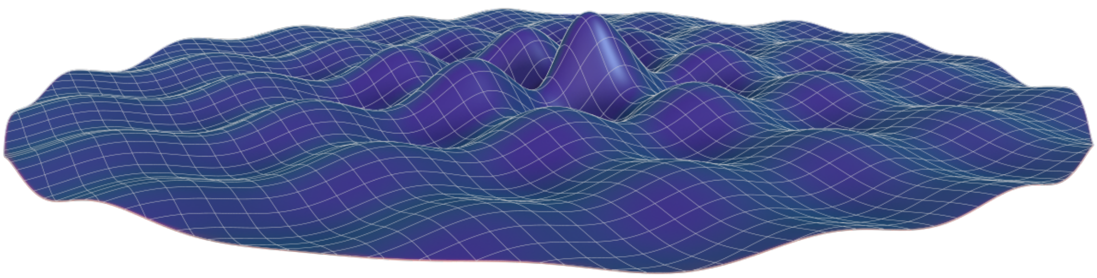
Linear Algebra

Jason Siefken



April 26, 2019

Contents

Introduction

Multivariable Calculus approaches the subject from a mathematical, but not overly technical, perspective. The key idea of calculus—chop things into little pieces and put them together again—is emphasized throughout.

Licensing

This book would not be possible without the long tradition of mathematical inquiry that came before. And like the ideas of mathematics, which are free for all to re-imagine, re-use, and re-purpose, so too is this book.

This book is a derived work of the Creative-Commons licensed *ISP Mathematics 281*[1] multivariable calculus textbook by Leonard Evans of Northwestern University. The problems from this textbook come from a variety of sources. Unannotated problems come from *ISP Mathematics 281*. Problems with a superscript OS come from *Calculus Volume 3*[2] from the Openstax project. The Openstax problems are licensed under a Creative Commons By-Attribution Non-Commercial Share Alike 4.0 license. If you require this textbook to be fully open-source, remove these problems.

[1] https://github.com/siefkenj/ISPMathematics281
[2] https://openstax.org/details/calculus-volume-3

Contributors

This book is a collaborative effort. The following people have contributed to its creation:

◦ Bailey Bjornstad ◦ Dylaan Cornish ◦ Julia Dierksheide ◦ Kathryn Riedel ◦ Barbara Siefken ◦ Maxwell Sigal ◦ Kim Soram ◦ Ieva Stakvileviciute ◦ Paul Strauss ◦ Andrew Wilson ◦

Part I

Linear Algebra I

Chapter 1

Preliminaries

1.1 Mathematical Notation

Mathematics is a sophisticated and precise language. We will start our linear algebra journey by learning some basic words.

Modern mathematics makes heavy use of *sets*. A set is an unordered collection of distinct objects. We won't try and pin it down more than this—our intuition about collections of objects will suffice.[1] We write a set with curly-braces { and } and list the objects inside. For instance

$$\{1,2,3\}.$$

This would be read aloud as "the set containing the elements 1, 2, and 3." Things in a set are called *elements*, and the symbol $\in$ is used to specify that something is an element of a set. In contrast, $\notin$ is used to specify something is not an element of a set. For example,

$$3 \in \{1,2,3\} \qquad 4 \notin \{1,2,3\}.$$

Sets can contain mixtures of objects, including other sets. For example,

$$\{1,2,a,\{-70,\infty\},x\}$$

is a perfectly valid set.

It is tradition to use capital letters to name sets. So we might say $A = \{6,7,12\}$ or $X = \{7\}$. However there are some special sets which already have names/symbols associated with them. The *empty set* is the set containing no elements and is written $\emptyset$ or $\{\}$. Note that $\{\emptyset\}$ is *not* the empty set—it is the set containing the empty set! It is also traditional to call elements of a set *points* regardless of whether you consider them "point-like" objects.

Operations on Sets

If the set A contains all the elements that the set B does, we call B a *subset* of A. Conversely, we call A a *superset* of B.

Definition 1.1.1 — Subset & Superset. The set B is a *subset* of the set A, written $B \subseteq A$, if for all $b \in B$ we also have $b \in A$. In this case, A is called a *superset* of B.[a]

[a] Some mathematicians use the symbol $\subset$ instead of $\subseteq$.

Some simple examples are $\{1,2,3\} \subseteq \{1,2,3,4\}$ and $\{1,2,3\} \subseteq \{1,2,3\}$. There's something funny about that last example, though. Those two sets are not only subsets/supersets of each other, they're *equal*. As surprising as it seems, we actually need to define what it means for two sets to be equal.

Definition 1.1.2 — Set Equality. The sets A and B are *equal*, written $A = B$, if $A \subseteq B$ and $B \subseteq A$.

[1] When you pursue more rigorous math, you rely on definitions to get yourself out of philosophical jams. For instance, with our definition of set, consider "the set of all sets that don't contain themselves." Such a set cannot exist! This is called *Russel's Paradox*, and shows that if we start talking about sets of sets, we may need more than intuition.

Having a definition of equality to lean on will help us when we need to prove things about sets.

■ **Example 1.1** Let A be the set of numbers that can be expressed as $2n$ for some whole number n, and let B be the set of numbers that can be expressed as $m+1$ where m is an odd whole number. We will show $A = B$.

First, let us show $A \subseteq B$. If $x \in A$ then $x = 2n$ for some whole number n. Therefore

$$x = 2n = 2(n-1)+1+1 = m+1$$

where $m = 2(n-1)+1$ is, by definition, an odd number. Thus $x \in B$, which proves $A \subseteq B$.

Now we will show $B \subseteq A$. Let $x \in B$. By definition, $x = m+1$ for some odd m and so by the definition of oddness, $m = 2k+1$ for some whole number k. Thus

$$\begin{aligned} x = m+1 &= (2k+1)+1 = 2k+2 \\ &= 2(k+1) = 2n, \end{aligned}$$

where $n = k+1$, and so $x \in A$. Since $A \subseteq B$ and $B \subseteq A$, by definition $A = B$. ■

Set-builder Notation

Specifying sets by listing all their elements can be a hassle, and if there are an infinite number of elements, it's impossible! Fortunately, *set-builder notation* solves these problems. If X is a set, we can define a subset

$$Y = \{a \in X \ : \ \text{some rule involving } a\},$$

which is read "Y is the set of a in X *such that* some rule involving a is true." If X is intuitive, we may omit it and simply write $Y = \{a \ : \ \text{some rule involving } a\}$.[2] You may equivalently use "|" instead of ":", writing $Y = \{a \mid \text{some rule involving } a\}$.

■ **Example 1.2** The set $\mathbb{Z}$ is the set of integers (positive, negative, and zero whole numbers). To define E as the even integers, we could write

$$E = \{n \in \mathbb{Z} \ : \ n = 2k \text{ for some } k \in \mathbb{Z}\}.$$

To define P as the set of positive integers, we could write

$$P = \{n \in \mathbb{Z} \ : \ n > 0\}.$$

■

There are also some common operations we can do with two sets.

Definition 1.1.3 — Intersections & Unions. Let A and B be sets. Then the *intersection* of A and B, written $A \cap B$, is defined by

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

[2] If you want to get technical, to make this notation unambiguous, you define a *universe of discourse*. That is, a set $\mathcal{U}$ containing every object you might want to talk about. Then $\{a \ : \ \text{some rule involving } a\}$ is short for $\{a \in \mathcal{U} \ : \ \text{some rule involving } a\}$

The *union* of A and B, written $A \cup B$, is defined by

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

For example, if $A = \{1,2,3\}$ and $B = \{-1,0,1,2\}$, then $A \cap B = \{1,2\}$ and $A \cup B = \{-1,0,1,2,3\}$. Set unions and intersections are *associative*, which means it doesn't matter how you apply parentheses to an expression involving just unions or just intersections. For example $(A \cup B) \cup C = A \cup (B \cup C)$, which means we can give an unambiguous meaning to an expression like $A \cup B \cup C$ (just put the parentheses wherever you like). But watch out, $(A \cup B) \cap C$ means something different than $A \cup (B \cap C)$!

Definition 1.1.4 — Set Subtraction. For sets A and B, the *set-wise difference* between A and B, written $A \backslash B$, is the set

$$A \backslash B = \{x : x \in A \text{ and } x \notin B\}.$$

Definition 1.1.5 — Cardinality. For a set A, the *cardinality* of A, written $|A|$ is the number of elements in A. If A contains infinitely many elements, we write $|A| = \infty$.

Some common sets have special notation:

$$\begin{aligned}
\emptyset &= \{\}, \text{ the empty set} \\
\mathbb{N} &= \{0,1,2,3,\ldots\} = \{\text{natural numbers}\} \\
\mathbb{Z} &= \{\ldots,-3,-2,-1,0,1,2,3,\ldots\} = \{\text{integers}\} \\
\mathbb{Q} &= \{\text{rational numbers}\} \\
\mathbb{R} &= \{\text{real numbers}\} \\
\mathbb{R}^n &= \{\text{vectors in } n\text{-dimensional Euclidean space}\}
\end{aligned}$$

Besides unions, there's another way to join sets together: *products*.

Definition 1.1.6 — Cartesian Product. Given two sets A and B, the *Cartesian product* (sometimes shortened to *product*) of the sets A and B is written $A \times B$ and defined to be

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

The Cartesian product of two sets is the set of all ordered pairs of elements from those sets. For example,

$$\{1,2\} \times \{1,2,3\} = \{(1,1),(1,2),(1,3),(2,1),(2,2),(2,3)\}.$$

You can repeat this operation more than once. $\mathbb{R} \times \mathbb{R} \times \mathbb{R}$ is the set of all triples of real numbers. Extending power notation, if you take the Cartesian product of a set with itself some number of times, you can represent it with an exponent. Thus, $\mathbb{R} \times \mathbb{R} \times \mathbb{R}$ can be written as $\mathbb{R}^3$, which is a set we've seen before.[3]

[3] If you're scratching your head saying, "I thought $\mathbb{R}^3$ was 3-dimensional Euclidean space. How do we know that's the same thing as triples of real numbers?" your mind is keen. This is a theorem of linear algebra.

Functions

You're probably used to seeing functions like $f(x) = x^2$, but it's worth reviewing some of the concepts and terminology associated with functions.

Definition 1.1.7 — Function. A *function* with *domain* the set A and *co-domain* the set B is an object that associates every point in the set A with *exactly one* point in the set B.

If a function f has domain A and co-domain[4] B, we notate this by writing $f : A \to B$. If we want to further specify what the function f actually is, we need to express how f associates each point in A to a point in B. This can be done with an equation. For example, we could define the function $f : \mathbb{R} \to \mathbb{R}$ by

$$f(x) = 2x,$$

which says that each real number gets associated to its double. We can notate the same thing using a special type of arrow: "$\mapsto$". Now we might write

$$f : \mathbb{R} \to \mathbb{R} \text{ where } x \mapsto 2x,$$

which is read "f is a function from $\mathbb{R}$ to $\mathbb{R}$ where $x \in \mathbb{R}$ gets mapped to $2x$."

Note that every point in the co-domain of a function doesn't need to get mapped to. For example $g : \mathbb{R} \to \mathbb{R}$ given by $g(x) = x^2$ outputs only non-negative numbers, but it is still valid to specify $\mathbb{R}$ as the co-domain. However, if we wanted to make a point of it, we are perfectly justified in writing $g : \mathbb{R} \to [0, \infty)$ when defining g.

Many common math operations give rise to functions. For example, $f(x) = \sqrt{x}$ is the familiar square root function. Sometimes, when we wish to talk about a function for which notation already exists, we will put a "$\cdot$" where we would normally put a variable. Thus, we might say, "$\sqrt{\cdot}$ is the square root function."[5]

Definition 1.1.8 — Range. The *range* of a function $f : A \to B$ is the set of all outputs of f. That is

$$\operatorname{range} f = \{y \in B \,:\, y = f(x) \text{ for some } x \in A\}.$$

Definition 1.1.9 — Image. Let $f : A \to B$ be a function. The *image* of a set $X \subseteq A$, written $f(X)$ is defined by

$$f(X) = \{y \in B \,:\, y = f(x) \text{ for some } x \in X\}.$$

We see that if $f : A \to B$, $\operatorname{range} f = f(A)$. In words, the range of f is the image of its domain. This language will become useful when we think of functions as transformations that move or bend space. If $f : \mathbb{R}^2 \to \mathbb{R}^2$ is a function that warps the Cartesian plane, then the image of X under f could be visualized by painting X on the Cartesian plane, warping the whole plane, and then looking at the resulting, painted shape.

Closely related to images, we have the idea of *restriction*. Suppose $f : \mathbb{R}^2 \to \mathbb{R}$ is defined by $f(x, y) = xy$, but we were only really interested in f on the unit circle, $\mathcal{C}$. In this case, we

---

[4] Some people use the word *range* interchangeably with co-domain; we will not.

[5] Since $\sqrt{x}$ is "the square root of the quantity x," it is technically a quantity and not a function. This is why we write $\cdot$ instead of x when we want to refer to the square root *function*.

might say f attains a maximum on $\mathcal{C}$, or f *restricted to* $\mathcal{C}$ attains a maximum, even though f itself is unbounded. This idea comes up often enough to deserve its own notation.

Definition 1.1.10 — Restriction. If $f : A \to B$ and $X \subseteq A$, the *restriction* of f to X is written $f\big|_X$ and is defined to be the function $g : X \to A$ where $x \mapsto f(x)$.

The last important function-related ideas for us are function composition and inverses. Given two functions $f : A \to B$ and $g : B \to C$, we can *compose* g and f to get a new function.

Definition 1.1.11 — Composition. Given two functions $f : A \to B$ and $g : B \to C$, the *composition* of g and f, written $g \circ f$, is the function $h : A \to C$ where $x \mapsto g(f(x))$.

Note that the composition $g \circ f$ has the domain of f and the co-domain of g. When a point is fed into $g \circ f$, it moves from $A \to B \to C$. The composition $g \circ f$ only makes sense because the outputs of f are allowed as inputs to g. If we wrote $f \circ g$, it wouldn't mean much, because g outputs points in C and f has no idea what to do with points in C.[6]

Inverses relate to composition and the *identity function*, the function that does nothing to its inputs.

Definition 1.1.12 — Identity Function. The *identity function* $\mathrm{id} : A \to A$ is defined by the relation

$$\mathrm{id}(x) = x$$

for all $x \in A$.

Notice that for every set, that set is the domain of an identity function. Since the domain and co-domain of a function are part of its definition, we don't want to confuse them. After all, $f : \{0, 1\} \to \{0, 1\}$ given by $f(x) = x^2$ is a different function from $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^2$. For the special case of the identity function, we sometimes write the domain of the function as a subscript. That is, for $\mathrm{id} : A \to A$ we'd write id_A so it doesn't get confused with $\mathrm{id} : B \to B$, which we'd write id_B.

Definition 1.1.13 — Inverse Function. Let $f : A \to B$ be a function. If there exists a function $g : B \to A$ such that

$$f \circ g = \mathrm{id}_B \qquad \text{and} \qquad g \circ f = \mathrm{id}_A,$$

we say f is *invertible* and we call g the *inverse* of f. If f is invertible, we notate its inverse by f^{-1}.

Inverses can be tricky some times. For example, consider $f(x) = x^2$ and $g(x) = \sqrt{x}$. Here $g \circ f(x) = \sqrt{x^2} = |x|$ and $f \circ g(x) = \sqrt{x}^2 = x$. What's the deal? Well, it's all about domains. $f : \mathbb{R} \to [0, \infty)$ and $g : [0, \infty) \to [0, \infty)$. So, the domain of $g \circ f$ is $\mathbb{R}$ and the domain of $f \circ g$ is $[0, \infty)$. The domains are different, and indeed f is not invertible. However, g *is* invertible, and $g^{-1} = f\big|_{[0,\infty)}$. If we only input non-negative numbers into f, then f exactly undoes what g did. This subtle domain trickery can cause us a lot of headaches if we're not used to thinking

[6] It seems a little backward to write $f : A \to B$, $g : B \to C$ and then write $g \circ f$ instead of $f \circ g$. You can thank Euler for that. He decided to write functions with their input on the right instead of the left. If we wrote functions backwards, like $((x)f)g$ for "g of f of x," they we could just *follow the arrows* and life would be simpler.

carefully, and many of our favorite functions that we're used to calling "inverse functions" are actually only inverses when paired with specific domains.

1.2 Proof

Mathematics has the highest standard of proof of any field. In the Platonic ideal of mathematics, we start from some basic assumptions, called *axioms*, that we have all agreed upon. Then from those axioms, using the rules of logic, we deduce *theorems*. Every single mathematical statement we make can be traced back from theorem to theorem and eventually to our initial axioms.

This is contrary to other disciplines, like physics. In physics, based on observation, we construct *laws*. Laws in physics are like axioms in mathematics, but they have an important difference—they can be disproven by observation. A mathematical axiom can never be disproven. One can certainly argue that an axiom is not *useful* or not *interesting*, but you cannot say it's *wrong*.[7] Of course, as human practitioners, we may misuse logic and be wrong ourselves, but that is no fault of the axioms.

But now, let's deviate from philosophical perfection and visit reality. In reality, *mathematics is a human pursuit to understand relationships between ideas and their consequences*. The key there is that *humans* do mathematics to *understand* relationships. If a theorem in math can ultimately be reduced to logical statements about axioms, but the argument is 100000 steps long, it doesn't help a human understand why something is true. Instead, a shorter argument that skips over some steps is more useful to us. And, indeed, most of our mathematics to date skips over some steps.[8]

We call a correct mathematical argument a *proof*. A proof starts from a set of assumptions, and following the rules of logic, arrives at a conclusion. Strictly speaking, a proof doesn't need to make sense or show motivation, applications, or examples. It just has to be a sequence of correct logical steps. However, for us, as humans studying mathematics, we prove things for two reasons: to understand why things are true and to avoid making mistakes.

Reconciling these two goals can be very hard for a novice mathematician. If you include *all* the steps, it won't help with understanding, but if you don't include enough steps, the argument may not be convincing and might contain mistakes. Even professionals struggle to balance these competing goals, and how you balance those goals depends on your audience—if you're trying to convince your math professor of something your proof will need to have more detail than if you were trying to convince your friend (mathematicians are very skeptical!).

Enough talk, let's go through a 2000-year-old example of a proof.

Theorem 1.2.1 There is no rational number p/q such that $(p/q)^2 = 2$.

[7] There are multiple ways to axiomatize geometry. In Euclidean geometry every pair of lines either coincides, intersects in exactly one place, or does not intersect. In spherical geometry, every pair of lines either coincides or intersects in exactly two places. Euclidean geometry is useful when your space looks flat. Spherical geometry is useful when your space is the surface of a sphere (like the Earth). Is one of these more *right* than the other? They're certainly contradictory.

[8] There are some projects to prove all of mathematics directly from the axioms using computer assistance. They've made progress, but there are still theorems in calculus that have not been reduced to the axioms. We believe that they *could be* reduced to the axioms, but no one has taken the time to do so.

Proof. If p/q is a rational number, it can be expressed in lowest terms. Suppose p/q is in lowest terms and $(p/q)^2 = 2$. Then $p^2 = 2q^2$ and so p^2 is even. Since p^2 is even, it must be that p is even, and so by definition, $p = 2m$ for some integer m. Now,

$$\frac{p^2}{q^2} = \frac{(2m)^2}{q^2} = \frac{4m^2}{q^2} = 2,$$

with the last equality following by assumption. Multiplying both sides by q^2 and dividing by 2 we arrive at the equation

$$2m^2 = q^2,$$

and so q^2 is even which means q is even. By definition, this means $q = 2n$ for some integer n. But now,

$$\frac{p}{q} = \frac{2m}{2n}$$

is not in lowest terms! This is a contradiction and so it cannot be that $(p/q)^2 = 2$. ■

This is nearly identical to the argument the ancient Greeks gave. It's elegant, beautiful, and convincing. But, if we look closer, it does skip some steps. For example, it relies on the fact that there is such a thing as *lowest terms*. This is something that would need to be proven—a priori, the conclusion of the proof could be that the assumption that p/q could be in lowest terms is false.

You will not, overnight, become a master at understanding what steps you can leave out and what steps you must show. However, with feedback, you'll get better. For a detailed guide on writing good proofs, please see Appendix A.

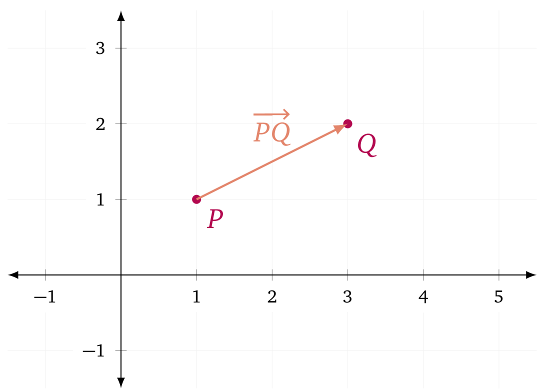
Chapter 2

Vectors

A *vector* is characterized by a *magnitude* and a *direction*, and using vectors, we can mathematically describe more situations than with regular numbers. For example, when driving a car, it may be sufficient to know your speed, which can be described by a single number, but the motion of an airplane must be described by a vector quantity—velocity—which takes into account its direction as well as its speed.

Ordinary numerical quantities are called *scalars* when we want to emphasize that they are not vectors.

Whereas numbers allow us to specify relationships between single quantities (put in twice as much flour as sugar), vectors will allow us to specify relationships between geometric objects in space.[1] If we have two points, $P = (1,1)$ and $Q = (3,2)$, we specify the *displacement* from P to Q as a vector.



We notate the displacement vector form P to Q by $\overrightarrow{PQ}$. The magnitude of $\overrightarrow{PQ}$ is given by the Pythagorean theorem to be $\sqrt{5}$ and its direction is specified by the directed line segment from P to Q.

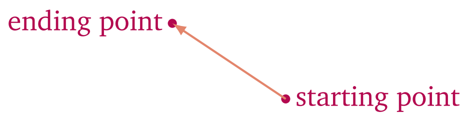
2.1 Vector Notation

There are many ways to represent vector quantities in writing. If we have two points, P and Q, we write $\overrightarrow{PQ}$ to represent the vector from P to Q. Absent of points, bold-faced letters or a letter with an arrow over it are the most common typographical representations of vectors. For example, $\vec{a}$ or $\mathbf{a}$ may both be used to represent the vector quantity named "a." In this book we will use $\vec{a}$ to represent a vector. The notation $||\vec{a}||$ represents the magnitude of the vector $\vec{a}$, which is sometimes called the *norm* or *length* of $\vec{a}$.

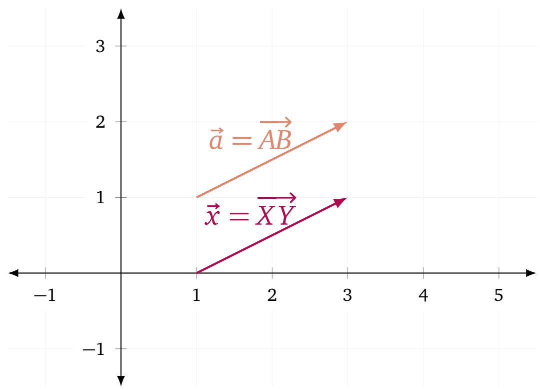
Definition 2.1.1 — Norm. The *norm* of a vector $\vec{a}$, notated $||\vec{a}||$, is the magnitude of $\vec{a}$.

Graphically, vectors are represented as directed line segments (a line segment with an arrow at one end), however, vectors are fundamentally directions and magnitudes. Since any directed line segment contains a starting point and an ending point, a directed line segment contains extra information that is not actually part of the vector.

[1] Though in this book we will treat vectors as objects in Euclidean space, but they are much more general. For instance, someone's internet browsing habits could be described by a vector—the topics they find most interesting might be the "direction" and the amount of time they browse might be the "magnitude."



For example, let $A = (1,1)$, $B = (3,2)$, $X = (1,0)$, and $Y = (3,1)$ and consider the vectors $\vec{a} = \overrightarrow{AB}$ and $\vec{x} = \overrightarrow{XY}$.



Are these vectors the same or different vectors? As directed line segments, they are different because they are at different locations in space. However, both $\vec{a}$ and $\vec{x}$ have the same magnitude and direction. Thus, $\vec{a} = \vec{x}$ despite the fact that $A \neq X$.[2]

Vectors and Points

The distinction between vectors and points is sometimes nebulous because they are so closely related to each other. A *point* in Euclidean space specifies an absolute position whereas a vector specifies a magnitude and direction. However, given a point P, one associates P with the vector $\vec{p} = \overrightarrow{OP}$, where O is the origin. Similarly, we associate the vector $\vec{v}$ with the point V so that $\overrightarrow{OV} = \vec{v}$. Thus, we have a way to unambiguously go back and forth between vectors and points.[3] As such, *we will treat vectors and points as interchangeable*.

