v}_2 \in T^{-1}(W)$; then $T(\vec{v}_1)$ and $T(\vec{v}_2)$ are elements of W. Now $T(\vec{v}_1 + \vec{v}_2) = T(\vec{v}_1) + T(\vec{v}_2)$, and since W is a subspace, this sum is also in W. Thus $\vec{v}_1 + \vec{v}_2 \in T^{-1}(W)$.

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

Example 4.41. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be the linear transformation given by $T\langle x, y\rangle = \langle x - y, y - z, z - x\rangle$. Let $W = \{\vec{0}\} \subset \mathbb{R}^m$ be the trivial subspace of $\mathbb{R}^m$; here, $\vec{0}$ means the origin $\langle 0, 0, 0\rangle$. The preimage is given by solving the equations

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

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

9. 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 $\ker(T)$:

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

Proposition 4.42. *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 $T(\vec{0}) = \vec{0}$; thus $\vec{0} \in \ker(T)$.

(S1) Let $\vec{v}_1, \vec{v}_2 \in \ker(T)$; this means that $T(\vec{v}_1) = T(\vec{v}_2) = \vec{0}$. Then $T(\vec{v}_1+\vec{v}_2) = T(\vec{v}_1) + T(\vec{v}_2) = \vec{0} + \vec{0} = \vec{0}$, so $\vec{v}_1 + \vec{v}_2 \in \ker(T)$.

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

Alternate Proof. Since $W = \{\vec{0}\}$ 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 4.40. □

Example 4.43. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be given by $T\langle x, y, z\rangle = \langle x, y, 0\rangle$. This is projection onto the xy-plane, and is linear. The kernel is the z-axis.

Proposition 4.44. *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then $\ker(T) = \{\vec{0}\}$ if and only if T is injective.*

Proof. We must show both sides of the implication. Recall that T is injective means that whenever $T(\vec{v}_1) = T(\vec{v}_2)$, we must have $\vec{v}_1 = \vec{v}_2$.

($\Rightarrow$) Suppose that $\ker(T) = \{\vec{0}\}$. Let $\vec{v}_1, \vec{v}_2 \in \mathbb{R}^n$ such that $T(\vec{v}_1) = T(\vec{v}_2)$; we wish to show that $\vec{v}_1 = \vec{v}_2$. Then $T(\vec{v}_1) - T(\vec{v}_2) = 0$, so $T(\vec{v}_1 - \vec{v}_2) = 0$, so $\vec{v}_1 - \vec{v}_2 \in \ker(T)$. Since $\ker(T) = \{\vec{0}\}$, we have $\vec{v}_1 - \vec{v}_2 = \vec{0}$, so $\vec{v}_1 = \vec{v}_2$. Therefore T is injective.

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

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

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

Proposition 4.45. *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Let $\vec{w} \in \mathbb{R}^m$ be in the image of T and let $\vec{v} \in \mathbb{R}^n$ such that $T(\vec{v}) = \vec{w}$. Then*

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

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

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

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

10. 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 with domains and identical ranges. We define the sum of these linear transformations to be the function $S + T$ given by adding pointwise:

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

Proposition 4.46. *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 **(T1)** and **(T2)**.

(T1) Let $\vec{v}_1, \vec{v}_2 \in \mathbb{R}^n$. Then

$$\begin{aligned}(S + T)(\vec{v}_1 + \vec{v}_2) &= S(\vec{v}_1 + \vec{v}_2) + T(\vec{v}_1 + \vec{v}_2)\\ &= S(\vec{v}_1) + S(\vec{v}_2) + T(\vec{v}_1) + T(\vec{v}_2)\\ &= S(\vec{v}_1) + T(\vec{v}_1) + S(\vec{v}_2) + T(\vec{v}_2)\\ &= (S + T)(\vec{v}_1) + (S + T)(\vec{v}_2).\end{aligned}$$

(T2) Let $\vec{v} \in \mathbb{R}^p$ and $a \in \mathbb{R}$. Then

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

□

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

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

Proposition 4.47. *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation and let $b \in \mathbb{R}$. Then $bT : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation.*

Proof. We verify properties **(T1)** and **(T2)**.

(T1) Let $\vec{v}_1, \vec{v}_2 \in \mathbb{R}^n$. Then

$$bT(\vec{v}_1 + \vec{v}_2) = b(T(\vec{v}_1) + T(\vec{v}_2)) = bT(\vec{v}_1) + aT(\vec{v}_2).$$

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

$$bT(a\vec{v}) = baT(\vec{v}) = abT(\vec{v}) = a(bT(\vec{v})).$$

□

Example 4.48. Let $S\langle x, y\rangle = \langle x - y, x + y\rangle$ and $T\langle x, y\rangle = \langle x + y, x - y\rangle$. Then $(S + T)\langle x, y\rangle = \langle 2x, 2x\rangle$. This has the effect of plummeting $\langle x, y\rangle$ vertically onto the line $y = 0$ and then stretching by a factor of 2.

11. 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

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

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

Proposition 4.49. *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 **(T1)** and **(T2)**.

(T1) Let $\vec{v}_1, \vec{v}_2 \in \mathbb{R}^p$. Then

$$T(S(\vec{v}_1 + \vec{v}_2)) = T(S(\vec{v}_1) + S(\vec{v}_2)) = T(S(\vec{v}_1)) + T(S(\vec{v}_2)).$$

(T2) Let $\vec{v} \in \mathbb{R}^p$ and $a \in \mathbb{R}$. Then

$$T(S(a\vec{v})) = T(aS(\vec{v})) = aT(S(\vec{v})).$$

□

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, $J(\vec{v}) = \vec{v}$ for all $\vec{v} \in \mathbb{R}^n$. 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 4.50. *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 4.51. *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 **(T1)** and **(T2)**.

(T1) Let $\vec{w}_1, \vec{w}_2 \in \mathbb{R}^m$. Since T is surjective, there exist $\vec{v}_1, \vec{v}_2 \in \mathbb{R}^n$ such that $T(\vec{v}_1) = \vec{w}_1$ and $T(\vec{v}_2) = \vec{w}_2$. Then $S(\vec{w}_1) = \vec{v}_1$ and $S(\vec{w}_2) = \vec{v}_2$. Now

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

(T2) Let $\vec{w} \in \mathbb{R}^m$ and $a \in \mathbb{R}$. There exists $\vec{v} \in \mathbb{R}^n$ such that $T(\vec{v}) = \vec{w}$. Then $S(\vec{w}) = \vec{v}$. Now

$$S(a\vec{w}) = S(aT(\vec{v})) = S(T(a\vec{v})) = a\vec{v} = aS(\vec{w}).$$

$\square$

Example 4.52. Let $S\langle x, y\rangle = \langle x - y, x + y\rangle$ and $T\langle x, y\rangle = \langle x + y, x - y\rangle$. Then $(T \circ S)\langle x, y\rangle = \langle 2x, -2y\rangle$. This has the effect of reflecting the plane across the x-axis and stretching by a factor of 2.

12. Exercises

Exercise 4.1. Let $\vec{v}_1 = \langle 3, -1\rangle$ and $\vec{v}_2 = \langle -2, 5\rangle$. Show that $\operatorname{span}\{\vec{v}_1, \vec{v}_2\} = \mathbb{R}^2$ by writing and arbitrary vector $\langle x, y\rangle \in \mathbb{R}^2$ as a linear combination of $\vec{v}_1$ and $\vec{v}_2$.

Exercise 4.2. Let $X = \{\vec{v}_1, \dots, \vec{v}_r\} \subset \mathbb{R}^n$. Let $\vec{w}_1, \vec{w}_2 \in \operatorname{span}(X)$.

(a) Let $t \in \mathbb{R}$. Show that $t(\vec{w}_1 - \vec{w}_2) + \vec{w}_1 \in \operatorname{span}(X)$.
(b) Conclude that the line through $\vec{w}_1$ and $\vec{w}_2$ is contained in $\operatorname{span}(X)$.

Exercise 4.3. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear transformation of the plane. Suppose that $T(\vec{e}_2) = aT(\vec{e}_1)$ for some $a \in \mathbb{R}$. Show that the image of T is a line through the origin in $\mathbb{R}^2$.

Exercise 4.4. Let $\vec{v}_1 = \langle 1, -2, 1\rangle$, $\vec{v}_2 = \langle -1, -2, 2\rangle$, and $\vec{v}_3 = \langle 5, -2, -1\rangle$, and let $X = \{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$.

(a) Show that X is linearly dependent set of vectors.
(b) Find a parametric equation for $\operatorname{span}X$.
(c) Find a general equation for $\operatorname{span}X$.

Exercise 4.5. Let $\vec{w} \in \mathbb{R}^n$ and set

$$V = \{\vec{v} \in \mathbb{R}^n \mid \vec{v} \cdot \vec{w} = 0\}.$$

Show that V is a vector space.

Exercise 4.6. Let V and W be vector spaces in $\mathbb{R}^n$. Show that $V \cap W$ is a vector space in $\mathbb{R}^n$.

Exercise 4.7. Let $X = \{\vec{v}_1, \dots, \vec{v}_r\}$, and let $V = \operatorname{span}X$. Show that X is a basis for V if and only if every vector in V can be written as a linear combination from X in a unique way; that is, $\sum_{i=1}^r a_i\vec{v}_i = \sum_{i=1}^r b_i\vec{v}_i$ implies $a_i = b_i$ for all i.

Exercise 4.8. Let $\vec{w} \in \mathbb{R}^n$ and define a function

$$T : \mathbb{R}^n \to \mathbb{R} \quad \text{by} \quad T(\vec{v}) = \vec{v} \cdot \vec{w}.$$

Show that T is a linear transformation.

Exercise 4.9. Let $\vec{w} \in \mathbb{R}^n$ and define a function

$$T : \mathbb{R}^n \to \mathbb{R}^n \quad \text{by} \quad T(\vec{v}) = \frac{\vec{v} \cdot \vec{w}}{\vec{w} \cdot \vec{w}}\vec{w}.$$

Show that T is a linear transformation, and interpret it geometrically.

Exercise 4.10. Let $\vec{w} \in \mathbb{R}^3$ and define a function

$$T : \mathbb{R}^3 \to \mathbb{R}^3 \quad \text{by} \quad T(\vec{v}) = \vec{v} \times \vec{w}.$$

Show that T is a linear transformation, and interpret it geometrically.

CHAPTER 5

Matrix Algebra

1. Matrices

1.1. Motivation. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. We use the notation $\vec{e}_j$ to denote the j^{th} standard basis vector in the domain $\mathbb{R}^n$ of T, and the notation $\vec{e}_i$ to denote the i^{th} standard basis vector in the range $\mathbb{R}^m$. Any ambiguity here can be resolved from the context.

We have seen that T is completely determined by its effect on the standard basis for $\mathbb{R}^n$, so that if we know $T(\vec{e}_j)$ for $j = 1, \ldots, n$, we understand T. Now $T(\vec{e}_j)$ is a vector in the range $\mathbb{R}^m$, and as such, is a linear combination of the standard basis vectors in $\mathbb{R}^m$. We let $a_{ij} \in \mathbb{R}$ denote the i^{th} component of $T(\vec{e}_j)$, so that

$$T(\vec{e}_j) = \sum_{i=1}^{m} a_{ij}\vec{e}_i.$$

Let us see more explicitly how the numbers a_{ij}, together with their manner of indexing, determine T.

Let $\vec{b} \in \mathbb{R}^n$ be an arbitrary vector, and call $T(\vec{b})$ the *destination* of $\vec{b}$ under T. To understand the transformation T, we wish to find a formula for $T(\vec{b})$. Since $\vec{b}$ is in the domain of T, it is a linear combination of the standard basis vectors in the domain $\mathbb{R}^n$. Thus $\vec{b} = \sum_{j=1}^{n} b_j\vec{e}_j$ for some real numbers $b_j \in \mathbb{R}$. Moreover, $T(\vec{b})$ is a linear combination of the standard basis vectors for $\mathbb{R}^m$, so $T(\vec{b}) = \sum_{i=1}^{m} c_i\vec{e}_i$ for some $c_i \in \mathbb{R}$. We wish to find a formula for the c_i's in terms of the b_j's and the $a_{ij}'s$. We compute

$$T(\vec{b}) = \sum_{j=1}^{n} b_j T(\vec{e}_j) = \sum_{j=1}^{n} b_j \left(\sum_{i=1}^{m} a_{ij}\vec{e}_i \right) = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_{ij}b_j \right)\vec{e}_i.$$

The final expression in the above equation reveals that the components of $T(\vec{b})$ are given by

$$c_i = \sum_{i=1}^{n} a_{ij}b_j = \langle a_{i1}, \ldots, a_{in} \rangle \cdot \langle b_1, \ldots b_n \rangle;$$

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

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.

1.2. 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^{th} row and the j^{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^{th} slot. You may ask, "why would we ever need to repeat the ij"? The reason is, the number in the ij^{th} slot is not always indexed by ij. For example, if A is a 2×3 matrix written as $A = [2]_{ij}$ and B is a 3×2 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^* whose rows are the columns of A and whose columns are the rows of A:

$$A^* = [a_{ji}]_{ij}.$$

Note that $(A^*)^* = A$. An $m \times n$ matrix is called *square* if $m = n$. A matrix A is *symmetric* if $A^* = A$; 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 $\vec{v}$ is a column vector, then $\vec{v}^*$ is a row vector. From now on, whenever we need to consider a vector from $\mathbb{R}^n$ as a matrix, we consider it to be a column vector.

Let A be an $m \times n$ matrix. Denote the i^{th} row of A by $A_{(i)}$ and the j^{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. For convenience when we have not already labeled the entries of a matrix, let A_{ij} denote the entry in the i^{th} row and j^{th} column

Let $\vec{v}_1, \dots, \vec{v}_n \in \mathbb{R}^m$. We consider these to be column vectors. Let

$$A = [\vec{v}_1 \mid \cdots \mid \vec{v}_n]$$

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

1.3. Matrix of a Linear Transformation. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Let i range from 1 to m and j range from 1 to n. For each standard basis vector $\vec{e}_j$ in the domain of T, its image $T(\vec{e}_j)$ under T is a linear combination of the standard basis vectors from $\mathbb{R}^m$ in a unique way. Thus there exist unique real numbers a_{ij} such that

$$T(\vec{e}_j) = \sum_{i=1}^{m} a_{ij}\vec{e}_i.$$

The *matrix of* T is the $m \times n$ matrix defined by

$$A_T = [a_{ij}]_{ij}.$$

Fixing j and letting i ranges over the rows of A_T, we see that the entries in the j^{th} column are the component of $T(\vec{e}_j)$; that is, the j^{th} column is the destination of $\vec{e}_j$ under the transformation T. Thus

$$A_T = [T(\vec{e}_1) \mid T(\vec{e}_2) \mid \cdots \mid T(\vec{e}_j)].$$

We emphasize, in words,

The columns of A_T are the destinations of the standard basis vectors .

Example 5.1. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation which stretches the x-axis by 2 and stretches the y-axis by 3. Find the matrix of T.

Solution. We find where T sends the standard basis vectors. We have $\vec{e}_1 \mapsto 2\vec{e}_1$ and $\vec{e}_2 \mapsto 3\vec{e}_2$. Thus the first column of A_T is $2\vec{e}_1$ and the second column is $3\vec{e}_2$:

$$A_T = [T(\vec{e}_1) \mid T(\vec{e})_2] = [2\vec{e}_1 \mid 3\vec{e}_2] = [\langle 2, 0\rangle \mid \langle 0, 3\rangle] = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}.$$

□

Example 5.2. Find the matrix of the linear transformation which rotates $\mathbb{R}^3$ by $90°$ around the y-axis.

Solution. Since the axis of rotation is fixed, we have $\vec{e}_2 \mapsto \vec{e}_2$. We assume the $90°$ means clockwise from the perspective of the positive y-axis. Thus $\vec{e}_1 \mapsto -\vec{e}_3$ and $\vec{e}_3 \mapsto \vec{e}_1$. This gives

$$A_T = [T(\vec{e}_1) \mid T(\vec{e})_2 \mid T(\vec{e}_3)] = [-\vec{e}_3 \mid \vec{e}_2 \mid \vec{e}_1] = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \end{bmatrix}.$$

□

1.4. 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.

