The Elements of Linear Algebra
Version 2018S

Theron J. Hitchman

February 27, 2018

Introduction: To The Student

So. I am writing this book. (Let's not pretend I am done. You can see the state it is in.) There are dozens (hundreds?) of introductory linear algebra books, so it is pretty reasonable to ask why I am putting in the effort, and, in the meantime, causing this much pain. I should explain.

I have taught linear algebra many times and I have liked some books, but never loved one. The closest match for what I wanted to teach is Strang's *Introduction to Linear Algebra*, and I am sure that people who have read that will see some influences here. But my students never seemed to connect with Prof. Strang's enthusiastic, stream-of-consciousness prose. And over time, I found that the things I need to emphasize for my students just don't match with that text, or any other.

In addition, most textbooks assume a certain class structure: lectures accompanied by weekly homework, with some exams. I don't want to run our course that way.

So this book is my solution. It is my attempt to make a thing which matches how I want our class to run.

It is important to read this book actively. If you haven't learned how to read a math text before, there are some key ideas:

Time Mathematics is often technical and tricky. It takes time to absorb. Plan to give yourself lots of time to read and think. And don't be surprised if you have to read some section more than once. (This is not a novel. As much as I see it as a story, it won't sweep you away.)

Examples In the interest of brevity, I have streamlined the exposition. In particular, there are no examples. **The point is that you should make your own.** This is so important a skill that it is basically a mathematical super-power. Whenever you come across an idea, if you understand it or not, you should make some very explicit examples and consider them carefully.

Questions As part of your *active* engagement with the text, you will find things that don't quite make sense, yet. This is normal. The mathematician's best approach then is to (1) write down a specific question or two about the confusing bit, and (2) talk to other people about it. You are fortunate that you have an instructor and classmates to talk to. Make lists of questions and try to get them answered!

The real beauty in linear algebra is the tight set of connections between algebra and geometry. I hope you enjoy it.

Chapter 1

Orientation and Preliminaries

1.1 The Three Viewpoints, Algebraically

Linear algebra is about solving systems of linear equations. It is also about the geometry of vectors, and it is also about matrices and transformations of one space into another. Somehow, it is all of those things all at the same time, because those are one and the same. Linear algebra is like a gem with many facets, and the beauty of the subject comes out when you learn how to see the light it casts in different directions.

Let's preview the most basic linear algebra situation and three useful geometric ways to look at it.

A Start: the Row Picture

Consider the following three examples. Each of the first two is a *system of two linear equations in two unknowns*. The third is a *system of three equations in two unknowns*. Because they are familiar, we are using x and y as the names of our unknowns, but we could use other symbols.

$$\left\{\begin{array}{rcrcr} x & + & 2y & = & 3 \\ x & + & y & = & 1 \end{array}\right. \tag{A}$$

$$\left\{\begin{array}{rcrcr} x & + & 2y & = & 3 \\ 3x & + & 6y & = & 4 \end{array}\right. \tag{B}$$

$$\begin{cases} x + 2y = 3 \\ x + y = 1 \\ 3x + 6y = 4 \end{cases} \tag{C}$$

These are typical examples, though they are "small." In applications, a system of linear equations might have thousands of equations, thousands of unknowns, or both!

Reading Exercise. *Make a few other examples of systems of linear equations. How are your examples different from those above? How are they the same?*

Clearly we are looking at equations, so we can ask the kinds of questions that mathematicians always ask when they see equations:

1. Does the system have any solutions at all? Can we decide before we do a lot of work?

2. If it has solutions, how many solutions are there? Is there a way to tell when the system will have a single unique solution?

3. What is the collection of all solutions? Is there a good way to understand this collection geometrically?

4. Is there a good algorithm for finding solutions? That is, is there an easily-followed-and-not-too-slow computational process for solving the system? Can we make a computer do the work for us?

5. If there is not a solution to the system, can we find an approximate solution? Is that a reasonable thing to compute? What should "approximate solution" mean in this setting?

Our first goal is to learn to answer these questions for any system of linear equations.

Reading Exercise. *Can you answer any of the questions in our list for any of the examples above? Or for any of your new examples?*

Consider how each of these systems is organized by being lined up in rows. Each row is a single equation, and that equation defines a collection of points in the xy-plane. Since the equation is linear, the collection happens to make a line. (That is why we call them linear equations.) So, we can make a geometric model of each system that involves the interaction of some

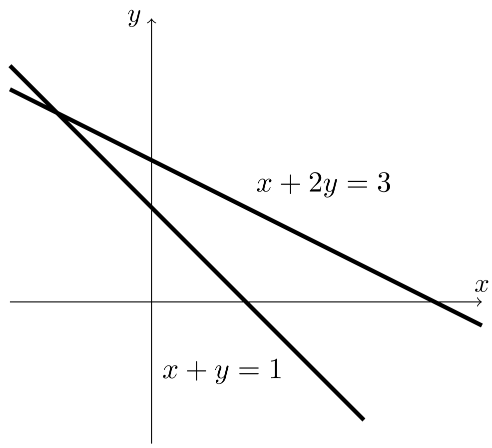


Figure 1.1: The row picture for system (A)

lines in the xy-plane. This geometric model is the *row picture* view of linear algebra.

Here is the row picture for the system (A). In this, we have labeled the two lines with their corresponding equations. The fact that these lines appear to meet gives us a way to talk about the possible solutions to the system (A).

Reading Exercise. *Make row pictures for the examples you designed above. Can you use these pictures to address any of our questions for your examples?*

A Second Look: the Column Picture

Notice that each of our systems is carefully set on the page so that the unknowns line up in columns, too. We can use that! Let's agree to bundle the coefficients together in columns so that they become objects. Just put each column of coefficients in a set of parentheses. We call these vertical-stacks-of-numbers objects *vectors*, and we reorganize each of our systems into a *linear combination of vectors equation.* The three examples above get reworked to look like this:

$$x\begin{pmatrix}1\\1\end{pmatrix} + y\begin{pmatrix}2\\1\end{pmatrix} = \begin{pmatrix}3\\1\end{pmatrix} \tag{A}$$

$$x\begin{pmatrix}1\\3\end{pmatrix} + y\begin{pmatrix}2\\6\end{pmatrix} = \begin{pmatrix}3\\4\end{pmatrix} \tag{B}$$

$$x\begin{pmatrix}1\\1\\3\end{pmatrix}+y\begin{pmatrix}2\\1\\6\end{pmatrix}=\begin{pmatrix}3\\1\\4\end{pmatrix} \tag{C}$$

First, we will have to make sense of the algebra of combining vectors like this. This will become the concept of *linear combination*. (We'll come back to the details, soon.) Note that in the first two examples, the vectors all have two entries, but in the third example, the vectors have three entries. These entries are called *coordinates*. Sometimes we talk about the *shape* of a vector, by which we mean the number of coordinates that vector has.

Then we can make a different sort of picture, as *column picture*. Let us do this for system B. We interpret a vector $\left(\begin{smallmatrix}a\\b\end{smallmatrix}\right)$ with two components as an arrow which goes from the origin to the point in the plane with Cartesian coordinates (a, b). Drawing all three of the vectors from our situation at once, we have the column picture for system (B).

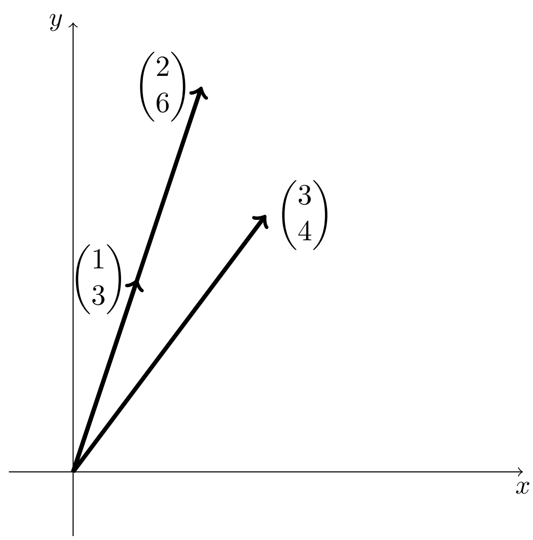


Figure 1.2: The column picture for system (B)

Notice that the two vectors from the left-hand side of our equation point in the same direction, but the one from the right-hand side points in another direction. This can help us reason about our equation.

Reading Exercise. *Translate your examples of systems of equations into the form of a linear combination equation. Can you also draw their column pictures?*

Often, a linear combination of vectors equation is written in this compact form

$$a_1 v_1 + a_2 v_2 + \cdots + a_k v_k = w,$$

where the symbols $a_1, \ldots, a_k$ represent numbers, and the symbols $v_1, \ldots, v_k$ and w represent vectors all having the same shape. This is visually much simpler than the full system.

A Third Look: Matrices and Transformations

As we passed from the system of linear equations to the linear combination of vectors equation, we managed to clean up our representation by getting rid of some messy duplication of structural symbols. Now, we will do it again. We take advantage of the alignment of rows and columns at the same time, and rewrite our equations like this, as *matrix-vector equations*:

$$\begin{pmatrix} 1 & 2 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \end{pmatrix} \tag{A}$$

$$\begin{pmatrix} 1 & 2 \\ 3 & 6 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 3 \\ 4 \end{pmatrix} \tag{B}$$

$$\begin{pmatrix} 1 & 2 \\ 1 & 1 \\ 3 & 6 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \\ 4 \end{pmatrix} \tag{C}$$

In each case, we have put the two unknowns into a vector, $\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)$. The new objects we have introduced are two-dimensional arrays of numbers, called *matrices*. Note that each matrix is constructed by taking the column vectors we found in the linear combination equation and setting them next to each other as columns. The first two examples,

$$\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 3 & 1 \\ 51 & 17 \end{pmatrix},$$

are 2×2 *square matrices* because they have two rows and two columns. The third example,

$$\begin{pmatrix} 1 & 2 \\ 1 & 1 \\ 3 & 6 \end{pmatrix},$$

is a 3×2 *matrix*, because it has three rows and two columns. In general, the number of rows corresponds to the number of equations, and the number of columns corresponds to the number of unknowns.

Reading Exercise. *Translate your examples into matrix-vector equations.*

Again, we have to figure out what the algebra of something like "a matrix times a vector" should mean, and sort out the geometry of that. The idea is that the matrix represents a kind of *function*, or *transformation*, that takes in vectors of a particular kind, and then gives you back vectors of a (possibly) different kind. In our third example, the matrix takes in vectors like $\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)$ with two components, and then gives back vectors with three components. Since we can represent vectors with two components by arrows in the plane, and vectors with three components by vectors in space, we can make a picture like Figure 1.3.

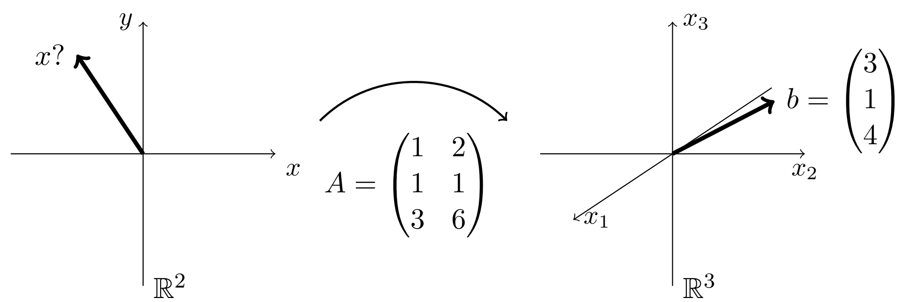


Figure 1.3: The transformational view of system (C)

Often, a matrix-vector equation is written in the ultra-compact form

$$Ax = b,$$

where A is a matrix, b is some known vector, and x is the unknown vector we seek. This form looks simple, because it hides all of the complexity in the abstract matrix and vector objects.

Exercises

We introduced many concepts and new words in this section, and most of them imprecisely. This should make you feel a bit uneasy, but we will start cleaning up by being more careful in the next section. For now, focus on the shapes and structures of the three (otherwise mysterious) algebraic representations introduced in this section.

Exercise 1. Systems of linear equations come up in all sorts of situations. Here is a typical one. When expanded out and rearranged a little, the equation of a circle in the plane takes the form

$$x^2 + y^2 + ax + by + c = 0.$$

Suppose that you have a circle which goes through the three points below. Set up the system of linear equations which should help you find the equation of your circle more exactly. (There is no need to solve the system.)

$$P = (2, 3), \qquad Q = (-4, 2), \qquad R = (7, 1)$$

How many equations do you have? How many unknowns do you have?

Translate your system of linear equations into a linear combination of vectors equation. What shape to the vectors take? Next, translate your system into a matrix-vector equation of the form $Ax = b$. What shape is the matrix A? What shape is the vector x? What shape is the vector b?

Exercise 2. Make an example of a system of two equations in three unknowns. (It's okay. Just pick something you find interesting.)

Translate your example into a linear combination of vectors equation. What shapes are your vectors? Now translate your system into a matrix-vector equation of the form $Ax = b$. What shape is the matrix A? What shape is the vector x? What shape is the vector b?

Exercise 3. Make up a matrix-vector equation $Ax = b$ that has a 4×2 matrix. That is, the matrix should have 4 rows and 2 columns.

Translate your matrix-vector equation into a linear combination of vectors equation. How many vectors does this have? What shape are they? Then translate your matrix-vector equation into a system of linear equations. How many equations do you have? How many unknowns are there?

Exercise 4. Make an example of a linear combination of vectors equation that has 5 vectors of the left-hand side of the equal sign, each of which is a stack of 3 numbers.

Translate your equation into a matrix-vector equation. What shape do all of the pieces have? Translate your equation into a system of linear equations. How many equations are there? How many unknowns are there?

Exercise 5. Think about the ideas discussed in this section. What questions do you have? What do you wonder about?

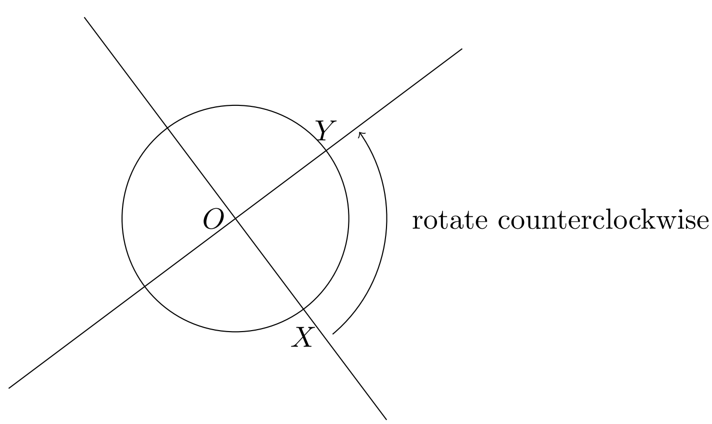
1.2 The Space $\mathbb{R}^m$: Points, Vectors, and Vectors

Our goal here is to introduce the fundamental object of linear algebra, the vector. The word has slightly different meanings depending on context, so we will to sort this out carefully first in two dimensions, where we can draw the best pictures. Then we will extend the ideas to the general situation.

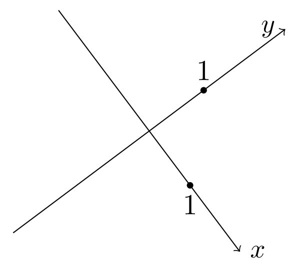
The Idea of a Vector

Let's recall the idea of *the plane* from classical geometry: the plane is like a flat sheet of drawing paper, which extends indefinitely in all directions without bound.

You have likely seen the idea of *Cartesian coordinates* on the plane before. To be clear, let's set things down carefully. In the plane, we choose a pair of perpendicular lines which meet at a point O. This special point is called the *origin*. Then, we choose a point X on one of the lines and draw the circle centered at O which passes through X. Note that this circle meets our two lines in two points each, four points total, one of which is the point X. Then we rotate from X around the circle by a quarter turn counterclockwise until we hit one of the points on the other line. We call this new point Y. Are you drawing with me? Here is my picture so far.



We call the line OX the *x-axis* and the line containing Y the *y-axis*. Here comes the amazing part: we declare the circle we used to be of *unit size*, and make the lines OX and OY into number lines! The important part is that the point O should represent 0 on both number lines, and the points X and Y should each represent 1 on their lines. So, instead of marking things with O, X, and Y, we put down marks where X and Y are and label them with 1's, and add little arrows marked with x and y near the positive "ends" of the lines OX and OY, respectively.



Note that above I have done something a bit silly and let the picture just fall on the paper in an unusual way. I really mean unusual as "not usual." The usual way arranges the lines on the paper to match our expected horizontal (x) and vertical (y) directions. This isn't actually required, but it is what everyone always does. The typical picture looks like Figure 1.4.

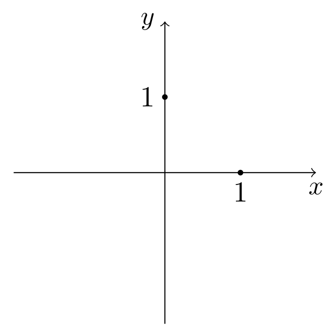


Figure 1.4: The Standard Cartesian Coordinate System

Now suppose we have some point in the plane, let's call it P. We can describe the location of P relative to our two lines in a simple way. First, we draw a line through P which is parallel to the y-axis and perpendicular to the x-axis. The foot of this perpendicular hits the x-axis at some point A. But this point A is part of the number line OX, so it has an associated real number, which we will call a. So the point A is instead marked with the label a from this number line.

Similarly, we draw a line through P parallel to the x-axis and perpendicular to the y-axis. The foot of this perpendicular hits the y-axis at some point B, which is part of the number line OY. We denote the number associated to B by b. Again, the point is labeled with the number b from the number line.

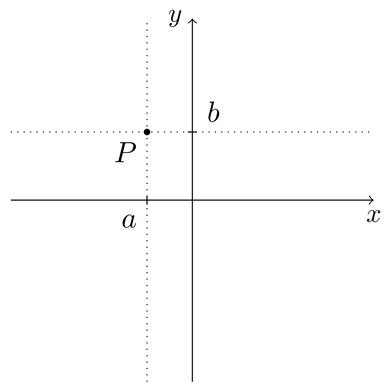


Figure 1.5: A point P and its coordinates (a, b).

So, to identify the point P, we can instead give the pair of numbers a and b. These numbers are called the *coordinates* of P. Of course, the order of the coordinates matters, so we make what we call an *ordered pair* (a, b) to keep things straight, where the x-coordinate comes first, and the y-coordinate comes second. Note that in Figure 1.5 the x-coordinate a is negative, but the y-coordinate b is postive.

This whole process is reversible, too. If we pick a pair of numbers c and d, in order, then we can find a point in the plane which corresponds, and we can do it unambiguously. Find the spot labeled c on the x-axis number line and construct a line perpendicular to the x-axis through this point. Similarly, find the spot labeled d on the y-axis and construct a line perpendicular to the y-axis through this point. The two lines you just drew will meet in exactly one point Q, and Q will have coordinates (c, d).

This setup of coordinates on the plane allows us to formalize a wonderful and useful idea from physics, too. Physicists use the concept of a *vector* to describe something (like the wind) which has both magnitude or size (like how fast the air is moving) and direction (which way the air is going). Usually, vectors are drawn as little arrows: the arrow has a direction, and it has a length which represents its magnitude. It is possible to draw vectors which have the same direction but different lengths, and vice versa.

We can use coordinates to describe vectors in the plane, too. Here's how: A physicist's vector v is some arrow in the plane. That arrow has an initial point P, called its *tail*, and a final point Q, called its *head*. We can write the coordinates of these points as $P = (a, b)$ and $Q = (c, d)$.

Then the coordinates of v are taken to be the numbers $c - a$ and $d - b$, which we interpret how much v acts in the directions parallel to the x-axis

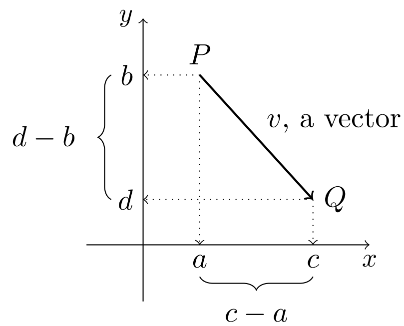


Figure 1.6: Coordinates for a physicist's vector v.

and the y-axis, respectively. Note that in Figure 1.6, the y-coordinate is negative, since $b > d$.

These coordinates have a hint of algebraic manipulation in them. Those subtractions line up almost like we could write $v = (c - a, d - b) = (c, d) - (a, b) = Q - P$. But v is a vector, and if we write it like $(c - a, d - b)$, it looks like the notation for a point. We should not do that because it could get confusing. Furthermore, that "equation" would mean that we are subtracting points and creating a vector, which is also weird. Still, there is something to it. We will return to this idea soon.

For now, let's focus on a bit of ambiguity in the physicist's idea of a vector. Where should that vector be? That is, given the coordinates of a vector in the plane, it is not clear where to draw it! We can slide a vector around the plane, and as long as we keep it parallel to the original, the coordinates won't change. So, unlike the coordinates of a point, the coordinates of a vector do not uniquely specify the vector.

The mathematician's special fix is this: we simply declare all our vectors to have their initial points, their tails, at the point O, the origin of our coordinate system. That curtails some of the (admittedly useful) freedom in the physics notion, but it also lets us be more clear.

It pays to keep in mind that the physicists conception of the vector V with coordinates $c-a$ and $d-b$ could be one of many different arrows, while the mathematician's vector v is the arrow from the point $O = (0, 0)$ to the point $(c - a, d - b)$.