2.2 Vector Arithmetic

Vectors provide a natural way to give directions. For example, suppose $\vec{e}_1$ points one mile eastwards and $\vec{e}_2$ points one mile northwards. Now, if you were standing at the origin and wanted to move to a location 3 miles east and 2 miles north, you might say: “Walk 3 times the length of $\vec{e}_1$ in the $\vec{e}_1$ direction and 2 times the length of $\vec{e}_2$ in the $\vec{e}_2$ direction.” Mathematically, we express this as

$$3\vec{e}_1 + 2\vec{e}_2.$$

[2] Some theories use *rooted vectors* instead of vectors as the fundamental object of study. A rooted vector represents a magnitude, direction, *and* a starting point. And, as rooted vectors, $\vec{a} \neq \vec{x}$ (from the example above). But for us, vectors will always be unrooted, even though our graphical representations of vectors might appear rooted.

[3] Mathematically, we say there is an *isomorphism* between vectors and points.

Of course, we've incidentally described a new vector. Namely, let P be the point at 3-east and 2-north. Then

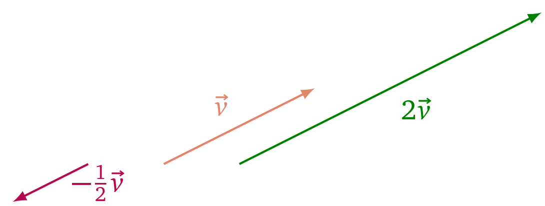
$$\overrightarrow{OP} = 3\vec{e}_1 + 2\vec{e}_2.$$

If the vector $\vec{r}$ points north but has a length of 10 miles, we have a similar formula:

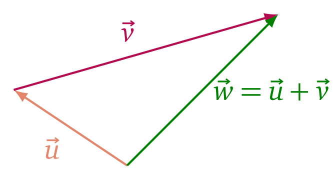
$$\overrightarrow{OP} = 3\vec{e}_1 + \tfrac{1}{5}\vec{r},$$

and we have the relationship $\vec{r} = 10\vec{e}_2$. Our notation here is very suggestive. Indeed, if we could make sense of what $\alpha\vec{v}$ is for any scalar α and vector $\vec{v}$, and we could make sense of what $\vec{v}+\vec{w}$ means for any vectors $\vec{v}$ and $\vec{w}$, we would be able to do algebra with vectors. We might even say we have *an algebra of vectors*.

Intuitively, for a vector $\vec{v}$ and a scalar $\alpha > 0$, the vector $\vec{w} = \alpha\vec{v}$ should point in the same direction as $\vec{v}$ but have magnitude scaled up by α. That is, $\|\vec{w}\| = \alpha\|\vec{v}\|$. Similarly, $-\vec{v}$ should be the vector of the same length as $\vec{v}$ but pointing in the exact opposite direction.



For two vectors $\vec{u}$ and $\vec{v}$, the sum $\vec{w} = \vec{u} + \vec{v}$ should be the displacement vector created by first displacing along $\vec{u}$ and then displacing along $\vec{v}$.



Now, there is one snag. What should $\vec{v} + (-\vec{v})$ be? Well, first we displace along $\vec{v}$ and then we displace in the exact opposite direction by the same amount. So, we have gone nowhere. This corresponds to a displacement with zero magnitude. But, what direction did we displace? Here we make a philosophical stand.

Definition 2.2.1 — Zero Vector. The *zero vector*, notated as $\vec{0}$, is the vector with no magnitude.

We will be pragmatic about the direction of the zero vector and say, *the zero vector does not have a well-defined direction*.[4] That means sometimes we consider the zero vector to point in every direction and sometimes we consider it to point in no directions. It depends on our mood—but we must never talk about *the* direction of the zero vector, since it's not defined.

We need the zero vector if we are to make precise mathematical sense of vector arithmetic. Further along this line of thinking, we can define precisely how vector arithmetic should behave.

[4] In the mathematically precise definition of vector, the idea of "magnitude" and "direction" are dropped. Instead, a set of vectors is defined to be a set over which you can reasonably define addition and scalar multiplication.

Specifically, if $\vec{u}$, $\vec{v}$, $\vec{w}$ are vectors and α and β are scalars, the following conditions should be satisfied:

$$(\vec{u}+\vec{v})+\vec{w}=\vec{u}+(\vec{v}+\vec{w}) \qquad \text{(Associativity)}$$
$$\vec{u}+\vec{v}=\vec{v}+\vec{u} \qquad \text{(Commutativity)}$$
$$\alpha(\vec{u}+\vec{v})=\alpha\vec{u}+\alpha\vec{v} \qquad \text{(Distributivity)}$$

and

$$(\alpha\beta)\vec{v}=\alpha(\beta\vec{v}) \qquad \text{(Associativity II)}$$
$$(\alpha+\beta)\vec{v}=\alpha\vec{v}+\beta\vec{v} \qquad \text{(Distributivity II)}$$

Indeed, if we intuitively think about vectors in flat (Euclidean) space, all of these properties are satisfied.[5] From now on, these properties of vector operations will be considered the *laws (or axioms) of vector arithmetic*.

We'll be talking about these vector operations (scalar multiplication and vector addition) a lot. So much so that the concept is worth naming.

Definition 2.2.2 — Linear Combination. A *linear combination* of the vectors $\vec{v}_1,\dots,\vec{v}_n$ is any vector expressible as

$$\alpha_1\vec{v}_1+\cdots+\alpha_n\vec{v}_n,$$

where $\alpha_1,\alpha_2,\dots,\alpha_n$ are scalars.

We've given laws for linear combinations of vectors, but what about for magnitudes of vectors? We'd like the magnitude (or norm) of a vector to obey the following laws.

$$||\vec{v}||\geq 0 \qquad \text{(Non-negativity)}$$
$$||\vec{v}||=0 \text{ only when } \vec{v}=\vec{0} \qquad \text{(Definiteness)}$$
$$||\alpha\vec{v}||=|\alpha|||\vec{v}|| \qquad \text{(Homogeneity)}$$
$$||\vec{v}+\vec{w}||\leq||\vec{v}||+||\vec{w}|| \qquad \text{(Triangle Inequality)}$$

for all $\vec{v}$, $\vec{w}$, and scalars α. Any function on vectors satisfying those four properties is called a *norm*, and our usual notion of length in three-dimensional space indeed obeys those properties.[6]

Homogeneity is a particularly special property of a norm. It allows us to easily create *unit vectors*.

Definition 2.2.3 — Unit Vector. A *unit vector* is a vector $\vec{u}$ satisfying $||\vec{u}||=1$.

Unit vectors are handy because if $\vec{u}$ is a unit vector, then $k\vec{u}$ has length $|k|$. Further, we can always turn a vector into a unit vector.

[5] If we deviate from flat space, some of these rules are no longer respected. Consider moving 100 miles north then 100 miles east on a sphere. Is this the same as moving 100 miles east and then 100 miles north?

[6] The Euclidean norm comes from the Pythagorean theorem $a^2+b^2=c^2$. However, by changing the exponent, we have a whole family of norms coming from the equations $|a|^p+|b|^p=|c|^p$.

■ **Example 2.1** The vector $\vec{v}/||\vec{v}||$ when $\vec{v} \neq \vec{0}$ is always a unit vector in the direction of $\vec{v}$. Computing,

$$\left\|\frac{\vec{v}}{||\vec{v}||}\right\| = \left|\frac{1}{||\vec{v}||}\right| ||\vec{v}|| = \frac{1}{||\vec{v}||}||\vec{v}|| = 1.$$

■

Exercises for 2.2

1 Find the displacement vector from the point $P(7,2,9)$ to the point $Q(-2,1,4)$ in the form $a\vec{e}_1 + b\vec{e}_2 + c\vec{e}_3$.

Solution: $-9\vec{e}_1 - 1\vec{e}_2 - 5\vec{e}_3$

2[OS] For the following exercises, consider points $P(-1,3)$, $Q(1,5)$, and $R(-3,7)$. Determine the requested vectors and express each of them i. in component form and ii. by using the standard unit vectors.

a) $\overrightarrow{PQ}$

b) $\overrightarrow{PR}$

c) $\overrightarrow{QP}$

d) $\overrightarrow{RP}$

Solution:

a) $\overrightarrow{PQ} = (2,2) = 2\vec{e}_1 + 2\vec{e}_2$.

b) $\overrightarrow{PR} = (-2,4) = -2\vec{e}_1 + 4\vec{e}_2$.

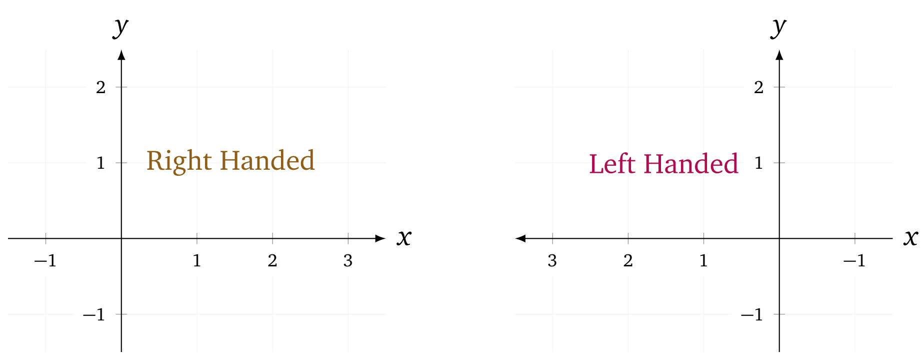
c) $\overrightarrow{QP} = (-2,-2) = -2\vec{e}_1 - 2\vec{e}_2$.

d) $\overrightarrow{RP} = (2,-4) = 2\vec{e}_1 - 4\vec{e}_2$.

2.3 Coordinates

A rectangular coordinate system in the plane is specified by choosing an origin O and then choosing two perpendicular axes meeting at the origin. These axes are chosen in some order so that we know which axis (usually the x-axis) comes first and which (usually the y-axis) second. Note that there are many different coordinate systems which could be used although we often draw pictures as if there were only one.

In physics, one often has to think carefully about the coordinate system because choosing one appropriately may greatly simplify the analysis. Note that axes for coordinate systems are usually drawn with *right-hand orientation*, where the right angle from the positive x-axis to the positive y-axis is in the counter-clockwise direction. However, it would be equally valid to use the *left-hand orientation* in which that angle is in the clockwise direction. One can easily switch the orientation of a coordinate system by reversing one of the axes.[7]



For any coordinate system, there are special vectors associated with it. For the plane, the vector pointing one unit along the positive x-axis is called $\vec{e}_1$ and the vector pointing one unit along the positive y-axis is called $\vec{e}_2$. The vectors $\vec{e}_1$ and $\vec{e}_2$ are called the *standard basis* vectors for $\mathbb{R}^2$.

Notice that every point (or vector) in the plane can be represented as a linear combination of $\vec{e}_1$ and $\vec{e}_2$, and the vector $\alpha\vec{e}_1+\beta\vec{e}_2$ is the vector $\overrightarrow{OP}$ where $P=(\alpha,\beta)$. Now, to state an intuitive fact: if $\vec{w}$ is a vector in the plane, *there is only one way to write a vector as a linear combination of $\vec{e}_1$ and $\vec{e}_2$*. This means, if $\vec{w}=\alpha\vec{e}_1+\beta\vec{e}_2$, the pair (α,β) captures all information[8] about $\vec{w}$.

For a vector $\vec{w}=\alpha\vec{e}_1+\beta\vec{e}_2$, we call the pair (α,β) the *coordinates* of the vector $\vec{w}$. There are many equivalent notations used to represent a vector in coordinates.

[7] The concept of orientation is quite fascinating and it arises in mathematics, physics, chemistry, and even biology in many interesting ways. Note that almost all of us base our intuitive concept of orientation on our inborn notion of "right" versus "left".

[8] Maybe you already knew this because the point (α,β) is described by the pair of numbers (α,β), duh! But consider, what would we do if we didn't know about coordinates at all? One approach is to *define* coordinates in terms of vectors, which is really what we're doing.

(α, β)	parenthesis
$\langle\alpha, \beta\rangle$	angle brackets
$\begin{bmatrix}\alpha & \beta\end{bmatrix}$	square brackets in a row (a row matrix)
$\begin{bmatrix}\alpha \\ \beta\end{bmatrix}$	square brackets in a column (a column matrix)

Given what we now know about representing vectors and their equivalency with points, we can dissect the notation $\mathbb{R}^2$. On the one hand, $\mathbb{R}^2$ is the set of vectors in two-dimensional Euclidean space. On the other hand $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ is the set of all pairs of real numbers. Via the use of coordinates, we know these concepts represent the same thing! Further, since vectors in $\mathbb{R}^2$ are equivalent to their representation in coordinates, we will often write

$$\vec{v} = \begin{bmatrix}\alpha \\ \beta\end{bmatrix}$$

as a shorthand for $\vec{v} = \alpha\vec{e}_1 + \beta\vec{e}_2$.

Breaking vectors into coordinates, and in particular, viewing vectors as linear combinations of the standard basis vectors, allows us to solve problems that were difficult before. For instance, suppose we have vectors $\vec{v}$ and $\vec{w}$. How can we compute $\|\vec{v} + \vec{w}\|$? With coordinates, it's easy.

■ **Example 2.2** Suppose $\vec{v} = \alpha_1\vec{e}_1 + \beta_1\vec{e}_2$ and $\vec{w} = \alpha_2\vec{e}_1 + \beta_2\vec{e}_2$. By the laws of vector arithmetic we have

$$\vec{v} + \vec{w} = (\alpha_1\vec{e}_1 + \beta_1\vec{e}_2) + (\alpha_2\vec{e}_1 + \beta_2\vec{e}_2) = (\alpha_1 + \alpha_2)\vec{e}_1 + (\beta_1 + \beta_2)\vec{e}_2.$$

Now, since $\vec{e}_1$ and $\vec{e}_2$ are orthogonal to each other, the Pythagorean theorem gives

$$\|\vec{v} + \vec{w}\| = \sqrt{(\alpha_1 + \alpha_2)^2 + (\beta_1 + \beta_2)^2}.$$

■

Writing things in terms of the standard basis allowed us to make easy work of computing $\|\vec{v} + \vec{w}\|$ in Example 2.2. We can use the laws of vector arithmetic to produce rules for working with components.

The rules are are likely familiar:

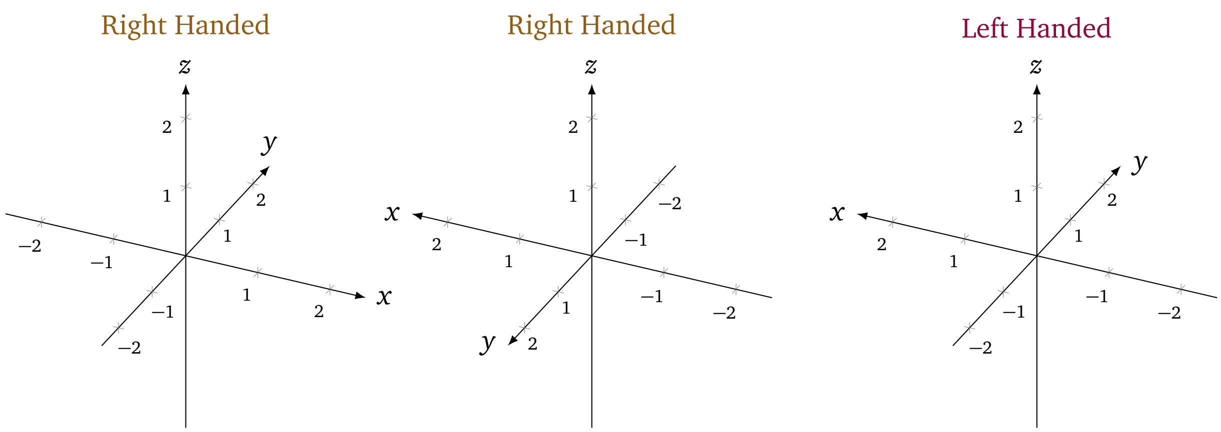
$$\begin{bmatrix}a \\ b\end{bmatrix} + \begin{bmatrix}\alpha \\ \beta\end{bmatrix} = \begin{bmatrix}a + \alpha \\ b + \beta\end{bmatrix} \qquad \text{and} \qquad \alpha\begin{bmatrix}a \\ b\end{bmatrix} = \begin{bmatrix}\alpha a \\ \alpha b\end{bmatrix}.$$

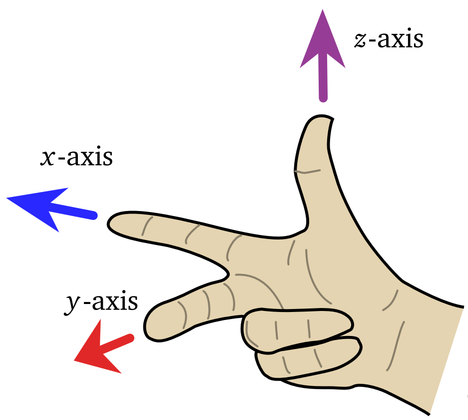
Exercise 2.3 Prove the rules for adding the component representation of vectors and multiplying the component representation of vectors directly from the laws of vector arithmetic.

Armed with these rules, we will be able to tackle sophisticated vector problems.

Three-dimensional Coordinates

In three-dimensional space, the story is very similar. Again, we imagine three perpendicular axes, the x, y, and z axes. To draw consistent pictures, we have a notion of a right-handed three-dimensional coordinate system given by the *right-hand rule*.





[9]

We now have three standard basis vectors, $\vec{e}_1$, $\vec{e}_2$, and $\vec{e}_3$, each pointing one unit in the positive direction of their respective axes. Any vector in three-dimensional space can be represented in exactly one way as a linear combination $\alpha\vec{e}_1 + \beta\vec{e}_2 + \gamma\vec{e}_3$. Thus, vectors in three-dimensional space, notated $\mathbb{R}^3$, are synonymous with triplets (α, β, γ) of real numbers. With some clever geometry, we deduce

$$\|\alpha\vec{e}_1 + \beta\vec{e}_2 + \gamma\vec{e}_3\| = \sqrt{\alpha^2 + \beta^2 + \gamma^2}.$$

Historically, three-dimensional space has been studied a lot and there are several notations for the standard basis vectors still in use.

The following is a non-exhaustive list.

$\hat{\mathbf{x}}$	$\hat{\mathbf{y}}$	$\hat{\mathbf{z}}$
$\hat{\imath}$	$\hat{\jmath}$	$\hat{k}$
$\mathbf{i}$	$\mathbf{j}$	$\mathbf{k}$
$\vec{e}_1$	$\vec{e}_2$	$\vec{e}_3$

Keep these notations in the back of your mind. You might see them in other classes.

[9] Image credit: Acdx, from Wikipedia https://en.wikipedia.org/wiki/Cross_product

Higher dimensions

One can't progress very far in the study of science and mathematics without encountering a need for higher dimensional "vectors." For example, physicists have known since Einstein that the physical universe is best thought of as a four-dimensional entity called spacetime in which time plays a role close to that of the three spatial coordinates. Since we don't have any way to deal with $\mathbb{R}^n$ intuitively, we must proceed by analogy with two and three dimensions. The easiest way to proceed is to generalize the idea of a standard basis. From there, we can represent vectors in $\mathbb{R}^n$ as n-tuples of real numbers. We then define

$$\|(x_1, x_2, \ldots, x_n)\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}.$$

We've now unified our theory of vectors across all integer dimensions $n > 0$. The case $n = 1$ yields "geometry" on a line, the cases $n = 2$ and $n = 3$ geometry in the plane and in space, and the case $n = 4$ yields the geometry of "4-vectors" which are used in the special theory of relativity. Larger values of n are used in a variety of contexts, some of which we shall encounter later.

Exercises for 2.3

1 Find $\|\vec{a}\|$, $5\vec{a}-2\vec{b}$, and $-3\vec{b}$ for each of the following vector pairs.

a) $\vec{a}=2\vec{e}_1+3\vec{e}_2$, $\vec{b}=4\vec{e}_1-9\vec{e}_2$

b) $\vec{a}=(1,2,-1)$, $\vec{b}=(2,-1,0)$

Solution:

a) $\|\vec{a}\|=\sqrt{13}$, $5\vec{a}-2\vec{b}=2\vec{e}_1+33\vec{e}_2$, $-3\vec{b}=-12\vec{e}_1+27\vec{e}_2$.

b) $\|\vec{a}\|=\sqrt{6}$, $5\vec{a}-2\vec{b}=(1,12,-5)$, $-3\vec{b}=(-6,3,0)$.

2 Let $P=(7,2,9)$ and $Q=(-2,1,4)$. Find $\overrightarrow{PQ}$ as a linear combination of $\vec{e}_1$, $\vec{e}_2$, and $\vec{e}_3$.

Solution: $\overrightarrow{PQ}=-9\vec{e}_1-\vec{e}_2-5\vec{e}_3$.

3 Find unit vectors with the same direction as the vectors a) $(-4,-2)$, b) $3\vec{e}_1+5\vec{e}_2$, c) $(1,3,-2)$.

Solution:

a) $\frac{1}{\sqrt{5}}(-2,-1)$

b) $\frac{1}{\sqrt{34}}(3,5)$

c) $\frac{1}{\sqrt{14}}(1,3,-2)$

4 Show by direct calculation that the rule $\overrightarrow{AB}+\overrightarrow{BC}+\overrightarrow{CA}=\vec{0}$ holds for the three points $A(2,1,0)$, $B(-4,1,3)$, and $C(0,12,0)$. Can you prove the general rule for any three points in space?

Solution: In the specific case, $\overrightarrow{AB}=(-6,0,3)$, $\overrightarrow{BC}=(4,11,-3)$, $\overrightarrow{CA}=(2,-11,0)$, so that

$$\overrightarrow{AB}+\overrightarrow{BC}+\overrightarrow{CA}=\begin{bmatrix}-6+4+2\\0+11-11\\3-3+0\end{bmatrix}=\begin{bmatrix}0\\0\\0\end{bmatrix}$$

More generally, suppose that we have the points $A(a_1,a_2,a_3)$, $B(b_1,b_2,b_3)$, and $C(c_1,c_2,c_3)$. Then $\overrightarrow{AB}=(b_1-a_1,b_2-a_2,b_3-a_3)$, $\overrightarrow{BC}=(c_1-b_1,c_2-b_2,c_3-b_3)$, and $\overrightarrow{CA}=(a_1-c_1,a_2-c_2,a_3-c_3)$. Thus,

$$\overrightarrow{AB}+\overrightarrow{BC}+\overrightarrow{CA}=\begin{bmatrix}(b_1-a_1)+(c_1-b_1)+(a_1-c_1)\\(b_2-a_2)+(c_2-b_2)+(a_2-c_2)\\(b_3-a_3)+(c_3-b_3)+(a_3-c_3)\end{bmatrix}=\begin{bmatrix}0\\0\\0\end{bmatrix}$$

5 If an airplane flies with apparent velocity $\vec{v}_a$ relative to air, and the wind velocity is denoted $\vec{w}$, then the planes true velocity relative to the ground, is $\vec{v}_g=\vec{v}_a+\vec{w}$. Draw the diagram to assure yourself of this.

a) A farmer wishes to fly his crop duster at 80 km/h north over his fields. If the weather vane atop the barn shows easterly winds at 10 km/h, what should his apparent velocity, $\vec{v}_a$ be?

b) What if the wind were northeasterly? Southeasterly?

Solution:

a) $10\sqrt{65}$ km/h, $\tan^{-1}\left(\frac{1}{8}\right)^\circ$ east of north.

b) for a northeasterly wind: $2\sqrt{\left(40+\frac{5\sqrt{2}}{2}\right)^2+\frac{25}{2}}$ at $\tan^{-1}\left(\frac{5\sqrt{2}}{80+5\sqrt{2}}\right)^\circ$ east of north

for a southeasterly wind: $2\sqrt{\left(40-\frac{5\sqrt{2}}{2}\right)^2+\frac{25}{2}}$ at $\tan^{-1}\left(\frac{5\sqrt{2}}{80-5\sqrt{2}}\right)^\circ$ east of north

6 Suppose a right handed coordinate system has been set up in space. What happens to the orientation of the coordinate system if you make the following changes?

a) Change the direction of one axis.

b) Change the direction of two axes.

c) Change the direction of all three axes.

d) Interchange the x and y-axes.