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: $Z_{m\times n} = [0]_{ij}$.

Proposition 5.3 (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. □

The definitions of matrix addition and scalar multiplication have been designed so that they correspond to the analogous operations for linear transformations.

Proposition 5.4. *Let $S : \mathbb{R}^n \to \mathbb{R}^m$ and $T : \mathbb{R}^n \to \mathbb{R}^m$. Then*

$$A_{S+T} = A_S + A_T \quad \textit{and} \quad aA_T = A_{aT}.$$

Remark. These follows fairly immediately from linearity. □

Complicated linear transformations can be broken down into simpler, more easily understood transformations. Sometimes we use sums and scalar products, but more significantly, we can decompose a transformation into a composition of other transformations.

For example, we may rotate space around some axis, then stretch it in various directions, then reflect it across some plane, then again rotate it around another axis. We would like to compute the the destinations of the standard basis vectors of this composition. Thus we wish to define matrix multiplication so that the product of matrices of transformations produces the matrix of the composition.

1.5. Matrix Multiplication. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ and $S : \mathbb{R}^p \to \mathbb{R}^n$; we intend to compose these, first applying S, and then T. Let i range from 1 to m, j range from 1 to n, and k range from 1 to p. Let a_{ij} denote the i^{th} component of $T(\vec{e}_j)$, and b_{jk} denote the j^{th} component of $S(\vec{e}_k)$. Finally, let c_{ik} denote the i^{th} component of $(T \circ S)(\vec{e}_k)$. The matrix of T is $[a_{ij}]_{ij}$, the matrix of S is $[b_{jk}]_{jk}$, the matrix of $T \circ S$ is $[c_{ik}]_{ik}$, and

$$T(\vec{e}_j) = \sum_{i=1}^{m} a_{ij}\vec{e}_i; \quad S(\vec{e}_k) = \sum_{j=1}^{n} b_{jk}\vec{e}_j; \quad \text{and} \quad (T \circ S)(\vec{e}_k) = \sum_{i=1}^{k} c_{ik}\vec{e}_i.$$

We wish to find a formula the cd_{ik}'s in terms of the $a'_{ij}s$ and the b_{jk}'s. Compute

$$(T \circ S)(\vec{e}_k) = \sum_{j=1}^{n} b_{jk} T(\vec{e}_j) = \sum_{j=1}^{n} b_{jk} \sum_{i=1}^{m} a_{ij}\vec{e}_i = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_{ij} b_{jk} \right) \vec{e}_i.$$

Thus

$$c_{ik} = \sum_{j=1}^{n} a_{ij} b_{jk}.$$

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

$$AB = [c_{ik}]_{ik}, \qquad \text{where} \qquad c_{ik} = \sum_{j=1}^{n} a_{ij} b_{jk}.$$

From this definition, we have

$$c_{ik} = A_{(i)} B^{(k)}.$$

The i^{th} row of A may be viewed as a row vector, so its transpose is a column vector. Then c_{ik} is the dot product of $A^*_{(i)}$ and $B^{(k)}$. That is, the $(ik)^{\text{th}}$ entry of AB is computed by dotting the i^{th} row of A with the k^{th} column of B.

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

Example 5.5. Let

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \qquad \text{and} \qquad B = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix},$$

then

$$AB = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} 9 & 8 \\ 7 & 6 \\ 5 & 4 \end{bmatrix} = \begin{bmatrix} 1 \cdot 9 + 2 \cdot 7 + 3 \cdot 5 & 1 \cdot 8 + 2 \cdot 6 + 3 \cdot 4 \\ 4 \cdot 9 + 5 \cdot 7 + 6 \cdot 5 & 4 \cdot 8 + 5 \cdot 6 + 6 \cdot 4 \end{bmatrix} = \begin{bmatrix} 38 & 32 \\ 101 & 86 \end{bmatrix}.$$

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 δ_{ij} is the "Kronecker delta" defined by

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

Proposition 5.6 (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)** $A I_n = A$;
- **(d)** $(A + C)B = AB + CB$;
- **(e)** $A(B + D) = AB + AD$;
- **(f)** $c(AB) = A(cB)$;
- **(g)** $(AB)^* = B^* A^*$
- **(h)** $(AB)_{(i)} = A_{(i)} B$;
- **(i)** $(AB)^{(k)} = AB^{(k)}$;
- **(j)** $(AB)^{(k)}_{(i)} = A_{(i)} B^{(k)}$.

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.

Example 5.7. Let

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 9 & 8 \\ 7 & 6 \end{bmatrix}.$$

Then

$$AB = \begin{bmatrix} 9+14 & 8+12 \\ 27+28 & 24+24 \end{bmatrix} = \begin{bmatrix} 23 & 20 \\ 55 & 48 \end{bmatrix}, \text{ but } BA = \begin{bmatrix} 9+24 & 18+32 \\ 7+18 & 14+24 \end{bmatrix} = \begin{bmatrix} 33 & 50 \\ 25 & 38 \end{bmatrix}.$$

Let $\vec{x} = \langle x_1, \ldots, x_n \rangle$ be a vector in $\mathbb{R}^n$. We view $\vec{x}$ 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 $A\vec{x}$ is defined to be an $m \times 1$ matrix:

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

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

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

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

1.6. Matrices and Linear Transformations. To complete our geometric interpretation of the product of a matrix and a column vector, we first prove that this operation is linear.

Proposition 5.8. *Let A be an $m \times n$ matrix. Then $A\vec{e}_j = A^{(j)}$, where $\vec{e}_j$ is the j^{th} standard basis vector in $\mathbb{R}^n$.*

Proof. Since $\vec{e}_j = \langle 0, \ldots, 1, \ldots, 0 \rangle$, with 1 in the j^{th} slot, we have

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

□

Proposition 5.9. *Let A be an $m \times n$ matrix, and let $\vec{x}, \vec{y} \in \mathbb{R}^n$. Then*

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

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

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

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

$$\begin{aligned} A(a\vec{x}) &= a x_1 A^{(1)} + \cdots + a x_n A^{(n)} \\ &= a(x_1 A^{(1)} + \cdots + x_n A^{(n)}) \\ &= a(A\vec{x}). \end{aligned}$$

□

Proposition 5.10. *Let A be an $m \times n$ matrix. Define a function*

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

Then T is a linear transformation.

Proof. This is immediate from the previous proposition. □

Proposition 5.11. *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Define a matrix*

$$A_T = [T(\vec{e}_1) \mid \cdots \mid T(\vec{e}_n)].$$

Then

(a) $T(\vec{x}) = A_T\vec{x}$ *for all* $\vec{x} \in \mathbb{R}^n$*;*
(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 5.12. Find the matrix R_θ of the linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ which rotates the plane by an angle of θ radians.

Solution. We only need to discover what T does to the standard basis vectors. We see that $T(\vec{e}_1) = \langle \cos\theta, \sin\theta \rangle$ and $T(\vec{e}_2) = \langle -\sin\theta, \cos\theta \rangle$. 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 5.13. Find the matrix F_θ 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 θ.

Solution. We see that

$$T(\vec{e}_1) = \langle \cos 2\theta, \sin 2\theta \rangle$$

and that

$$T(\vec{e}_2) = -\left\langle \cos\left(2\theta + \frac{\pi}{2}\right), \sin\left(2\theta + \frac{\pi}{2}\right) \right\rangle = \langle \sin 2\theta, -\cos 2\theta \rangle.$$

Thus

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

□

1.7. Matrices and Compositions of Linear Transformations. We now investigate 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 $T \circ S(\vec{x}) = T(S(\vec{x}))$ is a linear transformation.

Proposition 5.14. *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ and $S : \mathbb{R}_p \to \mathbb{R}_n$ be linear transformations. Then*

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

Remark. This means that the matrix associated to a composition of transformations is the product of the associated matrices. We defined matrix multiplication in the way we did precisely for this to be true, and it was demonstrated in the derivation. ☐

Proposition 5.15. *Let A be an $m \times n$ matrix and let B be an $n \times p$ matrix. Then*

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

Remark. This means that the transformation associated to a product of matrices is the composition of the associated transformations. This is true because the matrix of $T_A \circ T_B$ is computed to be AB, and the matrix completely determines the transformation. ☐

Example 5.16. 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,

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

This is another example where matrix multiplication is not commutative.

Example 5.17. Show that the composition of rotations is a rotation whose angle is the sum of the original angles.

Solution. We compute with matrices:

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

☐

1.8. 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 $J_n(\vec{v}) = \vec{v}$. 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. Here, we let $J = J_n$ be the identity transformation on $\mathbb{R}^n$, and let $I = I_n$ be the identity $n \times n$ matrix.

If T is invertible, then $T \circ T^{-1} = T^{-1} \circ T = J$, so $A_T A_{T^{-1}} = A_{T^{-1}} A_T = A_J = I$.

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).

Proposition 5.18 (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. To prove **(a)**, suppose that B and C are inverses for A. Then $B = BI = B(AC) = (BA)C = IC = C$.

To prove **(b)**, note that $(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AIA^{-1} = AA^{-1} = I$. Similarly, $(B^{-1}A^{-1})(AB) = I$, so AB is invertible with inverse $B^{-1}A^{-1}$.

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

The proof of **(d)** is postponed for now. □

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

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 $AX = I$.

2. Gaussian Elimination

2.1. 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^{th} row by c and add to i^{th} row
$c\mathsf{R}_i$	Type **D**	Multiply i^{th} row by c
$\mathsf{R}_i \leftrightarrow \mathsf{R}_j$	Type **P**	Swap the i^{th} row and the j^{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 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 why, 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 $a_1 \vec{e}_1 + \cdots + a_{i-1} Q^{(i-1)} - \vec{e}_i$ 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 $E_1, \dots, E_r$ such that

$$X = E_r \cdots E_1 AX = E_r \cdots E_1 I = E_r \cdots E_1.$$

Example 5.19. Let
$$A = \begin{bmatrix} 1 & 2 & -1 \\ 2 & 3 & -2 \\ -2 & -4 & 1 \end{bmatrix}.$$
Write A as a product of elementary invertible matrices, and find A^{-1}.

Solution. We represented the equation $AX = I$ with an *augmented matrix* $[A \mid I]$, and apply row operations which correspond to multiplying both sides of this equation by the same elementary invertible matrix. When we attain $[I \mid B]$, we know that $B = A^{-1}$. Keeping track of the row operations used allows us to reconstruct A^{-1} as a product of elementary invertible matrices; then A is the product, in the reverse order, of the inverses of the matrices.

$$\begin{aligned}
\left[\begin{array}{ccc|ccc} 1 & 2 & -1 & 1 & 0 & 1 \\ 2 & 3 & -2 & 0 & 1 & 0 \\ -2 & -4 & 1 & 0 & 0 & 1 \end{array}\right] \xrightarrow[\mathsf{R_3+2R_1}]{\mathsf{R_2-2R_1}} & \left[\begin{array}{ccc|ccc} 1 & 2 & -1 & 1 & 0 & 1 \\ 0 & -1 & 0 & -2 & 1 & 0 \\ 0 & 0 & -1 & 2 & 0 & 1 \end{array}\right] \\
\xrightarrow[\mathsf{-R_3}]{\mathsf{-R_2}} & \left[\begin{array}{ccc|ccc} 1 & 2 & -1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 2 & -1 & 0 \\ 0 & 0 & 1 & -2 & 0 & -1 \end{array}\right] \\
\xrightarrow[\mathsf{R_1-2R_2}]{\mathsf{R_1+R_3}} & \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -5 & 2 & -1 \\ 0 & 1 & 0 & 2 & -1 & 0 \\ 0 & 0 & 1 & -2 & 0 & -1 \end{array}\right]
\end{aligned}$$

If E_1 through E_6 are the matrices corresponding to these row operations, then
$$A^{-1} = E_6E_5E_4E_3E_2E_1 = \begin{bmatrix} -5 & 2 & -1 \\ 2 & -1 & 0 \\ -2 & 0 & -1 \end{bmatrix},$$
so $A = E_1^{-1}E_2^{-1}E_3^{-1}E_4^{-1}E_5^{-1}E_6^{-1}$, which we write as
$$A = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$
□

If A and X are square and $AX = I$, then we can write A as a product of elementary invertible matrices. Therefore A is itself invertible, so if $AB = I$, we automatically have $BA = I$.

2.2. 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.

2.3. 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 $a = \langle a_1, \dots, a_n \rangle$, $x = \langle x_1, \dots, x_n \rangle$, and $q = \langle \frac{b_1}{a_1}, 0, \dots, 0 \rangle$. The above equation becomes $\vec{a} \cdot \vec{x} = \vec{a} \cdot \vec{q}$, or

$$(\vec{x} - \vec{q}) \cdot \vec{a} = 0.$$

We recognize this as the equation of a hyperplane in $\mathbb{R}^n$ through the tip of vector $\vec{q}$ with normal vector $\vec{a}$.

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 row reduction 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 $\vec{x} = \langle x_1, \dots, x_n \rangle$ and $\vec{b} = \langle b_1, \dots, b_m \rangle$, we see that the solution set of the matrix equation $A\vec{x} = \vec{b}$ is exactly the solution set of the system of equations. This matrix equation, stated in terms of linear transformations, is $T_A(\vec{x}) = \vec{b}$; solving means finding $T_A^{-1}(\vec{b})$, the preimage of the point $\vec{b}$ under the linear transformation T_A.

Our approach to the problem uses matrices; we seek column vectors $\vec{x}$ such that $A\vec{x} = \vec{b}$.

A *general solution* is the set of all such column vectors $\vec{x}$.

A *particular solution* is a specific such column vector $\vec{x}$.

The system is called *homogeneous* if $b_i = 0$ for $i = 1, \dots, m$. 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 $\vec{b}$ is in the image of T_A, say $T_A(\vec{v}) = \vec{b}$ for some $\vec{v} \in \mathbb{R}^n$, then $T^{-1}(\vec{b}) = \vec{v} + \ker(T_A)$. If T_A is injective, then $\ker(T_A)$ consists of a single point (the origin), so $T^{-1}(\vec{b}) = \{\vec{v}\}$. Otherwise, $\ker(T_A)$ is a nontrivial subspace, so $\vec{v} + \ker(T_A)$ at least one line, and possibly a plane or more. Thus there are three possibilities:

(1) there are no solutions ($\vec{b}$ 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

$$\vec{v}_0 + c_1\vec{v}_1 + \cdots + c_k\vec{v}_k,$$

where $\vec{v}_0, \dots, \vec{v}_k$ are vectors which span the kernel of T_A, $c_1, \dots, c_k$ are free scalars, $\vec{v}_0$ is a particular solution to $A\vec{x} = \vec{b}$, and $c_1\vec{v}_1 + \cdots + c_k\vec{v}_k$ is the general solution to the homogeneous equation $A\vec{x} = 0$ (the kernel of T_A).

Suppose that there exists an invertible matrix E such that the matrix EA has a particularly nice form. Then $A\vec{x} = \vec{b} \Rightarrow EA\vec{x} = E\vec{b}$; since E is invertible, we have $EA\vec{x} = E\vec{b} \Rightarrow E^{-1}EA\vec{x} = E^{-1}E\vec{b} \Rightarrow A\vec{x} = \vec{b}$. Thus the solution set of $A\vec{x} = \vec{b}$ is exactly the solution set of $EA\vec{x} = E\vec{b}$, so is suffices to find the solution set of $EA\vec{x} = E\vec{b}$.

The nice form we refer to here is known as reduced row echelon form.