Now we have circled back around to a muddle. If a mathematician's vector is always based at O, we only need to specify where the head of the vector is. . . which is just a point. So, how is a vector supposed to be different from a point, again?

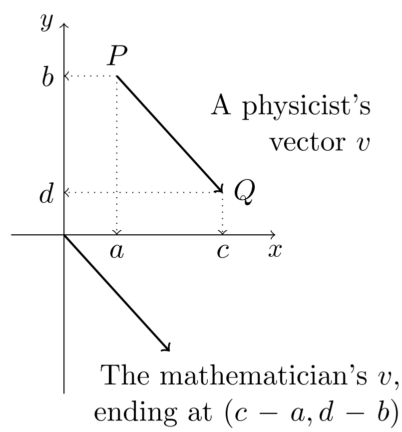


Figure 1.7: The physicist's vector vs. the mathematician's vector.

This confusion of three different, shifting, partially-overlapping interpretations causes some trouble to the new learner. Professionals tend to pass back and forth between these and use them flexibly to get results. Once you have gotten used to the ideas, you will, too. You should watch out for these instances where the words point and vector get interchanged. If they cause you trouble, remember that we have three different things, which are closely related.

For now, the simplest way to handle things is like this:

- Ignore the physicist's version of the word vector as much as possible.
- A point is a location in the plane, and represented by coordinates in the form of an ordered pair of numbers (a, b).
- A vector is an arrow based at the origin, which can be specified by the coordinates of its head. To keep this separate from the idea of a point, we will write it differently, with the numbers stacked vertically like this: $\left(\begin{smallmatrix} a \\ b \end{smallmatrix}\right)$.
- Always remember that for each point in the plane, there is a unique mathematician's vector which corresponds, and vice versa.

With this in mind, we make our first official definition.

Definition 1. A 2-vector is a vertical stack of 2 real numbers, like so:

$$v = \begin{pmatrix} a \\ b \end{pmatrix}.$$

The collection of all such 2-vectors is called *the plane*, and written with this notation: $\mathbb{R}^2$.

The notation $\mathbb{R}^2$ is often read "arr-two," and many people use that language interchangeably with "the plane." Also, most of the time we will just say "vector," rather than "2-vector."

Vector Algebra

Let's return to that glimpse of subtraction we saw in Figure 1.6. We saw there that for points $P = (a, b)$ and $Q = (c, d)$, the vector v from P to Q has coordinates $c - a$ and $d - b$. This looks almost like we subtracted the points to get $Q - P = v$. Can we use that? The weird part is that it mixes up points and vectors. So, we change viewpoints, and instead think of P and Q as (mathematician's) vectors. To keep things clear, let's introduce new labels.

$$p = \begin{pmatrix} a \\ b \end{pmatrix}, \quad q = \begin{pmatrix} c \\ d \end{pmatrix}$$

If we put these together on the plane with the physicist's vector v from p to q and the mathematician's v we see a wonderful triangle, and an extra vector. So we see a way to talk about subtracting vectors: Given two vectors

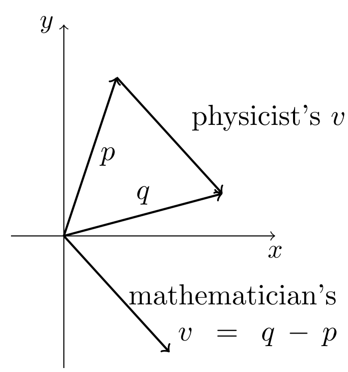


Figure 1.8: Subtraction of vectors

p and q as above, their *difference* is the vector

$$V = Q - P = \begin{pmatrix} c \\ d \end{pmatrix} - \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} c - a \\ d - b \end{pmatrix}.$$

Geometrically, we draw the arrow from p to q and then translate it down so that its tail is at the origin $O = (0, 0)$. Of course, the order of p and q in this operation matters. If we switch them, we get an arrow pointing in the opposite direction.

If we can subtract vectors, surely we can add vectors. How would that work? Algebraically, if $v = q - p$, then we expect $q = v + p$ by rearranging. That would mean

$$q = \begin{pmatrix} c \\ d \end{pmatrix} = v + p = \begin{pmatrix} c-a \\ d-b \end{pmatrix} + \begin{pmatrix} a \\ b \end{pmatrix},$$

which all fits. It looks like addition should be defined coordinate-by-coordinate.

Definition 2 (Addition of Vectors)**.** Let $p = \left(\begin{smallmatrix} a \\ b \end{smallmatrix}\right)$ and $q = \left(\begin{smallmatrix} c \\ d \end{smallmatrix}\right)$ be two vectors in $\mathbb{R}^2$. Their sum is the vector

$$p + q = \begin{pmatrix} a+c \\ b+d \end{pmatrix}.$$

Theorem 3. Addition of vectors satisfies the same rules as addition of real numbers:

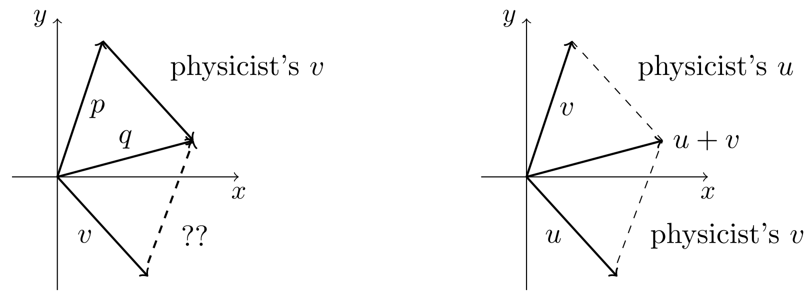
- when adding more than two vectors, it doesn't matter which operations you do first: $(p+q)+r = p+(q+r)$;
- one can add in either order $p + q = q + p$;
- the vector 0 corresponding to the origin O is a "zero" since $p + 0 = 0 + p = p$;
- for each vector p, there is an opposite vector $-p$ so that $p + (-p) = 0$.

Reading Exercise. *Remember that you are supposed to read actively. You can draw all of these pictures and try out all of these things with specific examples that you invent. You should check the statements in Theorem 3 by making examples and working out the details. Can you also draw the pictures which go with your examples?*

But what about subtracting geometrically? In Figure 1.8, I have a strong desire to complete the figure by joining the loose end of v to the rest of the figure. If we draw the arrow from the head of v to the head of q, we get Figure 1.9a.

What should the label on the dashed vector in Figure 1.9a be? Just as the physicist's v and the mathematician's v are parallel, this new vector is parallel to the mathematician's vector p. So the dashed vector must be a physicist's version of p. Then we see $q = v + p$.

Now we know how to add geometrically: to add two vectors u and v, translate v until its tail is on the head of u, then draw a new vector $u + v$ as the vector from the tail of u to the head of this translated v. It just



(a) Subtraction of vectors (b) Geometric Addition of Vectors

Figure 1.9: Some algebra of vectors

repurposes the structure of Figure 1.9a. This is called the *parallelogram rule* for addition of vectors.

There is another useful operation on vectors called *scalar multiplication*. The terminology comes from physics (again) where a *scalar* quantity is just a number, and not a vector. So "scalar multiplication" means to multiply a vector by a scalar.

Definition 4 (Scalar Multiplication for vectors)**.** Let $p = \left(\begin{smallmatrix} a \\ b \end{smallmatrix}\right)$ be a vector in $\mathbb{R}^2$, and let λ be a real number. Then the *scalar multiple* λp is defined to be

$$\lambda p = \begin{pmatrix} \lambda a \\ \lambda b \end{pmatrix}$$

If you have never seen the symbol λ before, it is an old Greek letter pronounced "lamb-duh." It is traditional to use it in linear algebra in lots of places. Welcome to the λ-club. Oh, there are other such letters, like μ, which is pronounced "myoo."

Again, this operation has some important similarities to the familiar multiplication of numbers, but because it combines a scalar (a number) with a vector (not a number, exactly) to produce another vector (again, not a number) things are a little different.

Theorem 5. Suppose that p and q are vectors, and λ and μ are numbers. Scalar multiplication has the following properties:

- Scalar multiplication distributes over vector addition:
  $\lambda(p + q) = \lambda p + \lambda q$;
- Scalar multiplication distributes over scalar addition:
  $(\lambda + \mu)p = \lambda p + \mu q$;

- Scalar multiplication and regular multiplication can be done in either order: $\lambda(\mu p) = (\lambda\mu)p$;
- if $\lambda = 0$, then $\lambda p = 0p = 0$ is the zero vector.
- if n is a counting number, then np is the same thing as adding together n copies of p. In particular, $1p = p$.

This Theorem, like the last one, says that some natural properties you hope will still work really do still work. When you study **Modern Algebra**, making lists of these kinds of properties will be really useful. So far, we have collected up the properties that describe a *vector space*.

What is the geometry of scalar multiplication? It corresponds to stretching (or shrinking) a vector, without changing its direction. Let λ be a non-zero number. Since

$$\lambda p = \lambda \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \lambda a \\ \lambda b \end{pmatrix},$$

we see that the ratio of the two coordinates of a vector doesn't change under scalar multiplication. This means that p and λp point in the same direction. One can see this by considering similar triangles with sides parallel to the x-axis, the y-axis, and p.

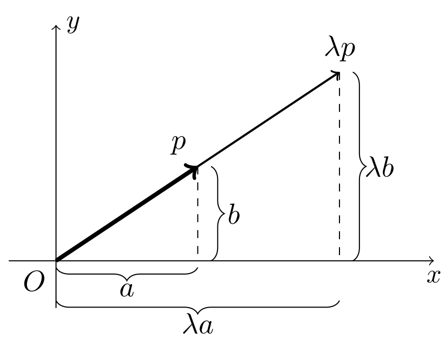


Figure 1.10: Similar triangles and scalar multiplication, $\lambda > 1$

The triangles in Figure 1.10 are similar: their corresponding horizontal and vertical sides are parallel, and those pairs of sides have a common ratio. We learn that p and λp lie in the same line.

This is important! Later we will describe lines in the plane, and we have just discovered how scalar multiplication relates to those lines which pass through the origin, O.

By the way, this picture helps explain the terminology. The vector λp is a rescaled version of p. A *scalar* is a thing which *scales* vectors.

The General Case

Now we can give the fully general definition.

Definition 6. Let m be a counting number. We define an m-vector to be a vertical stack of m real numbers, like so:

$$u = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_m \end{pmatrix}.$$

The individual entries u_i of u are called its *components*. The collection of all m-vectors is called *m-space*, and denoted $\mathbb{R}^m$.

The symbol $\mathbb{R}^m$ is usually read as "arr-em."

Definition 7 (Linear Combinations of vectors)**.** Suppose that u and v are two m-vectors. Their *sum* $u + v$ is defined by adding the individual components in corresponding positions. If λ is a number, then the *scalar multiple* λu is defined by multiplying each of the components of u by the number λ.

Suppose that $v_1, v_2, \dots, v_k$ are all m-vectors, and that $a_1, a_2, \dots, a_k$ are all real numbers. Then the vector

$$a_1 v_1 + a_2 v_2 + \cdots + a_k v_k$$

is called a *linear combination of the vectors $v_1, v_2, \dots, v_k$ with weights $a_1, a_2, \dots, a_k$.*

It is much harder to think through the geometry of $\mathbb{R}^m$ when m is large, but the algebra works in much the same way.

Theorem 8. *Fix a natural number m. The results of Theorem 3 and Theorem 5 about the algebra of vectors hold for vectors in $\mathbb{R}^m$, too.*

Now that we have a little bit of algebraic structure for vectors, we can form equations.

Definition 9. An equation of the form $\lambda_1 u_1 + \dots \lambda_n u_n = w$, where all of the vectors u_i and w are known, but the scalars λ_i are unknown, is called a *linear combination of vectors equation*. A *solution* to such an equation is a collection of scalars which make the equation true.

Exercises

Exercise 6. Write down three distinct points in the plane in proper notation. Plot those three points on a single diagram.

Exercise 7. Write down three distinct vectors in the plane in proper notation. Your three points from this task should NOT match any of the three points from the previous task. Plot those three vectors on a single diagram.

Exercise 8. Find the sum of your three vectors from the last exercise. Then, choose some order of those three vectors so that they are u_1, u_2 and u_3, and compute the linear combination

$$3u_1 - 2u_2 + (1/2)u_3.$$

Exercise 9. Let's consider the vectors $u = \left(\begin{smallmatrix} 2 \\ 1 \end{smallmatrix}\right)$, $v = \left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right)$, and $w = \left(\begin{smallmatrix} -3 \\ 1 \end{smallmatrix}\right)$. Compute all of the vectors in this list:

$$\frac{u+v}{2}, v-u, v-w, u+\left(\frac{v-u}{3}\right), u+\left(\frac{3(v-u)}{4}\right)$$

Then make a single diagram which contains u, v, w and all of those vectors from the list, plotted as accurately as you can.

What do you notice? Is anything interesting going on?

Exercise 10. Consider the vector $u = \left(\begin{smallmatrix} -4 \\ 2 \end{smallmatrix}\right)$. Find a vector v which has the property that $u + v$ is the zero vector, or explain why this is not possible.

Exercise 11. For now, keep $u = \left(\begin{smallmatrix} -4 \\ 2 \end{smallmatrix}\right)$. Let $v = \left(\begin{smallmatrix} 3 \\ 5 \end{smallmatrix}\right)$. How many solutions does the linear combination of vectors equation $\lambda u = v$ have?

How many solutions does the linear combination of vectors equation $\lambda u + \mu v = 0$ have? (Here, treat 0 as the zero vector.)

Exercise 12. We still use the notation u for the vector $u = \left(\begin{smallmatrix} -4 \\ 2 \end{smallmatrix}\right)$, but now use v for the vector $v = \left(\begin{smallmatrix} 28 \\ -14 \end{smallmatrix}\right)$. How many solutions does the linear combination of vectors equation $\lambda u = v$ have?

How many solutions does the linear combination of vectors equation $\lambda u + \mu v = 0$ have? (Again, treat 0 as the zero vector.)

Exercise 13. Find the midpoint between the points $P = (4, -2)$ and $Q = (3, 5)$. Then find the two points which divide the segment PQ into thirds.

How can vectors make this simpler than it first appears?

Challenge 14. Suppose you are given three points in the plane. Let's call them P, Q, and R. How can you use vectors to (quickly) determine if these three points are collinear?

Exercise 15. Write down four 3-vectors of your choosing, where none has any coordinate equal to 0. Call these vectors u_1, u_2, u_3, and u_4. Now choose any four non-zero scalars you like, call them λ_1, λ_2, λ_3 and λ_4. Compute the linear combination

$$\lambda_1 u_1 + \lambda_2 u_2 + \lambda_3 u_3 + \lambda_4 u_4.$$

1.3 The Idea of a Subset

As our study progresses, we will encounter many examples of collections of things. Often those collections will have other interesting collections inside of them. We will also encounter this phenomenon in reverse, where a collection which interests us actually lies inside some other collection, which is somehow larger. It will be convenient for us to have some basic language for these relationships, so we take that up now.

A Set and its Elements

The way a mathematician talks about a collection of things is to use the word *set*. What is a *set*? Well, a *set* is a collection of things. (There are some really deep intellectual waters nearby this concept, but we will avoid those this semester. The basic idea will be enough to get us through.) What is important is that a set has *elements*.

In fact, the only thing we want to formalize here is the idea of membership. Suppose that we have two mathematical objects. The first, S, is required to be a set. The second, a, need not be a set, though it might be. We will say that *a is an element of S* and write the notation

$$a \in S$$

when a happens to be one of the things in the collection S.

A simple example will go far. Suppose that S is the collection consisting of the numbers 1, 2, 3 and 4. Then S is a set, and the numbers 1, 2, 3, and 4 are the elements of S. But the number 5 is not an element of S, because it is not part of the collection.

It will help us to have notation for describing sets. The standard way to do it is to describe a set using curly braces, like so:

$$S = \{1, 2, 3, 4\}$$

The above sentence should be read "S is the set consisting of the elements 1, 2, 3, and 4." Here, the curly braces are a visual signal of the beginning and the end of the description of the set. They also serve as a visual metaphor: mathematicians tend to think of a set as a type of container which holds other things. Because we know that 3 is an element of this set S, but 5 is not an element of S, we write

$$3 \in S \text{ and } 5 \notin S,$$

where the slash through the $\in$ symbol changes the meaning from "is an element of" to "is not an element of."

Sometimes it is convenient to list all of the elements in a set, but often it is not. In those cases, we use a modification of the notation above. First, we set up some notation for the possible elements, then we write a vertical bar, and then we write down a description telling what has to be true for that object before the bar to be an element. Take note that this description might be something written in an sentence, or something written with mathematical symbols.

Again, some examples will help. The set of all real numbers, the set of integers, the set of 2d vectors, the set of positive real numbers, and the set of points in the plane which lie on a standard parabola can be written as in these examples below.

$$\begin{aligned}
\mathbb{R} &= \{x \mid x \text{ is a real number}\} \\
\mathbb{Z} &= \{x \in \mathbb{R} \mid x \text{ is an integer}\} \\
\mathbb{R}^2 &= \{v = \left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right) \mid x, y \in \mathbb{R}\} \\
P &= \{x \in \mathbb{R} \mid x > 0\} \\
Q &= \{p = (x, y) \mid x, y \in \mathbb{R} \text{ and } y = x^2\}
\end{aligned}$$

The description to the right of the vertical bar is very important for deciding if something is an element of the set or not. It gives exactly the criterion for testing. If your object makes the statement true, it is an element of the set. If your object makes the statement false, then it is not an element of the set. For example, it should be clear that the following statements are true:

$$(5, 10) \notin Q, \quad (-3, 9) \in Q.$$

Subsets

We can build on the relationship between a set and its elements to consider a kind of containment relationship between two sets.

Definition 10. Let S and T be two sets. We say that S *is a subset of* T when for each element x of S, we have that x is also an element of T. If S is a subset of T, we will use the notation

$$S \subset T.$$

In our list of examples above, we have several subset relationships. For example, since every integer is a real number, we know that $\mathbb{Z} \subset \mathbb{R}$. Similarly, $P \subset \mathbb{R}$. But there are integers which are not positive numbers, so it is not the case that $\mathbb{Z}$ is a subset of P.

Exercises

Exercise 16. Use our notation to write down a description of the set E which consists of all even integers.

Exercise 17. Rewrite the set definition below as a sentence in plain English.

$$C = \left\{ v = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \middle| \, x_1^2 + x_2^2 = 2 \right\}$$

It turns out that C is a subset of some familiar set. Which one is it?

Exercise 18. Let C be the set from the last exercise. Find three examples of vectors in $\mathbb{R}^2$ which are elements of the set C.

Then find three example of vectors in $\mathbb{R}^2$ which are not elements of C.

Exercise 19. Let $v \in \mathbb{R}^2$ be the vector $v = \left(\begin{smallmatrix} 1 \\ -2 \end{smallmatrix}\right)$. We are interested in the subset of $\mathbb{R}^2$ which consists of vectors w so that v and w make a right angle. First, write out a description of this set in proper notation. Then try to describe what this set looks like in common terms.

Exercise 20. Repeat the last exercise, but instead work in $\mathbb{R}^3$ and with the vector v below.

$$v = \begin{pmatrix} 3 \\ -1 \\ 2 \end{pmatrix}$$

1.4 The Dot Product and Geometry

A lot of the basic geometry of $\mathbb{R}^n$ is captured by a mysterious object called the *dot product*. First, we will show how this can be understood in $\mathbb{R}^2$, and then we will generalize everything to $\mathbb{R}^n$ for $n \geq 2$.

Lengths and Angles in $\mathbb{R}^2$

Given two vectors, their dot product is a real number. This number is a strange measure of how much the vectors are alike, and captures information about both lengths and angles. The formal definition is this.

Definition 11 (The Dot Product in $\mathbb{R}^2$)**.** Let $u = \left(\begin{smallmatrix} u_1 \\ u_2 \end{smallmatrix}\right)$ and $v = \left(\begin{smallmatrix} v_1 \\ v_2 \end{smallmatrix}\right)$ be vectors in $\mathbb{R}^2$. The *dot product* of u and v is the number

$$u \cdot v = u_1 v_1 + u_2 v_2.$$

The clearest geometric information we can pull out of the dot product is about lengths.

Theorem 12 (Lengths in $\mathbb{R}^2$)**.** Let $u = \left(\begin{smallmatrix} a \\ b \end{smallmatrix}\right) \in \mathbb{R}^2$ be a vector. The length of u is equal to the square root of the dot product of u with itself. That is, the length of u is

$$\sqrt{u \cdot u} = \sqrt{a^2 + b^2}$$

Proof. The key comes from considering Figure 1.11. There we see the vector u in $\mathbb{R}^2$ is the hypotenuse of a right triangle having its legs parallel to the two coordinate axes.

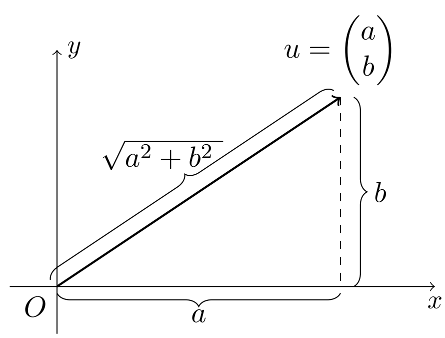


Figure 1.11: The Norm of a vector

These two legs then have lengths which are equal to our two coordinates, respectively. So we can compute the length of the hypotenuse (our vector) by the Pythagorean Theorem, which completes our proof. □

The length of a vector is a useful concept, but mathematicians often use another name for it. New learners often find this confusing, so beware!