Solution:

a) the orientation changes.

b) the orientation stays the same.

c) the orientation changes.

d) the orientation changes.

7 In body-centered crystals, a large atom (assumed spherical) is surrounded by eight smaller atoms. If the structure is placed in a "box" that just contains the large atom, the smaller atoms each occupy a corner. If the central atom has radius R, what is the greatest atomic radii the smaller atoms may have?

Solution: $(2-\sqrt{3})R$

8 Prove that the diagonals of a parallelogram bisect each other. (Hint: show that the position vectors from the origin to the midpoints are equal).

Solution: Suppose that the parallelogram is defined by two vectors $\vec{a}$ and $\vec{b}$. Then the first diagonal (which originates from the origin) is given by $\vec{d}_1 = \vec{a} + \vec{b}$ and the second diagonal (which originates from the tip of $\vec{b}$) is given by $\vec{d}_2 = \vec{a} - \vec{b}$. Then, the position of the midpoint of the first diagonal is given by $\vec{d}_1/2 = 1/2(\vec{a} + \vec{b})$ and the position of the midpoint of the second diagonal is given by $\vec{b} + \vec{d}_2/2 = \vec{b} + 1/2(\vec{a} - \vec{b}) = 1/2(\vec{a} + \vec{b})$. Thus, the position vectors of the midpoints are equal, and so the diagonals must bisect each other.

2.4 Lines and Planes

With a handle on vectors, we can now use them to describe some common geometric objects: lines and planes.

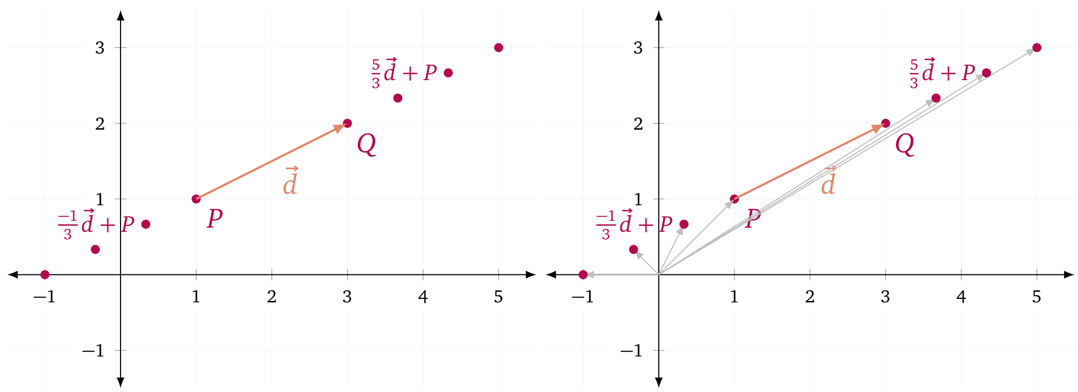
Lines

Consider for a moment the line ℓ through the points P and Q. When $P, Q \in \mathbb{R}^2$, we can describe ℓ with an equation of the form $y = mx + b$ (provided it isn't a vertical line), but if $P, Q \in \mathbb{R}^3$, it's much harder to describe ℓ with an equation. Using vectors provides an easier way.

Let $\vec{d} = \overrightarrow{PQ}$ and consider the set of points (or vectors) $\vec{x}$ that can be expressed as

$$\vec{x} = t\vec{d} + P$$

for $t \in \mathbb{R}$. Geometrically, this is the set of all points we get by starting at P and displacing by some multiple of $\vec{d}$. This is a line!



We simultaneously interpret this line as a set of points (the points that make up the line) and as a set of vectors rooted at the origin (the vectors pointing from the origin to the line). Note that sometimes we draw vectors as directed line segments from the origin. Other times, drawing drawing vectors as line segments makes it hard to see what is going on, and so it is better to draw each vector by marking only its ending point.

The line ℓ described above can be written in set-builder notation as:

$$\ell = \{\vec{x} \,:\, \vec{x} = t\vec{d} + P \text{ for some } t \in \mathbb{R}\}.$$

Notice that in set-builder notation, we write "for some $t \in \mathbb{R}$." Make sure you understand why replacing "for some $t \in \mathbb{R}$" with "for all $t \in \mathbb{R}$" would be incorrect.

Writing lines with set-builder notation all the time can be overkill, so we will allow ourselves to describe lines in a shorthand called *vector form*.[10]

[10] $y = mx + b$ form of a line is also shorthand. The line ℓ described by the equation $y = mx + b$ is actually the set $\{(x, y) \in \mathbb{R}^2 \,:\, y = mx + b\}$.

Definition 2.4.1 — Vector form of a Line. Let ℓ be a line and let $\vec{d}$ and $\vec{p}$ be vectors such that $\ell = \{\vec{x} : \vec{x} = t\vec{d} + \vec{p} \text{ for some } t \in \mathbb{R}\}$. Then, ℓ may be described in *vector form* as

$$\vec{x} = t\vec{d} + \vec{p}.$$

We call $\vec{d}$ a *direction vector* for ℓ and the equation $\vec{x} = t\vec{d} + \vec{p}$ a *vector equation* or *vector form* of ℓ.

Note that if $\vec{x} = t\vec{d} + \vec{p}$ is the vector equation of a line ℓ, by setting $t = 0$ we necessarily have $\vec{p} \in \ell$. The converse is true, too. If $\vec{q} \in \ell$ and $\vec{d}$ is a direction for ℓ, then ℓ may be expressed in vector form as $\vec{x} = t\vec{d} + \vec{p}$.

The direction of a line is easily obtained by finding the displacement vector between two points on the line. Thus, given a line in another form, computing its vector form is straightforward.

■ **Example 2.4** Find vector form of the line $\ell \subseteq \mathbb{R}^2$ with equation $y = 2x + 3$.

First, we find two points on the line. By guess-and-check we see $P = (0,3)$ and $Q = (1,5)$ are on ℓ. Thus, a direction vector for ℓ is given by

$$\vec{d} = (1,5) - (0,3) = (1,2).$$

We may now write the vector equation of ℓ as

$$\vec{x} = t\vec{d} + P$$

or, in components,

$$\begin{bmatrix} x \\ y \end{bmatrix} = t\begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 0 \\ 3 \end{bmatrix}.$$

■

It's important to note that when we write a line in vector form, it is a *specific shorthand* notation. If we augment the notation, we no longer have written a line in "vector form."

■ **Example 2.5** Let ℓ and let $\vec{d}$ be a direction vector for ℓ and let $\vec{p} \in \ell$ be a point on ℓ. Writing

$$\vec{x} = t\vec{d} + \vec{p}$$

or

$$\vec{x} = t\vec{d} + \vec{p} \quad \text{where} \quad t \in \mathbb{R}$$

specifies ℓ in vector form and are both shorthands for $\{\vec{x} : \vec{x} = t\vec{d} + \vec{p} \text{ for some } t \in \mathbb{R}\}$. But,

$$\vec{x} = t\vec{d} + \vec{p} \quad \text{for some} \quad t \in \mathbb{R}$$

and

$$\vec{x} = t\vec{d} + \vec{p} \quad \text{for all} \quad t \in \mathbb{R}$$

are logical statements about the vectors $\vec{x}$, $\vec{d}$, and $\vec{p}$. These statements are either true or false; they do *not* specify ℓ in vector form.

Similarly, the statement

$$\ell = t\vec{d} + \vec{p}$$

is mathematically nonsensical and does not specify ℓ in vector form. (On the left is a *set* and on the right is a *vector*!)

■

The downside of writing lines in vector form is that there are multiple direction vectors and multiple points for every line. Thus, merely by looking at the vector equation for two lines, it can be hard to tell if they're equal.

For example,

$$\begin{bmatrix} x \\ y \end{bmatrix} = t\begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 0 \\ 3 \end{bmatrix}, \qquad \begin{bmatrix} x \\ y \end{bmatrix} = t\begin{bmatrix} 2 \\ 4 \end{bmatrix} + \begin{bmatrix} 0 \\ 3 \end{bmatrix}, \quad \text{and} \quad \begin{bmatrix} x \\ y \end{bmatrix} = t\begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 1 \\ 5 \end{bmatrix}$$

all represent the same line. In the second equation, the direction is parallel but scaled, and in the third equation, a different point on the line was chosen.

In vector form, the variable t is called the *parameter variable*. It is an instance of a *dummy variable*; that is, it is mostly there as a placeholder. Remember, vector form is shorthand for a set described in set-builder notation.

Let $\vec{d}_1, \vec{d}_2 \neq \vec{0}$ and $\vec{p}_1, \vec{p}_2$ be vectors and define the lines

$$\ell_1 = \{\vec{x} \,:\, \vec{x} = t\vec{d}_1 + \vec{p}_1 \text{ for some } t \in \mathbb{R}\}$$

$$\ell_2 = \{\vec{x} \,:\, \vec{x} = t\vec{d}_2 + \vec{p}_2 \text{ for some } t \in \mathbb{R}\}.$$

These lines have vector equations $\vec{x} = t\vec{d}_1 + \vec{p}_1$ and $\vec{x} = t\vec{d}_2 + \vec{p}_2$. However, declaring that $\ell_1 = \ell_2$ if and only if $t\vec{d}_1 + \vec{p}_1 = t\vec{d}_2 + \vec{p}_2$ does *not* make sense. Instead $\ell_1 = \ell_2$ if $\ell_1 \subseteq \ell_2$ and $\ell_2 \subseteq \ell_1$. If $\vec{x} \in \ell_1$ then $\vec{x} = t\vec{d}_1 + \vec{p}_1$ for some $t \in \mathbb{R}$. If $\vec{x} \in \ell_2$ then $\vec{x} = t\vec{d}_2 + \vec{p}_2$ for some *possibly different* $t \in \mathbb{R}$. This can get confusing really quickly. The easiest solution is to use different parameter variables if we want to compare lines in vector form.

■ **Example 2.6** Determine if the lines ℓ_1 and ℓ_2, represented in vector form by the equations

$$\vec{x} = t\begin{bmatrix} 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 2 \\ 1 \end{bmatrix} \qquad \text{and} \qquad \vec{x} = t\begin{bmatrix} 2 \\ 2 \end{bmatrix} + \begin{bmatrix} 4 \\ 3 \end{bmatrix}$$

are the same line. To determine this, we need to figure out if $\vec{x} \in \ell_1$ implies $\vec{x} \in \ell_2$ and if $\vec{x} \in \ell_2$ implies $\vec{x} \in \ell_1$.

If $\vec{x} \in \ell_1$, then $\vec{x} = t\begin{bmatrix} 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ for some $t \in \mathbb{R}$. If $\vec{x} \in \ell_2$, then $\vec{x} = s\begin{bmatrix} 2 \\ 2 \end{bmatrix} + \begin{bmatrix} 4 \\ 3 \end{bmatrix}$ for some $s \in \mathbb{R}$. Thus if

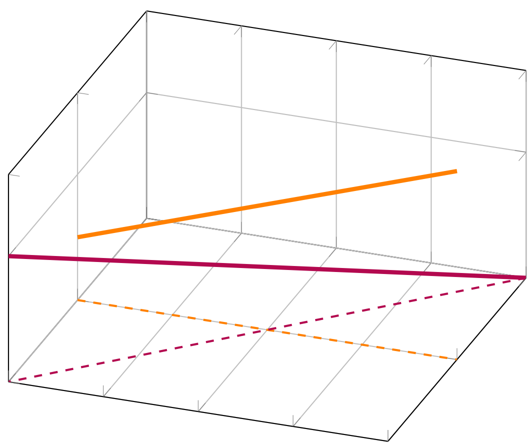
$$t\begin{bmatrix} 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \vec{x} = s\begin{bmatrix} 2 \\ 2 \end{bmatrix} + \begin{bmatrix} 4 \\ 3 \end{bmatrix}$$

always has a solution, $\ell_1 = \ell_2$. Moving everything to one side we see

$$\begin{aligned}\vec{0} = \begin{bmatrix}4\\3\end{bmatrix} - \begin{bmatrix}2\\1\end{bmatrix} + s\begin{bmatrix}2\\2\end{bmatrix} - t\begin{bmatrix}1\\1\end{bmatrix} &= \begin{bmatrix}2\\2\end{bmatrix} + s\begin{bmatrix}2\\2\end{bmatrix} - t\begin{bmatrix}1\\1\end{bmatrix}\\ &= (s+1)\begin{bmatrix}2\\2\end{bmatrix} - \tfrac{t}{2}\begin{bmatrix}2\\2\end{bmatrix}\\ &= (s+1-\tfrac{t}{2})\begin{bmatrix}2\\2\end{bmatrix}.\end{aligned}$$

This has a solution whenever $0 = s + 1 - t/2$. Since for every $t \in \mathbb{R}$ we can find an $s \in \mathbb{R}$ and for every $s \in \mathbb{R}$ we can find a $t \in \mathbb{R}$ satisfying this equation, we know $\ell_1 = \ell_2$. ■

The geometry of lines in space is a bit more complicated than that of lines in the plane. Lines in the plane either intersect or are parallel. In space, we have to be a bit more careful about what we mean by "parallel lines," since lines with entirely different directions can still fail to intersect.[11]



■ **Example 2.7** Consider the lines described by

$$\begin{aligned}\vec{x} &= t(1,3,-2) + (1,2,1)\\ \vec{x} &= t(-2,-6,4) + (3,1,0).\end{aligned}$$

They have parallel directions since $(-2,-6,4) = -2(1,3,-2)$. Hence, in this case, we say the lines are *parallel*. (How can we be sure the lines are not the same?) ■

■ **Example 2.8** Consider the lines described by

$$\begin{aligned}\vec{x} &= t(1,3,-2) + (1,2,1)\\ \vec{x} &= t(0,2,3) + (0,3,9).\end{aligned}$$

They are not parallel because neither of the direction vectors is a multiple of the other. They may or may not intersect. (If they don't, we say the lines are *skew*.) How can we find out?

[11] Recall that in Euclidean geometry two lines are defined to be parallel if they coincide or never intersect.

Mirroring our earlier approach, we can set their equations equal and see if we can solve for the point of intersection *after ensuring we give their parametric variables different names*. We'll keep one parametric variable named t and name the other one s. Thus, we want

$$\vec{x} = t(1,3,-2) + (1,2,1) = s(0,2,3) + (0,3,9),$$

which after collecting terms yields

$$(t+1, 3t+2, -2t+1) = (0, 2s+3, 3s+9).$$

Picking out the components yields three equations

$$\begin{aligned} t+1 &= 0 \\ 3t+2 &= 2s+3 \\ -2t+1 &= 3s+9 \end{aligned}$$

in 2 unknowns s and t. This is an *overdetermined* system, and it may or may not have a consistent solution. The first two equations yield $t = -1$ and $s = -2$. Putting these values in the last equation yields $(-2)(-1) + 1 = 3(-2) + 9$, which is indeed true. Hence, the equations are consistent, and the lines intersect. To find the point of intersection, put $t = -1$ in the equation for the first line (or $s = -2$ in that for the second) to obtain $(0, -1, 3)$. ■

Planes

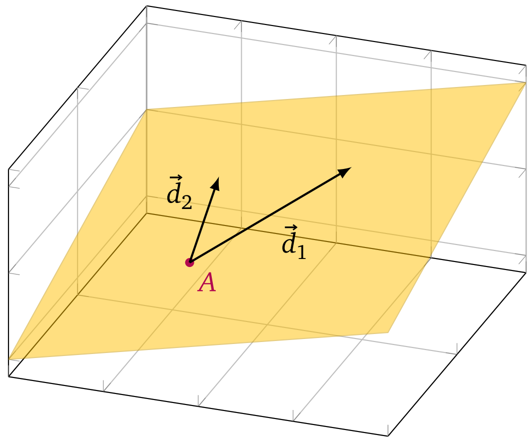
Any two distinct points define a line. To define a plane, we need three points. But there's a caveat: the three points cannot be on the same line, otherwise they'd define a line and not a plane. Let $A, B, C \in \mathbb{R}^3$ be three points that are not collinear and let $\mathcal{P}$ be the plane that passes through A, B, and C.

Just like lines, planes have direction vectors. For $\mathcal{P}$, both $\vec{d}_1 = \overrightarrow{AB}$ and $\vec{d}_2 = \overrightarrow{AC}$ are direction vectors for $\mathcal{P}$. Of course, $\vec{d}_1$, $\vec{d}_2$ and their multiples are not the only direction vectors for $\mathcal{P}$. There are infinitely many more, including $\vec{d}_1 + \vec{d}_2$, and $\vec{d}_1 - 7\vec{d}_2$, and so on. However, since a plane is a *two*-dimensional object, we only need two different direction vectors to describe it.

Again like lines, planes have a vector form. $\mathcal{P}$ can be written in vector form as

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = t\vec{d}_1 + s\vec{d}_2 + A.$$

Vector form of $\mathcal{P}$ is not unique. Any two different directions in $\mathcal{P}$ suffice for defining $\mathcal{P}$ in vector form.



Definition 2.4.2 — Vector form of a plane. Let $\mathcal{P}$ be a plane and let $\vec{d}_1$, $\vec{d}_2$, and $\vec{p}$ be vectors such that $\mathcal{P} = \{\vec{x} \,:\, \vec{x} = t\vec{d}_1 + s\vec{d}_2 + \vec{p}$ for some $t, s \in \mathbb{R}\}$. We then say the equation

$$\vec{x} = t\vec{d}_1 + s\vec{d}_2 + \vec{p}$$

describes $\mathcal{P}$ in *vector form*. The vectors $\vec{d}_1$ and $\vec{d}_2$ are called *direction vectors* for the plane $\mathcal{P}$.

■ **Example 2.9** Describe the plane $\mathcal{P} \subseteq \mathbb{R}^3$ with equation $z = 2x + y + 3$ in vector form.

To describe $\mathcal{P}$ in vector form, we need a point on $\mathcal{P}$ and two direction vectors for $\mathcal{P}$. By guess-and-check, we see the points

$$A = \begin{bmatrix} 0 \\ 0 \\ 3 \end{bmatrix} \qquad B = \begin{bmatrix} 1 \\ 0 \\ 5 \end{bmatrix} \qquad C = \begin{bmatrix} 0 \\ 1 \\ 4 \end{bmatrix}$$

are all on $\mathcal{P}$. Thus

$$\vec{d}_1 = B - A = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} \qquad \text{and} \qquad \vec{d}_2 = C - A = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

are both direction vectors for $\mathcal{P}$. Therefore, we can express $\mathcal{P}$ in vector form as

$$\vec{x} = t\vec{d}_1 + s\vec{d}_2 + A = t\begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} + s\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 3 \end{bmatrix}.$$

■

Exercises for 2.4

1 Give, in vector form, an equation for the line that

a) passes through $(0,0,0)$ and is parallel to $\vec{v} = 3\vec{e}_1 + 4\vec{e}_2 + 5\vec{e}_3$,

b) passes through $(1,2,3)$ and $(4,-1,2)$,

c) passes through $(1,1)$ and is orthogonal to $\vec{v} = (3,1)$,

d) passes through $(9,-2,3)$ and $(1,2,3)$.

Solution:

a) $r(t) = 3t\vec{e}_1 + 4t\vec{e}_2 + 5t\vec{e}_3$

b) $r(t) = (1+3t)\vec{e}_1 + (2-3t)\vec{e}_2 + (3-t)\vec{e}_3$

c) $r(t) = (1-t)\vec{e}_1 + (1+3t)\vec{e}_2$

d) $r(t) = (1+8t)\vec{e}_1 + (2-4t)\vec{e}_2 + 3\vec{e}_3$.

2 Determine if the lines with the following vector equations intersect:

$$\vec{p}(t) = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} + t \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} \qquad \vec{r}(t) = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} + t \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}.$$

Solution: The lines intersect if they have at least one common point between them. Therefore, if we can find $t, s \in \mathbb{R}$ so that $p(t) = r(s)$, then the lines will intersect. Thus if $p(t) = r(s)$ then

$$\begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} + t \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} + s \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}.$$

Therefore, we want to find $t, s \in \mathbb{R}$ so that

$$\begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} = s \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} - t \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}.$$

In particular, it must be true that $-t = -2$, which says that $t = 2$. Thus, by substitution

$$\begin{bmatrix} 5 \\ 0 \\ 3 \end{bmatrix} = s \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}.$$

There are no selections of $s \in \mathbb{R}$ for which this is true, which says that there is no pair $t, s \in \mathbb{R}$ so that $p(t) = r(s)$. Therefore the lines do not intersect.

3 Find an equation for the plane with the given normal vector and containing the given point:

a) $\vec{n} = (2,-1,3)$, $P(1,2,0)$

b) $\vec{n} = (1,0,3)$, $P(2,4,5)$

Solution:

a) $2(x-1) - (y-2) + 3z = 0$.

b) $(x-2) + 3(z-5) = 0$.

4 Write an equation for the plane that

a) passes through $(1,4,3)$ and is orthogonal to the line with equation $\vec{r}(t) = (1+t, 2+4t, t)$,

b) passes through the origin and is parallel to the plane with equation $3x + 4y - 5z = -1$,

c) passes through $(0,0,0)$, $(1,-2,8)$, and $(-2,-1,3)$.

Solution:

a) $x + 4y + z = 20$

b) $3x + 4y - 5z = 0$

c) $2x - 19y - 5z = 0$.

5 Find a vector (parametric) equation for the line of intersection of the planes with equations $2x + 3y - z = 1$ and $x - y - z = 0$.

Solution: $\vec{r}(t) = (1-4t, t, 1-5t)$ is one solution.

6 Find the angle between the normals to the following planes:

a) the planes with equations $x + 2y - z = 2$ and $2x - y + 3z = 1$,

b) the plane with equation $2x + 3y - 5z = 0$ and the plane containing the points $(1,3,-2)$, $(5,1,3)$, and $(1,0,1)$.

Solution:

a) $\theta = \cos^{-1}\left(\frac{-3}{2\sqrt{21}}\right)$

b) $\theta = \cos^{-1}\left(\frac{14}{\sqrt{38}\sqrt{41}}\right)$

7[G] Two of these points are on the same side of the plane $3x+2y-2z=0$. Which two are they? $P(10,3,2)$, $Q(-2,5,3)$, $R(-2,5,1)$.

Solution: The normal vector $\vec{n}=(3,2,-2)$ points towards one side of the plane. If any point lies on this side of the plane, then it will form an angle with the normal vector that is less than 90°. If any point lies on the opposite side of the plane, then it will form an angle with the normal vector that is larger than 90°. So we should calculate the angles each of P,Q and R make with $\vec{n}$ using the dot-product.

a) Considering P, we find that $\theta = \cos^{-1}\left(\frac{32}{\sqrt{113}\sqrt{17}}\right) < 90°$.

b) Considering Q, we find that $\theta = \cos^{-1}\left(\frac{-2}{\sqrt{38}\sqrt{17}}\right) > 90°$.

c) Considering R, we find that $\theta = \cos^{-1}\left(\frac{2}{\sqrt{30}\sqrt{17}}\right) < 90°$.

Since P and R both make angles with $\vec{n}$ that are larger than 90°, they must lie on the same side of the plane.

2.5 Geometry & Sets

Using vectors, we can describe more than just lines and planes—we can describe all sorts of geometric objects.

Recall that vector form of a line is actually a shorthand. When we write $\vec{x} = t\vec{d} + \vec{p}$ to describe the line ℓ, what we mean is

$$\ell = \{\vec{x} \,:\, \vec{x} = t\vec{d} + \vec{p} \text{ for some } t \in \mathbb{R}\}.$$

We could notate a portion of this line by restricting t. For instance, consider the ray R and the line segment S:

$$R = \{\vec{x} \,:\, \vec{x} = t\vec{d} + \vec{p} \text{ for some } t \geq 0\}$$
$$S = \{\vec{x} \,:\, \vec{x} = t\vec{d} + \vec{p} \text{ for some } t \in [0,2]\}$$

XXX Figure

We can also make polygons by adding restrictions to the vector form of a plane. Let $\vec{a} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and $\vec{b} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ and consider the unit square U and the parallelogram P defined by

$$U = \{\vec{x} \,:\, \vec{x} = t\vec{e}_1 + s\vec{e}_2 \text{ for some } t, s \in [0,1]\}$$
$$P = \{\vec{x} \,:\, \vec{x} = t\vec{a} + s\vec{b} \text{ for some } t \in [0,1] \text{ and } s \in [-1,1]\}$$

XXX Figure

Each set so far is a set of linear combinations, and we have made different shapes by restricting the coefficients of those linear combinations. There are two ways of restricting linear combinations that arise up often enough to get their own names.