2.4. Reduced Row Echelon Form. Row operations are used to put matrices into standard forms.

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 pivot in the column is zero, permute with the first available lower row so that the diagonal entry is nonzero (use P). If this is impossible, continue to the next column.
(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 $A\vec{x} = \vec{b}$, form the augmented matrix $[A \mid \vec{b}]$ and work on A and $\vec{b}$ 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 5.20. 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

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 2 \\ 3 \\ 0 \\ 4 \end{bmatrix} + x_2 \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} 0 \\ 0 \\ 0 \\ -5 \\ 1 \\ 0 \end{bmatrix}$$

Notice that this writes the solution space as the image of a parametric transformation; in this example, we let $P = (x_1, x_2, x_3, x_4, x_5)$ be a variable point, $P_0 = (1, 0, 2, 3, 0, 5)$, $\vec{v} = \langle -2, 1, 0, 0, 0, 0 \rangle$, and $\vec{w} = \langle 0, 0, 0, -5, 1, 0 \rangle$. With $r = x_2$ and $s = x_5$ (which are free to vary over all of $\mathbb{R}$), we have

$$P = P_0 + r\vec{v} + s\vec{w}.$$

2.5. Solution Method. Let $A = [a_{ij}]_{ij}$ be an $m \times n$ matrix, $\vec{x} = [x_1, \ldots, x_n]^*$ an n-dimensional variable column vector, and $\vec{b} = [b_1, \ldots, b_m]^*$ an m-dimensional constant column vector. Then the solution set of the matrix equation $A\vec{x} = \vec{b}$ 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 $\vec{c} = O\vec{c}$. Then the solution set of $A\vec{x} = \vec{b}$ is equal to the solution set of $OA\vec{x} = O\vec{b}$, i.e., $Q\vec{x} = \vec{c}$.

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 $\vec{d} = U\vec{b}$. Then the solution set of $A\vec{x} = \vec{b}$ is equal to the solution set of $UA\vec{x} = U\vec{b}$, i.e., $R\vec{x} = \vec{d}$. We describe how to read off the general solution from the matrix equation $R\vec{x} = \vec{d}$.

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

$$\vec{v}_0 + c_1\vec{v}_1 + \cdots + c_k\vec{v}_k,$$

where k is the number of free variables; we have $k = n - r$, where r is the number of nonzero rows.

The vector $\vec{v}_0$ is the particular solution obtained by setting the free variables equal to 0 and solving for the basic variables.

The vectors $\vec{v}_i$ are found by replacing $\vec{d}$ by the zero vector, setting the i^{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) Multiply each free column by -1.
(3) Insert a zero row at row i for every free variable x_i.
(4) Add $\vec{e}_i$ 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.

Example 5.21. Solve the system of linear equations

$$\begin{aligned} x + 2z - w &= 3 \\ y + 2z + 3w &= 5 \\ 2x + 3y + 11z &= 7 \end{aligned}$$

Solution. Let A be the matrix of coefficients and $\vec{b}$ the column vector of constants. We form the augmented matrix $[A \mid \vec{b}]$ and row reduce it. Then we use solution readoff.

$$\begin{bmatrix} 1 & 0 & 2 & -1 & | & 3 \\ 0 & 1 & 2 & 3 & | & 5 \\ 2 & 3 & 11 & 0 & | & 7 \end{bmatrix} \xrightarrow{\mathsf{R_3-2R_1}} \begin{bmatrix} 1 & 0 & 2 & -1 & | & 3 \\ 0 & 1 & 2 & 3 & | & 5 \\ 0 & 3 & 7 & 2 & | & 1 \end{bmatrix}$$
$$\xrightarrow{\mathsf{R_3-3R_2}} \begin{bmatrix} 1 & 0 & 2 & -1 & | & 3 \\ 0 & 1 & 2 & 3 & | & 5 \\ 0 & 0 & 1 & -7 & | & -14 \end{bmatrix}$$
$$\xrightarrow{\mathsf{R_2-2R_3}} \begin{bmatrix} 1 & 0 & 2 & -1 & | & 3 \\ 0 & 1 & 0 & 17 & | & 33 \\ 0 & 0 & 1 & -7 & | & -14 \end{bmatrix}$$
$$\xrightarrow{\mathsf{R_1-2R_3}} \begin{bmatrix} 1 & 0 & 0 & 13 & | & 31 \\ 0 & 1 & 0 & 17 & | & 33 \\ 0 & 0 & 1 & -7 & | & -14 \end{bmatrix}.$$

Columns 1,2, and 3 are basic, and column 4 is free, so the free variable is x_4. The adjusted readoff matrix is

$$\begin{bmatrix} 1 & 0 & 0 & -13 & | & 31 \\ 0 & 1 & 0 & -17 & | & 33 \\ 0 & 0 & 1 & -7 & | & -14 \\ 0 & 0 & 0 & 1 & | & 0 \end{bmatrix}.$$

The solution is

$$\begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} = \begin{bmatrix} 31 \\ 33 \\ -14 \\ 0 \end{bmatrix} + w \begin{bmatrix} -13 \\ -17 \\ -7 \\ 1 \end{bmatrix}.$$

The solution is a line, which we may write parametrically as

$$(x, y, z, w) = (31, 33, -14, 0) - t\langle 13, 17, 7, -1\rangle.$$

□

2.6. 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 $A\vec{x} = \vec{b}$, form the augmented matrix $[A \mid \vec{b}]$ and work A and $\vec{b}$ 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, $\vec{b}$ be the column vector of values, and $\vec{x}$ be the column vector of variables. The matrix equation $A\vec{x} = \vec{b}$ 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

$$\left[\begin{array}{ccc|c} 1 & 2 & 2 & -7 \\ 3 & 6 & 0 & 9 \\ -2 & -4 & -1 & -1 \end{array}\right].$$

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 $\mathbf{P} : \mathsf{R}_i \leftrightarrow \mathsf{R}_j$).
(3) Eliminate all entries below this one in order (use $\mathbf{E} : \mathsf{R}_i + c\mathsf{R}_j$).

In our example, the row operations $\mathsf{R}_2 - 3\mathsf{R}_1$ and $\mathsf{R}_3 + 2\mathsf{R}_1$ produce

$$\left[\begin{array}{ccc|c} 1 & 2 & 2 & -7 \\ 0 & 0 & -6 & 30 \\ 0 & 0 & 3 & -15 \end{array}\right].$$

This completes the requirements for column one. Now do column two via $\mathsf{R}_3 + \frac{1}{2}\mathsf{R}_2$ to obtain

$$\left[\begin{array}{ccc|c} 1 & 2 & 2 & -7 \\ 0 & 0 & -6 & 30 \\ 0 & 0 & 0 & 0 \end{array}\right];$$

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 $\mathbf{D} : c\mathsf{R}_i$).
(2) Starting from the right, working upward then leftward, make all entries above a pivot equal to zero (use $\mathbf{E} : \mathsf{R}_i + c\mathsf{R}_j$).

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 $\vec{e}_i$ 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

$$\begin{aligned} x_1 - 2x_2 &= 3 \\ x_3 &= -5 \\ 0 &= 0 \end{aligned}$$

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

$$x = \begin{bmatrix} 3 \\ 0 \\ -5 \end{bmatrix} + x_2 \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}.$$

3. Geometric Interpretations

3.1. 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.

Reconsider the system of linear equations

$$\begin{aligned} 3x_1 - 4x_2 &= 11; \\ x_1 + 2x_2 &= 7. \end{aligned}$$

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:

$$A = \begin{bmatrix} 3 & -4 \\ 1 & 2 \end{bmatrix}; \quad \vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}; \quad \text{and } \vec{b} = \begin{bmatrix} 11 \\ 7 \end{bmatrix}.$$

We see that solving the system of equations is equivalent to solving the matrix equation

$$A\vec{x} = \vec{b}.$$

But A corresponds to a linear transformation T_A; thus we seek the preimage of $\vec{b}$ under the linear transformation T_A.

How do these two geometric interpretations coincide?

3.2. Component Functions. Let $f : \mathbb{R}^n \to \mathbb{R}^m$. For each $i \in \{1, \dots, m\}$, define a function

$$f_i : \mathbb{R}^n \to \mathbb{R} \quad \text{by } f_i(\vec{v}) = \operatorname{proj}_{\vec{e}_i} f(\vec{v});$$

this is called the i^{th} *component function* of f.

Example 5.22. Let $f : \mathbb{R} \to \mathbb{R}^2$ be given by $f(t) = \langle \cos t, \sin t \rangle$. 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 $f_1, \dots, f_m : \mathbb{R}^n \to \mathbb{R}$, we construct a function $f : \mathbb{R}^n \to \mathbb{R}^m$ by defining $f(\vec{v}) = \langle f_1(\vec{v}), \dots, f_m(\vec{v}) \rangle$.

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 $\langle x_1, x_2 \rangle$ 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 $f(x) = \langle f_1(x), f_2(x) \rangle$; that is, $f(x_1, x_2) = \langle 3x_1 - 4x_2, x_1 + 2x_2 \rangle$. 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 $\vec{v}$ such that $f(v) = (11, 7)$, which is the same as saying that we wish to discover the set $f^{-1}\langle 11, 7 \rangle = f_1^{-1}(11) \cap f_2^{-1}(7)$.

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 $T\langle x_1, x_2 \rangle = \langle 3x_1 - 4x_2, x_1 + 2x_2 \rangle$. What is the effect of T on the standard basis, and what is the matrix associated to T?

We have $T(\vec{e}_1) = T\langle 1, 0 \rangle = \langle 3, 1 \rangle$ and $T(\vec{e}_2) = T\langle 0, 1 \rangle = \langle -4, 2 \rangle$. Thus the matrix which corresponds to T is

$$A = \begin{bmatrix} 3 & -4 \\ 1 & 2 \end{bmatrix},$$

and finding the preimage of $\langle 11, 7 \rangle$ under T is equivalent to solving the matrix equation

$$A\vec{x} = \vec{b}, \quad \text{where } \vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \text{ and } \vec{b} = \begin{bmatrix} 11 \\ 7 \end{bmatrix}.$$

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 $\vec{x}$ of variables, and an $m \times 1$ column vector $\vec{b}$ of values. Solving the system is equivalent to solving the matrix equation $A\vec{x} = \vec{b}$. 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 $\vec{b}$ under the linear transformation T.

We may also view this as follows. Recall that if $f : A \to B$ is any function, and $C, D \subset B$, 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

$$H_i = \{(y_1, \dots, y_m) \in \mathbb{R}^m \mid y_i = b_i\}.$$

Then

$$\{\vec{b}\} = \cap_{i=1}^m H_i.$$

Let L_i be the hyperplane in $\mathbb{R}^n$ (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

$$X = \cap_{i=1}^m T^{-1}(H_i) = T^{-1}(\cap_{i=1}^m H_i) = T^{-1}(\vec{b}).$$

3.3. Geometric Interpretation of the Solution Process. We have the matrix equation $A\vec{x} = \vec{b}$, 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 $\vec{b}$ is a linear combination of these destinations, in which case there is a solution to the equation.

When row reducing the augmented matrix $[A \mid \vec{b}]$, we are in theory multiplying both sides of the equation $A\vec{x} = \vec{b}$ 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, $\vec{b}$ 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 $\vec{b}$ is a linear combination of the destinations of the standard basis vectors.

4. Exercises

Exercise 5.1. 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 5.2. 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

$$\begin{aligned} 3x_1 - 2x_2 &= 3 \\ x_1 + 4x_2 &= 5 \end{aligned}$$

Exercise 5.3. 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 5.4. 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 $\vec{b}$ of values, and the column vector $\vec{x}$ of variables.
(b) Find the matrix A^{-1}.
(c) Use **(b)** to solve the matrix equation $A\vec{x} = \vec{b}$.

Exercise 5.5. 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, $\vec{b}$ the column vector of values, and $\vec{x}$ the column vector of variables. Use Gaussian elimination (forward elimination, backward elimination, and solution readoff) to find the general solution to this system.

Exercise 5.6. 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 $\vec{b} = \langle 2, -5, 1 \rangle \in \mathbb{R}^3$. Viewing this as a column vector, we have

$$\vec{b} = \begin{bmatrix} 2 \\ -5 \\ 1 \end{bmatrix}.$$

(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 $\vec{d} = U\vec{b}$ and use solution readoff of find the preimage under T of the vector $\langle 2, -5, 1 \rangle$.

(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 5.7. 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 $([x, y, z]^*$ into A_T).

(c) Find $\ker(T) = \{(x, y, z) \in \mathbb{R}^3 \mid T(x, y, z) = (0, 0)\}$.

Exercise 5.8. Each 3×3 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 5.9. 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 6

Matrix Geometry

Abstract. In this chapter, we use matrix techniques to find tests for linear independence and spanning, and to find bases for the kernel and the image of a linear transformation.

We define and study direct sums and perpendicular spaces, eventually studying to four fundamental spaces of a matrix to decompose a space into the direct sum of a subspace and its perpendicular space.

Finally, we cut loose from the ambient space by slightly generalizing our definition of linear transformation to include those whose domain and codomain are arbitrary vector spaces. We see how to describe linear transformations on arbitrary vector spaces using matrices, written with respect to the some basis for the domain and range.

1. Matrix Techniques

1.1. Basic and Free Columns. Let A be an $m \times n$ matrix. Let U be a matrix such that $R = UA$ is in reduced row echelon form.

The basic columns of R consist of distinct standard basis vectors in $\mathbb{R}^m$, and as such, they are linearly independent. On the other hand, it is clear that the free columns of R are linear combinations of the preceding basic columns.

We say that $A^{(j)}$ is a basic column if $R^{(j)}$ is a basic column.

We say that $A^{(j)}$ is a free column if $R^{(j)}$ is a free column.

We now claim that the basic columns of A are linearly independent, and that the free columns of A are linear combinations of the preceding basic columns. This follows from the next proposition.

Proposition 6.1. *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation, and let $X \subset \mathbb{R}^n$.*

- **(a)** *If X is linearly independent and T is injective, then $T(X)$ is linearly independent.*
- **(b)** *If $\vec{v}$ is a linear combination from X, then $T(\vec{v})$ is a linear combination from $T(X)$.*

Proof. Suppose that X is linearly independent and T is injective. Let

$$\sum_{i=1}^{k} a_i T(\vec{x}_i) = \vec{0}$$

be an arbitrary linear dependence relation from $T(X)$. Since T is linear, $T(\sum_{i=1}^k a_i \vec{x}_i) = \vec{0}$. Thus $\sum_{i=1}^k a_i \vec{x}_i \in \ker(T)$, and since T is injective, $\ker(T) = \{\vec{0}\}$. Therefore $\sum_{i=1}^k a_i \vec{x}_i = 0$, and since X is linearly independent, $a_i = 0$ for all i. Thus $T(X)$ is linearly independent.

Suppose that $\vec{v} = \sum_{i=1}^k a_i \vec{x}_i$ is an arbitrary linear combination from X. Then $T(\vec{v}) = \sum_{i=1}^k a_i T(\vec{x}_i)$ is a linear combination from $T(X)$. $\square$

To apply this to our situation where $UA = R$, we realize that U is invertible, so $R = U^{-1}A$. The columns of R are the column vectors $R^{(j)} = U^{-1}A^{(j)}$.

Let $X \subset \mathbb{R}^m$ denote the set of basic columns of R, and let $T : \mathbb{R}^m \to \mathbb{R}^m$ be the linear transformation corresponding to U^{-1}. Then the set of basic columns of A is $T(X)$, and since X is independent, so is $T(X)$. Moreover, the free columns of R are linear combinations from X, so the free columns of A are linear combinations from $T(X)$.

1.2. Rank and Nullity. Let A be an $m \times n$ matrix. The *rank* of A, denoted $\mathrm{rank}(A)$, is the number of basic columns. The *nullity* of A, denoted $\mathrm{null}(A)$, is the number of free columns. Clearly, since A has n columns, $n = \mathrm{rank}(A) + \mathrm{null}(A)$.

Proposition 6.2. (Rank Plus Nullity Theorem)
Let A be an $m \times n$ matrix, and let $T : \mathbb{R}^n \to \mathbb{R}^m$ be the corresponding linear transformation. Then

- **(a)** *the basic columns of A form a basis for the image of T;*
- **(b)** *the adjusted free columns of A form a basis for the kernel of T;*
- **(c)** $\dim(\mathrm{img}(T)) = \mathrm{rank}(A)$*;*
- **(d)** $\dim(\ker(T)) = \mathrm{null}(A)$*;*
- **(e)** $\dim(\mathbb{R}^n) = \dim(\mathrm{img}(T)) + \dim(\ker(T))$.