Definition 13 (Norm, Unit Vector)**.** Let $u = \left(\begin{smallmatrix} a \\ b \end{smallmatrix}\right)$ be a vector in $\mathbb{R}^2$. The *norm* of u is the number

$$\|u\| = \sqrt{u \cdot u} = \sqrt{a^2 + b^2}\,.$$

We say that u is a *unit vector* when it has a norm of $\|u\| = 1$.

Theorem 14 (The Norm and Scalar Multiplication)**.** Let λ be a scalar, and let $u \in \mathbb{R}^2$ be a vector. Then

$$\|\lambda u\| = |\lambda|\,\|u\|$$

This conforms to our basic ideas about scalar multiplication. If we rescale a vector, we are really just changing its length by that factor. (And then there is the worry over the sign of the scalar, because lengths must not be negative.)

Proof. Suppose that the vector is $u = \left(\begin{smallmatrix} a \\ b \end{smallmatrix}\right)$. We can do the following straightforward computation:

$$\begin{aligned} \|\lambda u\| &= \left\| \lambda \begin{pmatrix} a \\ b \end{pmatrix} \right\| = \left\| \begin{pmatrix} \lambda a \\ \lambda b \end{pmatrix} \right\| = \sqrt{(\lambda a)^2 + (\lambda b)^2} \\ &= \sqrt{\lambda^2(a^2+b^2)} = |\lambda| \sqrt{a^2+b^2} \\ &= |\lambda|\,\|u\| \end{aligned}$$

Since the first and last are equal, we have established the desired result. $\square$

In a terrible linguistic collision, the most useful role for the norm of a vector is in the process of *normalizing* that vector. Beware! Mathematics has several words that are overused! Norm, normal, normalize, normalized and other forms of this are a prime example. To *normalize* a vector u, we multiply it by the scalar $\|u\|^{-1}$, and thus produce a new vector $u/\|u\|$ which points in the same direction as u, but has norm equal to 1. This is because we can use our observation just above about how norms interact with scalar multiplication to compute:

$$\left\| \|u\|^{-1} u \right\| = \|u\|^{-1} \|u\| = 1.$$

That last multiplication is just a multiplication of numbers. Note that all of that only makes sense as long as u is not the zero vector. If u is the zero vector, its norm is $\|u\| = 0$. Since it makes no sense to divide by zero, it is not possible to normalize the zero vector.

To sum up: given a non-zero vector u, it is possible to normalize it, which means to replace it with the unit vector $u/\|u\|$.

Let's turn out attention to measuring angles between vectors. To do this properly, we will need a little trigonometry. The three necessary facts are collected in the next theorem as a refresher.

Theorem 15 (Some Trigonometry)**.** The following statements are true.

1. Each point on the unit circle can be represented as a point of the form $(\cos(\alpha), \sin(\alpha))$ for some angle α between 0 and 2π, so the corresponding vector can be written as $p = \begin{pmatrix} \cos(\alpha) \\ \sin(\alpha) \end{pmatrix}$.

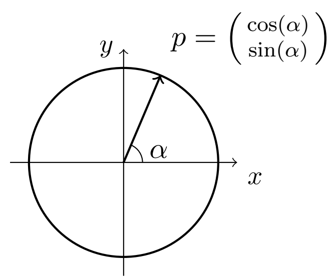


Figure 1.12: Trigonometry for points on the unit circle

2. The function $\theta \mapsto \cos(\theta)$ associates to each angle in the interval $[0, \pi]$ a unique real number in the interval $[-1, 1]$, and vice versa. In particular, this function has a sensible inverse function $\arccos : [-1, 1] \to [0, \pi]$. For us, this means that it to measure an angle, we can get away with instead finding the cosine of that angle.

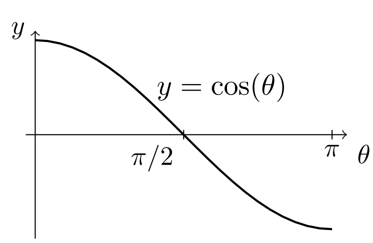


Figure 1.13: Part of the graph of $y = \cos(\theta)$

3. There is an identity on trigonometric functions that helps us deal with the the cosine of a difference of angles: If α and β are two angles, then
$$\cos(\beta - \alpha) = \cos(\alpha)\cos(\beta) + \sin(\alpha)\sin(\beta).$$

You might recognize all of these facts from a trigonometry or pre-calculus class. In any case, you should take them as true. It would take us too far out of our studies to establish them now.

Theorem 16 (Angles in $\mathbb{R}^2$)**.** Let $u, v \in \mathbb{R}^2$ be two vectors. Then the angle θ between u and v is
$$\theta = \arccos\left(\frac{u \cdot v}{\|u\|\,\|v\|}\right).$$

Proof. So, we consider the angle between two vectors, u and v, in $\mathbb{R}^2$. As they are mathematician's vectors, they are both based at the origin, and naturally form an angle there. We will now apply each of the facts from Theorem 15.

Note that the angle between these vectors does not depend at all on their lengths. If we change one, or both, of the vectors by rescaling them, that will not change the directions involved, and hence will not change the angle we seek. We can get to a more uniform set up for our task by normalizing u and v. Thus, we will instead consider the unit vectors $u/\|u\|$ and $v/\|v\|$, and look for the angle θ between them as in Figure 1.14.

Since our new vectors are unit vectors, we can represent them with trigonometric functions:
$$\frac{u}{\|u\|} = \begin{pmatrix} \cos(\alpha) \\ \sin(\alpha) \end{pmatrix}, \qquad \frac{v}{\|v\|} = \begin{pmatrix} \cos(\beta) \\ \sin(\beta) \end{pmatrix}.$$
Let's suppose that the angle β that v makes with the x-axis is larger than the angle α that u makes with the x-axis. Then we want to find the angle $\theta = \beta - \alpha$.

By angle difference formula for cosine, we have
$$\begin{aligned} \cos(\theta) &= \cos(\beta - \alpha) \\ &= \cos(\alpha)\cos(\beta) + \sin(\alpha)\sin(\beta) \\ &= \left(\frac{u}{\|u\|}\right) \cdot \left(\frac{v}{\|v\|}\right) \\ &= \frac{u \cdot v}{\|u\|\,\|v\|}. \end{aligned}$$
Now, we apply the function arccos to both sides to get the theorem. $\square$

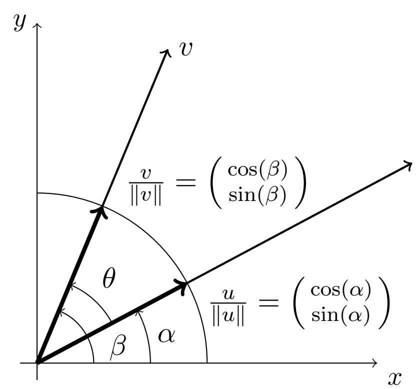


Figure 1.14: Finding the angle between two vectors.

As our study progresses, we will not have much use for measuring particular angles, but it will be very important to us to understand situations when two vectors make an angle of $\pi/2$ radians (i.e. $90°$, a right angle). It is common in geometry to call such vectors *perpendicular*. It is a fact that $\cos(\pi/2) = 0$, so that non-zero vectors u and v make an angle of $\pi/2$ when

$$0 = \cos(\pi/2) = \frac{u \cdot v}{\|u\| \, \|v\|}.$$

By clearing out the denominator of this fraction, we see that u and v make an angle of $\pi/2$ exactly when $u \cdot v = 0$. In another instance of cluttering the vocabulary list for math students, in a linear algebra context such vectors are called by yet another term.

Definition 17 (Orthogonal Vectors)**.** We say that two vectors u and v in $\mathbb{R}^2$ are *orthogonal* when $u \cdot v = 0$.

The Dot Product in $\mathbb{R}^n$

Now we will give the general definition and results. Fortunately, everything works the same.

Definition 18 (The Dot Product)**.** Let u and v be vectors in $\mathbb{R}^n$ with coordinates labeled like those below.

$$u = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}, \quad v = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$$

The *dot product* of u and v is the number

$$u \cdot v = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n.$$

Once we have the dot product available, we can work by analogy to define all of the other new words of this section for vectors in $\mathbb{R}^n$, too.

Definition 19 (Geometry in $\mathbb{R}^n$)**.** Let $u, v \in \mathbb{R}^n$ be vectors.

- The *norm of* u is $\|u\| = \sqrt{u \cdot u}$.
- The vector u is called a *unit vector* when $\|u\| = 1$.
- The *angle between* u *and* v is the number

$$\theta = \arccos\left(\frac{u \cdot v}{\|u\|\,\|v\|}\right).$$

- We say that u and v are *orthogonal* when $u \cdot v = 0$.

The next two theorems are not hard to prove, but they are tedious. In each case, one has to argue one coordinate at a time, using a relevant property for real numbers. It is best to just check them for some examples until you understand and believe them.

Theorem 20 (Algebra of the Dot Product)**.** Let u, v, and w be vectors in $\mathbb{R}^n$ and let λ and μ be scalars. Then

- The dot product is *symmetric*: $u \cdot v = v \cdot u$;
- The dot product *distributes* over linear combinations

$$u \cdot (\lambda v + \mu w) = \lambda(u \cdot v) + \mu(u \cdot w)$$

Theorem 21 (Algebra of the Norm)**.** Let $u \in \mathbb{R}^n$ be a vector, and let λ be a scalar. Then

- $\|u\| \geq 0$.
- $\|u\| = 0$ if, and only if, u is the zero vector.
- $\|\lambda u\| = |\lambda|\,\|u\|$

Exercises

Exercise 21. Choose three different vectors in $\mathbb{R}^2$ which have neither of their components equal to zero. Call these vectors u, v, and w.

a) Compute the norms of u, v, and w.
b) Compute the dot products $u \cdot v$, $v \cdot w$, and $u \cdot w$.
c) Find unit vectors u', v', and w' which point in the same directions as u, v, and w, respectively.
d) Find the angles between each of the pairs, u and v, u and w, v and w in radians.

Exercise 22. Fix some vector $u \in \mathbb{R}^2$. Draw a picture of u in the plane, and then shade the region of the plane which contains vectors v so that $u \cdot v > 0$.

Exercise 23. This task continues our quest for understanding the sign of a dot product geometrically.

a) Find an example of two v and w in $\mathbb{R}^2$ so that $\left(\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right) \cdot v = 0$ and $\left(\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right) \cdot w = 0$, or explain why such an example is not possible.
b) Let $v = \left(\begin{smallmatrix} 3 \\ -1 \end{smallmatrix}\right)$. Find an example of a pair of 2-vectors u and w such that $v \cdot u < 0$ and $v \cdot w < 0$ and $w \cdot u = 0$, or explain why no such pair of vectors can exist.
c) Find an example of three 2-vectors u, v, and w so that $u \cdot v < 0$ and $u \cdot w < 0$ and $v \cdot w < 0$, or explain why no such example exists.

Exercise 24. What shape is the set of solutions $\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)$ to the equation

$$\begin{pmatrix} 3 \\ 7 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} = 5?$$

That is, if we look at all possible vectors $\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)$ which make the equation true, what shape does this make in the plane? Draw this shape.

What happens if we change the vector $\left(\begin{smallmatrix} 3 \\ 7 \end{smallmatrix}\right)$ to some other vector? What happens if we change the number 5 to some other number?

Exercise 25. a) Find an example of a number c so that the equation

$$\begin{pmatrix} 1 \\ -1 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} = c$$

has the vector $\left(\begin{smallmatrix} 4 \\ 7 \end{smallmatrix}\right)$ as a solution, or explain why no such number exists.

b) Let $v = \left(\begin{smallmatrix} 2 \\ 1 \end{smallmatrix}\right)$ and $w = \left(\begin{smallmatrix} -3 \\ 4 \end{smallmatrix}\right)$. Find an example of a number c so that

$$v \cdot \begin{pmatrix} 1 \\ -1 \end{pmatrix} = c \quad \text{and} \quad w \cdot \begin{pmatrix} 1 \\ -1 \end{pmatrix} = c,$$

or explain why this is not possible.

c) Let $P = \left(\begin{smallmatrix} -3 \\ 4 \end{smallmatrix}\right)$. Find an example of numbers c and d so that

$$\begin{pmatrix} 2 \\ -1 \end{pmatrix} \cdot P = c \quad \text{and} \quad \begin{pmatrix} 1 \\ -1 \end{pmatrix} \cdot P = d,$$

or explain why no such example is possible.

Exercise 26. Write down two vectors in $\mathbb{R}^3$ which have no coordinates equal to zero. Call them u and v. Find the following things:

- The dot product $u \cdot v$;
- The norms of u and v;
- unit vectors which point in the same directions as u and v, respectively; and
- the angle between u and v.

1.5 Lines, Especially in $\mathbb{R}^2$

We will need a good understanding of lines in $\mathbb{R}^n$. We shall introduce lines with a geometric definition using vectors, and then show how to describe a line as the image of a parametric curve. Finally, we will focus on just the plane $\mathbb{R}^2$, and see how to describe a line as the set of solutions to an equation.

Lines as Parametrized Objects

Any pair of points should define a line. In $\mathbb{R}^n$, we can do this using vectors. It will be fruitful to regularly conflate a point with the (mathematician's) vector having the point as its head, so we will do that throughout this section.

Definition 22 (Lines in $\mathbb{R}^n$)**.** Suppose that P and Q are two distinct vectors in $\mathbb{R}^n$. We say that a point X *lies on the line through P and Q* when there exists scalars λ and μ which are not both zero so that

$$\lambda(X - P) = \mu(X - Q).$$

The *line through P and Q* is the set of all such points, that is, it is the set

$$\begin{aligned}&\{X \in \mathbb{R}^n \mid X \text{ lies on the line through } P \text{ and } Q\}\\ &= \{X \in \mathbb{R}^n \mid \lambda(X - P) = \mu(X - Q), \text{for some } \lambda \text{ and } \mu \text{ where } \lambda\mu \neq 0\}\end{aligned}$$

Of course, we want P and Q to both lie on the line through P and Q. Fortunately, this is true.

Theorem 23. Let P and Q be distinct vectors in $\mathbb{R}^n$. Then both P and Q lie on the line through P and Q.

Proof. We will show that P lies on the line through P and Q. We must check the condition from the definition with X replaced by P. This means we want to know about the vectors $X - P$ and $X - Q$, when $X = P$. But $P - P = 0$, so we can just choose $\lambda = 1$ and $\mu = 0$, and get the equation

$$\lambda(X - P) = 1(P - P) = 0(P - Q) = \mu(X - Q).$$

The proof that Q lies on the line through P and Q is basically the same, with some of the letters moved around. I will leave it to you to check the details of that case. $\square$

Theorem 24 (Only One Direction Matters)**.** Let P and Q be two distinct vectors in $\mathbb{R}^n$. A vector $X \in \mathbb{R}^n$ lies on the line through P and Q if, and only if, there exists a scalar t so that $X - P = t(Q - P)$.

Proof. First suppose that X lies on the line through P and Q. Then, by definition there exists a pair of scalars λ and μ so that

$$\lambda(X - P) = \mu(X - Q). \tag{1.1}$$

Notice that we cannot have $\lambda = \mu$. The reason is that if $\lambda = \mu \neq 0$, then we can cancel them and deduce that $X - P = X - Q$, which means that $P = Q$. But we have explicitly chosen P and Q to be distinct. So, we will proceed knowing that λ and μ are different.

Rearranging equation (1.1) , we see that

$$(\lambda - \mu)X - \lambda P = -\mu Q.$$

By adding μP to both sides and gently regrouping things, we find

$$\begin{aligned}
(\lambda - \mu)X - \lambda P + \mu P &= -\mu Q + \mu P \\
(\lambda - \mu)X - (\lambda - \mu)P &= -\mu(Q - P) \\
(\lambda - \mu)(X - P) &= -\mu(Q - P)
\end{aligned}$$

Since $\lambda - \mu \neq 0$, we can divide by it, and we see that

$$X - P = \frac{-\mu}{\lambda - \mu}(Q - P).$$

So we can choose $t = \dfrac{-\mu}{\lambda - \mu}$ to satisfy the condition of the theorem.

Now suppose that X is a point so that $X - P = t(Q - P)$ for some scalar t. We must find the scalars λ and μ which satisfy equation (1.1). Since $X - P = t(Q - P)$, we have that

$$X = tQ + (1 - t)P \tag{1.2}$$

This means that $X - Q = (t - 1)(Q - P)$. So if we choose $\lambda = t - 1$ and $\mu = t$, we have

$$\begin{aligned}
\lambda(X - P) &= (t - 1)(X - P) \\
&= (t - 1)t(Q - P) \\
&= t(X - Q) \\
&= \mu(X - Q).
\end{aligned}$$

Therefore, X lies on the line through P and Q. $\square$

Equation (1.2) is a beautifully symmetric description for points X which lie on the line through P and Q. Many authors, especially those who want to do lots of geometry, emphasize this description.

Also, notice the importance this theorem places on the vector $Q - P$. This vector is useful enough for dealing with the line that it has a special name.

Definition 25 (Direction Vector)**.** Let P and Q be two distinct vectors in $\mathbb{R}^n$. The vector $Q-P$ is called a *direction vector* for the line through P and Q.

The order of the points P and Q really doesn't matter for defining a line, so we can see that $P-Q$ is a direction vector, too. It just points in the opposite direction. Furthermore, a line has lots of points on it, and any pair of them is good enough to describe the line. Choosing a different pair for P and Q will lead to different direction vectors, but only different by rescaling. All of the vectors will point in the same direction. So keep in mind that a line has lots of direction vectors, but only one direction.

Corollary 26 (Parametric form of a Line)**.** Let P and Q be two distinct vectors in $\mathbb{R}^n$. The line through P and Q is the set

$$\{X \in \mathbb{R}^n \mid X = P + t(Q-P) \text{ for some scalar } t\}.$$

Proof. This condition is a simple rearrangement of the one in Theorem 24. Just add P to both sides of the description there. ☐

Corollary 27 (Lines through the origin)**.** A line in $\mathbb{R}^n$ can be written in the form

$$\{X \in \mathbb{R}^n \mid X = tV\},$$

where V is some fixed vector.

Proof. Essentially, this is the case where $P = 0$. We replace $Q = Q - P$ by V. ☐

Note that the vector $X = P + t(Q-P)$ changes as t changes. We often think of this as defining a function of the form $\gamma : t \mapsto P + t(Q-P)$. (By the way, γ is a version of a Greek letter and is read "gamma.") Usually, the best way is to think of the number t, called the *parameter*, as representing time. As the time t changes, the vector $X = \gamma(t)$ moves around in $\mathbb{R}^n$. In the case here, the point $X = \gamma(t)$ traces out the shape of a line as time moves on.

If you prefer to think about the whole vector $\gamma(t)$, rather than just its head, it helps to imagine that vector sweeping through space as time goes on. The heads of the vectors still trace out the line, but the tails of the vectors all stay at the origin.

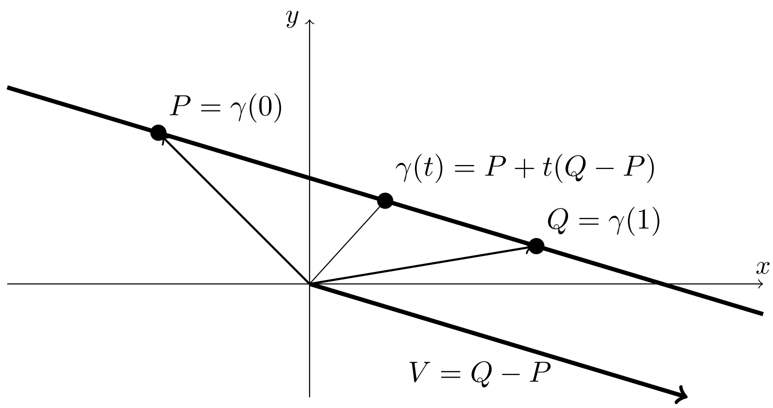


Figure 1.15: parametric line and a direction vector

You might see parametric functions in other places (like a course on calculus), but written differently. It is common to break apart a vector description into a system of component functions. We will have occasion to use this, too, so let's see how it is done. Begin with a parametric line

$$\gamma(t) = P + t(Q - P).$$

Write out the vectors $\gamma(t)$, P and Q as stacks of components,

$$\gamma(t) = \begin{pmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{pmatrix} \qquad P = \begin{pmatrix} P_1 \\ P_2 \\ \vdots \\ P_n \end{pmatrix} \qquad Q = \begin{pmatrix} Q_1 \\ Q_2 \\ \vdots \\ Q_n \end{pmatrix}$$

and unpack the definition using the algebra of linear combinations.

$$\begin{pmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{pmatrix} = \begin{pmatrix} P_1 \\ P_2 \\ \vdots \\ P_m \end{pmatrix} + t \begin{pmatrix} Q_1 - P_1 \\ Q_2 - P_2 \\ \vdots \\ Q_n - P_n \end{pmatrix} = \begin{pmatrix} P_1 + t(Q_1 - P_1) \\ P_2 + t(Q_2 - P_2) \\ \vdots \\ P_n + t(Q_n - P_n) \end{pmatrix}$$

Now, just read off each component, one at a time, to make a system of functions. It sounds like cheating, but you basically just erase the parentheses and then group with a big curly brace on the left instead.

$$\left\{\begin{array}{lcl} x_1 & = & P_1 + t(Q_1 - P_1) \\ x_2 & = & P_2 + t(Q_2 - P_2) \\ \vdots & & \quad\quad\vdots \\ x_n & = & P_n + t(Q_n - P_n) \end{array}\right.$$

Sometimes people will write $x_i(t)$ with the parameter t explicitly present, and sometimes they will write just x_i and leave it as understood that x_i is a function of t. I have left the t's out of the final expression because it was convenient.