Definition 2.5.1 — Non-negative Linear Combination. Let $\vec{v}_1, \ldots, \vec{v}_n$ be vectors. The vector

$$\vec{x} = \alpha_1\vec{v}_1 + \cdots + \alpha_n\vec{v}_n.$$

is a *non-negative linear combination* of $\vec{v}_1, \ldots, \vec{v}_n$ if $\alpha_1, \ldots, \alpha_n \geq 0$.

Definition 2.5.2 — Convex Linear Combination. Let $\vec{v}_1, \ldots, \vec{v}_n$ be vectors. The vector

$$\vec{x} = \alpha_1\vec{v}_1 + \cdots + \alpha_n\vec{v}_n.$$

is a *convex linear combination* of $\vec{v}_1, \ldots, \vec{v}_n$ if $\alpha_1, \ldots, \alpha_n \geq 0$ and $\alpha_1 + \cdots + \alpha_n = 1$.

You can think of a non-negative linear combinations as vector you can arrive at by only displacing "forward" along your vectors.

Convex linear combinations can be thought of as weighted averages of vectors (the average of $\vec{v}_1, \ldots, \vec{v}_n$ would be the convex linear combination with coefficients $\alpha_i = \frac{1}{n}$). A convex linear combination of two vectors gives a point on the line segment connecting them.

■ **Example 2.10** Let $\vec{a}=\begin{bmatrix}2\\1\end{bmatrix}$ and $\vec{b}=\begin{bmatrix}-1\\1\end{bmatrix}$ and define

$$\begin{aligned} A &= \{\vec{x} \,:\, \vec{x} \text{ is a convex linear combination of } \vec{a} \text{ and } \vec{b}\} \\ &= \{\vec{x} \,:\, \vec{x}=\alpha\vec{a}+(1-\alpha)\vec{b} \text{ for some } \alpha\in[0,1]\}. \end{aligned}$$

Draw A.

XXX Finish

XXX Figure ■

Set Operations

We can also use set operations to define geometric objects. For instance, let H be the upper half-plane and C be the unit circle. That is

$$\begin{aligned} H &= \{\vec{x}\in\mathbb{R}^2 \,:\, \vec{x}=\alpha\vec{e}_1+\beta\vec{e}_2 \text{ for some } \alpha,\beta\in\mathbb{R} \text{ with } \alpha\geq 0\} \\ C &= \{\vec{x}\in\mathbb{R}^2 \,:\, \|\vec{x}\|=1\}. \end{aligned}$$

We can define now define the semi-circle $S=C\cap H$ using a set intersection.

XXX Figure

But what if we wanted to describe three identical circles but centered in different locations? We could define three circles S_1, S_2, and S_3 and consider their union $S_1\cup S_2\cup S_3$, or we could use the operation of *set addition* or the *sum of sets*.[12]

Definition 2.5.3 — Set Addition. Let A and B be sets of vectors. The *set sum* of A and B, denoted $A+B$, is the set

$$A+B=\{\vec{x} \,:\, \vec{x}=\vec{a}+\vec{b} \text{ for some } \vec{a}\in A \text{ and } \vec{b}\in B\}.$$

Set sums are very different than regular sums despite using the same symbol, "+".[13] However, they are very useful. Let $C=\{\vec{x}\in\mathbb{R}^2 \,:\, \|\vec{x}\|=1\}$ be the unit circle centered at the origin, and consider the sets

$$X=C+\{\vec{e}_2\} \qquad Y=C+\{3\vec{e}_1,\vec{e}_2\} \qquad Z=C+\{\vec{0},\vec{e}_1,\vec{e}_2\}.$$

Rewriting, we see $X=\{\vec{x}+\vec{e}_2 \,:\, \|\vec{x}\|=1\}$ is just C translated by $\vec{e}_2$. Similarly, $Y=\{\vec{x}+\vec{v} \,:\, \|\vec{x}\|=1 \text{ and } \vec{v}=3\vec{e}_1 \text{ or } \vec{v}=\vec{e}_2\}=(C+\{3\vec{e}_1\})\cup(C+\{\vec{e}_1\})$, and so Y is the union of two translated copies of C.[14]

XXX Figure

Set addition allows us to easily create parallel lines and planes by translation. For example, consider the lines ℓ_1 and ℓ_2 given in vector form as $\vec{x}=t\vec{d}$ and $\vec{x}=t\vec{d}+\vec{p}$, respectively, where

[12] Recall that the result of $a+b$ is called the *sum* of a and b, and the mathematical operation performed to get the sum is called *addition*

[13] For example, $A+\{\}=\{\}$, which might seem counterintuitive for an "addition" operation.

[14] If you want to stretch your mind, consider what $C+C$ is as a set.

$\vec{d} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and $\vec{p} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$. These lines differ from each other by translation, and using the idea of set addition we can write

$$\ell_2 = \ell_1 + \{\vec{p}\}.$$

XXX Figure (with lots of copies of $\vec{p}$ translating ℓ_1).

Note that it would be incorrect to write "$\ell_2 = \ell_1 + \vec{p}$". Because ℓ_1 is a set and $\vec{p}$ is not a set, "$\ell_1 + \vec{p}$" does not make mathematical sense.

As we pursue our study of linear algebra, we will come to see that lines and planes through the origin are especially important (and will have special notation). Set addition will allow us to describe all lines and planes as translations of lines and planes through the origin.

2.6 Span

Let $\vec{u} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\vec{v} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$. Can the vectors $\vec{w} = \begin{bmatrix} 2 \\ 5 \end{bmatrix}$ be obtained as a linear combination of $\vec{u}$ and $\vec{v}$?

By drawing a picture, the answer appears to be *yes*.

XXX Figure

Algebraically, we can use the definition of *linear combination* to set up a system of equations. We know $\vec{w}$ can be expressed as a linear combination of $\vec{u}$ and $\vec{v}$ if and only if the vector equation

$$\vec{w} = \begin{bmatrix} 2 \\ 5 \end{bmatrix} = \alpha \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \beta \begin{bmatrix} -1 \\ 2 \end{bmatrix} = \alpha\vec{u} + \beta\vec{v}$$

has a solution. By inspection, we see $\alpha = 3$ and $\beta = 1$ solve this equation.

After initial success, we might be tempted to ask the following: *what are all the locations in $\mathbb{R}^2$ that can be obtained as a linear combination of $\vec{u}$ and $\vec{v}$?* Geometrically, it appears any location can be reached. To verify this algebraically, consider the vector equation

$$\vec{x} = \begin{bmatrix} x \\ y \end{bmatrix} = \alpha \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \beta \begin{bmatrix} -1 \\ 2 \end{bmatrix} = \alpha\vec{u} + \beta\vec{v}. \tag{2.1}$$

Here $\vec{x}$ represents an arbitrary point in $\mathbb{R}^2$. Thus, if equation (2.1) always has a solution,[15] any vector in $\mathbb{R}^2$ can be obtained as a linear combination of α and β.

We can solve this equation for α and β by considering the equations arising from the first and second coordinates. Namely,

$$\begin{aligned} x &= \alpha + \beta \\ y &= \alpha - 2\beta. \end{aligned}$$

Subtracting the second equation from the first, we get $x - y = 3\beta$ and so $\beta = (x - y)/3$. Plugging β into the first equation and solving, we get $\alpha = (2x + y)/3$. Thus, equation (2.1) *always* has a solution. Namely,

$$\begin{aligned} \alpha &= \tfrac{1}{3}(2x + y) \\ \beta &= \tfrac{1}{3}(x - y). \end{aligned}$$

[15] The official terminology would be to say that the equations is always *consistent*.

There is a formal term for the set of vectors that can be obtained as linear combinations of others: *span*.

Definition 2.6.1 — Span. Let $\mathcal{X}$ be a set of vectors. The *span* of $\mathcal{X}$, written $\operatorname{span}\mathcal{X}$, is the set of all linear combinations of vectors in $\mathcal{X}$. Formally,

$$\operatorname{span}\mathcal{X} = \{\vec{x} \,:\, \vec{x} = \alpha_1\vec{v}_1 + \cdots + \alpha_n\vec{v}_n \text{ for some } \vec{v}_1,\ldots,\vec{v}_n \in \mathcal{X} \text{ and scalars } \alpha_1,\ldots,\alpha_n\}.$$

Further, we define $\operatorname{span}\emptyset = \{\vec{0}\}$.

We just showed above that $\operatorname{span}\left\{\begin{bmatrix}1\\1\end{bmatrix},\begin{bmatrix}-1\\2\end{bmatrix}\right\} = \mathbb{R}^2$.

■ **Example 2.11** Let $\vec{u} = \begin{bmatrix}-1\\2\end{bmatrix}$ and $\vec{v} = \begin{bmatrix}1\\-2\end{bmatrix}$. Find $\operatorname{span}\{\vec{u},\vec{v}\}$.

XXX Finish ■

The objects that arise from spans are familiar. If $\vec{v} \neq \vec{0}$, then $\operatorname{span}\{\vec{v}\}$ is the line through the origin with direction vector $\vec{v}$. If $\vec{v},\vec{w} \neq \vec{0}$ and aren't parallel, $\operatorname{span}\{\vec{v},\vec{w}\}$ is a plane through the origin. In fact, vector form of a line or a plane is nothing more than a *translated span*.

Representing Lines & Planes as Translated Spans

Consider the line ℓ given in vector form by

$$\vec{x} = t\vec{d} + \vec{0}.$$

The line ℓ passes through the origin, and if we unravel its definition, we see

$$\ell = \{\vec{x} \,:\, \vec{x} = t\vec{d} + \vec{0} \text{ for some } t \in \mathbb{R}\} = \{\vec{x} \,:\, \vec{x} = t\vec{d} \text{ for some } t \in \mathbb{R}\} = \operatorname{span}\{\vec{d}\}.$$

Similarly, if $\mathcal{P}$ is a plane given in vector form by

$$\vec{x} = t\vec{d}_1 + s\vec{d}_2 + \vec{0},$$

then

$$\mathcal{P} = \{\vec{x} \,:\, \vec{x} = t\vec{d}_1 + s\vec{d}_2 \text{ for some } t,s \in \mathbb{R}\} = \operatorname{span}\{\vec{d}_1,\vec{d}_2\}.$$

If the "$\vec{p}$" in our vector form is $\vec{0}$, then that vector form actually defines a span. This means, if you accept that every line/plane through the origin has a vector form, then every line/plane through the origin can be written as a span. Conversely, if $X = \operatorname{span}\{\vec{v}_1,\ldots,\vec{v}_n\}$ is a span, we know $\vec{0} = 0\vec{v}_1 + \cdots + 0\vec{v}_n \in X$, and so every span contains the origin.

As it turns out, spans exactly describe points, lines, planes, and volumes[16] through the origin.

■ **Example 2.12** The line $\ell_1 \subseteq \mathbb{R}^2$ is described by the equation $x + 2y = 0$ and the line $\ell_2 \subseteq \mathbb{R}^2$ is described by the equation $4x - 2y = 6$. If possible, describe ℓ_1 and ℓ_2 using spans.

XXX Finish ■

[16] We use the word *volume* to indicate the higher-dimensional analogue of a plane.

However, not all points, lines, planes, and volumes pass through the origin and so we can't describe every such object directly as a span. But, if we *translate* a span using set addition we can describe objects which don't pass through the origin.

■ **Example 2.13** Recall $\ell_2 \subseteq \mathbb{R}^2$ is the line described by the equation $4x - 2y = 6$. Describe ℓ_2 as a translated span.

XXX Finish ■

We can now see translated spans provide an alternative notation to vector form for specifying lines and planes. If X is described in vector form by

$$\vec{x} = t\vec{d}_1 + s\vec{d}_2 + \vec{p},$$

then

$$X = \operatorname{span}\{\vec{d}_1, \vec{d}_2\} + \{\vec{p}\}.$$

2.7 Linear Independence

Let

$$\vec{u} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \qquad \vec{v} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \qquad \vec{w} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}.$$

Since $\vec{w} = \vec{u} + \vec{v}$, we know that $\vec{w} \in \operatorname{span}\{\vec{u}, \vec{v}\}$. Geometrically, this is also clear because $\operatorname{span}\{\vec{u}, \vec{v}\}$ is the xy-plane in $\mathbb{R}^3$ and $\vec{w}$ lies on that plane.

What about $\operatorname{span}\{\vec{u}, \vec{v}, \vec{w}\}$? Intuitively, since $\vec{w}$ is already a linear combination of $\vec{u}$ and $\vec{v}$, we can't get anywhere *new* by taking linear combinations of $\vec{u}$, $\vec{v}$, and $\vec{w}$ compared to linear combinations of just $\vec{u}$ and $\vec{v}$. So $\operatorname{span}\{\vec{u}, \vec{v}\} = \operatorname{span}\{\vec{u}, \vec{v}, \vec{w}\}$.

Can we prove this from the definitions? Yes! Suppose $\vec{r} \in \operatorname{span}\{\vec{u}, \vec{v}, \vec{w}\}$. By definition,

$$\vec{r} = \alpha\vec{u} + \beta\vec{v} + \gamma\vec{w}$$

for some $\alpha, \beta, \gamma \in \mathbb{R}$. Since $\vec{w} = \vec{u} + \vec{v}$, we see

$$\vec{r} = \alpha\vec{u} + \beta\vec{v} + \gamma(\vec{u} + \vec{v}) = (\alpha + \gamma)\vec{u} + (\beta + \gamma)\vec{v} \in \operatorname{span}\{\vec{u}, \vec{v}\}.$$

Thus, $\operatorname{span}\{\vec{u}, \vec{v}, \vec{w}\} \subseteq \operatorname{span}\{\vec{u}, \vec{v}\}$. Conversely, if $\vec{s} \in \operatorname{span}\{\vec{u}, \vec{v}\}$, by definition,

$$\vec{s} = a\vec{u} + b\vec{v} = a\vec{u} + b\vec{v} + 0\vec{w}$$

for some $a, b \in \mathbb{R}$, and so $\vec{s} \in \operatorname{span}\{\vec{u}, \vec{v}, \vec{w}\}$. Thus $\operatorname{span}\{\vec{u}, \vec{v}\} \subseteq \operatorname{span}\{\vec{u}, \vec{v}, \vec{w}\}$. We conclude $\operatorname{span}\{\vec{u}, \vec{v}\} = \operatorname{span}\{\vec{u}, \vec{v}, \vec{w}\}$.

In this case, $\vec{w}$ was a redundant vector—it wasn't needed for the span.

Definition 2.7.1 — Linear Independence & Dependence (Geometric). The vectors $\vec{v}_1, \dots, \vec{v}_n$ are called *linearly dependent* if for at least one i,

$$\vec{v}_i \in \operatorname{span}\{\vec{v}_1, \dots, \vec{v}_{i-1}, \vec{v}_{i+1}, \dots, \vec{v}_n\}.$$

If there is no such i, the vectors $\vec{v}_1, \dots, \vec{v}_n$ are called *linearly independent*.

We will also refer to sets of vectors (for example $\{\vec{v}_1, \dots, \vec{v}_n\}$) as being linearly independent or linearly dependent. For technical reasons, we didn't state the definition in terms of sets.[17]

Definition 2.7.1 says that the vectors $\vec{v}_1, \dots, \vec{v}_n$ are linearly dependent if you can remove at least one vector without changing the span. In other words, $\vec{v}_1, \dots, \vec{v}_n$ are linearly dependent *if there is a redundant vector*.

■ **Example 2.14** Let $\vec{u} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$, $\vec{v} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$, and $\vec{w} = \begin{bmatrix} 4 \\ 5 \end{bmatrix}$. Determine whether $\{\vec{u}, \vec{v}, \vec{w}\}$ is linearly independent or linearly dependent.

XXX Finish ■

■ **Example 2.15** Determine whether the planes ... (given in vector form) are the same.

XXX Finish ■

Suppose the vectors $\vec{u}$, $\vec{v}$, and $\vec{w}$ satisfy

$$\vec{w} = \vec{u} + \vec{v}. \tag{2.2}$$

The set $\{\vec{u}, \vec{v}, \vec{w}\}$ is linearly dependent since $\vec{w} \in \operatorname{span}\{\vec{u}, \vec{v}\}$. But we can think of this another way. In particular, equation (2.2) can be rearranged to get

$$\vec{0} = \vec{u} + \vec{v} - \vec{w}.$$

Here we have expressed $\vec{0}$ as a non-trivial linear combination of $\vec{u}$, $\vec{v}$, and $\vec{w}$. That is, we have written $\vec{0}$ as a linear combination without all zero coefficients.

Definition 2.7.2 — Trivial Linear Combination. A linear combination $\alpha_1\vec{v}_1 + \cdots + \alpha_n\vec{v}_n$ is called *trivial* if $\alpha_1 = \cdots = \alpha_n = 0$. If at least one $\alpha_i \neq 0$, the linear combination is called *non-trivial*.

We can always write $\vec{0}$ as a linear combination of vectors if we let all the coefficients be zero, but it turns out we can only write $\vec{0}$ as a *non-trivial* linear combination of vectors if those vectors are linearly dependent. This is the basis for another definition of linear independence and dependence.

Definition 2.7.3 — Linear Independence & Linear Dependence (Algebraic). The vectors $\vec{v}_1, \dots, \vec{v}_n$ are called *linearly independent* if for the only linear combination satisfying

$$\vec{0} = \alpha_1\vec{v}_1 + \cdots + \alpha_n\vec{v}_n$$

is the trivial linear combination (where $\alpha_1 = \cdots = \alpha_n = 0$).

The idea of a "redundant vector" coming from the geometric definition of linear dependence is geometrically intuitive, but it can be hard to work with. After all, checking independence with this definition involves verifying for every vector that it is not in the span of the others. The algebraic definition on the other hand is less obvious, but checking independence or dependence of a set involves reasoning about solutions to just one equation.

[17] The issue is, every element of a set is unique. Clearly, the vectors $\vec{v}$ and $\vec{v}$ are linearly dependent, but $\{\vec{v}, \vec{v}\} = \{\vec{v}\}$, and so $\{\vec{v}, \vec{v}\}$ is technically a linearly independent set. This issue would be resolved by talking about *multisets* instead of sets, but it isn't worth the hassle.

Theorem 2.7.4 The geometric and algebraic definitions of linear independence are equivalent.

Proof. To show the two definitions are equivalent, we need to show that geometric $\implies$ algebraic and algebraic $\implies$ geometric.

(geometric $\implies$ algebraic) Suppose $\vec{v}_1, \ldots, \vec{v}_n$ are linearly dependent by the geometric definition. That means that for some i, we have

$$\vec{v}_i \in \operatorname{span}\{\vec{v}_1, \ldots, \vec{v}_{i-1}, \vec{v}_{i+1}, \ldots, \vec{v}_n\}.$$

Fix such an i. Then, by the definition of span we know

$$\vec{v}_i = \alpha_1\vec{v}_1 + \cdots \alpha_{i-1}\vec{v}_{i-1} + \alpha_{i+1}\vec{v}_{i+1} + \cdots + \alpha_n\vec{v}_n,$$

and so

$$\vec{0} = \alpha_1\vec{v}_1 + \cdots \alpha_{i-1}\vec{v}_{i-1} - \vec{v}_i + \alpha_{i+1}\vec{v}_{i+1} + \cdots + \alpha_n\vec{v}_n.$$

This must be a non-trivial linear combination because the coefficient of $\vec{v}_i$ is $-1 \neq 0$. Therefore, $\vec{v}_1, \ldots, \vec{v}_n$ is linearly dependent by the algebraic definition.

(geometric $\implies$ algebraic) Suppose $\vec{v}_1, \ldots, \vec{v}_n$ are linearly dependent by the algebraic definition. That means there exist $\alpha_1, \ldots, \alpha_n$, not all zero, so that

$$\vec{0} = \alpha_1\vec{v}_1 + \cdots + \alpha_n\vec{v}_n.$$

Fix i so that $\alpha_i \neq 0$ (why do we know there is such an i?). Rearranging we get

$$-\alpha_i\vec{v}_i = \alpha_1\vec{v}_1 + \cdots \alpha_{i-1}\vec{v}_{i-1} + \alpha_{i+1}\vec{v}_{i+1} + \cdots + \alpha_n\vec{v}_n,$$

and since $\alpha_i \neq 0$, we can multiply both sides by $\frac{-1}{\alpha_i}$ to get

$$\vec{v}_i = \tfrac{-\alpha_1}{\alpha_i}\vec{v}_1 + \cdots + \tfrac{-\alpha_{i-1}}{\alpha_i}\vec{v}_{i-1} + \tfrac{-\alpha_{i+1}}{\alpha_i}\vec{v}_{i+1} + \cdots + \tfrac{-\alpha_n}{\alpha_i}\vec{v}_n.$$

This shows that

$$\vec{v}_i \in \operatorname{span}\{\vec{v}_1, \ldots, \vec{v}_{i-1}, \vec{v}_{i+1}, \ldots, \vec{v}_n\},$$

and so $\vec{v}_1, \ldots, \vec{v}_n$ is linearly dependent by the geometric definition. ■

■ **Example 2.16** Let $\vec{u} = \begin{bmatrix}1\\2\end{bmatrix}$, $\vec{v} = \begin{bmatrix}2\\3\end{bmatrix}$, and $\vec{w} = \begin{bmatrix}4\\5\end{bmatrix}$. Use the algebraic definition of linear independence to determine whether $\{\vec{u}, \vec{v}, \vec{w}\}$ is linearly independent or dependent.

Notice that $\vec{u}$, $\vec{v}$, and $\vec{w}$ are the vectors from Example 2.14, so we already know that they are linearly dependent by the geometric definition of linear dependence.

XXX Finish ■

Linear Independence and Vector Form

Fix a vector $\vec{d}$ and consider the question: When is

$$\vec{x} = t\vec{d}$$

the vector form of a line? The answer is whenever $\vec{d} \neq \vec{0}$, a simple enough rule. What about for planes? Fix $\vec{d}_1$ and $\vec{d}_2$. When is

$$\vec{x} = t_1\vec{d}_1 + t_2\vec{d}_2$$

vector form of a plane? Here the rule is more complicated: $\vec{d}_1$ and $\vec{d}_2$ cannot be zero nor can they be parallel.

How about upping the dimension? Fix $\vec{d}_1$, $\vec{d}_2$, and $\vec{d}_3$. When is

$$\vec{x} = t_1\vec{d}_1 + t_2\vec{d}_2 + t_3\vec{d}_3$$

vector form of a *volume*? Coming up with such a rule seems hard, until we think about spans.

Recall that if $\vec{x} = t\vec{d}$ is vector form of a line, then that line is $\operatorname{span}\{\vec{d}\}$, and if $\vec{x} = t_1\vec{d}_1 + t_2\vec{d}_2$ is vector form of a plane, that plane is $\operatorname{span}\{\vec{d}_1, \vec{d}_2\}$. Similarly, if $\vec{x} = t_1\vec{d}_1 + t_2\vec{d}_2 + t_3\vec{d}_3$ is vector form of a volume, that volume must be $\operatorname{span}\{\vec{d}_1, \vec{d}_2, \vec{d}_3\}$. So we can focus on a (maybe) simpler question: When is $\operatorname{span}\{\vec{d}_1, \vec{d}_2, \vec{d}_3\}$ a volume?

This question is answered by considering whether or not $\{\vec{d}_1, \vec{d}_2, \vec{d}_3\}$ is linearly dependent. If $\{\vec{d}_1, \vec{d}_2, \vec{d}_3\}$ is linearly dependent, then $\operatorname{span}\{\vec{d}_1, \vec{d}_2, \vec{d}_3\}$ is either a (i) plane, if there is one redundant vector, (ii) a line, if there are two redundant vectors, or (iii) a point, if $\vec{d}_1 = \vec{d}_2 = \vec{d}_3 = \vec{0}$. If $\{\vec{d}_1, \vec{d}_2, \vec{d}_3\}$ is linearly independent, then $\operatorname{span}\{\vec{d}_1, \vec{d}_2, \vec{d}_3\}$ is a volume.

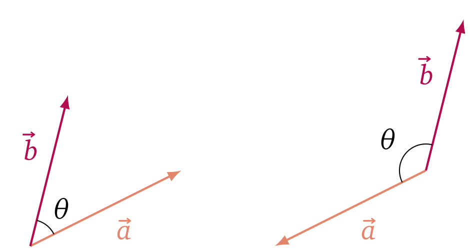
The moral of the story is that when you describe an object in vector form, the *direction vectors must be linearly independent*.