Proof. Since $\mathrm{span}(T(X)) = T(\mathrm{span}(X))$, and the columns of A are the destination of the standard basis vector under T, the image of T is the span of the columns of A. Every free column of A is a linear combination of the basic columns, so the basic columns span the image. Also, the basic columns are linearly independent, so they form a basis for $\mathrm{img}(T)$. We have seen that the adjusted free columns are span the kernel, and they are linearly independent, so they form a basis for the kernel. The rest follows. $\square$

1.3. Tests for Spanning and Linear Independence.

1.3.1. *Test for Linear Independence.* Let $X = \{\vec{w}_1, \ldots, \vec{w}_n\} \subset \mathbb{R}^m$.

If $n > m$, then X is dependent.

Form the $m \times n$ matrix $A = [\vec{w}_1 \mid \cdots \mid \vec{w}_n]$.

Reduce A; if $n = r = \mathrm{rank}(A)$, then X is independent, otherwise it is not.

This works because if $n = r$, every column is basic, so the set of columns is linearly independent.

1.3.2. *Test for Spanning.* Let $X = \{\vec{w}_1, \ldots, \vec{w}_n\} \subset \mathbb{R}^m$.

If $n < m$, then X does not span $\mathbb{R}^m$.

Form the $m \times n$ matrix $A = [\vec{w}_1 \mid \cdots \mid \vec{w}_n]$.

Reduce A; if $m = r = \mathrm{rank}(A)$, then X spans $\mathbb{R}^m$; otherwise it does not.

This works because if $m = r$, then there are m linearly independent vectors in the columns of A, and any set of m linearly independent vectors necessarily spans $\mathbb{R}^m$.

1.3.3. *Test for a Basis.* Let $X = \{\vec{w}_1, \ldots, \vec{w}_n\} \subset \mathbb{R}^m$.

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.

1.4. Finding a Basis for the Kernel and the Image. 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 6.3. Let $\vec{e}_1, \dots, \vec{e}_4$ be the standard basis vectors for $\mathbb{R}^4$. Let

$$\vec{v}_1 = \langle 2, -4, 4\rangle, \vec{v}_2 = \langle 1, -1, 3\rangle, \vec{v}_3 = \langle 3, -7, 5\rangle, \vec{v}_4 = \langle 0, 2, 5\rangle \in \mathbb{R}^3.$$

Let $T : \mathbb{R}^4 \to \mathbb{R}^3$ be the unique linear transformation given by $T(\vec{e}_i) = \vec{v}_i$. 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 $\{\langle -2, 1, 1, 0\rangle\}$ is a basis for $\ker(T)$, and $\{\langle 2, -4, 4\rangle, \langle 1, -1, 3\rangle, \langle 0, 2, 5\rangle\}$ is a basis for $\mathrm{img}(T)$, the image of T. □

Let $Y = \{\vec{v}_1, \dots, \vec{v}_n\} \in \mathbb{R}^m$. 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 $A = [\vec{v}_1 \mid \cdots \mid \vec{v}_n]$. 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 the span of X 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 $n \leq m$).

1.5. Retake on Linear Independence. Let $X = \{\vec{v}_1, \dots, \vec{v}_n\}$ be a subset of $\mathbb{R}^m$. Form the $m \times n$ matrix A by putting the vectors in columns:

$$A = [\vec{v}_1 \mid \cdots \mid \vec{v}_n].$$

If $\vec{v} = [x_1, \dots, x_n]^*$ is a column vector in $\mathbb{R}^n$, we have seen that

$$\begin{aligned} A\vec{x} &= x_1 A^{(1)} + \cdots + x_n A^{(n)} \\ &= x_1 \vec{v}_1 + \cdots + x_n \vec{v}_n; \end{aligned}$$

that is, $A\vec{x}$ is a linear combination of the columns of A. Now $A\vec{x} = 0$ has a solution other than $\vec{x} = (0, \dots, 0)$ if and only if there is a nontrivial dependence relation among the $\vec{v}_i$'s.

A nontrivial solution of $A\vec{x} = \vec{0}$ is an element of the kernel of the linear transformation corresponding to A. Thus, the vector are independent if any only if the kernel is trivial, which occurs when A has no free columns.

Example 6.4. Let $V = \mathbb{R}^3$ and let

$$\vec{v}_1 = \langle 1, 2, -3 \rangle, \vec{v}_2 = \langle 2, 0, 1 \rangle, \vec{v}_3 = \langle 4, -4, 9 \rangle \in \mathbb{R}^3.$$

Show that the set $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ 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. □

2. Direct and Perpendicular Sums

2.1. Sums and Intersections. 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 6.5. *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 $\vec{0} \in W_1$ and $\vec{0} \in W_2$, we see that $\vec{0} = \vec{0} + \vec{0} \in W_1 + W_2$.

(S1) Let $\vec{w}_1, \vec{w}_1' \in W_1$ and $\vec{w}_2, \vec{w}_2' \in W_2$ so that $\vec{w}_1 + \vec{w}_2$ and $\vec{w}_1' + \vec{w}_2'$ are arbitrary members of $W_1 + W_2$. Then $(\vec{w}_1 + \vec{w}_2) + (\vec{w}_1' + \vec{w}_2') = (\vec{w}_1 + \vec{w}_1') + (\vec{w}_2 + \vec{w}_2') \in W_1 + W_2$, by property **(S1)** of W_1 and W_2.

(S2) Let $\vec{w}_1 \in W_1$ and $\vec{w}_2 \in W_2$ so that $\vec{w}_1 + \vec{w}_2$ is an arbitrary member of $W_1 + W_2$ Let $a \in \mathbb{R}$. Then $a(\vec{w}_1 + \vec{w}_2) = a\vec{w}_1 + a\vec{w}_2 \in W_1 + W_2$, by property **(S2)** of W_1 and W_2. $\square$

It follows that any finite sum of subspaces is a subspace.

Proposition 6.6. *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 $\vec{v}_1, \vec{v}_2 \in W_1 \cap W_2$. Then $\vec{v}_1, \vec{v}_2 \in W_1$ and $\vec{v}_1, \vec{v}_2 \in W_2$, so $\vec{v}_1 + \vec{v}_2 \in W_1$ and $\vec{v}_1 + \vec{v}_2 \in W_2$, because both of these sets are subspaces. Thus $\vec{v}_1 + \vec{v}_2 \in W_1 \cap W_2$.

(S2) Let $\vec{v} \in W_1 \cap W_2$ and let $a \in \mathbb{R}$. Then $\vec{v} \in W_1$ and $\vec{v} \in W_2$, and since these are subspaces, we see that $a\vec{v} \in W_1$ and $a\vec{v} \in W_2$. Thus $a\vec{v} \in W_1 \cap W_2$.

Therefore $W_1 \cap W_2 \leq V$. $\square$

This argument generalizes so that the intersection of *any number* (even infinitely many) of subspaces is a subspace.

Recall from the theory of sets that if A and B are finite sets, then

$$|A \cup B| = |A| + |B| - |A \cap B|;$$

this is because the elements of the intersection are counted twice in $|A| + |B|$.

The next result can be viewed as an expression of this idea, only where the sets of bases for vector spaces.

Proposition 6.7. (Subspace Dimension Formula)
Let V be a finite dimensional vector space and let $W_1, W_2 \leq V$. Then

$$\dim(W_1 + W_2) = \dim(W_1) + \dim(W_2) - \dim(W_1 \cap W_2).$$

Proof. Let $\dim(W_1) = p$, $\dim(W_2) = q$, and $\dim(W_1 + W_2) = n$.

Let $X = \{\vec{x}_1, \ldots, \vec{x}_n\}$ be a basis for $W_1 \cap W_2$. We complete this to a basis $Y = \{\vec{x}_1, \ldots, \vec{x}_n, \vec{y}_1, \ldots, \vec{y}_{p-n}\}$ for W_1 and $Z = \{\vec{x}_1, \ldots, \vec{x}_n, \vec{z}_1, \ldots, \vec{z}_{q-n}\}$ for W_2. We see that $B = \{\vec{x}_1, \ldots, \vec{x}_n, \vec{y}_1, \ldots, \vec{y}_p, \vec{z}_1, \ldots, \vec{z}_q\}$ spans $W_1 + W_2$. 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

$$\sum_{i=1}^{n} a_i \vec{x}_i + \sum_{j=1}^{p-n} b_j \vec{y}_j + \sum_{k=1}^{p-n} c_k \vec{z}_k = 0.$$

Then

$$\sum_{j=1}^{p-n} b_j \vec{y}_j = -\sum_{i=1}^{n} a_i \vec{x}_j - \sum_{k=1}^{q-n} c_k \vec{z}_k.$$

The sum on the left is in W_1 and the sum on the right is in W_2, so the sum on the left is actually in $W_1 \cap W_2$. Thus we have $d_1, \ldots, d_n$ such that

$$\sum_{j=1}^{p-n} b_j \vec{y}_j = \sum_{i=1}^{n} d_i \vec{x}_i.$$

This is a dependence relation among the members of Y, which is a linearly independent set; thus

$$b_1, \ldots, b_p = 0.$$

Similarly the c_k's are all zero, whence the a_i's are all zero. Thus B is linearly independent. Therefore

$$\begin{aligned}
\dim(W_1 + W_2) &= |B| \\
&= n + (p - n) + (q - n) \\
&= p + q - n \\
&= \dim(W_1) + \dim(W_2) - \dim(W_1 \cap W_2),
\end{aligned}$$

because B is a basis for $W_1 + W_2$. □

Corollary 6.8. *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 6.9. 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 $U + W$.

2.2. Direct and Perpendicular Sums. Let V be a vector space and let $W_1, W_2 \leq V$. We say that V is a *direct sum* of W_1 and W_2, and write $V = W_1 \oplus W_2$, if

(D1) $V = W_1 + W_2$;
(D2) $W_1 \cap W_2 = \{\vec{0}\}$.

Example 6.10. Let $V = \mathbb{R}^3$, $W_1 = \text{span}\{\vec{e}_1, \vec{e}_2\}$ be the xy-plane, and $W_2 = \text{span}\{\vec{e}_3\}$. Then $V = W_1 \oplus W_2$.

Now let $W_3 = \text{span}\{\langle 1, 1, 1\rangle\}$. Then it is still the case that $V = W_1 \oplus W_3$, except in this case, the vectors of W_3 are not normal to those of W_1.

Let $U \leq \mathbb{R}^n$. The *perpendicular space* of U, or simply the *perp space* of U, is

$$U^\perp = \{\vec{v} \in \mathbb{R}^n \mid \vec{u} \cdot \vec{v} = 0 \text{ for all } \vec{u} \in U\}.$$

Proposition 6.11. *Let $U \leq \mathbb{R}^n$. Then*

(a) $U^\perp \leq \mathbb{R}^n$*;*
(b) $U \cap U^\perp = \{\vec{0}\}$*;*
(c) $U \subset (U^\perp)^\perp$.

Proof. Exercise. □

Result 6.12. *Let $W \leq \mathbb{R}^n$ and let $\vec{v} \in \mathbb{R}^n$. Then there exist unique vectors $\vec{w} \in W$ and $\vec{x} \in W^\perp$ such that $\vec{v} = \vec{w} + \vec{x}$.*

Proof. This will follow from the Gram-Schmidt Orthonormalization Process, which we may study later. It can also be concluded through an application of the four fundamental subspaces, which we pursue next. □

Result 6.13. *Let $U \leq \mathbb{R}^n$. Then*

$$\mathbb{R}^n = U \oplus U^\perp.$$

Proof. This follows from the previous Result. □

Result 6.14. *Let $U \leq \mathbb{R}^n$. Then $(U^\perp)^\perp = U$.*

Proof. This is a corollary of the previous Result and the previous Proposition. □

3. The Four Fundamental Subspaces of a Matrix

3.1. Transpose Transformations. 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^{th} row is the j^{th} column of A.

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 $T^*(\vec{w}) = A^*\vec{w}$, where

$$A = [T(\vec{e}_1) \mid \cdots \mid T(\vec{e}_n)].$$

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 $T_{A^*}(\vec{w}) = A^*(\vec{w})$. 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.

3.2. Column Spaces and Row Spaces. 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 $\text{col}(A)$.

The *row space* of A is the subspace of $\mathbb{R}^n$ spanned by the rows of A, and is denoted by $\text{row}(A)$.

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 $\text{col}(A)$, $\text{row}(A)$, $\ker(A)$, and $\ker(A^*)$.

The next proposition reviews what we have learned about bases for the kernel and image of the linear transformation corresponding to a matrix, rephrased in terms of the stages of Gaussian elimination. Only part **(j)** is completely new, and it is an optional technique to find all four fundamental subspaces with one Gaussian elimination.

Proposition 6.15. *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)** $\mathrm{col}(A) = \mathrm{row}(A^*)$;
- **(b)** $\mathrm{col}(A) = \mathrm{img}(T_A)$;
- **(c)** $\mathrm{row}(A) = \mathrm{img}(T_A^*)$;
- **(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)** $\mathrm{row}(A) = \mathrm{row}(B) = \mathrm{row}(C)$;
- **(h)** $\dim(\mathrm{col}(A)) = \dim(\mathrm{row}(A))$;
- **(i)** *the nonzero rows of B (or of C) form a basis for* $\mathrm{row}(A)$;
- **(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) $\mathrm{col}(A) = \mathrm{row}(A^*)$

This follows from the definition of transpose.

(b) $\mathrm{col}(A) = \mathrm{img}(T_A)$

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) $\mathrm{row}(A) = \mathrm{img}(T_A^*)$

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: $\dim(\ker(A)) = \dim(\ker(T_A)) = n - \dim(\mathrm{img}(T_A))$; since $\dim(\mathrm{img}(T_A))$ is the number of basic columns, $n - \dim(\mathrm{img}(T_A))$ 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) $\text{row}(A) = \text{row}(B) = \text{row}(C)$

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 $\text{row}(ED) \subset \text{row}(D)$. But E^{-1} is also an elementary invertible matrix, so $\text{row}(D) = \text{row}(E^{-1}ED) \subset \text{row}(ED)$, which shows that $\text{row}(ED) = \text{row}(D)$ and E does not change the row space.

Since $B = OA$ and O is a product of elementary invertible matrices, we see that $\text{row}(B) = \text{row}(OA) = \text{row}(A)$. Similarly, $\text{row}(C) = \text{row}(A)$.

(h) $\dim(\text{col}(A)) = \dim(\text{row}(A))$

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)**, $\dim(\text{col}(A)) = \dim(\text{col}(B))$. The dimension of $\text{col}(B)$ 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 $\text{row}(B)$. Thus $\dim(\text{col}(A)) = \dim(\text{col}(B)) = \dim(\text{row}(B)) = \dim(\text{row}(A))$.

(i) *the nonzero rows of B (or of C) form a basis for* $\text{row}(A)$

The nonzero rows of B (respectively C) form a basis for $\text{row}(B)$ (respectively $\text{row}(C)$). By **(g)**, $\text{row}(A) = \text{row}(B)$, 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 $r = \dim(\text{col}(A)) = \dim(\text{row}(A^*)) = \dim(\text{col}(A^*))$. Thus $\dim(\ker(A^*)) = m - \dim(\text{col}(A^*)) = m - r$.

Since O is invertible, its rows are linearly independent. Indeed, T_O is an isomorphism, so $\ker(O) = \{0\}$; thus $\dim(\text{row}(O)) = \dim(\text{col}(O)) = \dim(\mathbb{R}^m) - \dim(\ker(O)) = m$, since $\dim(\ker(O)) = 0$. Then $\text{row}(O) = \mathbb{R}^m$, 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 $\vec{x}^*$ is one of the last k rows of O, then $\vec{x}$ is one of the last k columns of O^*, and $A^*\vec{x}$ is one of the last k columns of B^*; that is, it is zero. Thus $\vec{x}$ is in the kernel of A^*. $\square$

3.3. Perpendicular Decompositions. We now show how to find the perpendicular space of a subspace.

Proposition 6.16. *Let A be an $m \times n$ matrix. Then*

(a) $\mathrm{row}(A) = \ker(A)^{\perp}$ *and* $\mathbb{R}^n = \mathrm{row}(A) \oplus \ker(A)$*;*

(b) $\mathrm{col}(A) = \ker(A^*)^{\perp}$ *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)**.

