Linear Algebra

Math 217: Saint Lawrence University

Gabriel Dorfsman-Hopkins

GABRIEL DORFSMAN-HOPKINS
SAINT LAWRENCE UNIVERSITY
DEPARTMENT OF MATHEMATICS, STATISTICS, AND COMPUTER SCIENCE

These are growing course notes developed for a first course in Linear Algebra at Saint Lawrence University in Canton, NY. They are written and maintained by Professor Gabriel Dorfsman-Hopkins. The notes have been typeset in LaTeXwith a template that was adapted from the Legrand Orange Book Template by Mathias Legrand (`legrand.mathias@gmail.com`) with modifications by Vel (`vel@latextemplates.com`).

First release, Spring 2023

Contents

1. Introduction

1.1 January 24, 2023

1.1.1 Functions

Almost any course in mathematics is centered around studying types of *functions*. For example, in *Calculus* we study the behavior of functions of a single variable, that is, functions whose input is a single real number and whose output is a single real number, looking especially closely at functions which are *continous* or *differentiable*.

■ **Example 1.1 — Functions of a single variable.** Consider the function

$$f(x) = 3x.$$

Its input is a real number, x, and the output is computed by multiplying the input by 3. To see what this function does to a real number, say, 11, we can compute:

$$f(11) = 3 \times 11 = 33.$$

Explicitly, f takes an input of eleven an *transforms it* into an output of 33. ■

■ **Example 1.2** Consider the function:

$$g(x) = x^2 - 2x + 1.$$

What does this function do to the number 2? ■

The study of calculus looks closely at these functions of a single variable, establishing concepts like *derivatives* and *integrals*, and connecting them to many real world questions and situations. A shorthand that we will adopt to describe a function f of a single variable is the following

$$f : \mathbb{R} \to \mathbb{R}$$

This can be read aloud as *f is a function from $\mathbb{R}$ to $\mathbb{R}$*. It signifies that f takes a real number (on the left of the arrow), and runs it through the arrow to produce another real number (on the right of the arrow). *Note: The set before the arrow is called the* ***domain*** *of the function. It is also sometimes called the* ***source****. The set after the arrow is called the* ***co-domain****. It is sometimes also called the* ***target****.*

In *Mulltivariable Calculus* we develop similar ideas, **but the types of functions we study are different**. In particular, we allow for functions which take more than one real number as an input. Allowing for mutli-variable inputs allows calculus to be applied to our multi-dimensional world, and vastly expands the applications of derivatives, integrals, and related ideas.

■ **Example 1.3 — Functions of 2 variables.** In multivariable calculus you may encounter a function like:

$$f(x,y) = x - y.$$

It takes as input a *pair* of real numbers (x,y), and outputs their difference. For example, to see what the function does to the pair of number $(5,2)$ we can compute:

$$f(5,2) = 5 - 2 = 3.$$

In partiuclar, f will *transform* the pair of numbers $(5,2)$ into the single number 3. ■

■ **Example 1.4 — Functions of 3 variables.** Consider the function of 3 variables:

$$f(x,y,z) = xyz + 1.$$

What does this function do to the triple $(1,2,3)$? ■

The *arrow notation* of a function introduced above carries over here as well. For example, if f is a function of two variables, (whose input is 2 real numbers) we may write:

$$f : \mathbb{R}^2 \to \mathbb{R},$$

which we read as *f is a function from $\mathbb{R}^2$ to $\mathbb{R}$*. Here $\mathbb{R}^2$ denotes the collection of *pairs of real numbers*. Similarly, if g is a function of 3 variables (like in Example 1.4), we may write

$$g : \mathbb{R}^3 \to \mathbb{R}.$$

Notice that for each function we've describe so far, the output is *1-dimensional*. That is, we may have a function into which takes multiple real numbers as an input, but in each case the output is *a single real number*.[1] But just as allowing a multi-dimensional input massively expanded the scope of calculus, allowing functions to have a multidimensional output can be very useful as well.