2.8 Dot Products

Let $\vec{a}$ and $\vec{b}$ be vectors. We assume they are placed so their starting points coincide. Let θ denote the *smaller* of the two angles between them, so $0 \le \theta \le \pi$. The *dot product* of $\vec{a}$ and $\vec{b}$ is defined to be

$$\vec{a} \cdot \vec{b} = \|\vec{a}\|\|\vec{b}\| \cos\theta.$$

We will call this the *geometric definition of the dot product*. The dot product is also sometimes called the *scalar product* because the result is a scalar. Note that $\vec{a} \cdot \vec{b} = 0$ when either $\vec{a}$ or $\vec{b}$ is zero or, more interestingly, if their directions are perpendicular.



Algebraically, we can define the dot product in terms of coordinates:

$$\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \cdot \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n.$$

We will call this the *algebraic definition of the dot product*.[18]

By switching between algebraic and geometric definitions, we can use the dot product to find quantities that are otherwise difficult to find.

■ **Example 2.17** Find the angle between the vectors $\vec{v} = (1,2,3)$ and $\vec{w} = (1,1,-2)$.

From the algebraic definition of the dot product, we know

$$\vec{v} \cdot \vec{w} = 1(1) + 2(1) + 3(-2) = -3.$$

From the geometric definition, we know

$$\vec{v} \cdot \vec{w} = \|\vec{v}\|\|\vec{w}\| \cos\theta = \sqrt{14}\sqrt{6}\cos\theta = 2\sqrt{21}\cos\theta.$$

Equating the two definitions of $\vec{v} \cdot \vec{w}$, we see

$$\cos\theta = \frac{-3}{2\sqrt{21}}$$

and so $\theta = \arccos\left(\frac{-3}{2\sqrt{21}}\right)$. ■

Recall that for vectors $\vec{a}$ and $\vec{b}$, the relationship $\vec{a} \cdot \vec{b} = 0$ can hold for two reasons: (i) either $\vec{a} = \vec{0}$, $\vec{b} = \vec{0}$, or both or (ii) $\vec{a}$ and $\vec{b}$ meet at 90°. Thus, the dot product can be used to tell if two vectors are perpendicular. There is some strangeness with the zero vector here, but it turns out this strangeness simplifies our lives mathematically.

Definition 2.8.1 — Orthogonal. The vectors $\vec{u}$ and $\vec{v}$ are *orthogonal* if $\vec{u} \cdot \vec{v} = 0$.

The definition of orthogonal encapsulates both the idea of two vectors forming a right angle and the idea of one of them being $\vec{0}$.

Before we continue, let's pin down the idea of one vector pointing in the *direction* of another. There are many ways we could define this idea, but we'll go with this one.

Definition 2.8.2 The vector $\vec{u}$ points in the *direction* of the vector $\vec{v}$ if $k\vec{u} = \vec{v}$ for some scalar k. The vector $\vec{u}$ points in the *positive direction* of $\vec{v}$ if $k\vec{u} = \vec{v}$ for some positive scalar k.

The vector $2\vec{e}_1$ points in the direction of $\vec{e}_1$ since $\frac{1}{2}(2\vec{e}_1) = \vec{e}_1$. Since $\frac{1}{2} > 0$, $2\vec{e}_1$ also points in the positive direction of $\vec{e}_1$. In contrast, $-\vec{e}_1$ points in the direction $\vec{e}_1$ but not the positive direction of $\vec{e}_1$.

This idea can be rephrased using using linear combinations.

[18] Philosophically, every object should have only one definition from which equivalent characterizations can be deduced as theorems. If you're bothered, pick your favorite definition to be the "true" definition and consider the other definition a theorem.

Definition 2.8.3 The vector $\vec{u}$ points in the *direction* of the vector $\vec{v}$ if $\vec{u} \in \operatorname{span}\{\vec{v}\}$.

An important use of the dot product is to determine two what extent two vectors point in the same direction.

Theorem 2.8.4 Let $\vec{u}$ and $\vec{v}$ be non-zero vectors. Then, $\vec{u}$ is in the direction of $\vec{v}$ if and only if

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

Proof. Let $\vec{u}$ and $\vec{v}$ be non-zero vectors. By the geometric definition of the dot product,

$$\frac{\vec{u}\cdot\vec{v}}{||\vec{u}||||\vec{v}||} = \frac{||\vec{u}||||\vec{v}||\cos\theta}{||\vec{u}||||\vec{v}||} = \cos\theta,$$

where θ is the angle between $\vec{u}$ and $\vec{v}$.

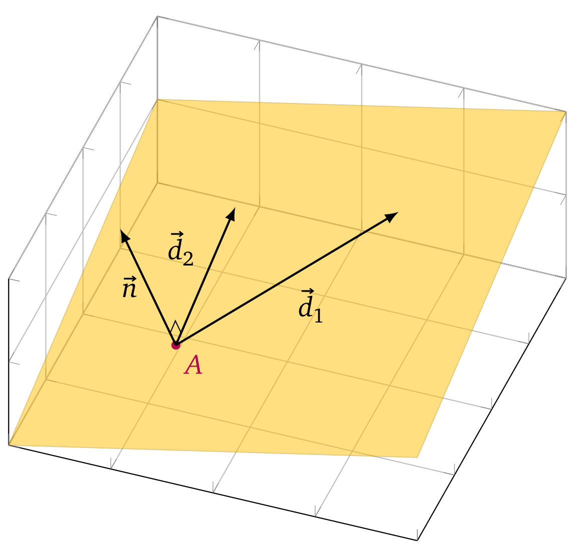
The vector $\vec{u}$ is in the direction of $\vec{v}$ if and only if $\theta = 0$ or π. But, for $\theta \in [0,\pi]$, we know $\cos\theta = \pm 1$ if and only if $\theta = 0$ or π. ■

■ **Example 2.18** Let $\vec{a} = \begin{bmatrix}1\\2\end{bmatrix}$, $\vec{b} = \begin{bmatrix}3\\3\end{bmatrix}$, $\vec{c} = \begin{bmatrix}2\\1\end{bmatrix}$, and $\vec{v} = \begin{bmatrix}3\\4\end{bmatrix}$. Which vector out of $\vec{a}$, $\vec{b}$, and $\vec{c}$ has a direction closest to the direction of $\vec{v}$?

XXX Finish ■

Normal Form of Lines and Planes

Since we will commonly be working in $\mathbb{R}^3$ there is another way to define a plane. Given any vector $\vec{n} \in \mathbb{R}^3$, we can consider the set $\mathcal{Q} \subseteq \mathbb{R}^3$ of vectors orthogonal to $\vec{n}$. If $\vec{n} = \vec{0}$, then $\mathcal{Q} = \mathbb{R}^3$. Otherwise, $\mathcal{Q}$ is a plane through the origin. In this case, $\vec{n}$ is called the *normal vector* of the plane $\mathcal{Q}$.



Definition 2.8.5 — Normal form of a plane. The plane $\mathcal{P}$ is described in *normal form* if for some $\vec{n}$ and $\vec{p}$, the equation
$$\vec{n}\cdot(\vec{x}-\vec{p})=0$$
if and only if $\vec{x}\in\mathcal{P}$. Equivalently, $\mathcal{P}$ is described in normal form if for some $\vec{n}$ and scalar $\alpha\in\mathbb{R}$ the equation
$$\vec{n}\cdot\vec{x}=\alpha$$
is satisfied if and only if $\vec{x}\in\mathcal{P}$. In either case, the vector $\vec{n}$ is call a *normal vector* for $\mathcal{P}$.

Normal form of a plane only exists in $\mathbb{R}^3$, but it is often useful.[19] The equivalence of the two ways to write a normal form of a plane is straight forward.
$$\vec{n}\cdot(\vec{x}-\vec{p})=0$$
if and only if
$$\vec{n}\cdot\vec{x}=\vec{n}\cdot\vec{p}=\alpha.$$
Since $\vec{n}$ and $\vec{p}$ are fixed, α is a constant. Expanding normal form in terms of coordinates we see
$$\vec{n}\cdot(\vec{x}-\vec{p})=\vec{n}\cdot\vec{x}-\alpha=n_xx+n_yy+n_zz-\alpha=0$$
and so
$$n_xx+n_yy+n_zz=\alpha \tag{2.3}$$
is another way to write a plane. Equation (2.3) is sometimes called *scalar form* of a plane. For us, it will not be important to distinguish between scalar and normal form.

It should be noted that like vector form of a plane, normal form of a plane is not unique. For example, the plane described by $\vec{n}\cdot(\vec{x}-\vec{p})=0$ is the same as the plane $(2\vec{n})\cdot(\vec{x}-\vec{p})=0$.

■ **Example 2.19** Find vector form and normal form of the plane $\mathcal{P}$ passing through the point $A=(1,0,0)$, $B=(0,1,0)$ and $C=(0,0,1)$.

To find vector form of $\mathcal{P}$, we need a point on the plane and two direction vectors. We have three points on the plane, so we can obtain two direction vectors by subtracting these points in different ways. Let
$$\vec{d}_1=\overrightarrow{AB}=\begin{bmatrix}-1\\1\\0\end{bmatrix}\qquad \vec{d}_2=\overrightarrow{AC}=\begin{bmatrix}-1\\0\\1\end{bmatrix}.$$
Using the point A, we may now write vector form of $\mathcal{P}$ as
$$\begin{bmatrix}x\\y\\z\end{bmatrix}=t\begin{bmatrix}-1\\1\\0\end{bmatrix}+s\begin{bmatrix}-1\\0\\1\end{bmatrix}+\begin{bmatrix}1\\0\\0\end{bmatrix}.$$

To write normal form we need to find a normal vector to $\mathcal{P}$. By symmetry, we can see that $\vec{n}=(1,1,1)$ is a normal vector to $\mathcal{P}$. If we weren't so insightful, we could also compute

[19] Just like $y=mx+b$ form of a line only exists in $\mathbb{R}^2$.

$\vec{d}_1 \times \vec{d}_2 = (1,1,1)$ to find a normal vector. Now, we may express $\mathcal{P}$ in normal form as

$$\begin{bmatrix}1\\1\\1\end{bmatrix} \cdot \left(\begin{bmatrix}x\\y\\z\end{bmatrix} - \begin{bmatrix}1\\0\\0\end{bmatrix} \right) = 0$$

or equivalently,

$$x + y + z = 1.$$

■

■ **Example 2.20** Find the line $\mathcal{P}_1 \cap \mathcal{P}_2$ where $\mathcal{P}_1$ is the plane given by the equation

$$x + y + z = 2$$

and $\mathcal{P}_2$ is the plane given by the equation

$$2x - y + z = 0.$$

Let $\ell = \mathcal{P}_1 \cap \mathcal{P}_2$. Since $\ell \subseteq \mathcal{P}_1$ and $\ell \subseteq \mathcal{P}_2$, every direction vector for ℓ is also a direction vector for $\mathcal{P}_1$ and $\mathcal{P}_2$.

Let $\vec{n}_1 = (1,1,1)$ be a normal vector for $\mathcal{P}_2$ and $\vec{n}_2 = (2,-1,1)$ be a normal vector for $\mathcal{P}_2$. If $\vec{d}$ is a direction vector for ℓ, then $\vec{n}_1 \cdot \vec{d} = 0$ and $\vec{n}_2 \cdot \vec{d} = 0$. Thus,

$$\vec{d} = \vec{n}_1 \times \vec{n}_2 = \begin{bmatrix}2\\1\\-3\end{bmatrix}$$

is a direction vector for ℓ. By guess and check we find that $\vec{p} = (0,1,1)$ satisfies $\vec{p} \in \mathcal{P}_1$ and $\vec{p} \in \mathcal{P}_2$ and so $\vec{p} \in \ell$. Thus, we may write ℓ in vector form as

$$\begin{bmatrix}x\\y\\z\end{bmatrix} = t\begin{bmatrix}2\\1\\-3\end{bmatrix} + \begin{bmatrix}0\\1\\1\end{bmatrix}.$$

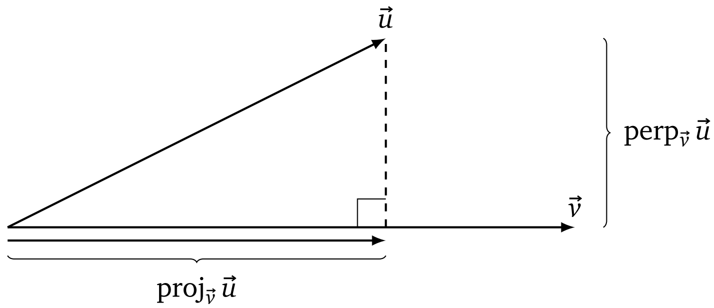
■

2.9 Projection

Another common vector operation is *projection*. Projection measures how much a vector points in the direction of another. This quantity is encoded as a vector. We make this definition mathematically precise as follows.

Definition 2.9.1 — Projection. For a vector $\vec{u}$ and a non-zero vector $\vec{v}$, the *projection* of $\vec{u}$ onto $\vec{v}$ is written as $\operatorname{proj}_{\vec{v}} \vec{u}$ and is a vector in the direction of $\vec{v}$ with the property that $\vec{u} - \operatorname{proj}_{\vec{v}} \vec{u}$ is orthogonal to $\vec{v}$.
The vector $\vec{u} - \operatorname{proj}_{\vec{v}} \vec{u}$ is called the *perpendicular component* of the projection of $\vec{u}$ onto $\vec{v}$ and is notated $\operatorname{perp}_{\vec{v}} \vec{u}$.

We can visualize projections with the following diagram.



From the picture, it appears that $\vec{u}$, $\operatorname{proj}_{\vec{v}} \vec{u}$, and $\operatorname{perp}_{\vec{v}} \vec{u}$ form a right triangle. Of course, we shouldn't trust the picture. We should verify this mathematically.

Theorem 2.9.2 If $\vec{u}$ and $\vec{v}$ are non-zero vectors, then $\vec{v}$, $\operatorname{proj}_{\vec{v}} \vec{u}$, and $\operatorname{perp}_{\vec{v}} \vec{u}$ form a (possibly degenerate) right triangle.

Proof. We need to verify that the sides $\operatorname{proj}_{\vec{v}} \vec{u}$ and $\operatorname{perp}_{\vec{v}} \vec{u}$ meet at a right angle and that the hypotenuse $\vec{u}$ meets the sides. That is, $\operatorname{perp}_{\vec{v}} \vec{u} + \operatorname{proj}_{\vec{v}} \vec{u} = \vec{u}$.

By the definition of projection, $\operatorname{perp}_{\vec{v}} \vec{u} = \vec{u} - \operatorname{proj}_{\vec{v}} \vec{u}$ is orthogonal to $\vec{v}$. Since $\operatorname{proj}_{\vec{v}} \vec{u}$ points in the direction of $\vec{v}$, we have $\operatorname{proj}_{\vec{v}} \vec{u} = k\vec{v}$ and so $\operatorname{perp}_{\vec{v}} \vec{u}$ is orthogonal to $\operatorname{proj}_{\vec{v}} \vec{u}$.

Finally, consider

$$\operatorname{perp}_{\vec{v}} \vec{u} + \operatorname{proj}_{\vec{v}} \vec{u} = (\vec{u} - \operatorname{proj}_{\vec{v}} \vec{u}) + \operatorname{proj}_{\vec{v}} \vec{u} = \vec{u},$$

so indeed the vectors form a right triangle. ■

Now that we've proved $\vec{u}$, $\operatorname{proj}_{\vec{v}} \vec{u}$, and $\operatorname{perp}_{\vec{v}} \vec{u}$ form a right triangle, we are free to use trigonometry to compute projections. If θ is the angle between $\vec{u}$ and $\vec{v}$ and $0 \leq \theta \leq \pi/2$, we know $\|\operatorname{proj}_{\vec{v}} \vec{u}\| = \|\vec{u}\| \cos\theta$. This means

$$\operatorname{proj}_{\vec{v}} \vec{u} = k\vec{v} = \|\vec{u}\| \cos\theta \frac{\vec{v}}{\|\vec{v}\|}$$

(Recall that $\vec{v}/\|\vec{v}\|$ is a unit vector in the direction of $\vec{v}$). But $\cos\theta$ appears in the formula for the dot product. Solving for $\cos\theta$ in the dot product formula, we see $\cos\theta = \frac{\vec{u}\cdot\vec{v}}{\|\vec{u}\|\|\vec{v}\|}$. Thus,

$$\operatorname{proj}_{\vec{v}} \vec{u} = \|\vec{u}\| \cos\theta \vec{v} = \frac{\vec{u}\cdot\vec{v}}{\|\vec{u}\|\|\vec{v}\|}\left(\frac{\vec{v}}{\|\vec{v}\|}\right) = \frac{\vec{u}\cdot\vec{v}}{\|\vec{v}\|^2}\vec{v}.$$

Upon close inspection, we see $\|\vec{v}\|^2 = \vec{v}\cdot\vec{v}$ (since $\cos 0 = 1$) and so we finally arrive at the formula

$$\operatorname{proj}_{\vec{v}} \vec{u} = \frac{\vec{u}\cdot\vec{v}}{\vec{v}\cdot\vec{v}}\vec{v}$$

Incredibly, if we use the algebraic definition of the dot product, we can compute a projection without computing cosine of anything!

Exercises for 2.9

1 In each case determine if the given pair of vectors is orthogonal.

a) $\vec{a} = 3\vec{e}_1 + 4\vec{e}_2$, $\vec{b} = -4\vec{e}_1 + 3\vec{e}_2$
b) $\vec{a} = (4, -1, 2)$, $\vec{b} = (3, 0, -6)$
c) $\vec{a} = 3\vec{e}_1 - 2\vec{e}_2$, $\vec{b} = -2\vec{e}_1 - 4\vec{e}_3$

Solution:

a) $\vec{a} \cdot \vec{b} = -12 + 12 = 0$, so $\vec{a}$ and $\vec{b}$ are orthogonal.
b) $\vec{a} \cdot \vec{b} = 12 + 0 - 12 = 0$, so $\vec{a}$ and $\vec{b}$ are orthogonal.
c) $\vec{a} \cdot \vec{b} = -6 + 8 = 2$, so $\vec{a}$ and $\vec{b}$ are not orthogonal.

2 Assume $\vec{u} = (a, b)$ is a non-zero plane vector. Show that $\vec{v} = (-b, a)$ is orthogonal to $\vec{u}$. By examining all possible signs for a and b, convince yourself that the 90 degree angle between $\vec{u}$ and $\vec{v}$ is in the counter-clockwise direction.
Solution: $\vec{u} \cdot \vec{v} = -ab + ba = 0$, so $\vec{u}$ and $\vec{v}$ are orthogonal.

3 The methane molecule, CH_4, has four Hydrogen atoms at the vertices of a regular tetrahedron and a Carbon atom at its center. Choose as the vertices of this tetrahedron the points $(0, 0, 0)$, $(1, 1, 0)$, $(1, 0, 1)$, and $(0, 1, 1)$.

a) Find the angle between two edges of the tetrahedron.
b) Find the bond angle between two Carbon–Hydrogen bonds.

Solution:

a) $60°$
b) $\cos^{-1}\left(\frac{1}{3}\right)^{\circ}$

4 An inclined plane makes an angle of 30 degrees with the horizontal. Use vectors and the dot product to find the projection of the gravitational acceleration vector $-g\vec{e}_2$ along a unit vector pointing along the inclined plane.
Solution: $\frac{-g\sqrt{3}}{4}\vec{e}_1 - \frac{g}{4}\vec{e}_2$

5 Derive the formula

$$||\vec{a} + \vec{b}||^2 = ||\vec{a}||^2 + ||\vec{b}||^2 + 2||\vec{a}||\,||\vec{b}|| \cos\theta.$$

How is it related to the Law of Cosines? A picture might help.
Solution:

$$\begin{aligned} ||\vec{a} + \vec{b}||^2 &= (\vec{a} + \vec{b}) \cdot (\vec{a} + \vec{b}) \\ &= \vec{a} \cdot \vec{a} + \vec{a} \cdot \vec{b} + \vec{b} \cdot \vec{a} + \vec{b} \cdot \vec{b} \\ &= ||\vec{a}||^2 + ||\vec{b}||^2 + 2\vec{a} \cdot \vec{b} \\ &= ||\vec{a}||^2 + ||\vec{b}||^2 + 2||\vec{a}||\,||\vec{b}|| \cos\theta \end{aligned}$$

6 Use the dot product to determine if the points $P(3, 1, 2)$, $Q(-1, 0, 2)$, and $R(11, 3, 2)$ are collinear.
Solution: P, Q and R are collinear if and only if the angle between $\overrightarrow{PQ}$ and $\overrightarrow{QR}$ is 180°. But $\overrightarrow{PQ} = (-4, -1, 0)$ and $\overrightarrow{QR} = (12, 3, 0)$. So $\overrightarrow{PQ} \cdot \overrightarrow{QR} = -51 = ||\overrightarrow{PQ}||\,||\overrightarrow{QR}|| \cos\theta = 51 \cos\theta$, which implies that $\cos\theta = -1$ and so $\theta = 180°$ and the points are collinear.

7[OS] For the following exercises, the vectors $\vec{u}$ and $\vec{v}$ are given. Find the projection $\vec{w} = \text{proj}_{\vec{u}}\,\vec{v}$ of vector $\vec{v}$ onto vector $\vec{u}$. Express your answer in component form.

a) $\vec{u} = 5\vec{e}_1 + 2\vec{e}_2$, $\vec{v} = 2\vec{e}_1 + 3\vec{e}_2$.
b) $\vec{u} = (4, 7)$, $\vec{v} = (3, 5)$
c) $\vec{u} = 3\vec{e}_1 + 2\vec{e}_3$, $\vec{v} = 2\vec{e}_2 + 4\vec{e}_3$
d) $\vec{u} = (4, 4, 0)$, $\vec{v} = (0, 4, 1)$

Solution: Generally, the projection of $\vec{v}$ onto $\vec{u}$ is given by

$$\text{proj}_{\vec{u}}\,\vec{v} = \frac{\vec{u} \cdot \vec{v}}{\vec{u} \cdot \vec{u}}\vec{u}.$$

Thus, using this formula, we can compute the following.

a) $\vec{w} = [16/29](5\vec{e}_1 + 2\vec{e}_2)$.
b) $\vec{w} = [23/65](-4, 7)$.
c) $\vec{w} = [8/13](3\vec{e}_1 + 2\vec{e}_3)$.

d) $\vec{w} = [1/2](4,4,0)$.

8[OS] Consider the vectors $\vec{u} = 4\vec{e}_1 - 3\vec{e}_2$ and $\vec{v} = 3\vec{e}_1 + 2\vec{e}_2$

a) Find the components of vector $\vec{w} = \operatorname{proj}_{\vec{u}} \vec{v}$ that represents the projection of $\vec{v}$ onto $\vec{u}$.

b) Write the decomposition $\vec{v} = \vec{w} + \vec{q}$ of vector $\vec{v}$ into the orthogonal components $\vec{w}$ and $\vec{q}$, where $\vec{w}$ is the projection of $\vec{v}$ onto $\vec{u}$ and $\vec{q}$ is a vector orthogonal to the direction of $\vec{u}$.

Solution:

a) $\vec{w} = [\vec{u}\cdot\vec{v}/\vec{u}\cdot\vec{u}]\vec{u}$. Therefore, we find that $\vec{w} = [6/25](4,-3)$, which is to say that the first component of $\vec{w}$ is $24/25$ and the second component of $\vec{w}$ is $-18/25$.

b) Begin by noting that $\vec{w} = \operatorname{proj}_{\vec{u}} \vec{v}$. Since $\vec{w}$ points in the same direction as $\vec{u}$, then it follows that $\vec{q}$ must be orthogonal to $\vec{w}$, which is to say that $\vec{q} = \operatorname{perp}_{\vec{u}} \vec{v} = \vec{v} - \operatorname{proj}_{\vec{u}} \vec{v}$. Then clearly $\vec{v} = \vec{w} + \vec{q} = \operatorname{proj}_{\vec{u}} \vec{v} + \vec{v} - \operatorname{proj}_{\vec{u}} \vec{v}$. Computation of $\vec{w}$ and $\vec{q}$ gives that

$$\vec{w} = \left(\frac{24}{25}, \frac{-18}{25}\right) \quad \vec{q} = \left(\frac{51}{25}, \frac{68}{25}\right)$$

9[OS] Consider the vectors $\vec{u} = 2\vec{e}_1 + 4\vec{e}_2$ and $\vec{v} = 4\vec{e}_1 + 2\vec{e}_3$

a) Find the components of vector $\vec{w} = \operatorname{proj}_{\vec{u}} \vec{v}$ that represents the projection of $\vec{v}$ onto $\vec{u}$.

b) Write the decomposition $\vec{v} = \vec{w} + \vec{q}$ of vector $\vec{v}$ into the orthogonal components $\vec{w}$ and $\vec{q}$, where $\vec{w}$ is the projection of $\vec{v}$ onto $\vec{u}$ and $\vec{q}$ is a vector orthogonal to the direction of $\vec{u}$.