By the Rank Plus Nullity Theorem, $\dim(\mathrm{col}(A)) + \dim(\ker(A)) = \dim(\mathbb{R}^n)$. Also, $\dim(\mathrm{row}(A)) = \dim(\mathrm{col}(A))$, so $\dim(\mathbb{R}^n) - \dim(\ker(A)) = \dim(\mathrm{row}(A))$.

The coordinates of $A\vec{x}$ are the dot products of the rows of A with the vector $\vec{x}$. If $\vec{x} \in \ker(A)$, the $A\vec{x} = \vec{0}$ (the zero vector). Thus each of the coordinates of $A\vec{x}$ 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. Thus $\mathrm{row}(A) \subset \ker(A)^{\perp}$. It follows that $\dim(\mathrm{row}(A)) \leq \dim(\ker(A)^{\perp}$.

Now $\ker(A) \cap \ker(A)^{\perp} = \{\vec{0}\}$, and $\ker(A) + \ker(A)^{\perp} \leq \mathbb{R}^n$, so $\dim(\ker(A)^{\perp} \leq \dim(\mathbb{R}^n) - \dim(\ker(A)) = \dim(\mathrm{row}(A))$.

Thus $\dim(\mathrm{row}(A)) = \dim(\ker(A)^{\perp}$, and since $\mathrm{row}(A) \subset \ker(A)^{\perp}$, they must be equal. That is, $\mathrm{row}(A) = \ker(A)^{\perp}$.

Since the row space of A is perpendicular to the kernel of A, we see that $\mathrm{row}(A) \cap \ker(A) = \{\vec{0}\}$, so $\dim(\mathrm{row}(A) \cap \ker(A)) = 0$. To summarize,

$$\begin{aligned}
\dim(\mathrm{row}(A) + \ker(A)) &= \dim(\mathrm{row}(A)) + \dim(\ker(A)) - \dim(\mathrm{row}(A) \cap \ker(A)) \\
&= \dim(\mathrm{row}(A)) + \dim(\ker(A)) + 0 \\
&= \dim(\mathrm{col}(A)) + \dim(\ker(A)) \\
&= \dim(\mathbb{R}^n).
\end{aligned}$$

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 6.17. *Let $U \leq \mathbb{R}^m$. Then $\mathbb{R}^m = U \oplus U^{\perp}$.*

Proof. Let $\{u_1, \ldots, u_r\}$ be a basis for U. Form the matrix

$$A = [u_1 \mid \cdots \mid u_r \mid 0 \mid \cdots \mid 0].$$

Then $\mathrm{col}(A) = U$, and $\mathrm{col}(A)^{\perp} = \ker(A^*)$ with $\mathbb{R}^m = \mathrm{col}(A) \oplus \ker(A^*)$. □

Example 6.18. 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 $U^{\perp}$.

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 $U^{\perp}$ is given by the last $m - r$ rows of O, where $r = \dim(U)$. □

4. Linear Transformations

4.1. Linear Transformations between Vector Spaces. Let V and W be vector spaces. A *linear transformation* between V and W is a function $T : V \to W$ satisfying

(T1) $T(\vec{v}_1 + \vec{v}_2) = T(\vec{v}_1) + T(\vec{v}_2)$, for all $\vec{v}_1, \vec{v}_2 \in V$;
(T2) $T(a\vec{v}) = aT(\vec{v})$, for all $\vec{v} \in V$ and $a \in \mathbb{R}$.

Thus, we now allow the domain to be any vector space in $\mathbb{R}^n$.

If $T : V \to W$ is a linear transformation, the *kernel* of T is

$$\ker(T) = \{\vec{v} \in V \mid T(\vec{v}) = \vec{0}\}.$$

The following facts remain true regarding linear transformations whose domain is any vector space. The proofs are virtually identical to those already given.

- The image of the zero vector is the zero vector.
- The image of a subspace in the codomain is a subspace of the domain.
- The preimage of a subspace in the domain is a subspace of the codomain.
- The composition of linear transformations is a linear transformation.
- The kernel is a subspace of the domain.
- The transformation is injective if and only if the kernel is trivial.

Proposition 6.19. *Let V and W be vector spaces. Let $X = \{\vec{v}_1, \ldots, \vec{v}_n\} \subset V$ be a basis for V. Let $Y = \{\vec{w}_1, \ldots, \vec{w}_n\} \subset W$. Then there exists a unique linear transformation $T : V \to W$ such that $T(\vec{v}_i) = \vec{w}_i$.*

Proof. For each $\vec{v} \in V$, there exist unique real numbers $a_1, \ldots, a_n$ such that $\vec{v} = \sum_{i=1}^n a_i\vec{v}_i$. Define $T(\vec{v}) = \sum_{i=1}^n a_i\vec{w}_i$. It is clear that $T(\vec{v}_i) = \vec{w}_i$, 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. □

Proposition 6.20. *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

$$a_1T(\vec{x}_1) + \cdots + a_nT(\vec{x}_n) = 0,$$

where $\vec{x}_i \in X$ and $a_i \in \mathbb{R}$, not all zero. Then $T(\sum_{i=1}^n a_i\vec{x}_i) = 0$, so $\sum_{i=1}^n a_i\vec{x}_i$ 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 $\vec{v} \in V$ such that $T(\vec{v}) = 0$. Since $\vec{v} \neq 0$, the set $\{\vec{v}\}$ is independent. But its image $T(\vec{v})$ is dependent. □

4.2. Images of Direct Sums. We examine the image of a linear transformation when the domain is expressed as a direct sum.

Proposition 6.21. *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.

$$\begin{aligned} \vec{w} \in T(U_1 + U_2) &\Leftrightarrow \vec{w} = T(\vec{u}_1 + \vec{u}_2) && \text{for some } \vec{u}_1 \in U_1, \vec{u}_2 \in U_2 \\ &\Leftrightarrow \vec{w} = T(\vec{u}_1) + T(\vec{u}_2) && \text{because } T \text{ is linear} \\ &\Leftrightarrow \vec{w} \in T(U_1) + T(U_2) && \text{by definition of image.} \end{aligned}$$

□

Proposition 6.22. *Let V be a vector space and let X be a basis for V.*
Let $Y_1 \subset X$ and let $Y_2 = X \setminus Y_1$. 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 6.23. *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 6.24. *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 $\dim(V) = \dim(U_1) + \dim(U_2)$.*

Proposition 6.25. *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 $Y_2 = X \setminus Y_1$. Let $U = \operatorname{span}(Y_2)$. Then by Proposition 6.22, $V = K \oplus U$. 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. □

4.3. Rank and Nullity. Let V be a finite dimensional vector space and let $T : V \to W$ be a linear transformation. Let $\mathrm{img}(T) = T(V)$ denote the image of T. The *rank* of T is the dimension of the image of T: rank $= \dim(\mathrm{img}(T))$. The *nullity* of T is the dimension of the kernel of T: nullity $= \dim(\ker(T))$.

Theorem 6.26. (Rank plus Nullity Theorem)
Let V be a finite dimensional vector space and let $T : V \to W$ be a linear transformation. Then $\dim(V) = \dim(\ker(T)) + \dim(\mathrm{img}(T))$.

Proof. Let $K = \ker(T)$. By Proposition 6.25 **(a)**, there exists a subspace $U \leq V$ such that $V = K \oplus U$. Thus $\dim(V) = \dim(K) + \dim(U)$. By Proposition 6.25 **(b)**, the linear transformation $T \restriction_U : U \to W$ is injective, so $\dim(T(U)) = \dim(U)$. Thus

$$\dim(V) = \dim(K) + \dim(U) = \dim(\ker(T)) + \dim(\mathrm{img}(T)).$$

□

Corollary 6.27. *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. □

Let $Y = \{\vec{v}_1, \ldots, \vec{v}_n\} \in \mathbb{R}^m$. 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 $A = [\vec{v}_1 \mid \cdots \mid \vec{v}_n]$. 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 $n \leq m$).

4.4. Isomorphisms. 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 6.28. *Let V be a vector space.*
Then $\mathrm{id}_V : V \to V$ is an isomorphism.

Proof. Clear. □

Proposition 6.29. *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 $\vec{w}_1, \vec{w}_2 \in W$. Since T is bijective, there exist unique elements $\vec{u}_1, \vec{u}_2 \in U$ such that $T(\vec{u}_1) = \vec{w}_1$ and $T(\vec{u}_2) = \vec{w}_2$. Now $T(\vec{u}_2 + \vec{u}_2) = T(\vec{u}_1) + T(\vec{u}_2) = \vec{w}_1 + \vec{w}_2$, so $T^{-1}(\vec{w}_1 + \vec{w}_2) = \vec{u}_1 + \vec{u}_2 = T^{-1}(\vec{w}_1) + T^{-1}(\vec{w}_2)$.

(T2) Let $\vec{w} \in W$ and $a \in \mathbb{R}$. There exists a unique element $\vec{u} \in U$ such that $T(\vec{u}) = \vec{w}$. Then $T(a\vec{u}) = aT(\vec{u}) = a\vec{w}$, so $T^{-1}(a\vec{w}) = a\vec{u} = aT^{-1}(\vec{w})$. □

Proposition 6.30. *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. □

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 6.31. *Let $T : V \to W$ be a linear transformation. Then T is an isomorphism if and only if for every basis X of V, $T(V)$ is a basis for W.*

Proof.

($\Rightarrow$) Suppose that T is an isomorphism, and let X be a basis for V. Then T is surjective, so $W = T(V) = T(\mathrm{span}(X)) = \mathrm{span}(T(X))$; that is, $T(X)$ spans W. Moreover, T is injective, so by Proposition 6.20, $T(X)$ is independent. Thus $T(X)$ is a basis for W.

($\Leftarrow$) Then either T is not surjective, or T is not injective. If T is not surjective and X is any basis for V, then $T(X)$ cannot span W, so $T(X)$ cannot be a basis for W. If T is not injective, then there exists a nontrivial kernel for T. We may take a basis for this kernel, and complete it to a basis for V. The image of this basis for V will contain the zero vector, and thus cannot be independent. □

In light of Proposition 6.19, we may construct an isomorphism between spaces by sending a basis to a basis.

Let V be a finite dimensional vector space of dimension n. An *ordered basis* for V is an ordered n-tuple $(\vec{x}_1, \ldots, \vec{x}_n) \in V^n$ of linearly independent vectors from V.

Note that if $(\vec{x}_1, \ldots, \vec{x}_n)$ is an ordered basis, then $X = \{\vec{x}_1, \ldots, \vec{x}_n\}$ 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 6.32. *Let V be a finite dimensional vector space of dimension n. Let $X = \{\vec{x}_1, \ldots, \vec{x}_n\}$ be an ordered basis for V. Define a linear transformation*

$$\Gamma_X : V \to \mathbb{R}^n \quad \textit{by} \quad \Gamma_X(\vec{x}_i) = \vec{e}_i.$$

Then Γ_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 $\vec{v} \in V$, then $\vec{v} = \sum_{i=1} a_i \vec{x}_i$ for some real numbers $a_1, \ldots, a_n$. Then

$$\Gamma_X(\vec{v}) = \sum_{i=1}^{n} a_i \Gamma_X(\vec{x}_i) = \sum_{i=1}^{n} \sum_{i=1}^{n} a_i \vec{e}_i = \langle a_1, \ldots, a_n \rangle;$$

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 6.19. It is an isomorphism by Proposition 6.31. □

Corollary 6.33. *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 $V \cong W$. □

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.

4.5. Computing Linear Transformations via Matrices. 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:

$$\begin{tikzcd} V \arrow[r, "T"] \arrow[d, "\Gamma_X"'] & W \arrow[d, "\Gamma_Y"] \\ \mathbb{R}^n \arrow[r, "S"'] & \mathbb{R}^m \end{tikzcd}$$

This says that to compute $T(\vec{v})$, it suffices to push $\vec{v}$ into $\mathbb{R}^n$ via $\vec{u} = \Gamma_X(\vec{v})$, compute $S(\vec{u})$, then pull this result back to W via Γ_Y.

But $S : \mathbb{R}^n \to \mathbb{R}^m$ corresponds to a matrix A, and we can compute $A\vec{u}$ by matrix multiplication. This also allows us to compute kernels, images, and so forth via matrices.

Example 6.34. Let $\vec{v}_1 = \langle 1,0,0,0\rangle, \vec{v}_2 = \langle 1,0,1,0\rangle, \vec{v}_3 = \langle 1,0,0,1\rangle \in \mathbb{R}^4$. Let V be the subspace of $\mathbb{R}^4$ spanned by $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$; these form a basis for V. Let $W = \mathbb{R}^2$ Let $\vec{w}_1 = \langle 1,2\rangle, \vec{w}_2 = \langle -1,0\rangle, \vec{w}_3 = \langle 3,2\rangle \in W$. Let $T : V \to W$ be the unique linear transformation given by $T(\vec{v}_i) = \vec{w}_i$. Find a basis for the kernel of T.

Solution. Let $\vec{e}_1, \vec{e}_2, \vec{e}_3$ be the standard basis vectors for $\mathbb{R}^3$. Let $S : V \to \mathbb{R}^3$ be given by $T(\vec{v}_i) = \vec{e}_i$. Then S is an isomorphism. Let $R : \mathbb{R}^3 \to \mathbb{R}^2$ be given by $T(\vec{e}_i) = \vec{w}_i$. 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 $\langle -7, -4, 1\rangle$.

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 $\langle -7, -4, 1\rangle$ back through S (find its preimage). This is $-7\langle 1,1,0,0\rangle - 4\langle 1,0,1,0\rangle + \langle 1,0,0,1\rangle = \langle -10, -7, -4, 1\rangle$. The kernel of T is the span of this vector. $\square$

5. Linear Operators

Let V be a vector space. A *linear operator* on V is a linear transformation $T : V \to V$.

If $S, T : V \to V$ are 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, so that TS is the transformation $T \circ S$.

This multiplication distributes over addition of operators:

$$T(S + R) = TS + TR; \quad (T + S)R = TR + SR.$$

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 $J = \mathrm{id}_V$.

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 $\vec{v} \in V$ is given by distributing $\vec{v}$ into the polynomial:

$$L(\vec{v}) = T^n(\vec{v}) + a_{n-1}T^{n-1}(\vec{v}) + \cdots + a_1 T(\vec{v}) + a_0\vec{v}.$$

If $J : V \to V$ is the identity transformation, we set $T^0 = J$. Thus we view a_0 as the linear transformation $a_0 J$, which stretches every vector by a factor of a_0.

In this way, we can form and factor polynomials such as

$$L = T^2 - 4T + 3 = (T - 3)(T - 1);$$

thus $L(\vec{v}) = T(T(\vec{v})) - 4T(\vec{v}) + 3\vec{v}$. Here, $(T - 3)(T - 1) = (T - 3J) \circ (T - J)$.

Let V be a vector space of dimension n, and let $X = \{\vec{x}_1, \ldots, \vec{x}_n\}$ be a basis for V. The linear transformation $\Gamma_X : V \to \mathbb{R}^n$ given by $\vec{x}_i \mapsto \vec{e}_i$ may be used to compute a linear transformation $T : V \to V$ by setting $S = \Gamma_X \circ T \circ \Gamma_X^{-1}$. In diagram form:

$$\begin{array}{ccc} V & \xrightarrow{T} & V \\ \Gamma_X \downarrow & & \downarrow \Gamma_X \\ \mathbb{R}^n & \xrightarrow[S]{} & \mathbb{R}^n \end{array}$$

Since $S : \mathbb{R}^n \to \mathbb{R}^n$, it has a matrix with respect to the standard basis.

6. Exercises

Exercise 6.1. Let $U \leq \mathbb{R}^n$. Show that

(a) $U^\perp \leq \mathbb{R}^n$;
(b) $U \cap U^\perp = \{0\}$;
(c) $U \subset U^{\perp\perp}$.

Exercise 6.2. 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 $T(\vec{v}) = A\vec{v}$.