■ **Example 1.5 — Analyzing Ocean Currents.** A group of oceanographers are measuring the movement of the water in the Atlantic, by studying where a collection of sensors start and end over the course of two weeks. They compile their data into a function C whose input is the GPS coordinates of a location in the Atlantic, and whose output of where the water at that location ends up 2 weeks later. For example,

$$C(40.47, -68.73) = (41.71, -64.07),$$

[1] You may recall that $\mathbb{R}$ can be thought of as a line, $\mathbb{R}^2$ as a plane, and $\mathbb{R}^3$ as 3-dimensional space. We will eventually adopt this notion of dimensionality, and explore it more carefully.

means that a drop of water whose GPS Coordinates are 40.47N 68.73W will move over the course of two weeks to the location 41.71N 64.07W. Observe that this is a function that takes as input two real numbers, and outputs *2 real numbers* as well! That is, both the input and the output are *2-dimensional*. In our arrow notation, we would write:

$$C : \mathbb{R}^2 \to \mathbb{R}^2.$$

TODO: put an image here!

■

■ **Example 1.6 — Casting Shadows.** Shadows are cast when a body in space blocks the sun from hitting the ground. If we'd like to study the shape of shadows mathematically, it is worth modelling shadows with a function, say S. Here:

$$S(\text{A point in space}) = \text{The spot on the ground where it casts a shadow.}$$

Modelling 3-dimensional space with $\mathbb{R}^3$ and the 2-dimensional ground with $\mathbb{R}^2$, this gives a function:

$$S : \mathbb{R}^3 \to \mathbb{R}^2.$$

In fact, this will be a *projection function*, a certain kind of *linear transformation* that we will study in TBA.

TODO: put an image here!

■

As we can see, functions with multivariable outputs are not hard to come up with, and model many different situations we would hope to study with mathematics. Let us begin by looking at a very special case:

1.1.2 Functions $\mathbb{R}^2 \to \mathbb{R}^2$

Suppose you wanted to describe a function $f : \mathbb{R}^2 \to \mathbb{R}^2$. How would you go about it? Both the input and output of f consist of pairs of numbers, so to be explicit with our notation, let's give the first $\mathbb{R}^2$ the coordinates (x, y), and the second $\mathbb{R}^2$ the coordinates (u, v). In particular, our function will look something like

$$f(x, y) = (u, v).$$

The function should be a rule so that, given a pair (x, y) of real numbers, we return with another pair of numbers, (u, v). In particular, we have to say what u is, and what v is. But each of these coordinates depend on both x and y, so in essense this is just *two functions* whose output is a real number:

$$u = u(x, y)$$

$$v = v(x, y).$$

■ **Slogan 1.1** To describe a function whose output is two real numbers, you can give 2 functions which output a single real number each.

Let's see how this works with an example.

■ **Example 1.7** Lets define a function $f : \mathbb{R}^2 \to \mathbb{R}^2$ via the rule $f(x,y) = (u,v)$ where:

$$u = u(x,y) = xy + 1,$$

$$v = v(x,y) = x + 2y^2$$

The input of this function is a pair of numbers (x,y), and the output is *another* pair of number (u,v). So, for example, if we feed the function the pair $(-1,3)$, we can compute:

$$u = u(-1,3) = -1 \times 3 + 1 = -3 + 1 = -2$$

$$v = v(-1,3) = -1 + 2 \times 3^2 = -1 + 18 = 17.$$

Therefore, this function transforms the pair $(-1,3)$ to the pair $(-2,17)$:

$$f(-1,3) = (-2,17).$$

■

■ **Example 1.8** Define a function $g : \mathbb{R}^2 \to \mathbb{R}^2$ which takes (x,y) to (u,v) via the rule

$$u = u(x,y) = 2x - 2y,$$

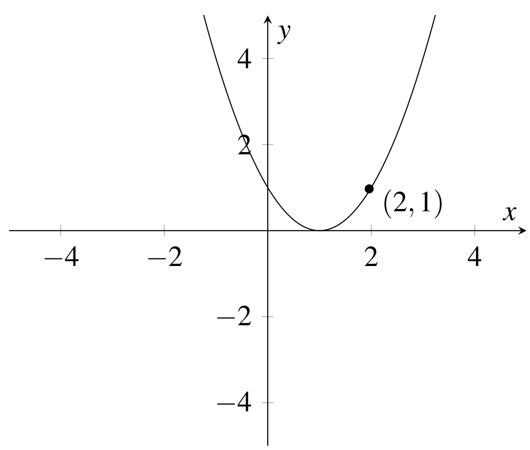
$$v = v(x,y) = \frac{1}{2}x + y.$$

Where does g take the point $(1,1)$? ■

It is often useful to think about a function as something that *moves* the point (x,y) to the point (u,v), and to emphasize this intuition, we will often refer a function $\mathbb{R}^2 \to \mathbb{R}^2$ as a *trnasformation of the plane*.