Implicit Description: The Equation of a Line in $\mathbb{R}^2$

Now we narrow our attention to the plane $\mathbb{R}^2$. We will see that in this case it is possible to also describe a line as the set of points satisfying a single, simple equation.

Theorem 28. Let P and Q be two distinct vectors in $\mathbb{R}^2$. Then there are numbers a, b, and c so that any point $X = \left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right) \in \mathbb{R}^2$ which lies on the line through P and Q must satisfy the equation

$$ax + by = c.$$

Proof. Let P and Q have components as follows:

$$P = \begin{pmatrix} P_1 \\ P_2 \end{pmatrix} \qquad Q = \begin{pmatrix} Q_1 \\ Q_2 \end{pmatrix}$$

By the discussion at the end of the last subsection, we can find a value of t so that

$$\left\{\begin{array}{lcl} x & = & P_1 + t(Q_1 - P_1) \\ y & = & P_2 + t(Q_2 - P_2). \end{array}\right.$$

Our goal is to eliminate t from these expressions and derive a single equation relating x and y. The mechanics of this is as follows: multiply the first equation through by $Q_2 - P_2$, multiply the second equation through by $-(Q_1 - P_1)$, and then add them. In a way, we are making a "linear combination of the equations." The result is

$$\begin{aligned} (Q_2 - P_2)x - (Q_1 - P_1)y &= (Q_2 - P_2)P_1 - (Q_1 - P_1)P_2 \\ &= Q_2P_1 - Q_1P_2 \end{aligned} \tag{1.3}$$

So, we choose $a = Q_2 - P_2$, $b = -(Q_1 - P_1)$ and $c = Q_2P_1 - Q_1P_2$. Then equation (1.3) reads as $ax + by = c$. This completes the proof. $\square$

Theorem 29. Fix numbers a, b, and c so that a and b are not both zero. Then the set of points $X = \left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)$ which satisfy the equation $ax + by = c$ is a line in $\mathbb{R}^2$.

Proof. We will work in the case where $a \neq 0$, and leave the case $a = 0$ for you to fill in.

The main idea is to turn the defining equation into a parametric description, and along the way find vectors to play the parts of P and $V = Q - P$.

We will pretend that y is our parameter by giving it a new name. So, introduce the parameter t and declare that $y = t$. Take the defining equation $ax + by = c$, isolate x and substitute in $y = t$ to remove reference to y. Then our information is represented by these two equations.

$$\begin{cases} x &= c/a - (b/a)\,t \\ y &= t \end{cases}$$

Rewrite this as a vector equation.

$$X = \begin{pmatrix} c/a \\ 0 \end{pmatrix} + t \begin{pmatrix} -b/a \\ 1 \end{pmatrix}$$

So if we choose P and Q as below, we have written exactly the parametric description of a line as in Corollary 26.

$$P = \begin{pmatrix} c/a \\ 0 \end{pmatrix} \qquad Q = \begin{pmatrix} c/a - b/a \\ 1 \end{pmatrix}$$

Therefore, our collection of points is exactly a line. This completes the proof. (You should go back and fill in how to think about the case when $a = 0$. Hint: if $a = 0$, what do you know about b?) $\square$

Looking carefully at the last two results together, we see that lines in $\mathbb{R}^2$ come with two different descriptions, a parametric description and an implicit description, but we can easily pass back and forth between them. In fact, the proofs of the last two results give us explicit methods for passing back and forth between them. In the first, we eliminate the parameter. In the second, we have to introduce a new one out of nowhere.

Theorem 30. Let a line in the plane be described as the set of solutions to the equation $ax + by = c$. A direction vector for this line is

$$V = \begin{pmatrix} -b/a \\ 1 \end{pmatrix}.$$

Proof. This follows directly from our work in the proof of Theorem 29. $\square$

While direction vectors are useful for lines in $\mathbb{R}^2$, in $\mathbb{R}^n$ we will have other, bigger objects defined by equations. In those cases, a single direction vector will not be sufficient. But we can keep things simple by changing perspective just a little.

Definition 31 (Normal Vector)**.** Suppose we have a line in $\mathbb{R}^2$ through points P and Q. A vector n is called a *normal vector* for this line if n is orthogonal to the direction vector $V = Q - P$. That is, n is a normal vector when $n \cdot (Q - P) = 0$.

Theorem 32. Let ℓ be a line in $\mathbb{R}^2$ defined as the set of solutions to the equation $ax + by = c$. Then one normal vector for ℓ is given by

$$n = \begin{pmatrix} a \\ b \end{pmatrix}.$$

Proof. By Theorem 30, the direction vector for ℓ is $V = \left(\begin{smallmatrix} -b/a \\ 1 \end{smallmatrix}\right)$. It is straightforward to check that

$$n \cdot V = \begin{pmatrix} a \\ b \end{pmatrix} \cdot \begin{pmatrix} -b/a \\ 1 \end{pmatrix} = -b + b = 0.$$

Therefore, n is a normal vector for ℓ. $\square$

Note that the normal vector is pretty handy for writing out the equation of a line. If we write

$$n = \begin{pmatrix} a \\ b \end{pmatrix} \qquad X = \begin{pmatrix} x \\ y \end{pmatrix}$$

then the equation $ax + by = c$ is the same thing as $n \cdot X = c$. This gives us a connection between the dot product and the linear equation which will be useful later.

Exercises

Exercise 27. Write down five different points which lie on the line described parametrically as:

$$t \mapsto \begin{pmatrix} -5 \\ 3 \end{pmatrix} + t \begin{pmatrix} 1 \\ -1/2 \end{pmatrix}.$$

Plot these points and the line.

Exercise 28. Write down a parametric description for the line which passes through the origin $O = (0, 0)$ and the point $S = (-5, 5)$.

Is this the *only* way to write down such a parametric description for that line?

Exercise 29. Write down a parametric description for the line which passes through the points below.

$$T = (\pi, 0), \qquad J = (0, -\pi)$$

Now find a different parametric description for that line. (*Hint: Can you find a way to write a parametric description that doesn't use the number* π*?*)

Exercise 30. For each of the conditions below, either find an example of a 2-vector Z so that the equation

$$\begin{pmatrix} 1 \\ 2 \end{pmatrix} + t \begin{pmatrix} -1 \\ 1 \end{pmatrix} = Z$$

has the given number of solutions, or explain why such an example is not possible.

a) exactly zero solutions;
b) exactly one solution;
c) exactly two solutions.

Exercise 31. For each of the conditions below, either find an example of a 2-vector Y so that the equation

$$\begin{pmatrix} -2/5 \\ 2 \end{pmatrix} + tY = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

has the given number of solutions, or explain why such an example is not possible.

a) exactly zero solutions;
b) exactly one solution;

c) exactly two solutions.

Exercise 32. What shape is the set of solutions $\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)$ to the equation

$$\begin{pmatrix} 3 \\ 7 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} = 5?$$

That is, if we look at all possible vectors $\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)$ which make the equation true, what shape does this make in the plane? Draw this shape.

What happens if we change the vector $\left(\begin{smallmatrix} 3 \\ 7 \end{smallmatrix}\right)$ to some other vector? What happens if we change the number 5 to some other number?

Exercise 33. We begin with a line described parametrically by

$$t \mapsto \begin{pmatrix} 6 \\ -\pi \end{pmatrix} + t \begin{pmatrix} 34 \\ -19/3 \end{pmatrix}.$$

a) Find a normal vector for this line.
b) Plot the line and the normal vector you found.
c) Find an equation for this line.

Exercise 34. We begin with a line described by the equation

$$-3x + y = 7.$$

a) Find a normal vector to the line
b) Plot this line and the normal vector you found.
c) Find a parametric description for this line.

Chapter 2

Rows and Solving Systems

Introduce the major goals of this chapter: solving systems of equations, completely, and studying things from the point of view of the rows

sample task: fitting conics and cubics to sets of points exactly.

2.1 Hyperplanes and the Row Picture

We now take up the task of learning to solve systems of linear equations. As a start, we shall be more careful about what one single equation is, and how it may be represented.

Definition 33 (Linear Equation)**.** A *linear equation the n unknowns x_1, x_2, ..., x_n* is an equation of the form

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b, \tag{2.1}$$

where each of the a_i's and b are scalars. The scalars a_i are called *coefficients*. Notice that some of the a_i's may be zero, in which case it looks like the corresponding terms are not present. In this case, it may still be necessary to remember that those unknowns matter, depending on the context.

A vector in R^n is said to be a *solution* of the equation when its components, in order, are substituted in for the unknowns x_i the equation becomes true.

Remark 34. A linear equation in n unknowns is one that can be rewritten using the dot product in $\mathbb{R}^n$. In particular, if we write the coefficients in a vector, and the unknowns in a vector like so

$$N = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \qquad X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix},$$

then Equation (2.1) becomes exactly $N \cdot X = b$.

Theorem 35. Let P be a solution to the linear Equation 2.1. Then the set of solutions to Equation (2.1) is equal to the the set

$$\mathcal{S} = \{X \in \mathbb{R}^n \mid N \cdot (X - P) = 0\},$$

where N is the vector of coefficients

$$N = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}.$$

Proof. Fix a solution P to Equation (2.1):

$$P = \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_n \end{pmatrix}.$$

Then by definition we have

$$N \cdot P = a_1p_1 + a_2p_2 + \cdots + a_np_n = b. \tag{2.2}$$

Now, suppose that X is some solution to the equation. Then $N \cdot X = b$, or

$$N \cdot X = a_1x_1 + a_2x_2 + \cdots + a_nx_n = b. \tag{2.3}$$

Using the left hand equalities of Equation (2.2) and Equation (2.3) and doing some rearranging, we see that

$$\begin{aligned} N \cdot X - N \cdot P &= a_1x_1 + a_2x_2 + \cdots + a_nx_n - (a_1p_1 + a_2p_2 + \cdots + a_np_n) \\ &= a_1(x_1 - p_1) + a_2(x_2 - p_2) + \cdots + a_n(x_n - p_n) \\ &= N \cdot (X - P). \end{aligned}$$

But using the right-hand equalities of those equations we see

$$N \cdot X - N \cdot P = b - b = 0.$$

So we conclude that if X is a solution to Equation (2.1), then $N \cdot (X - P) = 0$.

Now suppose that X is some vector for which $N \cdot (X - P) = 0$. We rearrange the equation like this:

$$\begin{aligned} N \cdot (X - P) &= 0 \\ N \cdot X - N \cdot P &= 0 \\ N \cdot X &= N \cdot P \end{aligned}$$

Now we substitute in from the right-hand equality in Equation (2.2) and expand out the dot product on the left to see

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b.$$

This means that X is a solution of Equation (2.1).

Since we have proved that an element of $\mathcal{S}$ is a solution of (2.1) and vice versa. The proof is complete. $\square$

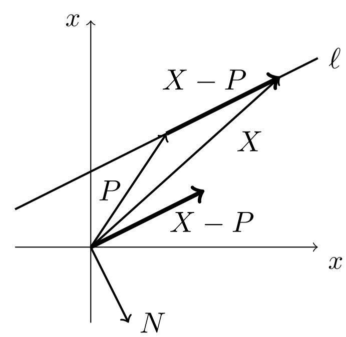


Figure 2.1: A hyperplane ℓ in $\mathbb{R}^2$ and a normal vector

This last theorem connects the idea of the set of solutions to a linear equation to a situation involving a dot product being zero, that is, of orthogonality. The basic picture is like the one in Figure 2.1. Note the right angle between $X - P$ and N there.

Definition 36 (Hyperplane, normal vector)**.** The set of all solutions to a linear equation in n unknowns is called a *hyperplane in* $\mathbb{R}^n$.

If the linear equation is written in dot product form $N \cdot X = b$, then the vector $N \in \mathbb{R}^n$ which consists of the coefficients of the equation is called a *normal vector* to the hyperplane.

More generally, any vector N which has the property that $N \cdot (P - Q) = 0$ for every pair of vectors P and Q which are in the hyperplane is called a *normal vector* to that hyperplane.

Theorem 37. Suppose that $N_1, N_2 \in \mathbb{R}^n$ and b_1, b_2 are scalars. If there is a non-zero scalar λ so that $N_1 = \lambda N_2$ and $b_1 = \lambda b_2$, then the two equations $N_1 \cdot X = b_1$ and $N_2 \cdot X = b_2$ give the same hyperplane.

Proof. We shall consider the following two hyperplanes.

$$\mathcal{H}_1 = \{X \in \mathbb{R}^n \mid N_1 \cdot X = b_1\}$$
$$\mathcal{H}_2 = \{X \in \mathbb{R}^n \mid N_2 \cdot X = b_2\}$$

Suppose that there is a non-zero scalar λ so that $N_1 = \lambda N_2$ and $b_1 = \lambda b_2$. If $X \in \mathcal{H}_1$, then $N_1 \cdot X = b_1$. By our hypothesis, this is the same as $\lambda N_2 \cdot X = \lambda b_2$. By cancelling the non-zero scalar λ we see that $X \in \mathcal{H}_2$.

Similarly, if $X \in \mathcal{H}_2$, then $N_2 \cdot X = b_2$. Multiplying by the non-zero scalar λ, we get that X satisfies $N_1 \cdot X = \lambda N_2 \cdot X = \lambda b_2 = b_1$, so X is an element of $\mathcal{H}_1$. $\square$

The converse statement is true, too. That is, if two hyperplanes are equal, then their equations are proportional. But a proof of that is surprisingly tricky. It will have to wait until after we develop some techniques. For now, we can do this little piece.

Definition 38 (Parallel Hyperplanes)**.** Two hyperplanes are called *parallel* when they do not intersect. That is, two hyperplanes are called parallel when there is no vector which satisfies both of their equations, and hence simultaneously is a solution to both linear equations.

Theorem 39. Given vectors $N_1, N_2 \in \mathbb{R}^n$ and scalars b_1, b_2, suppose that there is a scalar lambda so that $N_1 = \lambda N_2$, but $b_1 \neq \lambda b_2$. Then the hyperplanes

$$\mathcal{H}_1 = \{X \in \mathbb{R}^n \,|\, N_1 \cdot X = b_1\}$$
$$\mathcal{H}_2 = \{X \in \mathbb{R}^n \,|\, N_2 \cdot X = b_2\}$$

are parallel.

Proof. Suppose, to the contrary, that there is a vector X which is an element of both hyperplanes. Then X satisfies both of their equations. That means that

$$b_1 = N_1 \cdot X = \lambda N_2 \cdot X = \lambda b_2.$$

This directly contradicts the assumption that $b_1 \neq \lambda b_2$, so we conclude there is no vector X which is an element of both hyperplanes. Hence $\mathcal{H}_1$ and $\mathcal{H}_2$ are parallel. □

Hyperplanes in $\mathbb{R}^2$

A hyperplane in $\mathbb{R}^2$ is the set of solutions to a linear equation in exactly two unknowns.

$$\mathcal{H} = \left\{ X = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2 \,\middle|\, a_1x_1 + a_2x_2 = b \right\}$$

We saw in the last section of Chapter One that this describes a line. Usually, one writes just x and y in place of x_1 and x_2 in these cases, but either style of notation will do.

Hyperplanes in $\mathbb{R}^3$

What is a hyperplane in $\mathbb{R}^3$? Well, by definition, it is the set of solutions to a linear equation in three unknowns.

$$\mathcal{H} = \left\{ X \in \mathbb{R}^3 \,\middle|\, a_1x_1 + a_2x_2 + a_3x_3 = b \right\}$$

Theorem 40. A hyperplane in $\mathbb{R}^3$ may be given a parametric description as the set of points traced out by a parametric function of two variables of the form

$$s, t \mapsto P + sU + tV,$$

where $P, U, V \in \mathbb{R}^3$.

Before we prove this theorem, let's address the notion of a function of two parameters. This is a lot like the idea of the parametric description of a line in the plane, except we are allowed **two** different parameters, instead of just one. The geometric idea here is that you can imagine standing at a point P, and then being allowed to move in the direction of U as far as you like, or in the direction of V as far as you like, completely independently! In fact, you can take any combination of a motion in the direction of U and a motion in the direction of V. Now, just motion away from P in the direction of U would trace out a line.

$$s \mapsto P + sU.$$

But we are allowed more. For a particular value of s, we can think of $P + sU$ as a starting point, and then we can trace out the line

$$t \mapsto (P + sU) + tV.$$

Some imagining might help you see that by varying both of these choices the image of the parametric function is a plane in $\mathbb{R}^3$. So we see that hyperplanes in $\mathbb{R}^3$ are just what we would usually call a plane.

Proof. We will use x, y, and z as coordinate names in $\mathbb{R}^3$. Suppose the defining equation of our hyperplane is

$$ax + by + cz = d.$$

At least one of the three coefficients a, b, and c is not zero. By changing our perspective on which variable is "first," we can assume that $a \neq 0$.

We introduce two parameter names for two of the variables: $y = s$ and $z = t$. If we rearrange the first equation slightly and incorporate our two new equations, we can write

$$\begin{cases} x &= \frac{d}{a} - \frac{b}{a}s - \frac{c}{a}t \\ y &= \qquad\quad s \\ z &= \qquad\qquad\quad t \end{cases}$$

Now we reinterpret this as a vector equation.

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} d/a \\ 0 \\ 0 \end{pmatrix} + s \begin{pmatrix} -b/a \\ 1 \\ 0 \end{pmatrix} + t \begin{pmatrix} -c/a \\ 0 \\ 1 \end{pmatrix}$$

This is the form required by the theorem, so the proof is complete. □

So we see that it is possible to take a description of a hyperplane in $\mathbb{R}^3$ which uses an equation, and find a description that uses a parametric function. It is possible to work in the other direction, too.

Theorem 41. If P, U, V are vectors in $\mathbb{R}^3$, then the set of points traced out by the parametric function

$$s, t \mapsto P + sU + tV$$

is a hyperplane whose defining equation can be computed directly from P, U, and V.

Proof. Again, we will use x, y, and z as coordinate names in $\mathbb{R}^3$. Write each of the given vectors out in coordinates

$$P = \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix} \qquad U = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} \qquad V = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}$$

Then the parametric description means that

$$\begin{cases} x &= p_1 + u_1 s + v_1 t \\ y &= p_2 + u_2 s + v_2 t \\ z &= p_3 + u_3 s + v_3 t \end{cases}$$

Our method will be to eliminate the parameters s and t from the equations, slowly combining things until only one equation remains and it involves only x, y, and z.

First, consider the first two equations. We eliminate s by multiplying the first by u_2 and the second by $-u_1$ and adding. We obtain

$$u_2 x - u_1 y = u_2 p_1 - u_1 p_2 + (u_2 v_1 - u_1 v_2)t. \tag{2.4}$$

Next, consider the second and third equations. We will eliminate s from these by multiplying the second equation by u_3 and the third by $-u_2$ and adding. We obtain

$$u_3 y - u_2 z = u_3 p_2 - u_2 p_3 + (u_3 v_2 - u_2 v_3)t. \tag{2.5}$$

Note that we have replace the original three equations in the variable x, y, z, s, t by two equations in the variables x, y, z, t. We now do one more elimination to remove t. Multiply Equation (2.4) by $(u_3v_2 - u_2v_3)$ and multiply Equation (2.5) by $-(u_2v_1 - u_1v_2)$ and then add. We obtain

$$\begin{aligned}(u_3v_2 - u_2v_3)(u_2x - u_1y)&-(u_2v_1 - u_1v_2)(u_3y - u_2z) = \\ &(u_3v_2 - u_2v_3)(u_2p_1 - u_1p_2) - (u_2v_1 - u_1v_2)(u_3p_2 - u_2p_3)\end{aligned} \tag{2.6}$$

This is a linear equation in three unknowns, so the proof is complete. $\square$

That equation has lots of nasty terms in it when written in generality like this. It is usually not so scary in practice. It is possible to patiently clean up this equation a bit so it is slightly less frightening. (See the last exercise of this section.)

The Idea of a Row Picture

Now we are ready for the predominant object of our study.

Definition 42 (System of Linear Equations)**.** A *system of m linear equations in n unknowns* is a set of equations of the form

$$\begin{cases} a_{11}x_1 & + & a_{12}x_2 & + & \dots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \dots & + & a_{2n}x_n & = & b_2 \\ \vdots & & \vdots & & \ddots & & \vdots & & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \dots & + & a_{mn}x_n & = & b_m \end{cases}, \tag{2.7}$$

where the x_i's are the unknowns, and the a_{ij}'s are known scalars called *coefficients*.

A *solution* to this system of equations is a vector X which is a solution of each of the m different equations at the same time.

Sometimes, we use a bit of shorthand and refer to a system of m linear equations in n unknowns as an "$m \times n$ system."

Remark 43 (The Row Picture)**.** A solution of the system (2.7) should be a solution for each of its m constituent equations. Since the set of solutions of each single equation is represented by hyperplane in $\mathbb{R}^n$, we see that the set of solutions of the system of equations is represented by the set of common intersections of m different hyperplanes in $\mathbb{R}^n$.

The "picture" consisting of the m different hyperplanes in $\mathbb{R}^n$ is referred to as the *row picture* of the system (2.7).

For example, the row picture of a 2×2 system consists of two lines in the plane, and the row picture of a 5×13 system consists of 5 hyperplanes in $\mathbb{R}^{13}$.

Exercises

Exercise 35. Choose a linear equation in 3 unknowns. (Just pick one that looks interesting to you.) Make SageMath plot of the corresponding hyperplane in $\mathbb{R}^3$.