(a) Find a basis for $\mathrm{img}(T)$ and for $\ker(T)$.
(b) Find a basis for $\mathrm{img}(T)^\perp$ and for $\ker(T)^\perp$.

Exercise 6.3. Let U be the subspace of $\mathbb{R}^4$ spanned by the vectors
$$\vec{v}_1 = \langle 1, 0, -1, 1\rangle,\ \vec{v}_2 = \langle 2, 1, 1, 0\rangle,\ \text{and } \vec{v}_3 = \langle 0, -1, -3, 2\rangle.$$

(a) Find a basis for U.
(b) Find a basis for $U^\perp$.
(c) Find a matrix A such that $U = \ker(A)$.

Exercise 6.4. 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.

Exercise 6.5. 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 6.6. 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 6.7. 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 6.8. 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 6.9. Let $T : V \to W$ be a linear transformation.
Show that T is invertible if and only if T is bijective.

Exercise 6.10. 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 6.11. 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 6.12. 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 6.13. 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 Γ 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 6.14. 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

Orthogonality

Abstract. This chapter discusses the relationship between isomorphisms, which are transformations preserving algebraic structures, and isometries, which are transformations which preserve distance. We define the concept of an orthonormal basis, discover its role in this discussion, and describe the Gram-Schmidt process for taking a basis for a subspace of $\mathbb{R}^n$ and producing a orthonormal basis.

1. Isometries

1.1. Distance. Our study of $\mathbb{R}^n$ has, to this point, focused primarily on its algebraic structure. We have used dot product, cross product, and perpendicularity to understand some of the geometry of these spaces. At this point, we wish to study the relationship between linear algebra and distance.

Any element of the set $\mathbb{R}^n$ may be viewed as a point, or as a vector; a point is the tip of a vector whose representative arrow starts at the origin, and an vector is the difference between a point and the origin. So, let us restate our definition of distance in the notation of vectors.

Recall that the *norm* of a vector $\vec{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ is

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

We define a *distance function* on $\mathbb{R}^n$ by

$$d : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R} \quad \text{given by} \quad d(\vec{x}, \vec{y}) = \|\vec{x} - \vec{y}\|;$$

that is, the distance from $\vec{x}$ to $\vec{y}$ is the norm of the vector between their tips. If $\vec{x} = (x_1, \ldots, x_n)$ and $\vec{y} = (y_1, \ldots, y_n)$, we have

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

The distance function satisfies these properties:

- $d(\vec{x}, \vec{y}) \geq 0$ and $d(\vec{x}, \vec{y}) = 0$ if and only if $\vec{x} = \vec{y}$;
- $d(\vec{x}, \vec{y}) = d(\vec{y}, \vec{x})$;
- $d(\vec{x}, \vec{y}) + d(\vec{y}, \vec{z}) \geq d(\vec{x}, \vec{z})$.

We wish to investigate functions which preserve distance, and their relationship to linear algebra.

1.2. Isometries.

Definition 7.1. Let $A \subset \mathbb{R}^n$ and $B \subset \mathbb{R}^m$. An *isometry* (or *congruence*) from A to B is a surjective function

$$f : A \to B \quad \text{such that, for all } a_1, a_2 \in A, \text{ we have} \quad d(a_1, a_2) = d(f(a_1), f(a_2)).$$

If there exists an isometry from A to B, we say that A and B are *isometric*, or *congruent*.

An *isometry of A* is an isometry from A to itself.

For example, two triangles lying in a plane are congruent if and only if there exists an isometry from one to the other. Any to circles of the same radius are isometric. Any two lines are isometric, as are any two planes.

We begin with a few preliminary results regarding isometries in general.

Proposition 7.2. *Let $f : A \to B$ be an isometry. Then f is bijective.*

Proof. As part of the definition, we know f is surjective. Let $a_1, a_2 \in A$ such that $f(a_1) = f(a_2)$ Then $d(a_1, a_2) = d(f(a_1), f(a_2)) = 0$, so $a_1 = a_2$. □

Proposition 7.3. *Let $f : A \to B$ be an isometry. Then $f^{-1} : B \to A$ is an isometry.*

Proposition 7.4. *Let $f : A \to B$ and $g : B \to C$ be isometries. Then $g \circ f : A \to B$ is an isometry.*

We will focus on isometries of the form $f : \mathbb{R}^n \to \mathbb{R}^n$, or more generally, $f : V \to W$, where V and W are vector spaces. In this context, an isometry is a rigid motion of a line, plane, space, or higher dimensional space. Let us list some the isometries of low dimensional spaces. To understand these, we need the following proposition.

Isometries of $\mathbb{R}$ include the following.

- Identity
- Reflection through a point
- Translation by a number

For example, $f(x) = x - 5$ shifts the line left by 5, and $g(x) = -(x - 5) + 5$ reflects across 5.

Isometries of $\mathbb{R}^2$ include the following.

- Identity
- Rotation by an angle around a point
- Reflection across a fixed line
- Translation by a vector

Isometries of $\mathbb{R}^3$ include the following.

- Identity
- Rotation by an angle around an oriented line
- Reflection through a plane
- Translation by a vector

We postpone a precise and general definition of rotation and reflection, but we can at least define translation.

Definition 7.5. Let V be a vector space. A *translation* of V is a function

$$\tau : V \to V \quad \text{defined by} \quad \tau(\vec{x}) = \vec{x} + \vec{v},$$

where $\vec{v} \in V$. This function is known as *translation by* $\vec{v}$, and may be denoted by $\tau_{\vec{v}}$.

Proposition 7.6. *Let $\tau : V \to V$ be a translation. Then τ is an isometry.*

Proof. Suppose that the translation is given by $\tau(\vec{x}) = \vec{x} + \vec{v}$, where $\vec{v} \in V$. We note that since V is a vector space, $\tau(\vec{x}) \in V$.

Let $\vec{x}_1, \vec{x}_2 \in X$. Then

$$d(\vec{x}_1, \vec{x}_2) = \|\vec{x}_1 - \vec{x}_2\| = \|(\vec{x}_1 + \vec{v}) - (\vec{x}_2 + \vec{v})\| = \|\tau(\vec{x}_1) - \tau(\vec{x}_2)\| = d(\tau(\vec{x}_1), \tau(\vec{x}_2)).$$

Thus, τ is an isometry. □

Definition 7.7. Let V and W be vector spaces, and let $f : V \to W$. We say that f *fixes the origin* if $f(\vec{0}_V) = \vec{0}_W$.

We have previously made extensive use of the fact that every linear transformation fixes the origin.

Proposition 7.8. *Let V and W be vector spaces, and let $f : V \to W$ be an isometry that fixes the origin. Then, for all $\vec{v} \in V$, we have*

$$\|\vec{v}\| = \|f(\vec{v})\|.$$

Proof. The norm is the distance from the tip of $\vec{v}$ to the origin:

$$\|\vec{v}\| = d(\vec{v}, \vec{0}_V) = d(f(\vec{v}), f(\vec{0}_V)) = d(f(\vec{v}), \vec{0}_W) = \|f(\vec{v})\|.$$

□

Proposition 7.9. *Let V and W be vector spaces and let $f : V \to W$ be an isometry. Then there exists a unique translation $\tau : W \to W$ such that $\tau \circ f$ is an isometry which fixes the origin.*

Proof. Let $\vec{w} = -f(\vec{0})$ and set $\tau_{\vec{w}}(\vec{x}) = \vec{x} + \vec{w}$. Then $\tau_{\vec{w}} \circ f$ is the composition of isometries, and thus is an isometry. Also,

$$\tau_{\vec{w}} \circ f(\vec{0}) = \tau_{\vec{w}}(f(\vec{0})) = f(\vec{0}) + \vec{w} = f(\vec{0}) - f(\vec{0}) = \vec{0},$$

so $\tau_{\vec{w}} \circ f$ fixes the origin.

For uniqueness, suppose that $\vec{w} \in W$ and $\tau_{\vec{w}}$ is translation by $\vec{w}$ such that $\tau_{\vec{w}} \circ f$ fixes the origin. Then $\tau_{\vec{w}} \circ f(\vec{0}) = f(\vec{0}) + \vec{w} = \vec{0}$, which implies that $\vec{w} = -f(\vec{0})$. □

We will show that isometries not only preserve distance, but also preserve angles. Moreover, we will show that isometries are affine functions, and we will characterize the types of matrices which represent isometries.

1.3. Dilations. A concept related to isometry is that of dilation, the definition of which we give now.

Definition 7.10. Let $A \subset \mathbb{R}^n$ and $B \subset \mathbb{R}^m$. A *dilation* (or *similarity*) from A to B is a surjective function

$$f : A \to B \quad \text{such that, for all } a_1, a_2 \in A, \text{ we have} \quad \lambda d(a_1, a_2) = d(f(a_1), f(a_2)),$$

where $\lambda \in \mathbb{R}$ is a fixed positive real number, call the *dilation factor.* If there exists a dilation from A to B, we say that A and B are *similar.*

Proposition 7.11. *Let $f : A \to B$ be a dilation. Then f is bijective.*

Proof. As part of the definition, we know f is surjective. Let λ be the dilation factor of f. Let $a_1, a_2 \in A$ such that $f(a_1) = f(a_2)$ Then $\lambda d(a_1, a_2) = d(f(a_1), f(a_2)) = 0$. Since $\lambda > 0$, $d(a_1, a_2) = 0$, so $a_1 = a_2$. $\square$

Proposition 7.12. *Let $f : A \to B$ be an dilation. Then $f^{-1} : B \to A$ is an dilation.*

Proposition 7.13. *Let $f : A \to B$ and $g : B \to C$ be dilations. Then $g \circ f : A \to B$ is an dilation.*

Proposition 7.14. *Let V and W be vector spaces, and let $f : V \to W$ be a dilation with factor $\lambda \neq 1$. Then f has a unique fixed point.*

Idea of proof. The full proof requires the concepts of Cauchy sequences and complete metric spaces, however, the basic idea can be summarized.

First assume that $\lambda < 1$. Pick $x_1 \in \mathbb{R}^n$, let $x_2 = f(x_1)$, $x_3 = f(x_2)$, and so forth, to construct a sequence (x_n) whose terms get closer together. This sequence eventually converges.

Now, if $\lambda > 1$, note that f^{-1} is a dilation with factor $\dfrac{1}{\lambda} < 1$, which has a fixed point. The fixed point of f^{-1} is necessarily a fixed point of f.

If y_1 and y_2 are fixed points of f, then $\lambda d(y_1, y_2) = d(f(y_1), f(y_2)) = d(y_1, y_2)$; thus $\lambda = 1$, violating our premise. $\square$

It turns out that a dilation is completely determined by a fixed point and a dilation factor.

2. Orthogonal Transformations

Definition 7.15. Let V and W be vector spaces, and let $f : V \to W$. We say that f is *orthogonal* if f preserves dot products; that is, if

$$\vec{v}_1, \vec{v}_2 \in V \quad \Rightarrow \quad \vec{v}_1 \cdot \vec{v}_2 = f(\vec{v}_1) \cdot f(\vec{v}_2).$$

Proposition 7.16. *Let V and W be vector spaces, and let $f : V \to W$. Then f is an isometry which fixes the origin if and only if f is orthogonal.*

Proof.

($\Rightarrow$) Suppose that f is an isometry which fixes the origin. Recall that f preserves the norm of each vector in V, since

$$\|\vec{v}\| = d(\vec{v}, \vec{0}) = d(f(\vec{v}), f(\vec{0})) = d(f(\vec{v}), \vec{0}) = \|f(\vec{v})\|.$$

Let $\vec{v}_1, \vec{v}_2 \in V$. If two distances are the same, then so are their squares; thus

$$\|\vec{v}_1 - \vec{v}_2\|^2 = d(\vec{v}_1, \vec{v}_2)^2 = d(f(\vec{v}_1), f(\vec{v}_2))^2 = \|f(\vec{v}_1) - f(\vec{v}_2)\|^2$$

which, using that $\|\vec{x}\|^2 = \vec{x} \cdot \vec{x}$, implies that

$$(\vec{v}_1 - \vec{v}_2) \cdot (\vec{v}_1 - \vec{v}_2) = (f(\vec{v}_1) - f(\vec{v}_2)) \cdot (f(\vec{v}_1) - f(\vec{v}_2)).$$

Square both sides and expanding gives

$$\vec{v}_1 \cdot \vec{v}_1 - 2\vec{v}_1 \cdot \vec{v}_2 + \vec{v}_2 \cdot \vec{v}_2 = f(\vec{v}_1) \cdot f(\vec{v}_1) - 2f(\vec{v}_1) \cdot f(\vec{v}_2) + f(\vec{v}_2) \cdot f(\vec{v}_2).$$

Again, using that $\vec{x} \cdot \vec{x} = \|\vec{x}\|^2$, we get

$$\|\vec{v}_1\|^2 - 2\vec{v}_1 \cdot \vec{v}_2 + \|\vec{v}_2\|^2 = \|f(\vec{v}_1)\|^2 - 2f(\vec{v}_1) \cdot f(\vec{v}_2) + \|f(\vec{v}_2)\|^2.$$

But $\|\vec{x}\| = \|f(\vec{x})\|$, so

$$\|\vec{v}_1\|^2 - 2\vec{v}_1 \cdot \vec{v}_2 + \|\vec{v}_2\|^2 = \|\vec{v}_1\|^2 - 2f(\vec{v}_1) \cdot f(\vec{v}_2) + \|\vec{v}_2\|^2.$$

Cancelling and dividing by -2 produces the result,

$$\vec{v}_1 \cdot \vec{v}_2 = f(\vec{v}_1) \cdot f(\vec{v}_2).$$

($\Leftarrow$) Suppose that f is orthogonal. Then

$$\|f(\vec{0})\|^2 = (f(\vec{0}) \cdot f(\vec{0})) = \vec{0} \cdot \vec{0} = 0,$$

so $f(\vec{0}) = \vec{0}$, and f fixes the origin.

Now let $\vec{v}_1, \vec{v}_2 \in V$. We have

$$\begin{aligned} d(f(\vec{v}_1), f(\vec{v}_2))^2 &= \|f(\vec{v}_1) - f(\vec{v}_2)\|^2 \\ &= f(\vec{v}_1) \cdot f(\vec{v}_1) - 2f(\vec{v}_1) \cdot f(\vec{v}_2) + f(\vec{v}_2) \cdot f(\vec{v}_2) \\ &= \vec{v}_1 \cdot \vec{v}_1 - 2\vec{v}_1 \cdot \vec{v}_2 + \vec{v}_2 \cdot \vec{v}_2 \\ &= (\vec{v}_1 - \vec{v}_2) \cdot (\vec{v}_1 \cdot \vec{v}_2) \\ &= \|\vec{v}_1 - \vec{v}_2\|^2 \\ &= d(\vec{v}_1, \vec{v}_2)^2. \end{aligned}$$

Thus f is an isometry. $\square$

Proposition 7.17. *Let V and W be vector spaces, and let $f : V \to W$ be an isometry which fixes the origin. Then f is a linear transformation.*

Proof.

(T1) Let $\vec{v}_1, \vec{v}_2 \in V$. Let $\vec{x} = \vec{v}_1 + \vec{v}_2$. We wish to show that $f(\vec{x}) = f(\vec{v}_1) + f(\vec{v}_2)$; it suffices to show that the length L of $f(\vec{v}_1) + f(\vec{v}_2) - f(\vec{x})$ is zero. This length is

$$L = (f(\vec{v}_1) + f(\vec{v}_2) - f(\vec{x})) \cdot (f(\vec{v}_1) + f(\vec{v}_2) - f(\vec{x})).$$

By Proposition 7.16, we know that f is orthogonal, so expanding this dot product gives

$$\begin{aligned}
L &= f(\vec{v}_1) \cdot f(\vec{v}_1) + f(\vec{v}_2) \cdot f(\vec{v}_2) + f(\vec{x}) \cdot f(\vec{x}) \\
&\quad + 2f(\vec{v}_1) \cdot f(\vec{v}_2) - 2f(\vec{v}_1) \cdot f(\vec{x}) - 2f(\vec{v}_2) \cdot f(\vec{x}) \\
&= \vec{v}_1 \cdot \vec{v}_1 + \vec{v}_2 \cdot \vec{v}_2 + \vec{x}) \cdot \vec{x} \\
&\quad + 2\vec{v}_1 \cdot \vec{v}_2 - 2\vec{v}_1 \cdot \vec{x} - 2\vec{v}_2 \cdot \vec{x} \\
&= (\vec{v}_1 + \vec{v}_2 - \vec{x}) \cdot (\vec{v}_1 + \vec{v}_2 - \vec{x}) \\
&= \vec{0} \cdot \vec{0} \\
&= 0
\end{aligned}$$

Thus $f(\vec{v_1} + \vec{v}_2) = f(\vec{v}_2) + f(\vec{v}_2)$.