Solution:

a) $\vec{w} = [2/5](2,4,0)$, which is to say that the first component of $\vec{w}$ is $4/5$, the second component of $\vec{w}$ is $8/5$, and the last component of $\vec{w}$ is 0.

b) Computing $\vec{w} = \operatorname{proj}_{\vec{u}} \vec{v}$ and $\vec{q} = \vec{v} - \operatorname{proj}_{\vec{u}} \vec{v}$ gives

$$\vec{w} = \left(\frac{4}{5}, \frac{8}{5}, 0\right) \quad \vec{q} = \left(\frac{16}{5}, -\frac{8}{5}, 2\right)$$

2.10 The Cross Product

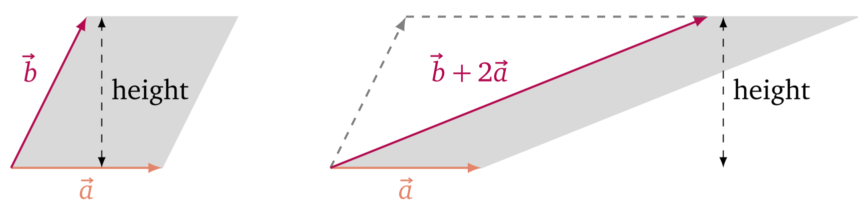
For vectors $\vec{a}$ and $\vec{b}$, the dot product $\vec{a}\cdot\vec{b}$ measures how close $\vec{a}$ and $\vec{b}$ are to being orthogonal. In contrast, the *cross product* of $\vec{a}$ and $\vec{b}$, written $\vec{a}\times\vec{b}$, will measure the *area* of the parallelogram whose sides are given by $\vec{a}$ and $\vec{b}$.

Let's explore this idea. Since the cross product is a *product*, we will demand it follow reasonable distribution laws[20]:

$$\begin{aligned}\vec{a}\times(\vec{b}+\vec{c}) &= \vec{a}\times\vec{b}+\vec{a}\times\vec{c}\\ (\vec{a}+\vec{b})\times\vec{c} &= \vec{a}\times\vec{c}+\vec{b}\times\vec{c}\\ (\alpha\vec{a})\times\vec{b} &= \alpha(\vec{a}\times\vec{b})\\ \vec{a}\times(\alpha\vec{b}) &= \alpha(\vec{a}\times\vec{b})\end{aligned}$$

for vectors $\vec{a}$, $\vec{b}$, $\vec{c}$ and scalars α.

Now, suppose $\vec{a}\times\vec{b}$ indeed encapsulates the area of the parallelogram with sides $\vec{a}$ and $\vec{b}$. If we slide the tip of $\vec{b}$ parallel to the vector $\vec{a}$, we should not change the area. Thus the cross product of $\vec{a}$ and $\vec{b}$ should be the same as that of $\vec{a}$ and $\vec{b}+\alpha\vec{a}$.



Using this invariance along with our distributive rules, we now see

$$\vec{a}\times\vec{b} = \vec{a}\times(\vec{b}+\alpha\vec{a}) = \vec{a}\times\vec{b}+\alpha(\vec{a}\times\vec{a}),$$

and so $\vec{a}\times\vec{a}=0$. We can apply this newly-found fact to the vector $\vec{a}+\vec{b}$ to deduce

$$\begin{aligned}0 = (\vec{a}+\vec{b})\times(\vec{a}+\vec{b}) &= \vec{a}\times\vec{a}+\vec{a}\times\vec{b}+\vec{b}\times\vec{a}+\vec{b}\times\vec{b}\\ &= 0+\vec{a}\times\vec{b}+\vec{b}\times\vec{a}+0\\ &= \vec{a}\times\vec{b}+\vec{b}\times\vec{a},\end{aligned}$$

and so

$$\vec{a}\times\vec{b} = -\vec{b}\times\vec{a}.$$

Products with this property are called *anti-commutative*. Now for an incredible fact of the universe: the result of the cross product of two vectors in $\mathbb{R}^3$ can be represented by another vector in $\mathbb{R}^3$ whose magnitude corresponds to the area of the parallelogram with sides $\vec{a}$ and $\vec{b}$.[21] Using trigonometry, we deduce

$$\|\vec{a}\times\vec{b}\| = \|\vec{a}\|\|\vec{b}\|\sin\theta$$

[20] The technical term for satisfying these laws is *bilinearity*.

[21] This is *only* true in $\mathbb{R}^3$. In $\mathbb{R}^4$ a product that produces area-like quantities does exist, but the output cannot be described by a vector. In higher dimensions, the cross product is called the *wedge product*.

where $0 \leq \theta \leq \pi$ is smaller of the two angles between $\vec{a}$ and $\vec{b}$. What remains to be seen is what direction $\vec{a} \times \vec{b}$ points in. For this, we use the standard basis for $\mathbb{R}^3$ as a launching point. Recall $\vec{e}_1$, $\vec{e}_2$, and $\vec{e}_3$ are all unit vectors and all orthogonal to each other. Thus, crossing any two of them must result in a unit vector. By convention,

$$\begin{aligned}\vec{e}_1 \times \vec{e}_2 &= \vec{e}_3,\\ \vec{e}_2 \times \vec{e}_3 &= \vec{e}_1,\\ \vec{e}_3 \times \vec{e}_1 &= \vec{e}_2.\end{aligned}$$

Let $\vec{a} = a_x\vec{e}_1 + a_y\vec{e}_2 + a_z\vec{e}_3$ and $\vec{b} = b_x\vec{e}_1 + b_y\vec{e}_2 + b_z\vec{e}_3$. Using the distributive laws of the cross product we see,

$$\begin{aligned}\vec{a} \times \vec{b} = (a_x\vec{e}_1 + a_y\vec{e}_2 + a_z\vec{e}_3) \times (b_x\vec{e}_1 + b_y\vec{e}_2 + b_z\vec{e}_3)\\ = \quad & a_x b_x\vec{e}_1 \times \vec{e}_1 + a_x b_y\vec{e}_1 \times \vec{e}_2 + a_x b_z\vec{e}_1 \times \vec{e}_3\\ & + a_y b_x\vec{e}_2 \times \vec{e}_1 + a_y b_y\vec{e}_2 \times \vec{e}_2 + a_y b_z\vec{e}_2 \times \vec{e}_3\\ & + a_z b_x\vec{e}_3 \times \vec{e}_1 + a_z b_y\vec{e}_3 \times \vec{e}_2 + a_z b_z\vec{e}_3 \times \vec{e}_3\\ = \quad & \vec{0} + a_x b_y\vec{e}_3 - a_x b_z\vec{e}_2\\ & - a_y b_x\vec{e}_3 + \vec{0} + a_y b_z\vec{e}_1\\ & + a_z b_x\vec{e}_2 - a_z b_y\vec{e}_1 + \vec{0}\end{aligned}$$

so

$$\vec{a} \times \vec{b} = (a_y b_z - a_z b_y)\vec{e}_1 - (a_x b_z - a_z b_x)\vec{e}_2 + (a_x b_y - a_y b_x)\vec{e}_3.$$

Exercise 2.21 Verify that $\|\vec{a} \times \vec{b}\| = \|\vec{a}\|\|\vec{b}\| \sin\theta$. (Hint: you can use $\vec{a} \cdot \vec{b} = \|\vec{a}\|\|\vec{b}\| \cos\theta$ to solve for θ and the proceed using components.)

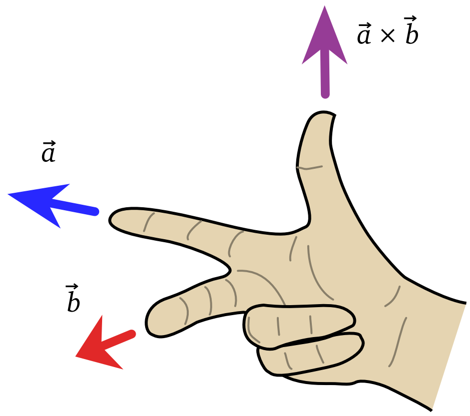
Now that we know what the cross product is and how to compute it, let's explore some of its incredible properties. First,

$$(\vec{a} \times \vec{b}) \cdot \vec{a} = (a_y b_z - a_z b_y)a_x - (a_x b_z - a_z b_x)a_y + (a_x b_y - a_y b_x)a_z = 0$$

and

$$(\vec{a} \times \vec{b}) \cdot \vec{b} = (a_y b_z - a_z b_y)b_x - (a_x b_z - a_z b_x)b_y + (a_x b_y - a_y b_x)b_z = 0.$$

Thus, $\vec{a} \times \vec{b}$ is orthogonal to both $\vec{a}$ and $\vec{b}$. Just based on this property, since the length of $\vec{a} \times \vec{b}$ is fixed, $\vec{a} \times \vec{b}$ can be one of two vectors in space. If we investigate further, we'll see that $\vec{a} \times \vec{b}$ is the vector that satisfies the *right-hand rule*.



[22]

A vector that encodes area, points orthogonally to others, and obeys the right-hand rule is handy indeed, and the cross product will be a useful tool for solving many problems.

[22] Image credit: Acdx, from Wikipedia https://en.wikipedia.org/wiki/Cross_product

Exercises for 2.10

1 Find $\vec{a} \times \vec{b}$ for the following pairs:

a) $\vec{a} = (4, -2, 0)$, $\vec{b} = (2, 1, -1)$
b) $\vec{a} = (3, 3, 0)$, $\vec{b} = (4, -3, 2)$
c) $\vec{a} = 2\vec{e}_1 + 3\vec{e}_2 + 4\vec{e}_3$, $\vec{b} = \vec{e}_1 - 3\vec{e}_2 + 4\vec{e}_3$.

Solution:

a) $(2, 4, 8)$
b) $(15, 6, -21)$
c) $24\vec{e}_1 - 4\vec{e}_2 - 9\vec{e}_3$

2 Use the cross product to find the areas of the following figures.

a) The parallelogram with vertices $(0, 0, 0)$, $(1, 1, 0)$, $(1, 2, 1)$ and $(0, 1, 1)$. (Perhaps you should first check that this is a parallelogram.)
b) The triangle with vertices $(1, 0, 0)$, $(0, 1, 0)$, and $(0, 0, 1)$.

Solution:

a) $\sqrt{3}$
b) $\frac{\sqrt{3}}{2}$

3 Prove that the cross product is not associative by calculating $\vec{a} \times (\vec{b} \times \vec{c})$ and $(\vec{a} \times \vec{b}) \times \vec{c}$ for $\vec{a} = \vec{e}_1$, $\vec{b} = \vec{c} = \vec{e}_2$.

Solution: $\vec{e}_1 \times (\vec{e}_2 \times \vec{e}_2) = \vec{e}_1 \times \vec{0} = \vec{0}$ while $(\vec{e}_1 \times \vec{e}_2) \times \vec{e}_2 = \vec{e}_3 \times \vec{e}_2 = -\vec{e}_1$

4 Verify the formula.

$$\|\vec{a} \times \vec{b}\|^2 = \|\vec{a}\|^2 \|\vec{b}\|^2 - (\vec{a} \cdot \vec{b})^2.$$

Hint: Use the definitions in terms of the sine and cosine of the included angle θ.

Solution:

$$\begin{aligned}
\|\vec{a} \times \vec{b}\|^2 &= (\|\vec{a}\|\|\vec{b}\| \sin\theta)^2 \\
&= \|\vec{a}\|^2\|\vec{b}\|^2 \sin^2\theta \\
&= \|\vec{a}\|^2\|\vec{b}\|^2 (1 - \cos^2\theta) \\
&= \|\vec{a}\|^2\|\vec{b}\|^2 - \|\vec{a}\|^2\|\vec{b}\|^2 \cos^2\theta \\
&= \|\vec{a}\|^2\|\vec{b}\|^2 - (\|\vec{a}\|\|\vec{b}\| \cos\theta)^2 \\
&= \|\vec{a}\|^2\|\vec{b}\|^2 - (\vec{a} \cdot \vec{b})^2.
\end{aligned}$$

Chapter 3

Geometry & Equations

Exploration Questions

3.1 Matrices

Talking about matrices now may seem like a non sequitur, but we will soon be using them as a notational device.

Definition 3.1.1 — Matrix. An $m \times n$ *matrix* is a rectangular array of numbers with m rows and n columns, usually surrounded by brackets.

The *dimensions*, *shape*, or *size* of a matrix refers to the number of rows and columns in a matrix and is always listed as #rows and then #columns. The most common way to specify the size of a matrix is by writing "rows × columns".

We are already familiar with certain matrices. When we write down a column vector like $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$, we are writing down a 2×1 matrix. It is a theorem that a vector $\vec{v} \in \mathbb{R}^n$ can be represented by an $n \times 1$ matrix.

We can also *index* a matrix—that is, refer to particular entries in the matrix. An $m \times n$ matrix A takes the form

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{bmatrix}.$$

Here, a_{ij} refers to the number in the ith row and jth column of A.[1] We can use *index notation* to define a matrix. For example, we can define a 2×3 matrix $B = [b_{ij}]$ where $b_{ij} = i + j$. In this case,

$$B = [b_{ij}] = \begin{bmatrix} 2 & 3 & 4 \\ 3 & 4 & 5 \end{bmatrix}.$$

Writing $B = [b_{ij}]$ is shorthand for "B is the matrix whose i,j entry is b_{ij}".

Matrices have no intrinsic meaning—*they are just boxes of numbers*. But, we can use matrices to represent things like vectors, coefficients of equations, grocery lists, etc..

Special Matrices

There are two special matrices that will come up often.

Definition 3.1.2 — The Identity Matrix. The $n \times n$ *identity matrix*, written $I_{n\times n}$ is the $n \times n$ matrix with ones along the diagonal and zeros everywhere else.

[1] It would be clearer to write $a_{i,j}$, but it is tradition to omit the comma.

Some examples are

$$I_{2\times 2} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \text{and} \qquad I_{3\times 3} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Identity matrices are always square, and when it is obvious from context what the size must be, we omit the subscript and simply write I.

Definition 3.1.3 — The Zero Matrix. The $m \times n$ *zero matrix*, written $0_{m\times n}$ is the matrix of all zeros.

Some examples are

$$0_{2\times 2} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \qquad \text{and} \qquad 0_{3\times 2} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

Again, when the size of a zero matrix is obvious from context, we omit the subscript and simply write 0.

Block Notation

Occasionally we want to create new matrices by stacking existing matrices. These are called *block* matrices. Typically, but not always, we draw lines to emphasize there is something special about the entry of a block matrix.

For example, if $M = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, then

$$\left[\begin{array}{c|c} 0_{2\times 2} & I_{2\times 2} \\ \hline I_{2\times 2} & M \end{array}\right] = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 2 \\ 0 & 1 & 3 & 4 \end{bmatrix}.$$

We can also read block notation backwards to define a matrix. For example, we can write

$$\left[\begin{array}{c|c} 7 & \vec{v} \\ \hline \vec{u} & M \end{array}\right] = \begin{bmatrix} 7 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 2 \\ 0 & 1 & 3 & 4 \end{bmatrix},$$

which defines the matrix $M = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 2 \\ 1 & 3 & 4 \end{bmatrix}$, the column vector $\vec{u} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ and the row vector $\vec{v} = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}$.

3.2 Systems of Linear Equations

Consider the vector equation

$$t\vec{u}+s\vec{v}+r\vec{w}=\vec{p} \qquad \text{where} \qquad \vec{u}=\begin{bmatrix}1\\2\\1\end{bmatrix},\ \vec{v}=\begin{bmatrix}2\\1\\-4\end{bmatrix},\ \vec{w}=\begin{bmatrix}-2\\-5\\1\end{bmatrix},\ \vec{p}=\begin{bmatrix}-15\\-21\\18\end{bmatrix}. \tag{3.1}$$

A *solution* to this equation is values of t, s, and r that make the equation true. There are many ways to find t, s, and r, but one way that always works is by equating components of each vector in the equation. By equating components in equation (3.1), we get the following system:

$$\begin{aligned} t+2s-2r&=-15 &\qquad \mathrm{row}_1\\ 2t+\ s-5r&=-21 &\qquad \mathrm{row}_2\\ t-4s+\ r&=18 &\qquad \mathrm{row}_3 \end{aligned} \tag{3.2}$$

The system (3.2) could be solve by *substitution*: solve the first equation for t; substitute t into the second two equation which then would contain s and r as the only unknowns; solve the second equation for s; substitute s into the last equation which now contains r as the only unknown; solve for r; work backwards plugging in r to get s, and finally plugging in r and s to get t.

We will instead solve system (3.2) by *elimination*.[2] Observe the following: if $A=B$ and $C=D$, then $A+\alpha C=B+\alpha D$ for any α. Using this fact, we can eliminate unknowns by summing equations rather than by substituting.

$$\begin{cases} t+2s-2r=-15\\ 2t+\ s-5r=-21\\ t-4s+\ r=18 \end{cases} \xrightarrow{\mathrm{row}_3\mapsto\mathrm{row}_3-\mathrm{row}_1} \begin{cases} t+2s-2r=-15\\ 2t+\ s-5r=-21\\ \quad -6s+3r=33 \end{cases}$$

$$\xrightarrow{\mathrm{row}_2\mapsto\mathrm{row}_2-2\mathrm{row}_1} \begin{cases} t+2s-2r=-15\\ \quad -3s-\ r=9\\ \quad -6s+3r=33 \end{cases}$$

$$\xrightarrow{\mathrm{row}_3\mapsto\mathrm{row}_3-2\mathrm{row}_3} \begin{cases} t+2s-2r=-15\\ \quad -3s-\ r=9\\ \qquad\quad 5r=15 \end{cases}$$

At this point, we have eliminated all but one unknown from the last equation. By inspection, we see that $r=3$; substituting r into the second equation gives $s=-4$; finally, substituting both r and s into the first equation gives $t=-1$.

The benefit of elimination over substitution is that elimination is *algorithmic* (that is, you could program a computer to do it) and can be made notationally convenient.

[2] Elimination is sometimes referred to as *Gaussian elimination* or *Gauss-Jordan elimination*.

Row Reduction

Recall system (3.2):

$$\begin{aligned} t + 2s - 2r &= -15 \qquad \text{row}_1 \\ 2t + \ \ s - 5r &= -21 \qquad \text{row}_2 \\ t - 4s + \ \ r &= 18 \qquad \text{row}_3 \end{aligned}$$

When performing elimination, there was a lot of redundant information. In particular, the variables and the "=" never changed—it was only the numbers that changed. To make our lives easier, we will write a this system as an *augmented matrix*.[3]

$$\begin{aligned} t + 2s - 2r &= -15 \\ 2t + \ \ s - 5r &= -21 \\ t - 4s + \ \ r &= 18 \end{aligned} \qquad \text{corresponds to} \qquad \left[\begin{array}{rrr|r} 1 & 2 & -2 & -15 \\ 2 & 1 & -5 & -21 \\ 1 & -4 & 1 & 18 \end{array}\right]$$

We use a vertical line in our matrix to separate coefficients of our unknowns from numbers on the right side of the "=". Written with matrices, elimination becomes easier to perform by hand.

$$\left[\begin{array}{rrr|r} 1 & 2 & -2 & -15 \\ 2 & 1 & -5 & -21 \\ 1 & -4 & 1 & 18 \end{array}\right] \xrightarrow{\text{row}_3 \mapsto \text{row}_3 - \text{row}_1} \left[\begin{array}{rrr|r} 1 & 2 & -2 & -15 \\ 2 & 1 & -5 & -21 \\ 0 & -6 & 3 & 33 \end{array}\right]$$

$$\xrightarrow{\text{row}_2 \mapsto \text{row}_2 - 2\text{row}_1} \left[\begin{array}{rrr|r} 1 & 2 & -2 & -15 \\ 0 & -3 & -1 & 9 \\ 0 & -6 & 3 & 33 \end{array}\right]$$

$$\xrightarrow{\text{row}_3 \mapsto \text{row}_3 - 2\text{row}_3} \left[\begin{array}{rrr|r} 1 & 2 & -2 & -15 \\ 0 & -3 & -1 & 9 \\ 0 & 0 & 5 & 15 \end{array}\right]$$

From here, we can solve the original system by substitution—the last row of the matrix corresponds to the equation $5r = 15$, just as before. Notice that if we multiplied the last row of the augmented matrix by $\frac{1}{5}$, we would be left with the equation $r = 3$ and could continue eliminating. Since working with augmented matrices is so fun, let's keep eliminating!

$$\left[\begin{array}{rrr|r} 1 & 2 & -2 & -15 \\ 0 & -3 & -1 & 9 \\ 0 & 0 & 5 & 15 \end{array}\right] \xrightarrow{\text{row}_3 \mapsto \frac{1}{5}\text{row}_3} \left[\begin{array}{rrr|r} 1 & 2 & -2 & -15 \\ 0 & -3 & -1 & 9 \\ 0 & 0 & 1 & 3 \end{array}\right] \xrightarrow{\text{row}_2 \mapsto \text{row}_2 + \text{row}_3} \left[\begin{array}{rrr|r} 1 & 2 & -2 & -15 \\ 0 & -3 & 0 & 12 \\ 0 & 0 & 1 & 3 \end{array}\right]$$

$$\xrightarrow{\text{row}_1 \mapsto \text{row}_1 + 2\text{row}_3} \left[\begin{array}{rrr|r} 1 & 2 & 0 & -9 \\ 0 & -3 & 0 & 12 \\ 0 & 0 & 1 & 3 \end{array}\right] \xrightarrow{\text{row}_2 \mapsto \frac{-1}{3}\text{row}_2} \left[\begin{array}{rrr|r} 1 & 2 & 0 & -9 \\ 0 & 1 & 0 & -4 \\ 0 & 0 & 1 & 3 \end{array}\right]$$

$$\xrightarrow{\text{row}_1 \mapsto \text{row}_1 - 2\text{row}_2} \left[\begin{array}{rrr|r} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -4 \\ 0 & 0 & 1 & 3 \end{array}\right]$$

[3] A *matrix* is just a box of numbers. An *augmented matrix* is a matrix with an extra column.

Now we have something truly special. If we turn the augmented matrix back into a system of equations, we have that

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -4 \\ 0 & 0 & 1 & 3 \end{array}\right] \qquad \text{corresponds to} \qquad \begin{aligned} t \qquad\quad &= -1 \\ s \quad &= -4 \\ r &= 3 \end{aligned}$$

The system's solution can be read right from the matrix!

The process of writing a system as an augmented matrix and manipulating the rows of the matrix until a simpler system is obtained is called *row reduction*.

The Row-reduction Algorithm

The beauty of row reduction is that it can be done to any matrix. Not only that, but there is an algorithm, called *Gauss's algorithm*,[4] that allows you to perform row reduction without the need for any creativity.

First, we present the possible operations used in row reduction.

Definition 3.2.1 — Elementary Row Operations. The *elementary row operations* are the three operations of

1. swapping two rows (written $\mathrm{row}_i \longleftrightarrow \mathrm{row}_j$);
2. multiplying a row by a non-zero scalar (written $\mathrm{row}_i \mapsto \alpha\,\mathrm{row}_i$); and
3. adding a multiple of one row to a different row (written $\mathrm{row}_i \mapsto \mathrm{row}_i + \alpha\,\mathrm{row}_j$).

Theorem 3.2.2 Every elementary row operation has an inverse which is also an elementary row operation. In other words, every elementary row operation can be undone.

Proof. We can explicitly write down the inverses of each elementary row operation. $\mathrm{row}_i \longleftrightarrow \mathrm{row}_j$ is its own inverse. Since $\alpha \neq 0$, we have that $\mathrm{row}_i \mapsto \alpha\,\mathrm{row}_i$ and $\mathrm{row}_i \mapsto \frac{1}{\alpha}\,\mathrm{row}_i$ are inverses. Finally, $\mathrm{row}_i \mapsto \mathrm{row}_i + \alpha\,\mathrm{row}_j$ and $\mathrm{row}_i \mapsto \mathrm{row}_i - \alpha\,\mathrm{row}_j$ are inverses. ∎

Definition 3.2.3 — Equivalent Systems. Two systems of equations, (X) and (Y), are called *equivalent*, written $(X) \sim (Y)$ if they have exactly the same solution(s).
Two matrices are called equivalent if their corresponding systems are equivalent.