How can you make a similar plot by hand? (Hint: how would you plot a line in the plane by hand?)

Exercise 36. Find a normal vector for each of the hyperplanes described below:
a) $\mathcal{H}_1 = \{X \in R^3 \mid 4x - 2y + \pi z = 12\}$
b) $\mathcal{H}_2 = \{P + sU + tV \in \mathbb{R}^3 \mid s, t \in \mathbb{R}\}$, where

$$P = \begin{pmatrix} 2 \\ 1 \\ -1 \end{pmatrix}, \quad U = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad V = \begin{pmatrix} -2 \\ -1 \\ 3 \end{pmatrix}.$$

Exercise 37. Suppose that the hyperplane $\mathcal{H}$ in $\mathbb{R}^4$ contains the point $P = (1, -2, 2, 1)$ and has normal vector N as below. Find an equation that describes $\mathcal{H}$.

Exercise 38. Suppose that the hyperplane $\mathcal{T}$ is parallel to the hyperplane $\mathcal{U}$ given below, and that $\mathcal{T}$ contains the point $Q = (0, 1, -4, 0)$. Find an equation for $\mathcal{T}$.

$$\mathcal{U} = \{X \in \mathbb{R}^4 \mid 5x_1 + x_2 - x_3 - 2x_4 = 1\}$$

Exercise 39. Consider the hyperplane $\mathcal{U}$ in the last exercise. Give a parametric description for $\mathcal{U}$.

Exercise 40. Consider the hyperplane $\mathcal{H}_2$ in Exercise 36(b). Find an equation which describes this hyperplane.

Exercise 41. Consider the three points $A = (1, 1, 1)$, $B = (2, 1, 3)$ and $C = (0, 1, 4)$ in $\mathbb{R}^3$. Find a way to describe the plane which goes through these three points. (Any description will do.)

Exercise 42. Clean up Equation (2.6). Can you make it look relatively nice?

2.2 Affine Subsets

Definition 44 (Affine Subset)**.** A set $\mathcal{A}$ in $\mathbb{R}^n$ is called an *affine subset* when for each pair of points P and Q in $\mathcal{A}$, if X is a point on the line through P and Q, X is an element of $\mathcal{A}$.

Theorem 45 (Hyperplanes are Affine)**.** Let $\mathcal{H}$ be a hyperplane in $\mathbb{R}^n$. Then $\mathcal{H}$ is an affine subset of $\mathbb{R}^n$.

Proof. Consider the hyperplane $\mathcal{H}$ described as

$$\mathcal{H} = \{X \in \mathbb{R}^n \mid a_1x_1 + \cdots + a_nx_n = b\}.$$

Suppose that $P, Q \in \mathbb{R}^n$ are elements of $\mathcal{H}$. The line through P and Q can be described as the set of all points of the form $P + t(Q - P)$, where t is a scalar. We must show that, for each t, the point $P + t(Q - P)$ is an element of $\mathcal{H}$.

If we write the components of P as p_i and the components of Q as q_i, then the fact that P and Q are elements of $\mathcal{H}$ means that the p_i's and q_i's satisfy the defining equation. That is,

$$a_1p_1 + \cdots + a_np_n = b \tag{2.8}$$
$$a_1q_1 + \cdots + a_nq_n = b. \tag{2.9}$$

Subtracting these, we see that

$$a_1(q_1 - p_1) + \cdots + a_n(q_n - p_n) = 0.$$

We can multiply this equation through by any number t to see that

$$a_1[t(q_1 - p - 1)] + \cdots + a_n[t(q_n - p_n)] = 0. \tag{2.10}$$

Adding Equation (2.8) to Equation (2.10), we see that

$$a_1\left[p_1 + t(q_1 - p_1)\right] + \cdots + a_n\left[p_n + t(q_n - p_n)\right] = b.$$

This means that the coordinates of the vector $P + t(Q - P)$ also satisfy the defining equation for $\mathcal{H}$.

We deduce that if P and Q are in $\mathcal{H}$, then so is the whole line through them. This completes the proof. $\square$

Theorem 46 (Solutions of Systems are Affine)**.** The solution set to a system of m linear equations in n unknowns is an affine subset of $\mathbb{R}^n$.

Proof. The solution set to a system of m linear equations in n unknowns is exactly the intersection of a set of m hyperplanes in $\mathbb{R}^n$. Let's call them $\mathcal{H}_1$, $\dots$, $\mathcal{H}_m$.

Suppose that P and Q are in the solution set. Then for each hyperplane $\mathcal{H}_i$, P and Q must be elements of $\mathcal{H}_i$. By the Theorem 45, the line through P and Q must also contained in $\mathcal{H}_i$. Since this is true for each of the hyperplanes, this line must be contained in the intersection of all of them. That is, the line through P and Q must be contained in the solution set.

By definition, this means that the solution set is an affine subset of $\mathbb{R}^n$. □

Corollary 47. A system of m linear equations in n unknowns can have only the following possibilities for the number of solutions: zero solutions, one solution, or infinitely many solutions.

Proof. Note that since the solution set to such a system is an affine subset, if it contains at least two points, then it must contain an entire line, too. That line has infinitely many points on it. So we are left with these options: the solution set has strictly fewer than two elements, or it has infinitely many elements. □

This result is a rough estimate. Later we shall want to know more about the shape of the solution set. It will matter if it is infinite like a line is, or infinite like a plane, or something "larger."

Definition 48. A system of m linear equations in n unknowns is called *homogeneous* when it has the form

$$\left\{\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \dots & + & a_{1n}x_n & = & 0 \\ a_{21}x_1 & + & a_{22}x_2 & + & \dots & + & a_{2n}x_n & = & 0 \\ \vdots & & \vdots & & \ddots & & \vdots & & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \dots & + & a_{mn}x_n & = & 0 \end{array}\right. ,$$

If we have a system which is not homogeneous, we call the system obtained by setting all the right-hand sides to 0 the *associated homogeneous system.*

Definition 49 (Parallel Affine Subsets)**.** Let $\mathcal{A}_1$ and $\mathcal{A}_2$ be two affine subsets of $\mathbb{R}^n$. We say that $\mathcal{A}_1$ and $\mathcal{A}_2$ are *parallel* when there exists a vector P such that

$$\mathcal{A}_2 = \{X + P \mid X \in \mathcal{A}_1\}.$$

Remark 50. Note the symmetry in the description of parallel affine subsets. If we can realize $\mathcal{A}_2$ as $\mathcal{A}_1$ but shifted over by P, then we can also realize $\mathcal{A}_1$ as just $\mathcal{A}_2$ but shifted over by $-P$.

Theorem 51. The solution set to a system of m linear equations in n unknowns is parallel to the solution set of the associated homogeneous system, when considered as affine subsets of $\mathbb{R}^n$.

In fact, the vector P required in the definition of parallel affine subsets can be chosen to be any solution to the original system.

Proof. We consider the system

$$\left\{\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \dots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \dots & + & a_{2n}x_n & = & b_2 \\ \vdots & & \vdots & & \ddots & & \vdots & & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \dots & + & a_{mn}x_n & = & b_m \end{array}\right., \tag{2.11}$$

and its associated homogeneous system

$$\left\{\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \dots & + & a_{1n}x_n & = & 0 \\ a_{21}x_1 & + & a_{22}x_2 & + & \dots & + & a_{2n}x_n & = & 0 \\ \vdots & & \vdots & & \ddots & & \vdots & & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \dots & + & a_{mn}x_n & = & 0 \end{array}\right.. \tag{2.12}$$

We shall denote the affine subset of $\mathbb{R}^n$ which consists of the solutions of (2.11) by $\mathcal{A}$. We shall denote the affine subset made of the solutions of (2.12) by $\mathcal{A}_0$. We must show that these two affine subsets are parallel.

Fix some solution P of the system (2.11). If we write the coordinates of P as p_i, we get

$$\left\{\begin{array}{ccccccccc} a_{11}p_1 & + & a_{12}p_2 & + & \dots & + & a_{1n}p_n & = & b_1 \\ a_{21}p_1 & + & a_{22}p_2 & + & \dots & + & a_{2n}p_n & = & b_2 \\ \vdots & & \vdots & & \ddots & & \vdots & & \vdots \\ a_{m1}p_1 & + & a_{m2}p_2 & + & \dots & + & a_{mn}p_n & = & b_m \end{array}\right.. \tag{2.13}$$

We want to show that $\mathcal{A}$ and $\mathcal{A}_0$ are parallel affine subsets, given by addition by P, so we must show that

$$\mathcal{A} = \{X + P \mid X \in \mathcal{A}_0\}.$$

This involves writing two short arguments: one that says each element of the set on the left is also an element of the set on the left, and vice versa.

Now if Y is any element of $\mathcal{A}$, that is a solution of the system (2.11), and we write the coordinates of Y as y_i's, we have

$$\left\{\begin{array}{ccccccccc} a_{11}y_1 & + & a_{12}y_2 & + & \dots & + & a_{1n}y_n & = & b_1 \\ a_{21}y_1 & + & a_{22}y_2 & + & \dots & + & a_{2n}y_n & = & b_2 \\ \vdots & & \vdots & & \ddots & & \vdots & & \vdots \\ a_{m1}y_1 & + & a_{m2}y_2 & + & \dots & + & a_{mn}y_n & = & b_m \end{array}\right. . \tag{2.14}$$

Next we shall consider the vector $X = Y - P$. We claim that X is an element of $\mathcal{A}_0$, that is, X is a solution to the homogeneous system Equation (2.12). This is straightforward to check, as we can just subtract the corresponding equations of (2.13) from those in (2.14). This gives us that the components of X satisfy

$$\left\{\begin{array}{ccccccccc} a_{11}(y_1 - p_1) & + & a_{12}(y_2 - p_2) & + & \dots & + & a_{1n}(y_n - p_n) & = & 0 \\ a_{21}(y_1 - p_1) & + & a_{22}(y_2 - p_2) & + & \dots & + & a_{2n}(y_n - p_n) & = & 0 \\ \vdots & & \vdots & & \ddots & & \vdots & & \vdots \\ a_{m1}(y_1 - p_1) & + & a_{m2}(y_2 - p_2) & + & \dots & + & a_{mn}(y_n - p_n) & = & 0 \end{array}\right. .$$

Therefore, X is a solution of the homogeneous equations.

So, every element of $\mathcal{A}$ can be written in the form $Y = X + P$, where $X = Y - P$ is an element of $\mathcal{A}_0$.

The other direction is a bit quicker. Suppose that X is a vector which represents a solution to (2.12). Then by adding these equations to those from (2.13) which say P is an element of $\mathcal{A}$, we see that

$$\left\{\begin{array}{ccccccccc} a_{11}(x_1 + p_1) & + & a_{12}(x_2 + p_2) & + & \dots & + & a_{1n}(x_n + p_n) & = & 0 \\ a_{21}(x_1 + p_1) & + & a_{22}(x_2 + p_2) & + & \dots & + & a_{2n}(x_n + p_n) & = & 0 \\ \vdots & & \vdots & & \ddots & & \vdots & & \vdots \\ a_{m1}(x_1 + p_1) & + & a_{m2}(x_2 + p_2) & + & \dots & + & a_{mn}(x_n + p_n) & = & 0 \end{array}\right. .$$

This means that $X + P$ is an element of $\mathcal{A}$.

Putting these together, we see that $\mathcal{A} = \{X + P \mid X \in \mathcal{A}_0\}$. Hence, $\mathcal{A}$ and $\mathcal{A}_0$ are parallel. $\square$

Exercises

NOTE: throughout this section, you may find it useful to use the computer to make plots.

Exercise 43. Make an example of a system of 2 linear equations in 3 unknowns which has infinitely many solutions. This system should NOT be homogeneous.

Then write down the associated homogeneous linear system.

Then make a fancy picture: your picture should contain the hyperplanes for the original system and the hyperplanes for the homogeneous system.

Can you identify the two solutions sets? Can you identify a vector P which helps you see that these two affine subsets are paralle?

Instructions: For each of the following exercises, you are to make an example of the indicated type of linear system of equations and make a row picture for it, or you are to say why it is not possible to make such a system.

Exercise 44. Find a system of 2 linear equations in 2 unknowns which has exactly one solution.

Exercise 45. Find a system of 2 linear equations in 2 unknowns which has no solutions.

Exercise 46. Find a system of 2 linear equations in 2 unknowns which has exactly three solutions.

Exercise 47. Find a system of 2 linear equations in 3 unknowns which has exactly one solution.

Exercise 48. Find a system of 2 linear equations in 3 unknowns which has no solutions.

Exercise 49. Find a system of 3 linear equations in 3 unknowns which has exactly one solution.

Exercise 50. Find a system of 3 linear equations in 3 unknowns which has a solution set which looks like a plane.

2.3 Row Operations

Our next goal is to learn how to solve systems of linear equations. That is, given a system of m linear equation in n unknowns, we want to find the affine subset of $\mathbb{R}^n$ which is the solution set of the system. We only begin here, as a full understanding of the process and its structure will take us until the end of the present chapter. The first step is to find simple ways to transform a system into a different system which is easier to solve. Of course, it is important to know that the system we start with and the system we end with have the same solution set.

Definition 52 (Equivalent Systems)**.** Two systems of m linear equations in n unknowns are called *equivalent* if they have the same affine subset of $\mathbb{R}^n$ as solution set. That is, two systems are equivalent if any solution of one of the systems is also a solution of the other solution.

The First Two Row Operations

We shall need a small set of useful things one can do to a system of equations that will produce an equivalent system, these changes are called *elementary row operations*. For reference, we shall use this generic system of m linear equations in n unknowns

$$\left\{\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \dots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \dots & + & a_{2n}x_n & = & b_2 \\ \vdots & & \vdots & & \ddots & & \vdots & & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \dots & + & a_{mn}x_n & = & b_m \end{array}\right. , \tag{2.15}$$

Definition 53. An *elementary row operation of type 1* consists of replacing a single equation, say the ith row,

$$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n = b_i$$

by a scalar multiple of that row, say

$$\lambda a_{i1}x_1 + \lambda a_{i2}x_2 + \cdots + \lambda a_{in}x_n = \lambda b_i$$

where λ is a non-zero scalar.

Remark 54. An elementary row operation of type 1 is reversible. To undo the operation "multiply row i by λ" use the operation "multiply row i by λ^{-1}".

Theorem 55. Making an elementary row operation of type 1 produces an equivalent system.

Proof. For a single equation, by Theorem 37, rescaling that equation does not change the hyperplane representing that equation. Since this one hyperplane is unchanged, the affine subset of solutions to the system is also unchanged, as the solution set is just where all of the hyperplanes for a system meet.

In essence, though an equation has changed, none of the hyperplanes have moved at all. □

Definition 56. An *elementary row operation of type 2* consists of swapping the positions of two rows. In particular, replace row i by row j, and at the same time, replace row j by row i. This is sometimes called a *row swap*, and one says "swap rows i and j."

Remark 57. An elementary row operation of type is reversible. To undo the operation "swap rows i and j" just repeat the operation.

Theorem 58. Making an elementary row operation of type 2 produces an equivalent system.

Proof. Again, though the ordering of things has changed, none of the actual hyperplanes involved has changed. Therefore, the affine subset which is the solution of the system is exactly the same. □

The Third Row Operation

The third operation is the most interesting. There are two ways to think about it, but the end effect is the same. As with operation 2, we will only involve two rows, so let's pick out two rows from our system, say row i and row j, and isolate them.

$$\left\{ \begin{array}{ccccccccc} a_{i1}x_1 & + & a_{i2}x_2 & + & \dots & + & a_{in}x_n & = & b_i \\ a_{j1}x_1 & + & a_{j2}x_2 & + & \dots & + & a_{jn}x_n & = & b_j \end{array} \right. , \tag{2.16}$$

Our goal is to eliminate the appearance of the variable x_1 in the second equation. To do so, we rearrange the first equation to isolate x_1

$$x_1 = \frac{b_i}{a_{i1}} - \left(\frac{a_{i2}}{a_{i1}} x_2 + \dots + \frac{a_{in}}{a_{i1}} x_n \right).$$

Now, substitute this in for x_1 in the second equation. Then the second equation becomes

$$a_{j1}\left(\frac{b_i}{a_{i1}} - \left(\frac{a_{i2}}{a_{i1}}x_2 + \cdots + \frac{a_{in}}{a_{i1}}x_n\right)\right) + a_{j2}x_2 + \cdots + a_{jn}x_n = b_j$$

which we can reorganize to keep like terms together so that it reads

$$\left(a_{j2} - \frac{a_{j1}}{a_{i1}}a_{i2}\right)x_2 + \cdots + \left(a_{jn} - \frac{a_{j1}}{a_{i1}}a_{in}\right)x_n = b_j - \frac{a_{j1}}{a_{i1}}b_i.$$

That is the end result. The substitution has eliminated the occurrence of the variable x_1 from the second equation. But look closely. An easier way to do the work is to notice that we have made a linear combination of the two equations: we have multiplied the ith equation by $\lambda = -a_{j1}/a_{i1}$ and then added it to the jth equation.

Definition 59. An *elementary row operation of type 3* consists adding a scalar multiple of one row to another. For example, multiply the ith row by a non-zero scalar multiple λ, and add the result to the jth equation. Every other row (not the jth) is left alone.

If we choose a variable x_k, and the scalar multiple λ is chosen to be $\lambda = -a_{jk}/a_{ik}$, this will have the effect of eliminating the occurrence of the variable x_k from row j.

Remark 60. An elementary row operation of type 3 is reversible. To undo the operation "add λ times row i to row j" use the operation "add $-\lambda$ times row i to row j".

Theorem 61. Making an elementary row operation of type 3 produces an equivalent system.

Proof. The only hyperplane which changes is the jth. But any vector which satisfies the original equations i and j, will still satisfy the new equations too, since we are just doing a substitution. □

If we put together the three theorems above, we get the following.

Theorem 62. If two systems of m linear equations in n unknowns are related by a sequence of elementary row operations, then those systems are equivalent.

Exercises

Exercise 51. Consider the following system of 2 linear equations in 2 unknowns. Make an example of an equivalent system which you think is easier to find a solution. Don't worry about finding the solution, yet. Just try to make it easier to solve.

$$\begin{cases} 2x & + & y & = & 7 \\ x & + & y & = & 5 \end{cases}$$

Exercise 52. Consider the following system of 2 linear equations in 2 unknowns. Make an example of an equivalent system which you think is easier to find a solution. Don't worry about finding the solution, yet. Just try to make it easier to solve.

$$\begin{cases} 9x & & & = & 7 \\ 3x & - & y & = & 5 \end{cases}$$

Exercise 53. Consider the following system of 3 linear equations in 3 unknowns. Make an example of an equivalent system which you think is easier to find a solution. Don't worry about finding the solution, yet. Just try to make it easier to solve.

$$\begin{cases} 2x & & & + & z & = & 7 \\ 3x & - & y & + & 2z & = & 5 \\ x & + & y & + & z & = & 0 \end{cases}$$

Exercise 54. Consider the following system of 3 linear equations in 3 unknowns. Use a sequence of row operations to make an equivalent system, and then find the solution set.

$$\begin{cases} 2x & & & + & z & = & 7 \\ 3x & - & y & + & 2z & = & 5 \\ x & + & y & + & z & = & 0 \end{cases}$$

Exercise 55. Consider the following system of 3 linear equations in 3 unknowns. Use a sequence of row operations to make an equivalent system, and then find the solution set.

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

Exercise 56. Consider the following system of 3 linear equations in 3 unknowns. Use a sequence of row operations to make an equivalent system, and then find the solution set.

$$\begin{cases} -x & + & 4y & + & z & = & 0 \\ 3x & + & y & + & 2z & = & 1 \\ x & + & 8y & + & 3z & = & 0 \end{cases}$$

Exercise 57. Think hard about the work you have done, and then write down two or three sentences answering the following question: What makes a system m linear equations in n unknowns "simple" from the point of view of finding a solution?

2.4 Upper Triangular Form and Backsolving

With row operations at our disposal, we have the tools necessary to solve systems of linear equations. The idea is to use the row operations to steadily transform a given system into one that has a form that is "easy" to solve. To make sense of this, we need a firm idea of what makes a system count as "easy to solve."

The Simplest Systems to Solve

As usual a system of m linear equations in n unknowns will be given in the generic notation below.

$$\left\{\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \dots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \dots & + & a_{2n}x_n & = & b_2 \\ \vdots & & \vdots & & \ddots & & \vdots & & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \dots & + & a_{mn}x_n & = & b_m \end{array}\right. . \tag{2.17}$$

We will restrict our attention to special systems in two ways, first, to square systems, and then to square systems with lots of zeros in them.

Definition 63 (Square Systems, Diagonal)**.** A system of linear equations is called a *square system* if $m = n$, that is, if the number of equations is equal to the number of unknowns. In a square system, the *diagonal* is the set of coefficients a_{ij} where $i = j$.

Definition 64 (Upper Triangular Systems)**.** A square system of n linear equations in n unknowns is said to be *upper triangular* if $a_{ij} = 0$ for each position with $i > j$. That is, a system is square if all of the coefficients below the diagonal are zero.

Note that an upper triangular $n \times n$ square system will have the form

$$\left\{\begin{array}{ccccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & a_{13}x_3 & + & \dots & + & a_{1n}x_n & = & b_1 \\ & & a_{22}x_2 & + & a_{23}x_3 & + & \dots & + & a_{2n}x_n & = & b_2 \\ & & & & a_{33}x_3 & + & \dots & + & a_{3n}x_n & = & b_3 \\ & & & & & & \ddots & & \vdots & & \vdots \\ & & & & & & & & a_{nn}x_n & = & b_n \end{array}\right. . \tag{2.18}$$

Theorem 65 (Backsolving an Upper Triangular Square System)**.** An upper triangular square system with non-zero coefficients on the diagonal has a unique solution.