(T2) Let $\vec{v} \in V$ and $a \in \mathbb{R}$. To show that $f(a\vec{v}) = af(\vec{v})$, we show that the length L of the difference is zero. Let

$$L = (f(a\vec{v}) - af(\vec{v})) \cdot (f(a\vec{v}) - af(\vec{v})).$$

Then

$$\begin{aligned}
L &= f(a\vec{v}) \cdot f(a\vec{v}) - 2f(a\vec{v}) \cdot (af(\vec{v})) + (af(\vec{v})) \cdot (af(\vec{v})) \\
&= f(a\vec{v}) \cdot f(a\vec{v}) - 2af(a\vec{v}) \cdot f(\vec{v}) + a^2(f(\vec{v}) \cdot f(\vec{v})) \\
&= (a\vec{v}) \cdot (a\vec{v}) - 2a(a\vec{v}) \cdot \vec{v} + a^2(\vec{v} \cdot \vec{v}) \\
&= a^2\vec{v} \cdot \vec{v} - 2a^2\vec{v} \cdot \vec{v} + a^2\vec{v}) \cdot \vec{v} \\
&= 0
\end{aligned}$$

So, $f(a\vec{v}) = af(\vec{v})$.

Therefore, f is linear. □

Definition 7.18. A *linear isometry* is an orthogonal linear transformation.

In light of what we have just shown, every isometry which fixes the origin is a linear isometry, and every linear isometry is an orthogonal isomorphism.

Let V and W be vector spaces, and let $f : V \to W$. Recall that f is an *affine transformation* if f is of the form $f = \tau \circ T$, where T is a linear transformation, and τ is a translation. Combining Proposition 7.9 and Proposition 7.17, we have the following theorem, which is a special case of the Mazur-Ulam Theorem.

Theorem 7.19. *Let $f : V \to W$ be an isometry. Then f is an affine transformation.*

3. Orthonormal Bases

Let $\vec{v}, \vec{w} \in \mathbb{R}^n$. Recall that we have said that $\vec{v}$ and $\vec{w}$ are *orthogonal* if $\vec{v} \cdot \vec{w} = 0$; this occurs exactly when $\vec{w}$ is perpendicular to $\vec{v}$. We now generalize this to sets of vectors.

Definition 7.20. A subset $X \subset \mathbb{R}^n$ is *orthogonal* if X is a set of nonzero vectors such that for every distinct $\vec{x}_1, \vec{x}_2 \in X$, we have $\vec{x}_1 \cdot \vec{x}_2 = 0$.

Proposition 7.21. *Let $X = \{\vec{x}_1, \dots, \vec{x}_r\} \subset \mathbb{R}^n$ be an orthogonal set of vectors. Then*

$$\vec{x}_i \cdot \vec{x}_j = \begin{cases} \|\vec{x}_j\|^2 & \text{if } i = j \ ; \\ 0 & \text{if } i \neq j \ . \end{cases}$$

Proof. The dot product is zero if $\vec{x}_i$ and $\vec{x}_j$ are distinct. If $i = j$, then $\vec{x}_i \cdot \vec{x}_j = \vec{x}_j \cdot \vec{x}_i = \|\vec{x}_j\|^2$. □

Proposition 7.22. *Let $X \subset \mathbb{R}^n$ be orthogonal. Then X is independent.*

Proof. Let $X = \{\vec{x}_1, \dots, \vec{x}_r\}$. Let $\sum_{i=1}^r a_i \vec{x}_i = \vec{0}$ be a dependence relation from X. Let j be between 1 and r, and take the dot product of both sides of the dependence relation with $\vec{x}_j$:

$$\left(\sum_{i=1}^r a_i \vec{x}_i \right) \cdot \vec{x}_j = \vec{0} \cdot \vec{x}_j = 0.$$

Since dot product is linear, this gives

$$\sum_{i=1}^r a_i (\vec{x}_i \cdot \vec{x}_j) = 0.$$

Since X is orthogonal, this becomes

$$a_i \vec{x}_j \cdot \vec{x}_j = a_i |\vec{x}_j| = 0.$$

Since $\vec{x}_j$ is nonzero, we conclude that $a_j = 0$. Since j was arbitrary, $a_i = 0$ for all i. This shows independence. □

Definition 7.23. Let $X \subset \mathbb{R}^n$. We say that X is an *orthonormal* set of vectors if X is orthogonal, and $\|\vec{x}\| = 1$ for all $\vec{x} \in X$.

Proposition 7.24. *Let $X\{\vec{x}_1, \dots, \vec{x}_r\} \subset \mathbb{R}^n$ be an orthonormal set of vectors. Then*

$$\vec{x}_i \cdot \vec{x}_j = \begin{cases} 1 & \text{if } i = j \ ; \\ 0 & \text{if } i \neq j \ . \end{cases}$$

Definition 7.25. Let $V \leq \mathbb{R}^n$ and let $X \subset \mathbb{R}^n$. We say that X is an *orthonormal basis* for V if X is an orthonormal set of vectors such that $\operatorname{span} X = V$.

Proposition 7.26. *Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear isometry, and let X is the standard basis for $\mathbb{R}^n$. Then $T(X)$ is an orthonormal basis for $\mathbb{R}^n$.*

Proof. A linear isometry is an orthogonal isomorphism; that is, it sends a basis to a basis, and preserves dot product. Thus for $\vec{x} \in X$, since X is the standard basis, $\|\vec{x}\| = 1$, so $\|T(\vec{x})\| = 1$. Moreover, if $\vec{x}_1, \vec{x}_2 \in X$ are distinct,

$$T(\vec{x}_1) \cdot T(\vec{x}_2) = \vec{x}_1 \cdot \vec{x}_2 = 0;$$

thus $T(X)$ is orthonormal. □

Proposition 7.27. *Let $T : V \to W$ be a linear isometry, and let X be an orthonormal basis for V. Then $T(X)$ is an orthonormal basis for W.*

Proof. A linear isometry is an orthogonal isomorphism; that is, it sends a basis to a basis, and preserves dot product; it also preserves norms. Thus for $\vec{x} \in X$, since X is an orthonormal basis, $\|\vec{x}\| = 1$, so $\|T(\vec{x})\| = 1$. Moreover, if $\vec{x}_1, \vec{x}_2 \in X$ are distinct,

$$T(\vec{x}_1) \cdot T(\vec{x}_2) = \vec{x}_1 \cdot \vec{x}_2 = 0;$$

thus $T(X)$ is orthonormal. □

Proposition 7.28. *Let $X = \{\vec{x}_1, \ldots, \vec{x}_r\} \subset \mathbb{R}^n$ be an orthonormal set. Let*

$$\vec{v}_1 = \sum_{i=1}^{r} a_i \vec{x}_i \quad \textit{and} \quad \vec{v}_2 = \sum_{i=1}^{r} b_i \vec{x}_i$$

be arbitrary vectors in $\operatorname{span}(X)$*. Then*

$$\vec{v}_1 \cdot \vec{v}_2 = \sum_{i=1}^{r} a_i b_i.$$

Proof. Since X is orthonormal, $\vec{x}_i \cdot \vec{x}_j$ equals zero, unless $i = j$, in which case it equals one. This, in addition to the fact that dot product is linear, allows us to compute:

$$\begin{aligned}
\vec{v}_1 \cdot \vec{v}_2 &= \left(\sum_{i=1}^{r} a_i \vec{x}_i\right) \cdot \left(\sum_{j=1}^{r} b_j \vec{x}_j\right) \\
&= \sum_{i=1}^{r} \left(a_i \vec{x}_i \cdot \left(\sum_{j=1}^{r} b_j \vec{x}_j\right)\right) \\
&= \sum_{i=1}^{r} \sum_{j=1}^{r} a_i b_j (\vec{x}_i \cdot \vec{x}_j) \\
&= \sum_{i=1}^{r} a_i b_i.
\end{aligned}$$

□

Proposition 7.29. *Let $T : V \to W$ be an isomorphism, and let X be an orthonormal basis for V. If $T(X)$ is an orthonormal basis for W, then T is an isometry.*

Proof. In light of Proposition 7.16, it suffices to show that T is orthogonal. Let

$$\vec{v}_1 = \sum_{i=1}^{r} a_i \vec{x}_i \quad \text{and} \quad \vec{v}_2 = \sum_{i=1}^{r} b_i \vec{x}_i$$

be arbitrary vectors in V. Then

$$T(\vec{v}_1) = \sum_{i=1}^{r} a_i T(\vec{x}_i) \quad \text{and} \quad \vec{v}_2 = \sum_{i=1}^{r} b_i T(\vec{x}_i).$$

Since X and $T(X)$ are orthonormal,

$$\vec{v}_1 \cdot \vec{v}_2 = \sum_{i=1}^{r} a_i b_i = T(\vec{v}_1) \cdot T(\vec{v}_2).$$

□

We describe the Gram-Schmidt process for taking a basis for a subspace of $\mathbb{R}^n$ and producing a orthonormal basis.

Proposition 7.30. *Let $V \leq \mathbb{R}^n$ and set $X = \{\vec{x}_1, \ldots, \vec{x}_m\} \subset V$ be an orthonormal basis for V. Then every for $v \in V$, we have*

$$\vec{v} = \sum_{i=1}^{m} (\vec{v} \cdot \vec{x}_i)\vec{x}_i.$$

Proof. Let $\vec{v} \in V$ and X is a basis for V, then $\vec{v}$ is a linear combination of the elements of X; that is,

$$\vec{v} = \sum_{i=1}^{m} a_i \vec{x}_i$$

for some $a_1, \ldots, a_m \in \mathbb{R}$. Let $\vec{x}_j \in X$. Taking the dot product of both sides of this equation with $\vec{x}_j$, we have

$$\begin{aligned} \vec{v} \cdot \vec{x}_j &= \left(\sum_{i=1}^{m} a_i \vec{x}_i \right) \cdot \vec{x}_j \\ &= a_j(\vec{x}_j \cdot \vec{x}_j) \quad (\text{ because } \vec{x}_i \cdot \vec{x}_j = 0 \text{ for } i \neq j) \\ &= a_j \quad (\text{because } |\vec{x}_j| = 1) \end{aligned}$$

This is all we needed to show. □

Proposition 7.31 (Gram-Schmidt Process)**.** *Let $V \leq \mathbb{R}^n$. Then V has an orthonormal basis.*

Proof. Let $Y = \{\vec{y}_1, \ldots, \vec{y}_m\}$ be a basis for V. Set $\vec{x}_1 = \vec{y}_1$. The *vector projection* of $\vec{y}_2$ onto $\vec{x}_1$ is $\frac{\vec{x}_1 \cdot \vec{y}_2}{|\vec{x}_1|^2}\vec{x}_1$. The difference between this and $\vec{y}_2$ is perpendicular to $\vec{x}_1$; thus let $\vec{x}_2 = \vec{y}_2 - \frac{\vec{x}_1 \cdot \vec{y}_2}{|\vec{x}_1|^2}\vec{x}_1$. Continuing in this way, inductively define

$$\vec{x}_k = \vec{y}_k - \left(\sum_{i=1}^{k-1} \frac{\vec{x}_i \cdot \vec{y}_k}{|\vec{x}_i|^2} \vec{x}_i \right).$$

Let $X = \{\vec{x}_1, \ldots, \vec{x}_m\}$. We claim that X is an orthogonal basis for V. It is easy to see that $\operatorname{span}(X) = \operatorname{span}(Y)$, 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 $\vec{x}_k \in X$ and select $j < k$. By the principle of induction, we may assume that $\vec{x}_i \cdot \vec{x}_j = 0$ for $i < k$; we wish to show that this implies that $\vec{x}_k \cdot \vec{x}_j = 0$. Thus compute:

$$\begin{aligned} \vec{x}_k \cdot \vec{x}_j &= \left[\vec{y}_k - \left(\sum_{i=1}^{k-1} \frac{\vec{x}_i \cdot \vec{y}_k}{|\vec{x}_i|^2} \vec{x}_i \right) \right] \cdot \vec{x}_j \\ &= \vec{y}_k \cdot \vec{x}_j - \left(\sum_{i=1}^{k-1} \frac{\vec{x}_i \cdot \vec{y}_k}{|\vec{x}_i|^2} \vec{x}_i \cdot \vec{x}_j \right) \\ &= \vec{y}_k \cdot \vec{x}_j - (\frac{\vec{x}_j \cdot \vec{y}_k}{|\vec{x}_j|} \vec{x}_j \cdot \vec{x}_j) \\ &= \vec{y}_k \cdot \vec{x}_j - \vec{x}_j \cdot \vec{y}_k \\ &= 0. \end{aligned}$$

Thus $\vec{x}_k \perp \vec{x}_j$. To obtain an orthonormal basis from X, divide each element $\vec{x} \in X$ by its length. □

4. Orthogonal Matrices

Definition 7.32. Let A be an $n \times n$ matrix. We say that A is *orthogonal* if

$$AA^* = A^*A = I.$$

Proposition 7.33. *Let A be an $n \times n$ matrix. Then A is orthogonal if and only if the columns of A form an orthonormal basis for $\mathbb{R}^n$.*

Proof. For any matrix C, let C_i denote the i^{th} row of C, and C^j denote the j^{th} column of C.

Let $B = A^*A$, and write $B = (b_{ij})_{ij}$. Then $b_{ij} = (A^*)_i \cdot A^j$; but the rows of A^* are the columns of A, so $b_{ij} = A^i \cdot A^j$.

Now $B = I$ i and only if

$$A^i \cdot A^j = \begin{cases} b_{ii} = 1 & \text{if } i = j \ ; \\ b_{ij} = 0 & \text{if } i \neq j \ . \end{cases}$$

So if A is orthonormal, $B = I$, so this equation holds, so the set of columns of A are orthonormal, and thus indepenent. Since there are n of them, they must span $\mathbb{R}^n$.

On the other hand, if the asetr of columns of A are an orthonormal set, the equation will hold, and it will be the case that $B = I$, so that A is orthogonal. □

Proposition 7.34. *Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation. Then T is an isometry if and only if the matrix corresponding to T is orthogonal.*

Proof. Let A be the matrix corresponding to T. Then

$$\begin{aligned} T \text{ is an isometry } &\Leftrightarrow T \text{ sends the standard basis to an orthonormal basis} \\ &\Leftrightarrow \text{the columns of } A \text{ are an orthonormal set} \\ &\Leftrightarrow A \text{ is orthogonal.} \end{aligned}$$

□

CHAPTER 8

Determinants and Eigenvectors

1. Transformations of a Vector Space

Let $T : \mathbb{R}^n \to \mathbb{R}^n$ 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^{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 $\vec{x} \in \mathbb{R}^n$, there is a unique solution to the equation $A\vec{x} = \vec{b}$, namely $\vec{x} = A^{-1}\vec{b}$, and this solution can be found by Gaussian elimination. In particular, if $\vec{x}_i$ is the unique solution to $A\vec{x} = \vec{e}_i$ for $i = 1, \dots, n$, then $A^{-1} = [\vec{x}_1 \mid \cdots \mid \vec{x}_n]$.