Theorem 3.2.4 Let (X) be a system of equation and let (Y) be the result of applying any number of elementary row operations to (X). Then, (X) and (Y) are equivalent systems.

Proof. Let (X) be a system of equations and let $R(X)$ be the result of applying a single row operation to (X). Because equivalence of systems is transitive,[5] it is sufficient to show $(X) \sim R(X)$.

[4] Gauss popularized this algorithm in the West, but it was known to the ancient Chinese long before Gauss's time.

[5] That is, if $(X) \sim (Y)$ and $(Y) \sim (Z)$, then $(X) \sim (Z)$

Note that every row operation obeys the *law of algebraic manipulation*. That is, a row operation takes a list of true statements (the equations) to another list of true statements (an new list of equations). As such, if the statement

$$\vec{x} \text{ is a solution to } (X)$$

is true, then the statement

$$\vec{x} \text{ is a solution to } R(X)$$

must also be true. Phrased compactly, the law of algebraic manipulation ensures that

$$\vec{x} \text{ is a solution to } (X) \qquad \text{implies} \qquad \vec{x} \text{ is a solution to } R(X).$$

To show $(X) \sim R(X)$, we need to prove that

$$\vec{x} \text{ is a solution to } R(X) \qquad \text{implies} \qquad \vec{x} \text{ is a solution to } (X).$$

Fix a row operation R and let R^{-1} denote its inverse. Since R^{-1} is also an elementary row operation, by the law of algebraic manipulation, we have that

$$\vec{x} \text{ is a solution to } (X) \qquad \text{implies} \qquad \vec{x} \text{ is a solution to } R^{-1}(X)$$

and so

$$\vec{x} \text{ is a solution to } R(X) \qquad \text{implies} \qquad \vec{x} \text{ is a solution to } R \circ R^{-1}(X) = (X).$$

Thus, $(X) \sim R(X)$. ■

Now that we have established that applying elementary row operations produces equivalent systems, we are ready to give a row-reduction algorithm.

Definition 3.2.5 — Row Reduction Algorithm. Let M be a matrix.

1. If M takes the form $M = [\vec{0}|M']$ (that is, its first column is all zeros), apply the algorithm to M';
2. if not, perform a row-swap so the upper-left entry of M is non-zero.
3. Perform the row operation $\mathrm{row}_1 \mapsto \frac{1}{\alpha}\mathrm{row}_1$ where α is the upper-left entry of M. This entry is referred to as a *pivot*.
4. Use the row operation $\mathrm{row}_i \mapsto \mathrm{row}_i - \beta\mathrm{row}_1$ to zero every entry below the pivot.
5. Now M has the form
$$M = \left[\begin{array}{c|c} 1 & ?? \\ \hline \vec{0} & M' \end{array}\right].$$
Apply the algorithm to M'.

The resulting matrix is now in *row echelon form*. To put the matrix in *reduced row echelon form*, additionally apply step 6.

6. Use the row operation $\text{row}_i \mapsto \text{row}_i + \alpha\text{row}_j$ to zero above each pivot.

Stated all at once, this algorithm can be hard to follow, but with practice it becomes straightforward.

■ **Example 3.1** Apply the row-reduction algorithm to the matrix

$$M = \begin{bmatrix} 0 & 0 & 0 & -2 & -2 \\ 0 & 1 & 2 & 3 & 2 \\ 0 & 2 & 4 & 5 & 3 \end{bmatrix}.$$

First notice that M starts with a column of zeros, so we will focus on the right side of M. We will draw a line to separate it.

$$M = \left[\begin{array}{c|cccc} 0 & 0 & 0 & -2 & -2 \\ 0 & 1 & 2 & 3 & 2 \\ 0 & 2 & 4 & 5 & 3 \end{array}\right]$$

Next, we perform a row swap to bring a non-zero entry to the upper left.

$$\left[\begin{array}{c|cccc} 0 & 0 & 0 & -2 & -2 \\ 0 & 1 & 2 & 3 & 2 \\ 0 & 2 & 4 & 5 & 3 \end{array}\right] \xrightarrow{\text{row}_1 \leftrightarrow \text{row}_2} \left[\begin{array}{c|cccc} 0 & 1 & 2 & 3 & 2 \\ 0 & 0 & 0 & -2 & -2 \\ 0 & 2 & 4 & 5 & 3 \end{array}\right]$$

The upper-left entry is already a 1, so we can use it to zero all entries below.

$$\left[\begin{array}{c|cccc} 0 & 1 & 2 & 3 & 2 \\ 0 & 0 & 0 & -2 & -2 \\ 0 & 2 & 4 & 5 & 3 \end{array}\right] \xrightarrow{\text{row}_3 \mapsto \text{row}_3 - 2\text{row}_1} \left[\begin{array}{c|cccc} 0 & 1 & 2 & 3 & 2 \\ 0 & 0 & 0 & -2 & -2 \\ 0 & 0 & 0 & -1 & -1 \end{array}\right]$$

Now we work on the submatrix.

$$\left[\begin{array}{cc|ccc} 0 & 1 & 2 & 3 & 2 \\ \hline 0 & 0 & 0 & -2 & -2 \\ 0 & 0 & 0 & -1 & -1 \end{array}\right]$$

Again, the submatrix has a first column of zeros, so we pass to a sub-submatrix.

$$\left[\begin{array}{ccc|cc} 0 & 1 & 2 & 3 & 2 \\ \hline 0 & 0 & 0 & -2 & -2 \\ 0 & 0 & 0 & -1 & -1 \end{array}\right]$$

Now we turn the upper left entry into a 1 and use that pivot to zero all entries below.

$$\left[\begin{array}{ccc|cc} 0 & 1 & 2 & 3 & 2 \\ \hline 0 & 0 & 0 & -2 & -2 \\ 0 & 0 & 0 & -1 & -1 \end{array}\right] \xrightarrow{\text{row}_2 \mapsto \frac{-1}{2}\text{row}_2} \left[\begin{array}{ccc|cc} 0 & 1 & 2 & 3 & 2 \\ \hline 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & -1 & -1 \end{array}\right] \xrightarrow{\text{row}_3 \mapsto \text{row}_3 + \text{row}_2} \left[\begin{array}{ccc|cc} 0 & 1 & 2 & 3 & 2 \\ \hline 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

The matrix is now in row echelon form. To put it in reduced row echelon form, we zero above each pivot.

$$\begin{bmatrix} 0 & 1 & 2 & 3 & 2 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \xrightarrow{\text{row}_1 \mapsto \text{row}_1 - 3\text{row}_2} \begin{bmatrix} 0 & 1 & 2 & 0 & -1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

■

The row-reduction algorithm applies to any matrix, whether augmented or not. However, whether or not the reduced form of a matrix has any *meaning* depends on where the matrix came from in the first place.

Geometry of Systems of Linear Equations

We encountered systems of linear equations when trying to answer questions about span and linear independence, but they are worth thinking about in their own right. Consider the system

$$\begin{aligned} x + 2y - 2z &= -15 \\ 2x + y - 5z &= -21 \\ x - 4y + z &= 18 \end{aligned} \tag{3.3}$$

A solution to this system is a tuple (x, y, z) that satisfies every equation simultaneously. Taken individually, each equation in this system represents a plane. Let

$$\begin{aligned} \mathcal{P}_1 &= \left\{(x, y, z) \in \mathbb{R}^3 \;:\; x + 2y - 2z = -15\right\} \\ \mathcal{P}_2 &= \left\{(x, y, z) \in \mathbb{R}^3 \;:\; 2x + y - 5z = -21\right\} \\ \mathcal{P}_3 &= \left\{(x, y, z) \in \mathbb{R}^3 \;:\; x - 4y + z = 18\right\} \end{aligned}$$

be the planes corresponding to the equations in system (3.3). We can think of solutions to system (3.3) as points in $\mathcal{P}_1 \cap \mathcal{P}_2 \cap \mathcal{P}_3$.

Geometrically, we now have an immediate intuition for the what the solution sets to systems with three equations and three unknowns look like. Three planes can either intersect in a plane (if they are all the same plane), a line, a point, or they may not have a common intersection.

XXX Figure

If a system of equations has only two unknowns, then the solution set must look like the intersection of lines. Namely, the solution set is a line, a point, or empty.

XXX Figure

This logic generalizes to higher dimensions: the set of solutions to a system of linear equations is always a "flat" object (a line, plane, volume, etc.), a point, or empty. This is captured in the following theorem, which we will prove later.

Theorem 3.2.6 Let (X) be a system of linear equations. Then (X) either has no solutions, one solution, or infinitely many solutions.

Free Variables

By now we are very familiar with the system

$$\begin{aligned} x + 2y - 2z &= -15 \\ 2x + y - 5z &= -21 \\ x - 4y + z &= 18 \end{aligned} \tag{3.4}$$

which has a solution $(x,y,z) = (-1,-4,3)$. When we use the row reduction algorithm on an augmented matrix, we get

$$\left[\begin{array}{rrr|r} 1 & 2 & -2 & -15 \\ 2 & 1 & -5 & -21 \\ 1 & -4 & 1 & 18 \end{array}\right] \quad \sim \quad \left[\begin{array}{rrr|r} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -4 \\ 0 & 0 & 1 & 3 \end{array}\right],$$

and we can read the unique solution directly from the matrix. But what happens when there isn't a unique solution?

Consider the system

$$\begin{aligned} x + 3y &= 2 \\ 2x + 6y &= 4 \end{aligned}. \tag{3.5}$$

When using an augmented matrix to solve this system, we run into an issue.

$$\left[\begin{array}{rr|r} 1 & 3 & 2 \\ 2 & 6 & 4 \end{array}\right] \quad \sim \quad \left[\begin{array}{rr|r} 1 & 3 & 2 \\ 0 & 0 & 0 \end{array}\right]$$

From the reduced row echelon form we're left with the equation $x + 3y = 2$, which isn't exactly a *solution*. Effectively, the original system had only one equation's worth of information, so we cannot solve for both x and y based on the original system. To get ourselves out of this pickle, we will use a notational trick: introduce the arbitrary equation $y = t$. Now, because we've already done row-reduction, we see

$$\begin{aligned} x + 3y &= 2 \\ 2x + 6y &= 4 \\ y &= t \end{aligned} \quad \sim \quad \begin{aligned} x + 3y &= 2 \\ y &= t \end{aligned}.$$

Here we've omitted the equation $0 = 0$ since it adds no information. We can write the solution to this system as

$$\vec{x} = \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2 - 3t \\ t \end{bmatrix} = t\begin{bmatrix} -3 \\ 1 \end{bmatrix} + \begin{bmatrix} 2 \\ 0 \end{bmatrix}.$$

Notice that t here stands for an arbitrary real number. And choice of t produces a valid solution to the original system (go ahead, pick some values for t and see what happens). We call t a *parameter* and y a *free variable*.[6] Notice further that

$$\vec{x} = t\begin{bmatrix} -3 \\ 1 \end{bmatrix} + \begin{bmatrix} 2 \\ 0 \end{bmatrix}$$

[6] We call y *free* because we may pick it to be anything we want and still produce a solution to the system.

is vector form of the line $x+3y=2$.

If a system of equations has infinitely many solutions, solving it will require picking a free variable. You have a lot of choice over which variables you pick to be free, but there is an algorithmic way to pick free variables that leaves no room for failure.

Definition 3.2.7 — Canonical Choice of Free Variables. Let (X) be a system of linear equations in the variables $x_1,\dots,x_n$, and let R be the corresponding row-reduced matrix. The variable x_i is a *canonical free variable* if the ith column of R does not contain a pivot.

By picking assigning parameters to the canonical free variables, you are guaranteed to be able to write down all solutions to a system of linear equations.

■ **Example 3.2** XXX Finish ■

XXX Finish. Complete solutions, etc.?

Consistent and Inconsistent Systems

XXX Finish

3.3 Subspaces & Bases

Lines or planes through the origin can be written as spans of their direction vectors. However, a line or plane that doesn't pass through the origin cannot be written as a span—it must be expressed as a *translated* span.

XXX Figure

There's something special about sets that can be expressed as (untranslated) spans. In particular, since a linear combination of linear combinations is still a linear combination, a span is *closed* with respect to linear combinations. That is, by taking linear combinations of vectors in a span, you cannot escape the span. In general, sets that have this property are called *subspaces*.

Definition 3.3.1 — Subspace. A subset $\mathcal{V}\subseteq\mathbb{R}^n$ is called a *subspace* if the following two properties are satisfied.

(i) $\vec v\in\mathcal{V}$ implies $\alpha\vec v\in\mathcal{V}$ for all α.

(ii) $\vec u,\vec v\in\mathcal{V}$ implies $\vec u+\vec v\in\mathcal{V}$.

The properties listed in the definition of a subspace are a succinct way of stating that a linear combinations of vectors in a subspace remain in that subspace.[7] Subspaces generalize the idea of *flat spaces through the origin*. They encompass lines, planes, volumes and more.

■ **Example 3.3** Let $\mathcal{V}\subseteq\mathbb{R}^2$ be the complete solution to $x+2y=0$. Show that $\mathcal{V}$ is a subspace.

XXX Finish ■

■ **Example 3.4** Let $\mathcal{W}\subseteq\mathbb{R}^2$ be the line expressed in vector form as

$$\vec x=t\begin{bmatrix}1\\2\end{bmatrix}+\begin{bmatrix}1\\1\end{bmatrix}.$$

[7] An even more succinct way of defining a subspace $\mathcal{V}$ is that if $\vec u,\vec v\in\mathcal{V}$, then $\alpha\vec u+\vec v\in\mathcal{V}$ for all scalars α.

Determine whether $\mathcal{W}$ is a subspace.

XXX Finish ■

As mentioned earlier, subspaces and spans are deeply connected by the following theorem.

Theorem 3.3.2 Every subspace is a span and every span is a subspace. More precisely, $\mathcal{V} \subseteq \mathbb{R}^n$ is a subspace if and only if $\mathcal{V} = \operatorname{span} \mathcal{X}$ for some set $\mathcal{X}$.

Proof. We will start by showing every span is a subspace. Fix $\mathcal{X} \subseteq \mathbb{R}^2$ and let $\mathcal{V} = \operatorname{span} \mathcal{X}$. By definition, $\vec{v} \in \mathcal{V}$ means that

$$\vec{v} = \sum \alpha_i \vec{x}_i$$

for some $\vec{x}_i \in \mathcal{X}$ and scalars α_i. It follows that

$$\alpha \vec{v} = \alpha \sum \alpha_i \vec{x}_i = \sum (\alpha \alpha_i) \vec{x}_i \in \operatorname{span} \mathcal{X} = \mathcal{V}.$$

Similarly, if $\vec{u} \in \mathcal{V}$, then

$$\vec{u} = \sum \beta_i \vec{y}_i$$

for some $\vec{y}_i \in \mathcal{X}$ and scalars β_i. It follows that

$$\vec{u} + \vec{v} = \sum \alpha_i \vec{x}_i + \sum \beta_i \vec{y}_i$$

is also a linear combination of vectors in $\mathcal{X}$, and so $\vec{u} + \vec{v} \in \operatorname{span} \mathcal{X} = \mathcal{V}$. Thus, $\mathcal{V}$ is a subspace.

Now we will prove that every subspace is a span. Let $\mathcal{V}$ be a subspace and consider $\mathcal{V}' = \operatorname{span} \mathcal{V}$. We know that $\mathcal{V} \subseteq \mathcal{V}'$. If we establish that $\mathcal{V}' \subseteq \mathcal{V}$, then $\mathcal{V} = \mathcal{V}' = \operatorname{span} \mathcal{V}$, which would complete the proof.

Fix $\vec{x} \in \mathcal{V}'$. By definition,

$$\vec{x} = \sum \alpha_i \vec{v}_i$$

for some $\vec{v}_i \in \mathcal{V}$ and scalars α_i. Observe that $\alpha_i \vec{v}_i \in \mathcal{V}$ since $\mathcal{V}$ is closed under scalar multiplication. Thus, $\alpha_1 \vec{v}_1 + \alpha_2 \vec{v}_2 \in \mathcal{V}$ because $\mathcal{V}$ is closed under sums. Continuing, $(\alpha_1 \vec{v}_1 + \alpha_2 \vec{v}_2) + \alpha_3 \vec{v}_3 \in \mathcal{V}$ because $\mathcal{V}$ is closed under sums. It follows that

$$\vec{x} = \sum \alpha_i \vec{v}_i = \Big(\big((\alpha_1 \vec{v}_1 + \alpha_2 \vec{v}_2) + \alpha_3 \vec{v}_3 \big) + \cdots + \alpha_{n-1} \vec{v}_{n-1} \Big) + \alpha_n \vec{v}_n \in \mathcal{V}.$$

Thus $\mathcal{V}' \subseteq \mathcal{V}$, which completes the proof. ■

Basis

Let $\vec{d} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and consider $\ell = \operatorname{span}\{\vec{d}\}$.

XXX Figure

We know that since ℓ is a subspace, $\ell = \operatorname{span} \ell$. However, the simplest descriptions of ℓ involve the span of only one vector.

Analogously, let $\mathcal{P} = \operatorname{span}\{\vec{d}_1, \vec{d}_2\}$ be the plane through the origin with direction vectors $\vec{d}_1$ and $\vec{d}_2$. There are many ways to write $\mathcal{P}$ as a span, but the simplest ones involve exactly two vectors. The simplest way to describe a subspace as a span is captured in the idea of *basis*.

Definition 3.3.3 — Basis. Let $\mathcal{V}$ be a subspace. A *basis* for $\mathcal{V}$ is a linearly independent set $\mathcal{B}$ such that $\mathcal{V} = \operatorname{span}\mathcal{B}$.

In short, a basis for a subspace is a linearly independent set that spans that subspace.

■ **Example 3.5** Let $\ell = \operatorname{span}\left\{\begin{bmatrix}1\\2\end{bmatrix}, \begin{bmatrix}-2\\-4\end{bmatrix}, \begin{bmatrix}1/2\\1\end{bmatrix}\right\}$. Find a basis for ℓ.

XXX Finish ■

Bases for subspaces are not unique. For a line through the origin, any non-zero direction vector serves as a basis. However, every basis must have the same size.

Theorem 3.3.4 Let $\mathcal{X}$ and $\mathcal{Y}$ both be bases for the subspace $\mathcal{V}$. Then $|\mathcal{X}| = |\mathcal{Y}|$. That is, $\mathcal{X}$ and $\mathcal{Y}$ have the same size.

Proof. Let $\mathcal{X} = \{\vec{x}_1, \dots, \vec{x}_m\}$ and $\mathcal{Y} = \{\vec{y}_1, \dots, \vec{y}_n\}$ both be bases for $\mathcal{V}$. Without loss of generality, assume $|\mathcal{X}| > |\mathcal{Y}|$. We will replace the elements of $\mathcal{X}$ with elements of $\mathcal{Y}$ one by one.

Consider the set $\mathcal{X} \cup \{\vec{y}_1\}$. Since $\vec{y}_1 \in \mathcal{V} = \operatorname{span}\mathcal{X}$, the set $\mathcal{X} \cup \{\vec{y}_1\}$ must be linearly dependent. Further, since $\mathcal{Y}$ is a basis, $\{\vec{y}_1\}$ is a linearly independent set. We conclude that there must exist some $0 \le r_1 \le m$ so that

$$\vec{x}_{r_1} \in \operatorname{span}(\mathcal{X} \cup \{y_1\}) \backslash \{x_{r_1}\}.$$

Let $\mathcal{B}_1 = (X \cup \{\vec{y}_1\}) \backslash \{\vec{x}_{r_1}\}$ and note that $|\mathcal{B}_1| = |\mathcal{X}|$ and that $\mathcal{V} = \operatorname{span}\mathcal{B}_1$.

Continuing, consider $\mathcal{B}_1 \cup \{\vec{y}_2\}$. Since $\{\vec{y}_1, \vec{y}_2\}$ is a linearly independent set, there must exist some $0 \le r_2 \le m$ with $r_2 \ne r_1$ so that

$$\vec{x}_{r_2} \in \operatorname{span}(\mathcal{B}_1 \cup \{y_2\}) \backslash \{\vec{x}_{r_2}\}.$$

Let $\mathcal{B}_2 = (\mathcal{B}_1 \cup \{\vec{y}_2\}) \backslash \{\vec{x}_{r_2}\}$.

Continuing in this way, construct the sets $\mathcal{B}_3, \dots, \mathcal{B}_n$ where

$$\mathcal{B}_n = \mathcal{Y} \cup (\mathcal{X} \backslash \{\vec{x}_{r_1}, \dots, \vec{x}_{r_n}\}).$$

Since $|\mathcal{X}| > |\mathcal{Y}|$, we know $\mathcal{Y} \subsetneq \mathcal{B}_n$. Since $\mathcal{Y}$ is a basis, we conclude that $\mathcal{B}_n$ must be linearly dependent. However, any dependency relationship in $\mathcal{B}_n$ corresponds to a dependency relationship in $\mathcal{X}$, which is a contradiction. ■

Since all bases for a particular subspace have the same size, we can refer to *the* size of *a* basis for a subspace—this number is special and is called the *dimension* of the subspace.

Definition 3.3.5 — Dimension. Let $\mathcal{V}$ be a subspace. The *dimension* of $\mathcal{V}$ is the size of a basis for $\mathcal{V}$.

Using this definition, the dimension of a plane is 2, of a line is 1, and of a point is 0.[8]

■ **Example 3.6** Find the dimension of $\mathbb{R}^2$.

Since $\{\vec{e}_1, \vec{e}_2\}$ is a basis for $\mathbb{R}^2$, we know $\mathbb{R}^2$ is two dimensional. ■

[8] The dimension of a line, plane, or point not through the origin is defined to be the dimension of the subspace obtained when translating it to the origin.

■ **Example 3.7** Let $\ell = \operatorname{span}\left\{\begin{bmatrix}1\\2\end{bmatrix}, \begin{bmatrix}-2\\-4\end{bmatrix}, \begin{bmatrix}1/2\\1\end{bmatrix}\right\}$.

In Example 3.5, we found XXX was a basis for ℓ.

XXX Finish ■

Definition 3.3.6 — Standard Basis for $\mathbb{R}^n$. The *standard basis* for $\mathbb{R}^n$ are the vectors $\vec{e}_1, \ldots, \vec{e}_n$ where

$$\vec{e}_1 = \begin{bmatrix}1\\0\\0\\\vdots\end{bmatrix} \qquad \vec{e}_2 = \begin{bmatrix}0\\1\\0\\\vdots\end{bmatrix} \qquad \vec{e}_3 = \begin{bmatrix}0\\0\\1\\\vdots\end{bmatrix} \qquad \cdots.$$

That is $\vec{e}_i$ is the vector with a 1 in its ith component and zeros elsewhere.

The notation $\vec{e}_i$ is context specific. If we say $\vec{e}_i \in \mathbb{R}^2$, then $\vec{e}_i$ must have exactly two components. If we say $\vec{e}_i \in \mathbb{R}^{45}$, then $\vec{e}_i$ must have 45 components.

3.4 Matrix Equations

Images

There is one more important way to create geometric objects, and that is using functions and the idea of *image*.

Definition 3.4.1 — Image. Let $f : A \to B$ be a function. The *image* of a set $X \subseteq A$, written $f(X)$ is defined by

$$f(X) = \{y \in B \;:\; y = f(x) \text{ for some } x \in X\}.$$

In plain language, the image of a set X under a function f is the set of all outputs of f when only points from X are input.

■ **Example 3.8** Let $f : \mathbb{R}^2 \to \mathbb{R}^2$ be defined as $f(x, y) = (2x, y)$ and let $C = \{\vec{x} \in \mathbb{R}^2 : \|\vec{x}\| = 1\}$ be the unit circle. Find $f(C)$.

The image $f(C)$ consists of all the points of C after f is applied. Since f doubles the x-coordinate, we $f(C)$ will be C stretched in the x-direction by 2.

XXX Figure ■

Images allow exotic transformations of sets. For example, consider

$$f(x, y) = ((x + y)\sin x, (x + y)\cos x),$$

and let $X = \{t\vec{e}_1 \;:\; t \geq 0\}$ be the positive x-axis. The image $f(X)$ is a spiral!

XXX Figure

Nearly any geometric figure can be made using images and cleverly chosen functions, and multi-variable calculus studies how arbitrary functions (via taking images) change geometry. For our purposes, we will stick mainly with down to earth functions: linear functions[9] and

[9] Linear functions have the property that they always take lines to lines.

translations. But, we can still get a lot of utility out of images even in the restricted case of linear functions and translations.