Proof. First, note that the last equation has the form $a_{nn}x_n = b_n$. Since the diagonal entry $a_{nn} \neq 0$, this has solution $x_n = b_n/a_{nn}$. This is the only value that x_n can take to satisfy the equation.

Now consider the case where we know all of the x_i's for $i \in \{k+1, \ldots, n\}$. If we look at the kth equation, we see

$$a_{kk}x_k + \text{terms we can compute} = b_k.$$

Those terms to the right of the diagonal are computed because we already know the x_i's for $i > k$ — we can just substitute the values. Again, because $a_{kk} \neq 0$, simple rearrangement will determine x_k, and that is the only value that will work.

So, now we see the process. First, find x_n. Then find x_{n-1} using the computed x_n. Then find x_{n-2} using the computed x_n and x_{n-1}. Continue in this way until at last we can find x_1. Our solution vector is then completely determined. □

Definition 66 (Backsolving)**.** The process in the proof above is called *backsolving.*

Systematic Simplification of Systems

So, how should we find the solution to a generic square system? The basic idea is to use row operations to transform the problem into an equivalent one which is upper triangular with non-zero coefficients on the diagonal. For small systems, that is for $n \times n$ systems with $n < 5$ or so, an experienced human being can usually just figure out a fairly efficient sequence of row operations which puts the system into triangular form. But as the systems get larger, this gets harder. Without some sort of organization and routine, it is possible to perform a row operation which undoes some of the simplification already done.

There is a process which can be used to systematically change a system into one which is in upper triangular form, which is called *row reduction*, or *Gaussian elimination*, or *Gauss-Jordan elimination*, depending on the exact details and the context. Here is the basic idea.

If all goes well, we only need to use type 3 row operations. Do them in the following order. Start with the upper left entry of the system – the 11 position. Use row operations of type 3 to eliminate the variable x_1 from equations 2 through n. Then move over and down to the 22 position. Use row operations of type 3 to eliminate the variable x_2 from equations 3 through n. We repeat this process over and over until we reach the nth equation, which should now have the form $\alpha x_n = \beta$. (Those numbers α and β probably won't be the original a_{nn} and b_n. They will change because you have performed other row operations to eliminate variables from this last equation.)

It is possible that along the way some diagonal coefficient a_{kk} ends up a zero at a stage where we need to use it to eliminate occurrences of x_k in later equations. If this happens, use a type 2 row equation to swap for some lower row with that corresponding coefficient not equal to zero. Then get back to the program of using type 3 operations.

There are more complications that can happen, so we are not quite done solving our problem. These kinds of difficulties occur very often when we start with a system that is not square, but they can even happen for square systems. We'll address these in the next section.

Exercises

Exercise 58. Make up an example of a 2×2 upper triangular square system of linear equations with non-zero coefficients on the diagonal. Find the solution by backsolving.

Exercise 59. Make up an example of a 3×3 upper triangular square system of linear equations with non-zero coefficients on the diagonal. Find the solution by backsolving.

Exercise 60. Make up an example of a 4×4 upper triangular square system of linear equations with non-zero coefficients on the diagonal. Find the solution by backsolving.

Exercise 61. Make up an example of a 2×2 square system of linear equations which is not upper triangular. Try to use row operations to make an equivalent system which is upper triangular. Keep track of the row operations you use.

Exercise 62. Make up an example of a 3×3 square system of linear equations which is not upper triangular. Try to use row operations to make an equivalent system which is upper triangular. Keep track of the row operations you use.

Exercise 63. Make up an example of a 4×4 square system of linear equations which is not upper triangular. Try to use row operations to make an equivalent system which is upper triangular. Keep track of the row operations you use.

Exercise 64. Consider the examples you made in the last three exercises. Did any of the upper triangular forms have a zero on the diagonal? Note that this means we can't use Theorem 65. How could you design a square system so that its associated upper triangular form has a zero on the diagonal, but this fact isn't obvious from the beginning?

Exercise 65. Can you see any difficulties with our row reduction process? What kinds of things might go wrong? Make a list.

2.5 Gauss-Jordan Elimination & RREF

Our goal now is to learn the complete process for solving any *homogeneous* system of m linear equations in n unknowns. We are opening up the problem a bit by allowing the system to be something other than a *square* system, but we will still keep some of the difficulties down by restricting our attention to the homogeneous case.

Matrix Form

Recall that a homogeneous system of m linear equations in n unknowns takes the following form:

$$\begin{cases} a_{11}x_1 & + & a_{12}x_2 & + & \dots & + & a_{1n}x_n & = & 0 \\ a_{21}x_1 & + & a_{22}x_2 & + & \dots & + & a_{2n}x_n & = & 0 \\ \vdots & & \vdots & & \ddots & & \vdots & & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \dots & + & a_{mn}x_n & = & 0 \end{cases}.$$

At this point, you have likely noticed that writing out a system is messy and has lots of repetition. The process we use to solve systems only depends on the coefficients, and it would be nicer to just track those, and not write out the x_i's and $+$'s and $=$'s all the time. If we keep everything lined up in rows and columns, then the structure is still there implicitly. Also, for a homogeneous system, the whole right hand side consists of zeros, so nothing interesting will happen there as we perform row operations. So, for now, as a notational device, we will write the system in *matrix-vector equation form*

$$Ax = 0,$$

where the symbol A denotes the $m \times n$ matrix

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

Now every elementary row operation becomes an operation done to the rows of the matrix A. This is the reason for the choice of terminology in "row operation."

The Algorithm: Version One

Let's remind ourselves of the process for finding a solution, but rewrite it in terms which make sense for the matrix form.

Remark 67 (Algorithms)**.** This is an example of an *algorithm*. Think of an algorithm is a computational process that is spelled out in enough detail that a computer can do it without human intervention. For example, long division is like this: you can master the routine of a cycle of different steps, and even if you don't know what those steps mean, you can compute a quotient of two numbers. We are now describing (an attempt) at an algorithm for finding the solution to a homogeneous system of linear equations.

Remark 68 (Recursion)**.** The cleanest way to describe the algorithm we want is to use *recursion*. This means that we give a description of how to solve the problem for an $m \times n$ matrix/system by reducing it to a smaller matrix/system (typically, at least one of m or n gets smaller, usually both.)

Our Solution Algorithm

Begin with an $m \times n$ homogeneous system, written in matrix form as $Ax = 0$.

Step One: If a_{11} is not zero, do nothing. If $a_{11} = 0$, then use a type 2 row operation to swap row 1 with row k, where k is the smallest index so that a_{k1} is not zero. This changes the matrix to a new one where a_{11} is not zero.

Step Two: Now we know that a_{11} is not zero, so we may freely divide by a_{11}. In order, use type 3 elementary row operations of the type "add $-a_{k1}/a_{11}$ times row 1 to row k", for $k = 2, 3, \dots m$. We have now changed the matrix so that it should have the form

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ 0 & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & a_{m2} & \dots & a_{mn} \end{pmatrix}.$$

Step Three: Go back to Step One but do the process with the smaller matrix

$$B = \begin{pmatrix} a_{22} & a_{23} & \dots & a_{2n} \\ a_{32} & a_{33} & \dots & a_{3n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m2} & a_{m3} & \dots & a_{mn} \end{pmatrix}.$$

Step Four: When the matrix under consideration is reduced something with only one row or one column, put everything back together. Your system should now look roughly upper triangular. Use backsolving to find the solution.

What Can Go Wrong?

There are two things that can go wrong with the above "algorithm" for homogeneous systems:

1. It might be possible that at some step of the process, our first column consists entirely of zeros! In this case, we cannot arrange that the top left entry is non-zero by making a row swap. And then we are stuck and cannot do the elimination operations in Step Two. Essentially, we cannot perform Step One, and so Step Two becomes impossible.

2. We have not described what happens "at the bottom" of the process. What do we do when the matrices have gotten as small as we can make them?

Let's take these up in reverse order. What kinds of shapes are possible "at the bottom" of the process? each pass through the three steps both reduces the number of columns and the number of rows. So the smallest possible cases look like these three: A is $k \times 1$, so it is a column vector; A is a 1×1 matrix; or, A is $1 \times k$, so it is a "row vector."

$$A = \begin{pmatrix} a \\ b \\ \vdots \\ z \end{pmatrix}, \quad A = (a), \quad A = \begin{pmatrix} a & b & \dots & z \end{pmatrix}$$

In the first case, where A is a single column, we simply perform the process one more time. It will lead to a new form which has a non-zero number as the top entry, but all of the entries below that will be zeros. This is the final form of things. Note that if we translate things back to a system of equations, this means we will have several equations at the bottom of our system that look like this:

$$0 * x_1 + 0 * x_2 + \cdots + 0 * x_n = 0.$$

Since that equation is always true, if a bit boring, we aren't bothered by it.

In the second case, we are happy. We just move on to Step Four.

In the third case we are still just done, but now that single row corresponds to an equation of the form

$$\alpha_k x_k + \alpha_{k+1} x_{k+1} + \cdots + \alpha_n x_n = \beta.$$

And this is fine. The trouble comes at Step Four, as it is not clear what to do with those extra variables. The key is that they are *free*! They can take any value whatsoever. The system of equations doesn't constrain them at all. We will fix this by introducing new parameters for them.

This concept of a free variable will help us solve the problem of being unable to complete Step One because of a column of all zeros, too. Essentially, where we are expecting to get an equation that tells us how to handle that variable, no equation comes to the rescue. This means that the variable is free to take any value at all.

What happens next is much like the case of turning an implicit description of a line into a parametric description of a line. Remember that this is the case where we start with

$$ax + by = c,$$

but then introduce a new parameter with a new equation, $y = t$, and rearrange like so:

$$\begin{cases} ax &= c - bt \\ y &= t \end{cases}$$

We will want to do this kind of thing, but with one parameter for each of the free variables.

A final comment is in order. Since we will want to do the backsolving portion of our process but not all of the variables will take on definite values (some will be free, and hence described by parameters), we will end up doing lots of more complicated substitutions and rearrangements than just "plug in a number here." We will clean this up by doing the work in advance: a substitution move is just a type 3 row operation, after all.

With all of this in our head, we can try to make a better version of the process for solving a homogeneous $m \times n$ system.

Gauss-Jordan Elimination

Let us describe the official algorithm for finding a solution the efficient way, just as we would ask a computer to do it. First, we want a definition.

Definition 69 (RREF, pivot column, pivot variable, free column, free variable)**.** A matrix is said to be in *reduced row echelon form* when the following hold.

- The first non-zero entry in each row is a 1. Such an entry is called a *pivot*. A column which contains a pivot is called a *pivot column*, and the corresponding variable is called a *pivot variable* for the associated system.
- All of the other entries in a pivot column beside the pivot are equal to zero.
- Any row of all zeros occurs below any row which has a non-zero entry.

The phrase *reduced row echelon form* is often abbreviated *RREF*. Any column in a matrix in RREF which is not a pivot column is called a *free column*, and the associated variable is called a *free variable* of the associated system.

There is an efficient algorithm for putting a matrix into RREF using elementary row operations. Using it well is the heart of computational technique in linear algebra. You want to know how to do this quickly for smallish matrices, and you want to know how to make a computer do this for any matrix.

Gauss-Jordan Elimination Algorithm

Given an $m \times n$ matrix A, one can find associated matrix R which is in reduced row echelon form by the following algorithm.

Step One (Forward Pass): We will produce a new matrix B which has many zeros in it by finding pivot positions and eliminating entries below those pivots.

Check for pivot or free column: Consider the top left entry a_{11}. If a_{11} is not zero, do nothing and move on to the next step.

If a_{11} is zero, but some entry below it in the same column is not zero, do a type 2 row operation – a row swap – to switch the places of the two rows. It is generally preferred to swap with the first non-zero row. If all of the entries in the first column are zero, declare the first column to be a free column, and start over but by considering the smaller matrix constructed by ignoring the first column.

We now may assume that a_{11} is not zero.

Eliminate below a pivot: Now we know that a_{11} is not zero, so we may freely divide by a_{11}. In order, use type 3 elementary row operations of the type "add $-a_{k1}/a_{11}$ times row 1 to row k", for

$k = 2, 3, \ldots m$. We have now changed the matrix so that it should have the form

$$A = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ 0 & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & a_{m2} & \ldots & a_{mn} \end{pmatrix}.$$

Move on: Go back to the first instruction of this step, but consider the smaller submatrix made by ignoring the first row and first column.

- The forward pass is done when you have reached the bottom row, or the final column. The matrix should now be in a form where each row has a first non-zero entry at a pivot position, and all entries below a pivot are zero.

Step Two (Backward Pass): We will now produce a new matrix C which has well-identified pivot positions, and has zeros in all other positions in those pivot columns.

- Start with the pivot position in the bottom-most row. Suppose that the l is the largest index of a row with a pivot position. For each k smaller than l, use a type 3 row operation of the form "add $-a_{kj}/a_{lj}$ times row l to row k" to eliminate the kj-position of the matrix. (Generally, one works from the bottom up.)
- Now that the pivot position under consideration has only zeros above and below it, move attention up and left to the next pivot position from the bottom. Repeat the last step.
- Stop after eliminating entries above every pivot position.

Step Three (Rescaling): We now produce the final RREF by making the pivots equal to 1.

- For each $j \in \{1, \ldots, m\}$, if row j has a non-zero entry, use a type 1 elementary row operation of "multiply row j by λ^{-1}," where λ is the value of the entry in the pivot position.

Theorem 70 (RREF)**.** Any matrix A can be transformed into one which is in reduced row echelon form by the Gauss-Jordan Elimination algorithm.

The Whole Process: Solving A Homogeneous System

With the RREF in our hands, we can now write write out the process for solving a homogeneous system of linear equations. It goes like this:

Given a homogeneous system of m linear equations in n unknowns, one finds the set of solutions like so:

1. Translate the system into matrix form $Ax = 0$. (The system is homogenous!)
2. Apply the Gauss-Jordan Elimination algorithm to find $R = \mathrm{rref}(A)$, the reduced row echelon form of A.
3. Make note of which columns and variables are pivot columns and which are free columns.
4. Translate the matrix-vector equation $Rx = 0$ back into a system of linear equations. This new system should have many zero coefficients.
5. Introduce a new parameter t_i and new equation $x_i = t_i$ for each of the free variables.
6. Move all of the terms involving parameters to the right hand side of the equals signs, essentially solving for x_i in equation i.
7. Line up the right-hand sides of these equations by the parameters, and write out a parametric description of the affine subspace which is a solution to the system.

Exercises

For each of the first four exercises, find the reduced row echelon form of the indicated matrix.

Exercise 66. The matrix A.

$$A = \begin{pmatrix} 2 & 1 & 0 \end{pmatrix}$$

Exercise 67. The matrix B.

$$B = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$$

Exercise 68. The matrix C.

$$C = \begin{pmatrix} 2 & 1 & 0 & 4 \\ 1 & 1 & 1 & 0 \\ 3 & 2 & 1 & 4 \end{pmatrix}$$

Exercise 69. The matrix D.

$$D = \begin{pmatrix} 0 & 1 & 2 & 1 & 3 \\ 0 & -1 & 2 & 3 & 1 \\ 1 & 1 & 1 & -1 & 1 \\ 4 & 0 & 0 & 1 & 1 \end{pmatrix}$$

For each of the next three exercises, find the affine subset of $\mathbb{R}^n$ which is the set of solutions of the system of m linear equations in n unknowns.

Exercise 70.

$$\left\{ \begin{array}{ccccccc} x & + & y & + & z & = & 0 \end{array} \right.$$

Exercise 71.

$$\left\{ \begin{array}{ccccccc} 2x & + & y & + & z & = & 0 \\ x & + & y & + & z & = & 0 \end{array} \right.$$

Exercise 72.

$$\left\{ \begin{array}{ccccccccccc} 3x_1 & + & x_2 & + & 4x_3 & + & x_4 & + & x_5 & = & 0 \\ x_1 & + & x_2 & + & 2x_3 & & & + & x_5 & = & 0 \\ -x_1 & - & x_2 & - & 2x_3 & - & 9x_4 & + & x_5 & = & 0 \end{array} \right.$$

Exercise 73. Design a homogeneous system of 3 linear equations in 3 unknowns that has 3 pivots and 1 free variable, or explain why this is not possible.

2.6 Solving a General System

We are now prepared to complete our solution to the main problem in linear algebra. This is a reasonable time to summarize our work, so let's recall what we know.

Recall that a system of m linear equations in n unknowns takes the following form:

$$\left\{\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \dots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \dots & + & a_{2n}x_n & = & b_2 \\ \vdots & & \vdots & & \ddots & & \vdots & & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \dots & + & a_{mn}x_n & = & b_2 \end{array}\right. . \tag{2.19}$$

The General Problem

Find the set of all possible solutions to the system (2.19).

Our Progress So Far

We first found that a good way to understand the solution set of an individual linear equation is as a hyperplane in $\mathbb{R}^n$. Then we saw that the solution set of a system of equations is naturally understood as an intersection of m hyperplanes, which makes it an affine subset of $\mathbb{R}^n$. This allows us to reframe our general problem geometrically as "Find a good description for an affine subset of $\mathbb{R}^n$ which is constructed as the intersection of m hyperplanes."

Let's denote the affine subset which is the solution set to system (2.19) by $\mathcal{S}$, and affine subset which is the solution set to the associated homogeneous system

$$\left\{\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \dots & + & a_{1n}x_n & = & 0 \\ a_{21}x_1 & + & a_{22}x_2 & + & \dots & + & a_{2n}x_n & = & 0 \\ \vdots & & \vdots & & \ddots & & \vdots & & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \dots & + & a_{mn}x_n & = & 0 \end{array}\right. \tag{2.20}$$

by $\mathcal{S}_0$. We saw that $\mathcal{S}$ and $\mathcal{S}_0$ are parallel as affine subsets of $\mathbb{R}^n$. Looking at the details of Theorem 51, we know that these two affine subsets of $\mathbb{R}^n$ are in fact related like this:

$$\mathcal{S} = \{X + P \mid X \in \mathcal{S}_0\},$$

where P is any single vector which is a solution to system (2.19). In essence, we have split our problem into two pieces:

1. Solve the associated homogeneous system (2.20);

2. Find any single solution to the original system (2.19).

We have already addressed the first of these. Using row operations we can take any homogeneous system and produce an equivalent, but much simpler, homogeneous system and then find the affine subset $\mathcal{S}_0$. So what remains is the second part. We must learn to find any single particular solution to a general system.

In practice, one does not solve the problem by considering these two parts completely independently. Rather, one considers a slightly modified version of the process we used to solve the homogeneous system which allows us to do both at once. Also, there is one more complication that can happen, so it pays to keep a sharp eye on the details.

The Method

Given the system of m linear equations in n unknowns, we first form the matrix-vector equation equivalent $Ax = b$, where

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}.$$

Then, recalling that only the raw numbers will matter and the structure sorts out everything else, we form an *augmented matrix*

$$(A \mid b) = \left(\begin{array}{cccc|c} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_m \end{array}\right).$$

To do this, simply tack on the vector b as an extra column to the matrix of coefficients A. It is helpful to keep this extra column separate, so we shall draw a vertical line between the original set of coefficients and this new column.

Next, perform row operations in the style of Gauss-Jordan Elimination. When doing so, only use the original matrix A when planning operations, but do the operations to the whole augmented matrix. The coefficients in the augmented column will change, but it is not our goal to find a pivot in that column. (In fact, a pivot in that column is a warning. See below!) This will eventually produce an augmented matrix of the form $(R \mid d)$, where the matrix R is in reduced row echelon form. Be sure to note which columns are pivot columns and which are free columns.

The matrix $(R \mid d)$ is the data we need to write down our solution. This comes in two steps, related to the two parts from our summary.

First, pretend the vector d is zero, and write out the solution to the homogeneous system $Rx = 0$ as we did in the last section. This will find the affine subset $\mathcal{S}_0$. Recall that this process makes a set of special vectors so that $\mathcal{S}_0$ is described as a set of linear combinations of those vectors.

Definition 71 (Primary Null Vectors)**.** The vectors produced by the Gauss-Jordan Elimination process applied to the system $Ax = 0$ are called *primary null vectors*.

Then, to find the solution P which should lie in $\mathcal{S}$ and realize the parallelism, set all of the free variables equal to zero, and solve the resulting equations from the whole system $(R \mid d)$. This will have the effect of putting the entries from d into the correct places in P, and filling the rest of P with zeros.

Definition 72 (Fundamental Solution)**.** We shall call the vector P obtained by the above process the *fundamental solution* of the system (2.19).

After we have collected all of the data, we write out the solution set in this form:

$$\mathcal{S} = \{P + t_1 v_1 + \dots t_k v_k \mid t_i \in \mathbb{R}\}$$

where P is the fundamental solution and the v_i's are the primary null vectors.

The Complication

Definition 73 (Consistent Systems, Inconsistent Systems)**.** A system of linear equations of the form (2.19) is called *consistent* if it has at least one solution. It is called *inconsistent* if it has no solutions at all, and hence the set $\mathcal{S}$ is empty.

How can we recognize when a system is inconsistent? How often does that happen? First, there are some cases where this will never be a problem.

Theorem 74 (Homogeneous Systems are Consistent)**.** A homogeneous system of m linear equations in n unknowns is consistent.