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

$$f(\vec{v}_1, \ldots, \vec{v}_{i-1}, \vec{v}_i + \vec{w}_i, \vec{v}_{i+1}, \ldots, \vec{v}_m)$$
$$= f(\vec{v}_1, \ldots, \vec{v}_{i-1}, \vec{v}_i, \vec{v}_{i+1}, \ldots, \vec{v}_m) + f(\vec{v}_1, \ldots, \vec{v}_{i-1}, \vec{w}_i, \vec{v}_{i+1}, \ldots, \vec{v}_m);$$

and

$$f(\vec{v}_1, \ldots, \vec{v}_{i-1}, a\vec{v}_i, \vec{v}_{i+1}, \ldots, \vec{v}_m) = af(\vec{v}_1, \ldots, \vec{v}_{i-1}, \vec{v}_i, \vec{v}_{i+1}, \ldots, \vec{v}_m).$$

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 $f(\vec{x}_{i_1}, \ldots, \vec{x}_{i_m})$, where the $\vec{x}_i$'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

$$f(\vec{v}_1, \ldots, \vec{v}_i, \ldots, \vec{v}_j, \ldots, \vec{v}_m) = -f(\vec{v}_1, \ldots, \vec{v}_j, \ldots, \vec{v}_i, \ldots, \vec{v}_m).$$

Let $f : V^m \to \mathbb{R}$ be alternating. Suppose that two positions of an n-tuple are the same, say $\vec{v}_i = \vec{v}_j$. Then switching them gives the same value for f; but it must also give the negative value, since f is alternating. Thus $f(\vec{v}_1, \ldots, \vec{v}_m) = 0$ whenever two positions are the same.

Example 8.1. Let $V = \mathbb{R}^2$ and let $f : \mathbb{R}^2 \to \mathbb{R}$ be given by $f(\vec{v}, \vec{w}) = ad - bc$, where $\vec{v} = \langle a, b\rangle$ and $\vec{w} = \langle c, d\rangle$. Then f is an alternating multilinear function.

Note that $f(\vec{e}_1, \vec{e}_2) = 1 \cdot 1 - 0 \cdot 0 = 1$.

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 $X = \{\vec{x}_1, \ldots, \vec{x}_n\}$ be a basis for V. Then f is completely determined by the value of $f(\vec{x}_1, \ldots, \vec{x}_n)$. To see this, pick an arbitrary ordered n-tuple $(\vec{v}_1, \ldots, \vec{v}_n)$. Write each of these as a linear combination of the vectors in X. Use multilinearity to break $f(\vec{v}_1, \ldots, \vec{v}_n)$ into a sum of things of the form $f(x_{i_1}, \ldots, x_{i_n})$. Use alternation to rearrange this into a sum of things of the form $\pm f(x_1, \ldots, x_n)$.

A function $f : V^m \to \mathbb{R}$ is called *normalized* with respect to an ordered basis $\{\vec{x}_1, \ldots, \vec{x}_n\}$ if $m = n$ and $f(\vec{x}_1, \ldots, \vec{x}_n) = 1$.

Proposition 8.2. *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 $(\vec{v}_1, \ldots, \vec{v}_n)$ 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 8.3. *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 1×1 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 the 1×1 matrix $[a]$ to be a.

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^{th} row and the j^{th} column. This matrix is called the ij^{th} *minor* of A.

Let $a'_{ij} = \det(A_{ij})$. This number is called the ij^{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^{th} row to expand along, the formula is

$$\det(A) = \sum_{j=1}^{n} (-1)^{j-1} a'_{ij}.$$

If we choose the j^{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)**:

$$\begin{aligned}
&\det[\vec{x}_1 \mid \cdots \mid \vec{x}_i + c\vec{x}_j \mid \cdots \mid \vec{x}_j \mid \cdots \mid \vec{x}_n] \\
&\quad = \det[\vec{x}_1 \mid \cdots \mid \vec{x}_i \mid \cdots \mid \vec{x}_j \mid \cdots \mid \vec{x}_n] + c\det[\vec{x}_1 \mid \cdots \mid \vec{x}_j \mid \cdots \mid \vec{x}_j \mid \cdots \mid \vec{x}_n] \\
&\qquad = \det[\vec{x}_1 \mid \cdots \mid \vec{x}_i \mid \cdots \mid \vec{x}_j \mid \cdots \mid \vec{x}_n] + 0.
\end{aligned}$$

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 $\vec{v}_1, \dots, \vec{v}_n \in \mathbb{R}^m$ is the set

$$\{t_1\vec{v}_1 + \cdots + t_n\vec{v}_n \mid t_i \in [0,1]\}.$$

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 $\vec{v}_n$ and the span of $\{\vec{v}_1, \dots, \vec{v}_{n-1}\}$, and the base is the $(n-1)$-box determined by $\vec{v}_1, \dots, \vec{v}_{n-1}$. Let $\text{vol}\{\vec{v}_1, \dots, \vec{v}_n\}$ 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,

$$\text{vol}(T_A(X)) = \pm\det(A)\text{vol}(X),$$

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. 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 $\vec{v} \in V$ such that $T(\vec{v}) = \lambda\vec{v}$ for some $\lambda \in \mathbb{R}$. The number λ is called an *eigenvalue* of T.

That is, a nonzero vector $\vec{v}$ is an eigenvector of T if and only if $T(\vec{v})$ is on the same line through the origin as $\vec{v}$, 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 $T_A(\vec{x}) = A\vec{x}$.

Proposition 8.4. *Let $T : V \to V$ be a linear transformation. Let $\vec{v} \in V$ be an eigenvector with eigenvalue λ. Let $a \in \mathbb{R}$. Then $a\vec{v}$ is an eigenvector with eigenvalue λ.*

Proof. We have $T(a\vec{v}) = aT(\vec{v}) = a\lambda\vec{v} = \lambda(a\vec{v})$. □

Example 8.5. 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 $T(\vec{e}_1) = 2\vec{e}_1$ and $T(\vec{e}_2) = 3\vec{e}_2$, 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 $\vec{v} = a\vec{e}_1 + b\vec{e}_2$, then $T(v) = 2a\vec{e}_1 + 3b\vec{e}_2$ is a scalar multiple of $\vec{v}$ if and only if either a or b is zero. Thus no other vectors are eigenvectors. □

Example 8.6. 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 λ. □

Example 8.7. Find the eigenvectors and eigenvalues of the linear transformation which rotates $\mathbb{R}^2$ by 90 degrees.

Solution. There are none. □

Example 8.8. Find the eigenvectors and eigenvalues of the linear transformation which reflects $\mathbb{R}^2$ across the y-axis.

Solution. Eigenvalue 1 corresponds to eigenvector $\vec{e}_1$. Eigenvalue -1 corresponds to eigenvector $\vec{e}_2$. □

7. Eigenspaces

Let $T : V \to V$ be a linear transformation with eigenvalue λ. The *eigenspace* of λ is the set containing zero and of all eigenvectors of T whose eigenvalue is λ:

$$\mathrm{eig}_\lambda(T) = \{\vec{v} \in V \mid T(\vec{v}) = \lambda\vec{v}\}.$$

Proposition 8.9. *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 $(T - a)(\vec{v}) = T(\vec{v}) - av$. Then $\vec{v} \in \mathrm{eig}_a(T)$ if and only if $(T - a)(\vec{v}) = \vec{0}$. 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 8.10. *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.** $\ker(T) = \{\vec{0}\}$*;*
- **vi.** $\mathrm{eig}_0(T) = \{\vec{0}\}$*;*
- **vii.** 0 *is not an eigenvalue of T.*

Let $T : V \to V$ be a linear transformation. The *total eigenspace* of T is

$$\mathrm{eig}(T) = \mathrm{span}\{\vec{v} \in V \mid \ \vec{v} \text{ is an eigenvector of } T\ \}.$$

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)** $U_i \cap U_j = \{\vec{0}\}$ whenever $i \neq j$.

In this case, we may write

$$V = \oplus_{i=1}^n U_i.$$

Proposition 8.11. *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 λ_i ranges from $i = 1, \dots, n$. Also, if v has eigenvalue λ_i, then it cannot also have a different eigenvalue λ_j. Thus the intersection of two of these eigenspaces is trivial. □

8. 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 $T(\vec{v}) = \lambda\vec{v}$, where λ is any real number. That is, we wish to solve

$$T(\vec{v}) - \lambda\vec{v} = \vec{0}.$$

Let us first try to find an appropriate λ.

If A is the matrix corresponding to T, then this equation becomes

$$A\vec{v} - \lambda I\vec{v} = \vec{0}.$$

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 λ. 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 λ 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 λ, one can find the corresponding eigenvectors by solving $(A - \lambda I)\vec{x} = \vec{0}$.

Example 8.12. 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 $\ker(A-2I) = \operatorname{span}\{\langle 1, 0, 1\rangle\}$ and $\ker(A+I) = \operatorname{span}\{\langle 1, -3, 4\rangle\}$. □

9. 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 Γ_X is the unique transformation from $V \to \mathbb{R}^n$ which sends x_i to e_i. Since Γ_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

$$\begin{tikzcd} V \arrow[r, "T"] \arrow[d, "\Gamma_X"'] & V \arrow[d, "\Gamma_X"] \\ \mathbb{R}^n \arrow[r, "\Gamma_X \circ T \circ \Gamma_X^{-1}"'] & \mathbb{R}^n \end{tikzcd}$$

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^{th} column of B is $(1, 0, 3, -2)$, then $T(x_4) = x_1 + 3x_3 - x_4$.

10. 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

$$\begin{tikzcd} \mathbb{R}^n \arrow[r, "A"] & \mathbb{R}^n \arrow[d, "C^{-1}"] \\ \mathbb{R}^n \arrow[u, "C"] \arrow[r, "C^{-1}AC"'] & \mathbb{R}^n \end{tikzcd}$$

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 8.13. 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 8.14. *Let $T : V \to V$ be a linear transformation.*
Let $\vec{v}_1, \ldots, \vec{v}_n \in V$ be eigenvectors with distinct eigenvalues.
Then $\{\vec{v}_1, \ldots, \vec{v}_n\}$ is independent.

Proof. Let d_i be the eigenvalue corresponding to $\vec{v}_i$. 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

$$\vec{v}_k = a_1\vec{v}_1 + \ldots a_{k-1}\vec{v}_{k-1},$$

where $\{\vec{v}_1, \ldots, \vec{v}_{k-1}\}$ is independent. Multiplying this equation by d_k gives

$$d_k\vec{v}_k = \sum_{i=1}^{k-1} a_{k-1}\vec{v}_{k-1},$$

but applying A gives

$$d_k\vec{v}_k = \sum_{i=1}^{k-1} a_i d_i \vec{v}_i.$$

Subtracting these gives

$$0 = \sum_{i=1}^{k-1} (d_k - d_i) a_i \vec{v}_i.$$

Since the d_i's are distinct, this is a nontrivial dependence relation, contradicting the fact that $\{\vec{v}_1, \ldots, \vec{v}_{k-1}\}$ is independent. □

Corollary 8.15. *Let $T : V \to V$ be a linear transformation, where* $\dim(V) = n$.
Let $\vec{v}_1, \ldots, \vec{v}_n \in V$ be eigenvectors with distinct eigenvalues.
Then

(a) *$\{\vec{v}_1, \ldots, \vec{v}_n\}$ is a basis for V;*
(b) $\mathrm{eig}(T) = V$*;*
(c) *$V = \oplus_{i=1}^{n} \mathrm{eig}_{\lambda_i}(T)$.*

11. 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 8.16. *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 $\vec{x}_i$ be the i^{th} column of C. Then

$$A\vec{x}_i = CDC^{-1}\vec{x}_i = CD\vec{e}_i = C(d_i\vec{e}_i) = d_iC\vec{e}_i = d_i\vec{x}_i.$$

Suppose that A has a basis of eigenvectors $X = \{\vec{x}_1, \dots, \vec{x}_n\}$ with corresponding eigenvalues $d_1, \dots, d_n$. Form the square matrix D with d_i's along the diagonal and 0 elsewhere. Let $C = [\vec{x}_1 \mid \cdots \mid \vec{x}_n]$. Then D is A written with respect to the basis X, so $D = C^{-1}AC$. □

Here is a criterion for diagonalizability.

Proposition 8.17. *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. □

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 8.18. *Let $B = C^{-1}AC$. Then $B^n = C^{-1}A^nC$.*

Proposition 8.19. *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 8.20. 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 $\vec{x}_1 = \langle -1, 0, 3\rangle$, $\vec{x}_2 = \langle 0, 1, 0\rangle$, and $\vec{x}_3 = \langle -1, 0, 1\rangle$. Let $C = [\vec{x}_1 \mid \vec{x}_2 \mid \vec{x}_3]$. 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). □

12. 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 $X = \{\vec{x}_1, \ldots, \vec{x}_n\}$ for V. If we know the value of T on each of the basis vectors $\vec{x}_i$, 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 8.21. *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 $B = C^{-1}AC$. $\square$

This proposition states that matrices of the same transformation with respect to different bases are conjugate. Diagrams help explain this; the transformation diagram

$$\begin{tikzcd}
\mathbb{R}^n \arrow[r, "\Gamma_X \circ T \circ \Gamma_X^{-1}"] & \mathbb{R}^n \\
V \arrow[u, "\Gamma_X"] \arrow[r, "T"] \arrow[d, "\Gamma_Y"'] & V \arrow[u, "\Gamma_X"'] \arrow[d, "\Gamma_Y"] \\
\mathbb{R}^n \arrow[r, "\Gamma_Y \circ T \circ \Gamma_Y^{-1}"'] & > \mathbb{R}^n
\end{tikzcd}$$

is converted into the matrix diagram

$$\begin{tikzcd}
\mathbb{R}^n \arrow[r, "A"] & \mathbb{R}^n \arrow[d, "C^{-1}"] \\
\mathbb{R}^n \arrow[u, "C"] \arrow[r, "B=C^{-1}AC"'] & \mathbb{R}^n
\end{tikzcd}$$

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 Γ_X.

Proposition 8.22. *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

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^{th} row by c and add to the i^{th} row
$D(i; c)$	Multiply the i^{th} row by c
$P(i, j)$	Swap the i^{th} row and the j^{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) Eliminate any zero rows.
(2) Insert a zero row at row i for every free variable x_i.
(3) Multiply each free column by -1.
(4) Add $\vec{e}_i$ to each free column.
(5) The particular solution is now the augmentation column.
(6) 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 $\operatorname{col}(A)$, the row space $\operatorname{row}(A)$, 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 $\operatorname{col}(A)$.
(F2) The nonzero rows of B or C are a basis for $\operatorname{row}(A)$.
(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 $\operatorname{row}(A^*)$.
(F6) The nonzero rows of B or C are a basis for $\operatorname{col}(A^*)$.

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 $X = \{\vec{w}_1, \ldots, \vec{w}_n\} \subset \mathbb{R}^m$ and let $W = \operatorname{span}(X)$.
Form the $m \times n$ matrix $A = [\vec{w}_1 \mid \cdots \mid \vec{w}_n]$.
Reduce A and apply **(F1)**; a basis for W is a basis for $\operatorname{col}(A)$.
Reduce A^* and apply **(F2)**; a basis for W is a basis for $\operatorname{row}(A^*)$.

5. Test for Linear Independence

Let $X = \{\vec{w}_1, \ldots, \vec{w}_n\} \subset \mathbb{R}^m$.
If $n > m$, then X is dependent.
Form the $m \times n$ matrix $A = [\vec{w}_1 \mid \cdots \mid \vec{w}_n]$.
Reduce A; if $n = r$, then X is independent, otherwise it is not.

6. Test for Spanning

Let $X = \{\vec{w}_1, \ldots, \vec{w}_n\} \subset \mathbb{R}^m$.
If $n < m$, then X does not span $\mathbb{R}^m$.
Form the $m \times n$ matrix $A = [\vec{w}_1 \mid \cdots \mid \vec{w}_n]$.
Reduce A; if $m = r$, then X spans $\mathbb{R}^m$; otherwise it does not.

7. Test for a Basis

Let $X = \{\vec{w}_1, \ldots, \vec{w}_n\} \subset \mathbb{R}^m$.
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^{th} row:

$$\det(A) = \sum_{j=1}^{n} (-1)^{j-1} a_{ij} \det(A_{ij}).$$

Along the j^{th} column:

$$\det(A) = \sum_{i=1}^{n} (-1)^{i-1} a_{ij} \det(A_{ij}).$$

Here, A_{ij} is the ij^{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.

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^{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^{nd} edition, Brooks/Cole Publishing Company (1987,1991)