Let $T = \{\vec{x} \in \mathbb{R}^2 \ : \ \vec{x} = \alpha\vec{e}_1 + \beta\vec{e}_2 \text{ where } \alpha, \beta \geq 0 \text{ and } \alpha + \beta \leq 1\}$ be a triangle and let $R : \mathbb{R}^2 \to \mathbb{R}^2$ be the function that rotates vectors counter-clockwise by $90°$. That is $R(x, y) = (-y, x)$. And, let $V : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $V(\vec{a}) = \vec{a} + \vec{e}_2$ be vertical translation. Then, we can start making some fun pictures!

XXX Figure (T, R(T), T, V of R(T), ...)

Chapter 4

Linear Transformations

Exploration Questions

Being able to picture multi-dimensional objects is invaluable during mathematical exploration. There are two main ways we visualize surfaces: perspective drawings and level-curves.

Chapter 5

Determinants

Exploration Questions

Integrals add things up, and iterated integrals are no different; it's just that instead of adding things up in one dimension, we add things up in multiple dimensions.

Chapter 6

Orthogonality

Exploration Questions

Part II

Linear Algebra II

Chapter 7

Vector Spaces

Exploration Questions

Being able to picture multi-dimensional objects is invaluable during mathematical exploration. There are two main ways we visualize surfaces: perspective drawings and level-curves.

Chapter 8

Linear Transformations

Exploration Questions

Being able to picture multi-dimensional objects is invaluable during mathematical exploration. There are two main ways we visualize surfaces: perspective drawings and level-curves.

Chapter 9

Matrices

Exploration Questions

Being able to picture multi-dimensional objects is invaluable during mathematical exploration. There are two main ways we visualize surfaces: perspective drawings and level-curves.

Chapter 10

Inner Product Spaces

Exploration Questions

Being able to picture multi-dimensional objects is invaluable during mathematical exploration. There are two main ways we visualize surfaces: perspective drawings and level-curves.

Appendix A

Proofs

Below are some guidelines to help you write proofs. The following rules apply whenever you write a proof.[1]

1. **The burden of communication lies on you, not on your reader.** It is your job to explain your thoughts; it is not your reader's job to guess them from a few hints. You are trying to convince a skeptical reader who doesn't believe you, so you need to argue with airtight logic in crystal clear language; otherwise the reader will continue to doubt. If you didn't write something on the paper, then (a) you didn't communicate it, (b) the reader didn't learn it, and (c) the grader has to assume you didn't know it in the first place.

2. **Tell the reader what you're proving.** The reader doesn't necessarily know or remember what "Theorem 2.13" is. Even a professor grading a stack of papers might lose track from time to time. Therefore, the statement you are proving should be on the same page as the beginning of your proof. For an exam this won't be a problem, of course, but on your homework, recopy the claim you are proving. This has the additional advantage that when you study for exams by reviewing your homework, you won't have to flip back in the notes/textbook to know what you were proving.

3. **Use English words.** Although there will usually be equations or mathematical statements in your proofs, use English sentences to connect them and display their logical relationships. If you look in your notes/textbook, you'll see that each proof consists mostly of English words.

4. **Use complete sentences.** If you wrote a history essay in sentence fragments, the reader would not understand what you meant; likewise, in mathematics you must use complete sentences with verbs to convey your logical train of thought.

   Some complete sentences can be written purely in mathematical symbols, such as equations (e.g., $a^3 = b^{-1}$), inequalities (e.g., $x < 5$), and other relations (like $5 \mid 10$ or $7 \in \mathbb{Z}$). These statements usually express a relationship between two mathematical *objects*, like numbers or sets. However, it is considered bad style to begin a sentence with symbols. A common phrase to use to avoid starting a sentence with mathematical symbols is "We see that..."

[1] This list is an adaptation of *The Elements of Style for Proofs* written by Anders Hendrickson of St. Norbert College and modified by Dana Ernst of Northern Arizona University.

5. **Show the logical connections among your sentences.** Use phrases like "Therefore" or "because" or "if..., then..." or "if and only if" to connect your sentences.

6. **Know the difference between statements and objects.** A mathematical object is a *thing*, a noun, such as a group, an element, a vector space, a number, an ordered pair, etc. Objects either exist or don't exist. Statements, on the other hand, are mathematical *sentences*: they can be true or false.

   When you see or write a cluster of math symbols, be sure you know whether it's an object (e.g., "$x^2 + 3$") or a statement (e.g., "$x^2 + 3 < 7$"). One way to tell is that every mathematical statement includes a verb, such as $=$, $\leq$, "divides", etc.

7. **"$=$" means equals.** Don't write $A = B$ unless you mean that A actually equals B. This rule seems obvious, but there is a great temptation to be sloppy. In calculus, for example, some people might write $f(x) = x^2 = 2x$ (which is false), when they really mean that "if $f(x) = x^2$, then $f'(x) = 2x$."

8. **Don't interchange $=$ and $\implies$.** The equals sign connects two *objects*, as in "$x^2 = b$"; the symbol "$\implies$" is an abbreviation for "implies" and connects two *statements*, as in "$ab = a \implies b = 1$." You should avoid using $\implies$ in your formal write-ups.

9. **Say exactly what you mean.** Just as the $=$ is sometimes abused, so too people sometimes write $A \in B$ when they mean $A \subseteq B$, or write $a_{ij} \in A$ when they mean that a_{ij} is an entry in matrix A. Mathematics is a very precise language, and there is a way to say exactly what you mean; find it and use it.

10. **Don't write anything unproven.** Every statement on your paper should be something you *know* to be true. The reader expects your proof to be a series of statements, each proven by the statements that came before it. If you ever need to write something you don't yet know is true, you *must* preface it with words like "assume," "suppose," or "if" (if you are temporarily assuming it), or with words like "we need to show that" or "we claim that" (if it is your goal). Otherwise the reader will think they have missed part of your proof.

11. **Write strings of equalities (or inequalities) in the proper order.** When your reader sees something like
$$A = B \leq C = D,$$
    he/she expects to understand easily why $A = B$, why $B \leq C$, and why $C = D$, and he/she expects the *point* of the entire line to be the more complicated fact that $A \leq D$. For example, if you were computing the distance d of the point $(12, 5)$ from the origin, you could write
$$d = \sqrt{12^2 + 5^2} = 13.$$
    In this string of equalities, the first equals sign is true by the Pythagorean theorem, the second is just arithmetic, and the *point* is that the first item equals the last item: $d = 13$.

    A common error is to write strings of equations in the wrong order. For example, if you were to write "$\sqrt{12^2 + 5^2} = 13 = d$", your reader would understand the first equals sign, would be baffled as to how we know $d = 13$, and would be utterly perplexed as to why you wanted or needed to go through 13 to prove that $\sqrt{12^2+5^2} = d$.

12. **Avoid circularity.** Be sure that no step in your proof makes use of the conclusion!

13. **Don't write the proof backwards.** Beginning students often attempt to write "proofs" like the following, which attempts to prove that $\tan^2(x) = \sec^2(x) - 1$:

$$\begin{aligned}
\tan^2(x) &= \sec^2(x) - 1 \\
\left(\frac{\sin(x)}{\cos(x)}\right)^2 &= \frac{1}{\cos^2(x)} - 1 \\
\frac{\sin^2(x)}{\cos^2(x)} &= \frac{1-\cos^2(x)}{\cos^2(x)} \\
\sin^2(x) &= 1 - \cos^2(x) \\
\sin^2(x) + \cos^2(x) &= 1 \\
1 &= 1
\end{aligned}$$

Notice what has happened here: the student *started* with the conclusion, and deduced the true statement "$1 = 1$." In other words, he/she has proved "If $\tan^2(x) = \sec^2(x) - 1$, then $1 = 1$," which is true but highly uninteresting.

Now this isn't a bad way of *finding* a proof. Working backwards from your goal often is a good strategy *on your scratch paper*, but when it's time to *write* your proof, you have to start with the hypotheses and work to the conclusion.

14. **Be concise.** Most students err by writing their proofs too short, so that the reader can't understand their logic. It is nevertheless quite possible to be too wordy, and if you find yourself writing a full-page essay, it's probably because you don't have a proof, but just an intuition. When you find a way to turn that intuition into a formal proof, it will be much shorter.

15. **Introduce every symbol you use.** If you use the letter "k," the reader should know exactly what k is. Good phrases for introducing symbols include "Let $n \in \mathbb{N}$," "Let k be the least integer such that...," "For every real number a...," and "Suppose that X is a counterexample."

16. **Use appropriate quantifiers (once).** When you introduce a variable $x \in S$, it must be clear to your reader whether you mean "for all $x \in S$" or just "for some $x \in S$." If you just say something like "$y = x^2$ where $x \in S$," the word "where" doesn't indicate whether you mean "for all" or "some."

Phrases indicating the quantifier "for all" include "Let $x \in S$"; "for all $x \in S$"; "for every $x \in S$"; "for each $x \in S$"; etc. Phrases indicating the quantifier "some" (or "there exists") include "for some $x \in S$"; "there exists an $x \in S$"; "for a suitable choice of $x \in S$"; etc.

On the other hand, don't introduce a variable more than once! Once you have said "Let $x \in S$," the letter x has its meaning defined. You don't *need* to say "for all $x \in S$" again, and you definitely should *not* say "let $x \in S$" again.

17. **Use a symbol to mean only one thing.** Once you use the letter x once, its meaning is fixed for the duration of your proof. You cannot use x to mean anything else.

18. **Don't "prove by example."** Most problems ask you to prove that something is true "for all"—You *cannot* prove this by giving a single example, or even a hundred. Your answer will need to be a logical argument that holds for *every example there possibly could be*.

19. **Write "Let $x = \ldots$," not "Let $\cdots = x$."** When you have an existing expression, say a^2, and you want to give it a new, simpler name like b, you should write "Let $b = a^2$," which means, "Let the new symbol b mean a^2." This convention makes it clear to the reader that b is the brand-new symbol and a^2 is the old expression he/she already understands.

    If you were to write it backwards, saying "Let $a^2 = b$," then your startled reader would ask, "What if $a^2 \neq b$?"

20. **Make your counterexamples concrete and specific.** Proofs need to be entirely general, but counterexamples should be concrete. When you provide an example or counterexample, make it as specific as possible. For a set, for example, you must name its elements, and for a function, you must give its rule. Do not say things like "θ could be one-to-one but not onto"; instead, provide an actual function θ that *is* one-to-one but not onto.

21. **Don't include examples in proofs.** Including an example very rarely adds anything to your proof. If your logic is sound, then it doesn't need an example to back it up. If your logic is bad, a dozen examples won't help it (see rule 18). There are only two valid reasons to include an example in a proof: if it is a *counterexample* disproving something, or if you are performing complicated manipulations in a general setting and the example is just to help the reader understand what you are saying.

22. **Use scratch paper.** Finding your proof will be a long, potentially messy process, full of false starts and dead ends. Do all that on scratch paper until you find a real proof, and only then break out your clean paper to write your final proof carefully. *Do not hand in your scratch work!*

    Only sentences that actually contribute to your proof should be part of the proof. Do not just perform a "brain dump," throwing everything you know onto the paper before showing the logical steps that prove the conclusion. *That is what scratch paper is for.*

Appendix B

Creative Commons License Fulltext

Creative Commons Legal Code

Attribution-ShareAlike 4.0 International

Official translations of this license are available in other languages.

Creative Commons Corporation ("Creative Commons") is not a law firm and does not provide legal services or legal advice. Distribution of Creative Commons public licenses does not create a lawyer-client or other relationship. Creative Commons makes its licenses and related information available on an as-is basis. Creative Commons gives no warranties regarding its licenses, any material licensed under their terms and conditions, or any related information. Creative Commons disclaims all liability for damages resulting from their use to the fullest extent possible.

Using Creative Commons Public Licenses

Creative Commons public licenses provide a standard set of terms and conditions that creators and other rights holders may use to share original works of authorship and other material subject to copyright and certain other rights specified in the public license below. The following considerations are for informational purposes only, are not exhaustive, and do not form part of our licenses.

> **Considerations for licensors:** Our public licenses are intended for use by those authorized to give the public permission to use material in ways otherwise restricted by copyright and certain other rights. Our licenses are irrevocable. Licensors should read and understand the terms and conditions of the license they choose before applying it. Licensors should also secure all rights necessary before applying our licenses so that the public can reuse the material as expected. Licensors should clearly mark any material not subject to the license. This includes other CC-licensed material, or material used under an exception or limitation to copyright. More considerations for licensors.
>
> **Considerations for the public:** By using one of our public licenses, a licensor grants the public permission to use the licensed material under specified terms and conditions. If the licensors permission is not necessary for any reason—for example, because of any applicable exception or limitation to copyrightthen that use

is not regulated by the license. Our licenses grant only permissions under copyright and certain other rights that a licensor has authority to grant. Use of the licensed material may still be restricted for other reasons, including because others have copyright or other rights in the material. A licensor may make special requests, such as asking that all changes be marked or described. Although not required by our licenses, you are encouraged to respect those requests where reasonable. More considerations for the public.

Creative Commons Attribution-ShareAlike 4.0 International Public License

By exercising the Licensed Rights (defined below), You accept and agree to be bound by the terms and conditions of this Creative Commons Attribution-ShareAlike 4.0 International Public License ("Public License"). To the extent this Public License may be interpreted as a contract, You are granted the Licensed Rights in consideration of Your acceptance of these terms and conditions, and the Licensor grants You such rights in consideration of benefits the Licensor receives from making the Licensed Material available under these terms and conditions.

Section 1 Definitions.

a. **Adapted Material** means material subject to Copyright and Similar Rights that is derived from or based upon the Licensed Material and in which the Licensed Material is translated, altered, arranged, transformed, or otherwise modified in a manner requiring permission under the Copyright and Similar Rights held by the Licensor. For purposes of this Public License, where the Licensed Material is a musical work, performance, or sound recording, Adapted Material is always produced where the Licensed Material is synched in timed relation with a moving image.

b. **Adapter's License** means the license You apply to Your Copyright and Similar Rights in Your contributions to Adapted Material in accordance with the terms and conditions of this Public License.

c. **BY-SA Compatible License** means a license listed at http://creativecommons.org/compatiblelicenses, approved by Creative Commons as essentially the equivalent of this Public License.

d. **Copyright and Similar Rights** means copyright and/or similar rights closely related to copyright including, without limitation, performance, broadcast, sound recording, and Sui Generis Database Rights, without regard to how the rights are labeled or categorized. For purposes of this Public License, the rights specified in Section 2(b)(1)-(2) are not Copyright and Similar Rights.

e. **Effective Technological Measures** means those measures that, in the absence of proper authority, may not be circumvented under laws fulfilling obligations under Article 11 of the WIPO Copyright Treaty adopted on December 20, 1996, and/or similar international agreements.

f. **Exceptions and Limitations** means fair use, fair dealing, and/or any other exception or limitation to Copyright and Similar Rights that applies to Your use of the Licensed Material.

g. **License Elements** means the license attributes listed in the name of a Creative Commons Public License. The License Elements of this Public License are Attribution and ShareAlike.

h. **Licensed Material** means the artistic or literary work, database, or other material to which the Licensor applied this Public License.

i. **Licensed Rights** means the rights granted to You subject to the terms and conditions of this Public License, which are limited to all Copyright and Similar Rights that apply to Your use of the Licensed Material and that the Licensor has authority to license.

j. **Licensor** means the individual(s) or entity(ies) granting rights under this Public License.

k. **Share** means to provide material to the public by any means or process that requires permission under the Licensed Rights, such as reproduction, public display, public performance, distribution, dissemination, communication, or importation, and to make material available to the public including in ways that members of the public may access the material from a place and at a time individually chosen by them.

l. **Sui Generis Database Rights** means rights other than copyright resulting from Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases, as amended and/or succeeded, as well as other essentially equivalent rights anywhere in the world.

m. **You** means the individual or entity exercising the Licensed Rights under this Public License. **Your** has a corresponding meaning.

Section 2 Scope.

a. **License grant**.

1. Subject to the terms and conditions of this Public License, the Licensor hereby grants You a worldwide, royalty-free, non-sublicensable, non-exclusive, irrevocable license to exercise the Licensed Rights in the Licensed Material to:
   A. reproduce and Share the Licensed Material, in whole or in part; and
   B. produce, reproduce, and Share Adapted Material.
2. Exceptions and Limitations. For the avoidance of doubt, where Exceptions and Limitations apply to Your use, this Public License does not apply, and You do not need to comply with its terms and conditions.
3. Term. The term of this Public License is specified in Section 6(a).
4. Media and formats; technical modifications allowed. The Licensor authorizes You to exercise the Licensed Rights in all media and formats whether now known or hereafter created, and to make technical modifications necessary to do so. The Licensor waives and/or agrees not to assert any right or authority to forbid You from making technical modifications necessary to exercise the Licensed Rights, including technical modifications necessary to circumvent Effective Technological Measures. For purposes of this Public License, simply making modifications authorized by this Section 2(a)(4) never produces Adapted Material.

5. Downstream recipients.
    A. Offer from the Licensor Licensed Material. Every recipient of the Licensed Material automatically receives an offer from the Licensor to exercise the Licensed Rights under the terms and conditions of this Public License.
    B. Additional offer from the Licensor Adapted Material. Every recipient of Adapted Material from You automatically receives an offer from the Licensor to exercise the Licensed Rights in the Adapted Material under the conditions of the Adapters License You apply.
    C. No downstream restrictions. You may not offer or impose any additional or different terms or conditions on, or apply any Effective Technological Measures to, the Licensed Material if doing so restricts exercise of the Licensed Rights by any recipient of the Licensed Material.
6. No endorsement. Nothing in this Public License constitutes or may be construed as permission to assert or imply that You are, or that Your use of the Licensed Material is, connected with, or sponsored, endorsed, or granted official status by, the Licensor or others designated to receive attribution as provided in Section 3(a)(1)(A)(i).

Other rights.

b. 1. Moral rights, such as the right of integrity, are not licensed under this Public License, nor are publicity, privacy, and/or other similar personality rights; however, to the extent possible, the Licensor waives and/or agrees not to assert any such rights held by the Licensor to the limited extent necessary to allow You to exercise the Licensed Rights, but not otherwise.
2. Patent and trademark rights are not licensed under this Public License.
3. To the extent possible, the Licensor waives any right to collect royalties from You for the exercise of the Licensed Rights, whether directly or through a collecting society under any voluntary or waivable statutory or compulsory licensing scheme. In all other cases the Licensor expressly reserves any right to collect such royalties.

Section 3 License Conditions.
Your exercise of the Licensed Rights is expressly made subject to the following conditions.

a. **Attribution**.

1. If You Share the Licensed Material (including in modified form), You must:
    A. retain the following if it is supplied by the Licensor with the Licensed Material:
        i. identification of the creator(s) of the Licensed Material and any others designated to receive attribution, in any reasonable manner requested by the Licensor (including by pseudonym if designated);
        ii. a copyright notice;
        iii. a notice that refers to this Public License;
        iv. a notice that refers to the disclaimer of warranties;

v. a URI or hyperlink to the Licensed Material to the extent reasonably practicable;

B. indicate if You modified the Licensed Material and retain an indication of any previous modifications; and

C. indicate the Licensed Material is licensed under this Public License, and include the text of, or the URI or hyperlink to, this Public License.

2. You may satisfy the conditions in Section 3(a)(1) in any reasonable manner based on the medium, means, and context in which You Share the Licensed Material. For example, it may be reasonable to satisfy the conditions by providing a URI or hyperlink to a resource that includes the required information.

3. If requested by the Licensor, You must remove any of the information required by Section 3(a)(1)(A) to the extent reasonably practicable.

b. **ShareAlike**.

In addition to the conditions in Section 3(a), if You Share Adapted Material You produce, the following conditions also apply.

1. The Adapters License You apply must be a Creative Commons license with the same License Elements, this version or later, or a BY-SA Compatible License.

2. You must include the text of, or the URI or hyperlink to, the Adapter's License You apply. You may satisfy this condition in any reasonable manner based on the medium, means, and context in which You Share Adapted Material.

3. You may not offer or impose any additional or different terms or conditions on, or apply any Effective Technological Measures to, Adapted Material that restrict exercise of the rights granted under the Adapter's License You apply.

Section 4 Sui Generis Database Rights.

Where the Licensed Rights include Sui Generis Database Rights that apply to Your use of the Licensed Material:

a. for the avoidance of doubt, Section 2(a)(1) grants You the right to extract, reuse, reproduce, and Share all or a substantial portion of the contents of the database;

b. if You include all or a substantial portion of the database contents in a database in which You have Sui Generis Database Rights, then the database in which You have Sui Generis Database Rights (but not its individual contents) is Adapted Material, including for purposes of Section 3(b); and

c. You must comply with the conditions in Section 3(a) if You Share all or a substantial portion of the contents of the database.

For the avoidance of doubt, this Section 4 supplements and does not replace Your obligations under this Public License where the Licensed Rights include other Copyright and Similar Rights.

Section 5 Disclaimer of Warranties and Limitation of Liability.

a. **Unless otherwise separately undertaken by the Licensor, to the extent possible, the Licensor offers the Licensed Material as-is and as-available, and makes no representations or warranties of any kind concerning the Licensed Material, whether express, implied, statutory, or other. This includes, without limitation, warranties of title, merchantability, fitness for a particular purpose, non-infringement, absence of latent or other defects, accuracy, or the presence or absence of errors, whether or not known or discoverable. Where disclaimers of warranties are not allowed in full or in part, this disclaimer may not apply to You.**

b. **To the extent possible, in no event will the Licensor be liable to You on any legal theory (including, without limitation, negligence) or otherwise for any direct, special, indirect, incidental, consequential, punitive, exemplary, or other losses, costs, expenses, or damages arising out of this Public License or use of the Licensed Material, even if the Licensor has been advised of the possibility of such losses, costs, expenses, or damages. Where a limitation of liability is not allowed in full or in part, this limitation may not apply to You.**

a. The disclaimer of warranties and limitation of liability provided above shall be interpreted in a manner that, to the extent possible, most closely approximates an absolute disclaimer and waiver of all liability.

Section 6 Term and Termination.

a. This Public License applies for the term of the Copyright and Similar Rights licensed here. However, if You fail to comply with this Public License, then Your rights under this Public License terminate automatically.

b. Where Your right to use the Licensed Material has terminated under Section 6(a), it reinstates:

   1. automatically as of the date the violation is cured, provided it is cured within 30 days of Your discovery of the violation; or
   2. upon express reinstatement by the Licensor.

   For the avoidance of doubt, this Section 6(b) does not affect any right the Licensor may have to seek remedies for Your violations of this Public License.

c. For the avoidance of doubt, the Licensor may also offer the Licensed Material under separate terms or conditions or stop distributing the Licensed Material at any time; however, doing so will not terminate this Public License.

d. Sections 1, 5, 6, 7, and 8 survive termination of this Public License.

Section 7 Other Terms and Conditions.

a. The Licensor shall not be bound by any additional or different terms or conditions communicated by You unless expressly agreed.

b. Any arrangements, understandings, or agreements regarding the Licensed Material not stated herein are separate from and independent of the terms and conditions of this Public License.

Section 8 Interpretation.

a. For the avoidance of doubt, this Public License does not, and shall not be interpreted to, reduce, limit, restrict, or impose conditions on any use of the Licensed Material that could lawfully be made without permission under this Public License.

b. To the extent possible, if any provision of this Public License is deemed unenforceable, it shall be automatically reformed to the minimum extent necessary to make it enforceable. If the provision cannot be reformed, it shall be severed from this Public License without affecting the enforceability of the remaining terms and conditions.

c. No term or condition of this Public License will be waived and no failure to comply consented to unless expressly agreed to by the Licensor.

d. Nothing in this Public License constitutes or may be interpreted as a limitation upon, or waiver of, any privileges and immunities that apply to the Licensor or You, including from the legal processes of any jurisdiction or authority.

Creative Commons is not a party to its public licenses. Notwithstanding, Creative Commons may elect to apply one of its public licenses to material it publishes and in those instances will be considered the Licensor. The text of the Creative Commons public licenses is dedicated to the public domain under the CC0 Public Domain Dedication. Except for the limited purpose of indicating that material is shared under a Creative Commons public license or as otherwise permitted by the Creative Commons policies published at creativecommons.org/policies, Creative Commons does not authorize the use of the trademark Creative Commons or any other trademark or logo of Creative Commons without its prior written consent including, without limitation, in connection with any unauthorized modifications to any of its public licenses or any other arrangements, understandings, or agreements concerning use of licensed material. For the avoidance of doubt, this paragraph does not form part of the public licenses.

Creative Commons may be contacted at http://creativecommons.org.