Proof. The zero vector in $\mathbb{R}^n$ is always a solution to a homogeneous system in n variables. $\square$

The key to recognizing an inconsistent system is to think carefully about our Gauss-Jordan Elimination process. The algorithm tells us how to find a solution, so if there is no solution, the algorithm must show us if it fails. And it will!

Theorem 75 (Recognizing Inconsistent Systems)**.** A system of m linear equations in n unknowns will be inconsistent when the corresponding augmented matrix has a reduced row echelon form with a pivot in the augmented column.

Proof. If there is a pivot in the augmented column of $(R \mid d)$, then the corresponding system of equations contains an equation of this form

$$0x_1 + 0x_2 + \dots 0x_n = d_k,$$

where d_k is not zero. But this equation clearly has no solutions. This means the whole system can have no solutions. □

So, the trick is as follows: do the method as outlined above. If at any point you see a whole row of zeros in the left hand portion of the augmented matrix, but a non-zero number in the augmented column, just stop. The system is inconsistent and has no solutions.

Exercises

In each of the following tasks, design an example of a system of m linear equations in n unknowns and then find the complete solution set, where m and n are as given. Your system should not be homogeneous. In each case, be sure to clearly identify the pivots, the free columns, the fundamental solution, and the primary null vectors.

Exercise 74. $m = 2$ and $n = 2$

Exercise 75. $m = 2$ and $n = 3$

Exercise 76. $m = 3$ and $n = 2$

Exercise 77. $m = 3$ and $n = 3$

Exercise 78. $m = 5$ and $n = 2$

Exercise 79. $m = 2$ and $n = 5$

Did you create any inconsistent systems?

Exercise 80. Find the complete solution set to this system of equations.

$$\left\{\begin{array}{rcrcrcr} 2x & + & 3y & + & z & = & 8 \\ 4x & + & 7y & + & 5z & = & 20 \\ & - & 2y & + & 2z & = & 0 \end{array}\right.$$

Chapter 3

Columns and Subspaces

Introduce the major goals of this chapter: understanding the structure of linear combinations and subspaces, and studying things from the point of view of the columns

sample task: points not fit by some conics? else?

3.1 Subspaces and the Column Picture

The Column Picture, Naïvely

Recall that we have recast the generic system of linear equations

$$\left\{\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \dots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \dots & + & a_{2n}x_n & = & b_2 \\ \vdots & & \vdots & & \ddots & & \vdots & & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \dots & + & a_{mn}x_n & = & b_2 \end{array}\right. \tag{3.1}$$

as an equation involving the linear combination of n vectors in $\mathbb{R}^m$, like so:

$$x_1 \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix} + x_2 \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix} + \cdots + x_n \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} \tag{3.2}$$

The basic questions about how to solve the system (3.1) are readily translated into similar questions about solving the equation (3.2).

The linear combination equation (3.2) is only one single equation, relating vectors in $\mathbb{R}^m$. So, one way to picture it is as follows: Imagine the n different vectors

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

all emanating from the origin. Let's make them all one color, say, red. Now make the subset of all of the vectors in $\mathbb{R}^m$ which can be expressed as a linear combination of those red vectors. That will make up some portion of the space $\mathbb{R}^m$, which we will also color red.

Now, in the same picture, place the vector b

$$\begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

and make it blue. Our basic questions about solvability are these: Is there a way to realize the blue vector as an element of the red subset we made before? If so, what are the coefficients that make the linear combination hit the blue vector b? Is there only one way to find a set of coefficients, or are there several ways?

In order to understand this interpretation based on columns, we need find ways to talk about the details of linear combinations with more clarity. This motivates the notion of a *subspace*.

The Idea of a Subspace

The fundamental construction in linear algebra is that of a linear combination. A subspace of $\mathbb{R}^n$ is a subset which interacts well with the operation of taking linear combinations.

Definition 76 (Subspace)**.** A subset $\mathcal{S}$ in $\mathbb{R}^n$ is called a *subspace* of $\mathbb{R}^n$ when the following two conditions hold:

1. If u and v are elements of $\mathcal{S}$ and α and β are scalars, then the linear combination $\alpha u + \beta v$ is also an element of $\mathcal{S}$.
2. The zero vector is an element of $\mathcal{S}$.

The first condition is the crucial one we want. The second condition is there for a technical reason, it disallows the set with no elements (called the *empty set*) from being a subspace. Each $\mathbb{R}^n$ has a pair of subspaces which are at the extremes for "size." These are easy to forget about!

Theorem 77 (Big and Small Subspaces)**.** Let n be a counting number.

- The set $\{0\}$ consisting of only the zero vector is a subspace of $\mathbb{R}^n$.
- The set $\mathbb{R}^n$ is a subspace of itself.

Proof. First, consider $\{0\}$. There is only one element of this set, namely the zero vector. Clearly the second condition is satisfied. And the first is satisfied because any linear combination of 0 with itself is still just 0:

$$\alpha 0 + \beta 0 = 0.$$

Hence $\{0\}$ is a subspace of $\mathbb{R}^n$.

Now, consider $\mathbb{R}^n$ as a subset of itself. Clearly 0 is an element of $\mathbb{R}^n$, so the second condition is satisfied. And for any pair of vectors in $\mathbb{R}^n$, any linear combination of those vectors is still in $\mathbb{R}^n$, so the first condition is satisfied, too. Hence $\mathbb{R}^n$ is a subspace of $\mathbb{R}^n$. □

Remark 78. The subspace $\{0\}$ of $\mathbb{R}^n$ is often called the *trivial subspace*, because it is the smallest one, and a bit uninteresting. At the other end, the whole space $\mathbb{R}^n$ is a subspace of itself, which is a quirk of the terminology we just have to tolerate.

Exercises

In each of the following, you will try to decide if a certain set is a subspace of the appropriate $\mathbb{R}^n$ or not. It might help you if you can find a way to draw the set.

Exercise 81. Let $\mathcal{C}$ be the set of all unit vectors in $\mathbb{R}^2$. Decide if $\mathcal{C}$ is a subspace or not, and explain your answer using the definition.

Exercise 82. Let $\mathcal{U}$ be the set of all vectors in $\mathbb{R}^2$ which either lie on the line through the origin and u_1 or on the line through the origin and u_2, where

$$u_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \qquad u_2 = \begin{pmatrix} -2 \\ 5 \end{pmatrix}.$$

Decide if $\mathcal{U}$ is a subspace or not, and explain your answer using the definition.

Exercise 83. Let $\mathcal{V}$ be the set of all vectors in $\mathbb{R}^2$ described below,

$$\mathcal{V} = \{X = au_1 + bu_2 \,|\, a, b \in \mathbb{R}\}$$

where the vectors u_1 and u_2 are as in the last exercise. Decide if $\mathcal{V}$ is a subspace or not, and explain your answer using the definition.

Exercise 84. Let $\mathcal{L}$ be the set described below.

$$\mathcal{L} = \left\{ \begin{pmatrix} a \\ b \end{pmatrix} \in \mathbb{R}^2 \,\middle|\, a \text{ and } b \text{ are integers} \right\}$$

Decide if $\mathcal{L}$ is a subspace or not, and explain your answer using the definition.

Exercise 85. Let $\mathcal{P}$ be the hyperplane in $\mathbb{R}^3$ described below.

$$\mathcal{P} = \left\{ \begin{pmatrix} a \\ b \\ c \end{pmatrix} \,\middle|\, a + b + c = 5 \right\}$$

Decide if $\mathcal{P}$ is a subspace or not, and explain your answer using the definition.

Exercise 86. Let $\mathcal{W}$ be the hyperplane in $\mathbb{R}^3$ described below.

$$\mathcal{W} = \left\{ \begin{pmatrix} a \\ b \\ c \end{pmatrix} \,\middle|\, a + 2b + c = 0 \right\}$$

Decide if $\mathcal{W}$ is a subspace or not, and explain your answer using the definition.

Exercise 87. Let $\mathcal{B}$ be the subset of $\mathbb{R}^2$ described as follows.

$$\mathcal{B} = \left\{ \begin{pmatrix} a \\ b \end{pmatrix} \middle| \, |b| \leq 2 \right\}$$

Decide if $\mathcal{B}$ is a subspace or not, and explain your answer using the definition.

3.2 Example Subspaces: Null Spaces

At this point, we have identified several different kinds of subsets of $\mathbb{R}^n$ which are useful from a linear algebra point of view: affine subsets, hyperplanes, and now subspaces. We now start to sort out the relationships between these three things. Along the way, we will introduce one of the primary ways to generate examples of subspaces: as the null space of a matrix.

Let's begin by looking at the connection between the idea of an affine subset and the idea of a subspace.

Theorem 79. Let $\mathcal{S}$ be a subspace of $\mathbb{R}^n$. Then $\mathcal{S}$ is an affine subset of $\mathbb{R}^n$.

Proof. Note that if $\mathcal{S}$ is the trivial subspace $\mathcal{S} = \{0\}$, then it has only one vector in it, so there is nothing to check in the definition of affine subset. So we may proceed by assuming that $\mathcal{S}$ has at least two vectors in it.

Let u and v be any two elements of the subspace $\mathcal{S}$. To show that $\mathcal{S}$ is an affine subset of $\mathbb{R}^n$, we must show that the whole line through u and v is also in $\mathcal{S}$.

But any point on the line through u and v can be written in the form

$$u + t(v - u) = (1 - t)u + tv,$$

where t is some real number. So we see that such a point is a linear combination of u and v. Since $\mathcal{S}$ is a subspace, this linear combination is also in $\mathcal{S}$. Therefore, $\mathcal{S}$ is also an affine subset of $\mathbb{R}^n$. □

Theorem 80. Let $\mathcal{A}$ be an affine subset of $\mathbb{R}^n$ which contains the origin. Then $\mathcal{A}$ is a subspace of $\mathbb{R}^n$.

Proof. Let u and v be two elements of $\mathcal{A}$. We must show that every linear combination $au + bv$ is also an element of $\mathcal{A}$.

Our approach will be to do this in stages. First we will find a line of points in $\mathcal{A}$, and then we will "sweep out" most of the linear combinations by taking lines through the origin and the points on this line. Finally, we will patch up the rest of the linear combinations by showing that the missing set (which will be a line) can be covered, too.

Step 1: $a + b \neq 0$

First note that the line ℓ' which passes through u and v is part of $\mathcal{A}$, because $\mathcal{A}$ is an affine subset. Therefore, each point of the form $u + t(v - u)$ is in $\mathcal{A}$. In particular, if we choose $t = \dfrac{b}{a+b}$, we see that

$$w = u + t(v - u) = u + \frac{b}{a+b}(v - u) = \frac{a}{a+b}u + \frac{b}{a+b}v$$

is an element of $\mathcal{A}$.

Since 0 is in $\mathcal{A}$, too, the line through 0 and w lies in $\mathcal{A}$. Therefore, each vector of the form λw is in $\mathcal{A}$. If we choose $\lambda = a + b$, we see that

$$(a+b)w = \left(\frac{a}{a+b}u + \frac{b}{a+b}v\right) = au + bv$$

is an element of $\mathcal{A}$. This completes the proof in the case that $a + b \neq 0$.

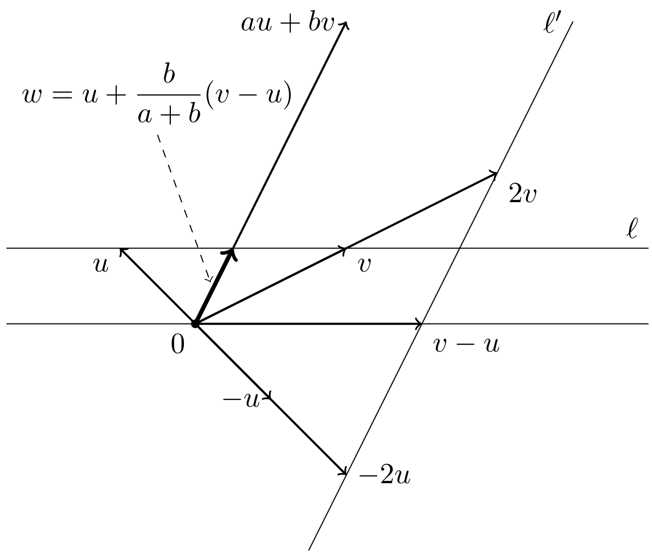


Figure 3.1: The geometry in our argument for Theorem 80.

You should note that we have definitely shown that $-2u = -2u + 0v$ and $2v = 0u + 2v$ are elements of $\mathcal{A}$.

Step 2: $a + b = 0$.

Now suppose $a + b = 0$. Then $au + bv = b(v - u)$. We will want to show that $v - u$ is an element of $\mathcal{A}$, and then the same reasoning as above about the line through this point will show that $au + bv = b(v - u$ lies in $\mathcal{A}$.

Since $-2u$ and $2v$ are elements of $\mathcal{A}$, so is the line which passes through them. So we see that every point of the form

$$-2u + s(2v - 2u) = (-2 - 2s)u + (2s)v$$

lies in $\mathcal{A}$. But if we choose $s = 1/2$, this tells us that $v - u$ is in $\mathcal{A}$.

Putting everything together, we get that all of the linear combinations $au + bv$ are elements of $\mathcal{A}$, so $\mathcal{A}$ is a subspace. $\square$

We have already seen that there is a connection between hyperplanes and affine subspaces. So it is natural to suspect that we can make a connection between hyperplanes and subspaces, too. That is the subject of the next few results. We will give the statements here, and leave it for the exercises to give arguments.

Remark 81 (Notation for Hyperplanes)**.** For a non-zero vector $N \in \mathbb{R}^n$ and a scalar $c \in \mathbb{R}$, we will use the following notation for the hyperplane with normal vector N and level c:

$$\mathcal{H}_{N,c} = \{X \in \mathbb{R}^n \mid N \cdot X = c\}.$$

That is, $\mathcal{H}_{N,c}$ will represent the set of all possible vectors X which are solutions to the linear equation $N \cdot X = c$.

Remember that $\mathcal{H}_{N,c}$ contains the zero vector (the origin) exactly when $c = 0$.

Theorem 82. The hyperplane $\mathcal{H}_{N,c}$ in $\mathbb{R}^n$ is a subspace if and only if $c = 0$. That is, a hyperplane $\mathcal{H}_{N,c}$ is a subspace if and only if it contains the origin.

Proof. You will work out a proof of this in Exercises 93 and 94. □

Theorem 83. A finite intersection of hyperplanes is a subspace if and only if that intersection contains the origin. That is, the intersection of hyperplanes $\mathcal{H}_{N_k,c_k}$ is a subspace if and only if each of the c_k's is zero.

Proof. Apply Theorem 82 several times. □

If we recall that the intersection of a finite number of hyperplanes is an affine subspace, we can restate this last result in a way that makes it a special case of Theorem 80.

Corollary 84. The affine subset of $\mathbb{R}^n$ which is the set of solutions to a homogeneous system of m linear equations in n unknowns is a subspace of $\mathbb{R}^n$.

Of course, we have already seen that a homogeneous system of m linear equations in n unknowns can be rewritten as a matrix-vector equation of the form $Ax = 0$, where A is the $m \times n$ matrix of coefficients. This motivates the following definition.

Definition 85 (Null Space of a Matrix)**.** Let A be an $m \times n$ matrix. The *null space of* A is the subspace of $\mathbb{R}^n$ which consists all solutions to the associated homogeneous equation $Ax = 0$. That is,

$$\text{null}(A) = \{x \in \mathbb{R}^n \mid Ax = 0\}.$$

Definition 86 (Annihilator Matrix)**.** Let S be a subspace of $\mathbb{R}^n$. A matrix A is called *an annihilator matrix for* $\mathcal{S}$ when $S = \text{null}(A)$.

Usually, a subspace has many different annihilator matrices.

Exercises

Exercise 88. Restate the definition of the null space of a matrix in a way that involves a linear-combination equation.

Exercise 89. Write down an example of subspace $\mathcal{T}$ which is just a line in $\mathbb{R}^2$ through the origin. Find two different annihilator matrices for $\mathcal{T}$.

Exercise 90. Find an annihilator matrix for the subspace $\mathcal{S}_1$ of $\mathbb{R}^3$ given below.

$$\mathcal{S}_1 = \{au + bv \mid a, b \in \mathbb{R}^3\},$$

where

$$u = \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix}, \quad v = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}.$$

You can take for granted that $\mathcal{S}_1$ is a subspace.

Exercise 91. Find an annihilator matrix for the subspace $\mathcal{S}_2$ of $\mathbb{R}^3$ given below.

$$\mathcal{S}_2 = \{au + bv \mid a, b \in \mathbb{R}^3\},$$

where

$$u = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \quad v = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}.$$

You can take for granted that $\mathcal{S}_2$ is a subspace.

Exercise 92. Can you describe the null space of this matrix A?

$$A = \begin{pmatrix} 1 & 2 & 0 & 0 & 2 \\ 0 & 0 & 1 & 0 & 9 \\ 0 & 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Exercise 93. Give an argument in support of Theorem 82 in the following manner: Assume that u and v are two elements in the hyperplane $\mathcal{H}_{N,0}$, then give a reason why $au + bv$ must also be an element of $\mathcal{H}_{N,0}$.

Exercise 94. Make an example of a hyperplane $\mathcal{H}_{N,c}$ in $\mathbb{R}^2$ where $c \neq 0$. Use the geometry your example to show why the hypothesis $c = 0$ is necessary by showing why $\mathcal{H}_{N,c}$ fails to be a subspace on both conditions of being a subspace.

Exercise 95. Make a diagram that helps to organize the logical connections between the ideas of affine subset, hyperplane, and subspace, given what we know right now. Is there anything else we should figure out?

3.3 Spans and Column Spaces

There is another way to describe subspaces, which is tied very closely to the idea of respecting linear combination. We'll explore that now, introduce the *column space* of a matrix, and tie this new idea to the basic question of when a system of equations is solvable.

Definition 87 (Spanning Set)**.** Suppose that $\mathcal{S}$ is a subset of $\mathbb{R}^n$. A set of vectors $\{v_1, \dots, v_k\}$, all chosen from $\mathcal{S}$, is called a *spanning set* for $\mathcal{S}$ when every vector in $\mathcal{S}$ can be written as a linear combination of the v_i's. That is, $\{v_1, \dots, v_k\}$ is a spanning set for $\mathcal{S}$ when for each vector b in $\mathcal{S}$, we can find some solution to the equation

$$a_1v_1 + a_2v_2 + \cdots + a_kv_k = b.$$

The big idea is that a collection of vectors is a spanning set for a subspace when those vectors are enough to describe all of the different possible directions one can move inside of the subspace. You should note that subspaces often have **many** different spanning sets.

Definition 88 (Span)**.** Let $\{v_1, \dots, v_k\}$ be a collection of vectors from $\mathbb{R}^n$. The *span* of this set is the collection of all vectors w which can be written as a linear combination of the v_i's. That is,

$$\operatorname{span}\{v_1, \dots, v_k\} = \{w = a_1v_1 + \dots a_kv_k \in \mathbb{R}^n \mid a_1, \dots, a_k \in \mathbb{R}\}.$$

This is read as "the span of v_1 through v_k."

Theorem 89. A span as constructed as in Definition 88 is a subspace of $\mathbb{R}^n$.

Proof. By the definition of a subspace, we must show two things. First, we must show that the zero vector is in our span. Then we must show that for any pair of vectors in the span, any linear combination of those vectors is also in the span.

First, let's show that 0 is an element of $\operatorname{span}\{v_1, \dots, v_k\}$. For this, choose all of the coefficients to be 0. Then we get the element

$$0 = 0v_1 + 0v_2 + \cdots + 0v_k \in \mathcal{S}.$$

Now, suppose that w_1 and w_2 are elements of $\mathcal{S}$, and that α and β are scalars. We must show that $\alpha w_1 + \beta w_2$ is an element of $\mathcal{S}$. Since $w_1 \in \mathcal{S}$, we must have scalars $a_1, \dots, a_k$ realizing this:

$$w_1 = a_1v_1 + \dots a_kv_k$$

Similarly, w_2 is in $\mathcal{S}$, so we must have scalars $b_1, \dots, b_k$ so that

$$w_2 = b_1 v_1 + \dots b_k v_k.$$

If we multiply these equations by α and β, respectively, and add, we see

$$\begin{aligned}\alpha w_1 + \beta w_2 &= \alpha\left(a_1 v_1 + \dots a_k v_k\right) + \beta\left(b_1 v_1 + \dots b_k v_k\right) \\ &= [\alpha a_1 + \beta b_1] v_1 + \dots + [\alpha a_k + \beta b_k] v_k.\end{aligned}$$

This realizes $\alpha w_1 + \beta w_2$ as an element of $\mathcal{S}$, and finishes the proof. $\square$

Definition 90 (Column Space)**.** Let A be an $m \times n$ matrix. The *column space of* A is the subspace of $\mathbb{R}^m$ constructed by taking the span of the columns. That is, if we think of A as a bundle of column vectors

$$A = \begin{pmatrix} | & | & & | \\ v_1 & v_2 & \dots & v_k \\ | & | & & | \end{pmatrix}$$

then

$$\begin{aligned}\operatorname{col}(A) &= \operatorname{span}\{v_1, v_2, \dots, v_k\} \\ &= \{b = x_1 v_1 + \dots + x_n v_n \in \mathbb{R}^m \mid x_1, x_2, \dots, x_n \in \mathbb{R}\}\end{aligned}$$

Theorem 91. The linear combination equation

$$x_1 v_1 + \dots + x_n v_n = b$$

has a solution if and only if b is an element of $\operatorname{span}\{v_1, \dots, v_k\}$.

Proof. This is just a restatement of the definition of the span with our perspective reversed. $\square$

Theorem 92. The column space of an $m \times n$ matrix A is the set of vectors $b \in \mathbb{R}^m$ such that $Ax = b$ has at least one solution.

Proof. This is just a restatement of the definition of the span in terms of the matrix-vector equation. $\square$

Exercise 96. Find a spanning set for the line in $\mathbb{R}^2$ given below:

$$\ell = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \middle| 4x - 5y = 0 \right\}$$

Exercise 97. Find a spanning set for a plane in $\mathbb{R}^3$ given below:

$$\mathcal{P} = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \middle| 4x - 5y + z = 0 \right\}$$

Exercise 98. Is this equation solvable?

$$\left\{ \begin{array}{rcrcrcr} 2x & + & 3y & + & z & = & 2 \\ 5x & - & 2y & + & z & = & 3 \\ x & - & 8y & - & z & = & 5 \end{array} \right.$$

How do you know?

Exercise 99. Consider the following system of linear equations. What would have to be true of the b_i's to make it solvable?

$$\left\{ \begin{array}{rcrcrcl} x & + & y & + & 3z & = & b_1 \\ 3x & - & 2y & - & z & = & b_2 \\ x & - & y & - & z & = & b_3 \end{array} \right.$$

Exercise 100. Consider the following linear combination equation. What would have to be true of the vector b to make this equation solvable?

$$x_1 \begin{pmatrix} 1 \\ -2 \end{pmatrix} + x_2 \begin{pmatrix} -3 \\ 6 \end{pmatrix} = b.$$

Exercise 101. Consider the following matrix-vector equation. What would have to be true of the vector b to make this equation solvable?

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \\ 11 & 13 & 15 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = b$$

Exercise 102. Find a vector which does not lie in subspace $\operatorname{span}\{u_1, u_2\}$, where u_1 and u_2 are the following vectors in $\mathbb{R}^3$.

$$u_1 = \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}, \quad u_2 = \begin{pmatrix} 4 \\ 1 \\ 7 \end{pmatrix}.$$

Exercise 103. Find a vector which does not lie in the column space of A.

$$A = \begin{pmatrix} 403 & -2344 & 78 & 34/7 \\ 76 & 1 & 0 & 23 \\ 45 & 3 & 56 & 12 \\ 31 & -2 & -56 & 11 \end{pmatrix}.$$

3.4 There and Back Again, a Vector's Story

We have introduce the concept of a subspace of $\mathbb{R}^n$, which is a piece of $\mathbb{R}^n$ that contains all of the linear combinations of its elements. Then we found that there are two interesting ways to describe subspaces, and each is associated to a matrix in some way.

Suppose that A is an $m \times n$ matrix. Let's recall our two situations.

The column space of A is the span of the columns of A, and hence a subspace of $\mathbb{R}^m$. It is the set of vectors b for which the equation $Ax = b$ has at least one solution.

The null space of A is the collection of vectors $x \in \mathbb{R}^n$ so that $Ax = 0$. That is, these are the vectors which are solutions to the homogeneous system of linear equations with coefficient matrix A.

If we wish to describe a subspace, we can make it as the column space of some matrix C, or we can make it the null space of some other matrix N, but in general, those two matrices C and N will be different. They will probably even have different shapes!

Here is a cool thing: given a subspace described as $\mathrm{null}(N)$, it is possible to also describe it as $\mathrm{col}(C)$ for some matrix C. That is, something described as a solution set can also be described as a span. Similarly, given a subspace described as $\mathrm{col}(C)$, it is possible to also describe it as $\mathrm{null}(N)$. That is, given something described as a span, we can also describe it as a solution set. Even better, the ways to do these translations are just the kinds of basic tools we have used before of eliminating or introducing new parameters.

From null space to column space: Finding null vectors

Suppose we have a subspace $\mathcal{S}$ which is given as the null space of a matrix N. Then

$$\mathcal{S} = \mathrm{null}(N) = \{x \in \mathbb{R}^n \mid Nx = 0\}$$

is the set of solutions of the homogeneous system of linear equations represented by $Nx = 0$. We can describe this solution set by our Gauss-Jordan elimination process: Put N into reduced row echelon form, identify the pivots and free columns, and then write down the primary null vectors $n_1, n_2, \ldots, n_k$. Then the solution set is

$$\mathcal{S} = \mathrm{span}\{n_1, n_2, \ldots, n_k\}.$$

In order to make this a column space, we just construct a matrix N which has these primary null vectors as columns:

$$N = \begin{pmatrix} | & | & & | \\ n_1 & n_2 & \dots & n_k \\ | & | & & | \end{pmatrix}.$$

Now by the definition of column space, it is the case that $\mathcal{S} = \operatorname{col}(N)$.

From column space to null space: Finding an annihilator

Next, suppose we have a subspace $\mathcal{S}$ which is described as a span, that is as a column space for some $m \times n$ matrix C:

$$\begin{aligned} \mathcal{S} = \operatorname{col}(C) &= \{b \in \mathbb{R}^m \mid b = Cy \text{ for some } y \in \mathbb{R}^n\} \\ &= \{b = x_1 c_1 + \dots x_n c_n \mid x_1, x_2, \dots, x_n \in \mathbb{R}\}, \end{aligned}$$

where the c_i's are the columns of the matrix

$$C = \begin{pmatrix} | & | & & | \\ c_1 & c_2 & \dots & c_n \\ | & | & & | \end{pmatrix}.$$

If b lies in $\mathcal{S}$, then the equation $Cx = b$ has at least one solution. In particular, the system of equations represented by $Cy = b$ is consistent. The thing is, we don't, yet, know what b is. So we will fill b with variables, and then see what must be true about those variables in order that $Cy = b$ is a consistent equation. Write the coefficients of $b \in \mathbb{R}^m$ as $x_1, x_2, \dots, x_m$. Form the augmented matrix

$$(C \,|\, b) = \left(\begin{array}{cccc|c} | & | & & | & | \\ c_1 & c_2 & \dots & c_n & x_i\text{'s} \\ | & | & & | & | \end{array}\right)$$

where the extra column has the variables in it. Then use Gauss-Jordan elimination to put C in reduced row echelon form. You should then have something that looks like

$$\left(\begin{array}{ccc|c} & R & & \text{mess} \\ 0 & \dots & 0 & \text{linear expression in the } x_i\text{'s} \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & 0 & \text{linear expression in the } x_i\text{'s} \end{array}\right).$$

All of those rows at the bottom are the key. Each one of them imposes a constraint on the system. In order that the system is consistent, we must have the linear expression in the extra column equal to zero. Collect those up! They are the equations that the coefficients of b must satisfy. This is how we find a homogeneous system of equations that describes $\mathcal{S}$. If we form the coefficient matrix of this homogeneous system, we have created the annihilator matrix N so that $\mathcal{S} = \text{null}(N)$.

In the first four exercises, you are given a subspace described as the solution set to a homogeneous system of linear equations. (This may be written in one of our other forms!) Find a way to write this subspace as a span. Then write down a matrix C which has the given subspace as its column space.

Exercise 104. The subspace $\mathcal{S}_1$ of $\mathbb{R}^5$ is the set of solutions of the homogeneous system of linear equations below:

$$\left\{\begin{array}{rcrcrcrcrcl} x_1 & - & 3x_2 & + & 2x_3 & - & x_4 & + & 2x_5 & = & 0 \\ 3x_1 & - & 9x_2 & + & 7x_3 & - & x_4 & + & 3x_5 & = & 0 \\ 2x_1 & - & 6x_2 & + & 7x_3 & + & 4x_4 & - & 5x_5 & = & 0 \end{array}\right.$$

Exercise 105. The subspace $\mathcal{S}_2$ of $\mathbb{R}^6$ is the null space of the matrix A below:

$$A = \begin{pmatrix} 1 & 2 & 1 & 2 & 3 & 1 \\ 2 & 4 & 3 & 7 & 7 & 4 \\ 1 & 2 & 2 & 5 & 5 & 6 \\ 3 & 6 & 6 & 15 & 14 & 15 \end{pmatrix}.$$

Exercise 106. The subspace $\mathcal{S}_3$ of $\mathbb{R}^3$ is the set of solutions of the linear combination of vectors equation below:

$$x_1 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + x_2 \begin{pmatrix} 45 \\ 17 \\ -32 \end{pmatrix} + x_3 \begin{pmatrix} 5 \\ -23 \\ -72 \end{pmatrix} = 0$$

Exercise 107. The subspace $\mathcal{S}_4$ of $\mathbb{R}^4$ is the null space of the matrix B below:

$$B = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 2 \\ 2 & 5 & 6 & 4 \\ 2 & 6 & 8 & 5 \end{pmatrix}.$$

In the next four exercises, you are given a subspace described as a span. (This may be written in one of our other forms.) Find a way to write this subspace as a solution set. Then write down the corresponding annihilator matrix.

Exercise 108. The subspace $\mathcal{S}_5$ of $\mathbb{R}^5$ is the span of the vectors u_i listed below:

$$u_1 = \begin{pmatrix} 1 \\ 2 \\ -1 \\ 3 \\ 4 \end{pmatrix}, \quad u_2 = \begin{pmatrix} 2 \\ 4 \\ -2 \\ 6 \\ 8 \end{pmatrix}, \quad u_3 = \begin{pmatrix} 1 \\ 3 \\ 2 \\ 2 \\ 6 \end{pmatrix}, \quad u_4 = \begin{pmatrix} 1 \\ 4 \\ 5 \\ 1 \\ 8 \end{pmatrix}, \quad u_5 = \begin{pmatrix} 2 \\ 7 \\ 3 \\ 3 \\ 9 \end{pmatrix}.$$

Exercise 109. The subspace $\mathcal{S}_6$ of $\mathbb{R}^5$ is the column space of the matrix F listed below:

$$F = \begin{pmatrix} 38 & 11 & 43 \\ -1 & 2 & 3 \\ 56 & 12 & 1 \\ 90 & 23 & 1 \\ -1 & 1 & 1 \end{pmatrix}$$

Exercise 110. The subspace $\mathcal{S}_7$ of $\mathbb{R}^3$ is the column space of the matrix G listed below:

$$G = \begin{pmatrix} 543 & 45 & -19 \\ 34 & 44 & 61 \\ -18 & -90 & 0 \end{pmatrix}$$

Exercise 111. The subspace $\mathcal{S}_8$ of $\mathbb{R}^4$ is the set of all vectors b such that the following homogeneous system of linear equations has at least one solution.

$$\begin{cases} x_1 + 3x_2 + 2x_3 = b_1 \\ -3x_1 - 9x_2 - 6x_3 = b_2 \\ 2x_1 + 7x_2 + 7x_3 = b_3 \\ -x_1 + -x_2 + 4x_3 = b_4 \end{cases}$$

3.5 Linear Independence

We have now described several subspaces of $\mathbb{R}^n$ as spans. Sometimes, the choice of spanning set is not optimal. It may be possible to remove a few vectors from the spanning set, and yet still span the same subspace. How can we deal with such a redundancy? Our next concept captures this phenomenon and gives us language for dealing with it.

Definition 93 (Linear Independence, Linear Dependence)**.** Suppose that $\{v_1, \dots, v_k\}$ is a set of vectors, all chosen from some $\mathbb{R}^n$. This set of vectors is called *linearly independent* when the only solution to the linear combination of vectors equation

$$x_1 v_1 + x_2 v_2 + \dots x_k v_k = 0 \tag{3.3}$$

is the *trivial solution* $x_1 = x_2 = \cdots = x_k = 0$.

The set of vectors is called *linearly dependent* otherwise. That is, we say that the set is linearly dependent when there exists some collection of scalars x_1, x_2, ..., x_k which are not all zero and are a solution to the equation (3.3).

Note that if some collection of vectors is linearly dependent, then at least one of the vectors can be written as a linear combination of the others. This is why that vector would be redundant when making a subspace as a span. For example, the equation

$$2\begin{pmatrix}1\\1\end{pmatrix} + 3\begin{pmatrix}-2\\-1\end{pmatrix} - 2\begin{pmatrix}-2\\-1/2\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix}$$

shows that the three vectors on the left form a linearly dependent set. This equation can be rewritten as

$$\begin{pmatrix}1\\1\end{pmatrix} + \frac{3}{2}\begin{pmatrix}-2\\-1\end{pmatrix} = \begin{pmatrix}-2\\-1/2\end{pmatrix},$$

which shows that the third vector can be written as a linear combination of the first two. So

$$\operatorname{Span}\left\{\begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}-2\\-1\end{pmatrix}, \begin{pmatrix}-2\\-1/2\end{pmatrix}\right\} = \operatorname{Span}\left\{\begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}-2\\-1\end{pmatrix}\right\}$$

This definition can be translated into our matrix language, too. See the exercises!

There are two standard tests to determine if a set of vectors is linearly dependent or linearly independent. Each has advantages depending on the situation.

A Test for Linear Independence: Casting Out

Given a set of vectors $\{v_1, \ldots, v_k\}$, the *Casting Out* algorithm tells us which of the v_i's we should keep to form a "smallest" linearly independent set which spans the same subspace. It can also be used as a simple test for linear independence or linear dependence.

Step One: Form the matrix C which has the v_i's as columns:

$$C = \begin{pmatrix} | & | & & | \\ v_1 & v_2 & \ldots & v_k \\ | & | & & | \end{pmatrix}$$

Step Two: Use Gauss-Jordan Elimination to put the matrix C into reduced row echelon form, R. (Actually, we mostly care about which columns are pivots columns and which are free. So you only need to do the forward pass, if you are pressed for time.)

Step Three: From the original set of vectors v_i, keep any of those that correspond to pivot columns, and "cast out" those that come from free columns.

Theorem 94. If the set $\{v_1, \ldots, v_k\}$ is linearly dependent, then the matrix C will have at least one free column, and the casting out algorithm will indicate at least one vector to remove.

If the set $\{v_1, \ldots, v_k\}$ is linearly independent, then the matrix C will have each of its columns as a pivot column, so the casting out algorithm will tell us to keep all of the v_i's.

A Test for Linear Independence: Row Algorithm

Given a set of vectors $\{v_1, \ldots, v_k\}$, the *Row* algorithm helps us to to form a "smallest" linearly independent set $\{w_1, \ldots, w_j\}$ which spans the same subspace. It can also be used as a simple test for linear dependence or linear independence.

Step One: Form the matrix A which has the v_i's as rows:

$$A = \begin{pmatrix} — & v_1 & — \\ & \vdots & \\ — & v_k & — \end{pmatrix}$$

Step Two: Use Gauss-Jordan Elimination to put the matrix A into reduced row echelon form, R.

Step Three: Take the non-zero rows of the RREF, R, and make vectors $w_1, \dots, w_j$ out of them. Often, the matrix R will have rows without pivots (rows of all zeros). In this case, $j < k$, so there are fewer w_i's than v_i's.

Theorem 95. If the set $\{v_1, \dots, v_k\}$ is linearly dependent, then the RREF of matrix A will have at least one row of all zeros (a row without a pivot).

If the set $\{v_1, \dots, v_k\}$ is linearly independent, then the RREF of matrix A will have a pivot in each of its rows. Note that the algorithm still usually produces a different set of vectors, but this time $j = k$, that is, there are exactly as many w_i's as v_i's.

Which Algorithm?

Take a moment to think over the two algorithms. If all you want is to test for linear independence or linear dependence, it does not matter much which you use.

Can you think of a situation in which you might prefer one algorithm or the other?

Exercises

Exercise 112. Rewrite the definition of a linearly independent set of vectors by translating the linear combination of vectors equation into a matrix-vector equation.

Exercise 113. Rewrite the definition of a linearly dependent set of vectors by translating the linear combination of vectors equation into a matrix-vector equation.

Exercise 114. Make an example of a set of linearly independent vectors in $\mathbb{R}^3$. How many vectors can you have?

Exercise 115. Make a collection of 5 different vectors in $\mathbb{R}^4$. Use the casting out algorithm to figure out if this set is linearly independent. If it is not, can you find a subset of your example which is linearly independent?

Exercise 116. Use the same example you made in the last exercise, but now use the row algorithm to produce a set of vectors which is linearly independent but has the same span.

Exercise 117. Make an example of a set of three vectors in $\mathbb{R}^4$ which is linearly independent, or explain why this is not possible.

Exercise 118. Make an example of a set of 5 linearly dependent vectors in $\mathbb{R}^5$, or explain why this is not possible.

Challenge 119. Why does the Casting Out algorithm work? Can you explain why Theorem 94 makes sense?

Challenge 120. Why does the row algorithm work? Can you explain why Theorem 95 makes sense?

3.6 Bases and Dimension

Here are some important, and related questions:

- How big is a subspace, anyway? In what way can we coherently talk about the size of some subspace of $\mathbb{R}^n$?

- It seems that describing a subspace as a span (like a column space) is a useful thing. But usually there are multiple ways to pick a spanning set for a given subspace. Is there some sense in which one spanning set might be better than another?

We'll now see how to answer these questions, in reverse order. First, to describe a subspace, we will want a set that satisfies a sort of Goldilocks principle. Our set should be big enough to be a spanning set, so that it describes all of the independent directions one can move in the subspace. But also small enough that it is linearly independent, so that it doesn't contain any redundant information.

Definition 96 (Basis)**.** Let $\mathcal{S}$ be a subspace of $\mathbb{R}^n$. A set of vectors $\{v_1, \dots, v_k\}$, all of them taken from $\mathcal{S}$, is called a *basis of* $\mathcal{S}$ if its both
- a linearly independent set, and
- a spanning set for $\mathcal{S}$.

By the way, the plural of "basis" is "bases." This is a constant trouble for new learners and spellers.

If a basis is the right kind of set to describe a subspace, we can use the size of the basis as a proxy for the size of the subspace. This is the linear algebra specific concept of dimension.

Definition 97 (Dimension)**.** Let $\mathcal{S}$ be a subspace of $\mathbb{R}^n$. The *dimension of* $\mathcal{S}$, denoted $\dim \mathcal{S}$, is the number of vectors in a basis of $\mathcal{S}$.

There is an important theorem to know here, but its proof is really tedious and annoying, so we'll skip that.

Theorem 98. Let $\mathcal{S}$ be a subspace of $\mathbb{R}^n$. Then $\mathcal{S}$ has a basis consisting of a finite set of vectors. Moreover, any two bases of $\mathcal{S}$ have the same number of elements. Thus, the dimension of $\mathcal{S}$ makes sense.

Exercises

Exercise 121. This tasks has two parts, each concerning the subspace of $\mathbb{R}^3$:

$$\mathcal{S}_0 = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \middle| \; z = 0 \right\}.$$

a) Is the following set of vectors a basis of the subspace $\mathcal{S}_0$?

$$\left\{ \begin{pmatrix} 3 \\ 4 \\ 0 \end{pmatrix}, \begin{pmatrix} -2 \\ -5 \\ 0 \end{pmatrix}, \begin{pmatrix} 5 \\ 5 \\ 0 \end{pmatrix} \right\}$$

Why or why not?

b) Is the following set of vectors a basis of the subspace $\mathcal{S}_0$?

$$\left\{ \begin{pmatrix} 4 \\ 2 \\ 0 \end{pmatrix} \right\}$$

Why or why not?

Exercise 122. This question has two parts.

a) Give three different examples of bases of $\mathbb{R}^3$. What is the dimension of $\mathbb{R}^3$?

b) What is the maximal dimension of a subspace of $\mathbb{R}^4$? How do you know?

Exercise 123. Consider the subspace of $\mathbb{R}^5$ defined as the solution set of the following homogeneous system of linear equations.

$$\begin{cases} x_1 + 2x_2 - 3x_3 + 2x_4 - 4x_5 = 0 \\ 2x_1 + 4x_2 - 5x_3 + x_4 - 6x_5 = 0 \\ 5x_1 + 10x_2 - 13x_3 + 4x_4 - 16x_5 = 0 \end{cases}$$

Find a basis for this subspace, and then determine its dimension.

Exercise 124. Consider the subspace of $\mathbb{R}^3$ defined as the span of the following set of vectors.

$$\left\{ \begin{pmatrix} -1 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 7 \\ 3 \\ 4 \end{pmatrix}, \begin{pmatrix} 2 \\ -4 \\ 2 \end{pmatrix}, \begin{pmatrix} 6 \\ 5 \\ 5 \end{pmatrix} \right\}$$

Find a basis for this subspace, and then determine its dimension.

Exercise 125. Find a basis for the null space of the following matrix. Then determine the dimension of this null space. (This null space lives in some $\mathbb{R}^n$. Which one?)

$$A = \begin{pmatrix} 5 & 3 & 2 & 0 \\ 1 & -1 & 3 & 2 \end{pmatrix}$$

Exercise 126. Find a basis for the span of the following four vectors in $\mathbb{R}^6$. Work in two ways: first, use the casting out algorithm to look for linear independence; then use the row algorithm.

$$v_1 = \begin{pmatrix} 101 \\ -35 \\ 2 \\ 17 \\ 9 \\ -4 \end{pmatrix}, v_2 = \begin{pmatrix} 1 \\ 3 \\ 88 \\ 47 \\ 6 \\ -2 \end{pmatrix}, v_3 = \begin{pmatrix} 296 \\ -126 \\ -610 \\ -278 \\ -15 \\ 2 \end{pmatrix}, v_4 = \begin{pmatrix} 4 \\ 9 \\ 17 \\ 2 \\ -35 \\ 101 \end{pmatrix},$$

How do the two results compare? Why might you want to do one process or the other?

Exercise 127. In general, is there a way to describe the dimension of the column space of a matrix? Make several examples and figure out what is going on.

Exercise 128. In general, is there a way to describe the dimension of the null space of a matrix? Make several examples and figure out what is going on.