1.1.3 Visualizing Transformations of the Plane

How do we visualize these types of functions? Since these will be central objects of study, let's start by spending some time developing techniques for how to think about and imagine a function from $\mathbb{R}^2$ to itself. Recall that in calculus you often visualize functions $g : \mathbb{R} \to \mathbb{R}$ using their graphs in the xy-plane. Here the x axis plays the role of the domain, and the y-axis the role of the co-domain, and the graph is generally a curve consisting of the points $(x, g(x))$. For example, the graph of the function $g(x) = x^2 - 2x + 1$ from Example 1.2 is below.



The fact that $f(2) = 1$ is captured by the fact that $(2,1)$ lies on the curve. A similar approach is used in multivariable functions, where now the domain is the entire xy-plane, and the co-domain is the z-axis. Then a function $f : \mathbb{R}^2 \to \mathbb{R}$ can be graphed in 3-dimensional space, coloring in the points $(x, y, f(x, y))$, generally giving rise to a surface in 3-dimensional space.

■ **Question 1.1** Can we take a similar approach to a function $f : \mathbb{R}^2 \to \mathbb{R}^2$? Why or why not?

Given the dimensional constraints, we have to come up with another way to represent a function $f : \mathbb{R}^2 \to \mathbb{R}^2$. One way to do this is to get to the heart of what a function really does: *it transforms a point in* $\mathbb{R}^2$ *to another point in* $\mathbb{R}^2$. In particular, we can think about such a function as *something that transforms the plane*, moving the points of the plane around.

■ **Slogan 1.2** Think about a function $\mathbb{R}^2 \to \mathbb{R}^2$ as something that moves around the points on a single plane. The input (x, y) is where the point starts, and the output $(u, v) = f(x, y)$ is where the point ends.

In fact, this is exactly what the function from Example 1.5 does, it keeps track of of where a drop of water in the Atlantic moves over the course of two weeks!

■ **Example 1.9** Let's visualizae the function from Example 1.7, which was function $f(x, y) = (u, v)$ where:

$$u = u(x, y) = xy + 1,$$
$$v = v(x, y) = x + 2y^2$$

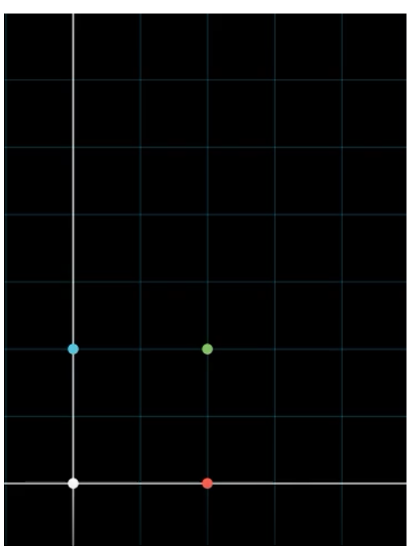
To get a sense of what kind of movement, let's keep track of what happens to a few points:

$$(0,0), (1,0), (0,1), (1,1).$$

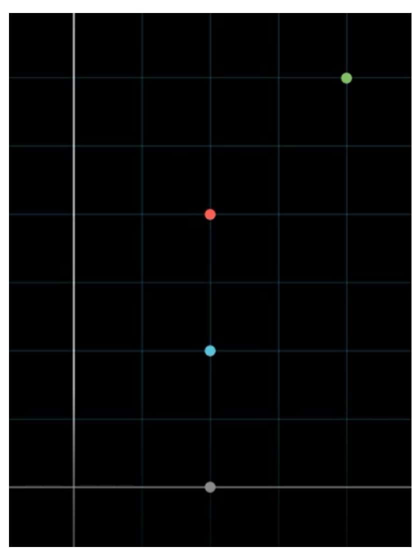
Using the formulas we can compute where f takes these points, just like in Example 1.7.

$$f(0,0) = (1,0), \quad f(1,0) = (1,2),$$
$$f(0,1) = (1,1), \quad f(1,1) = (2,3).$$

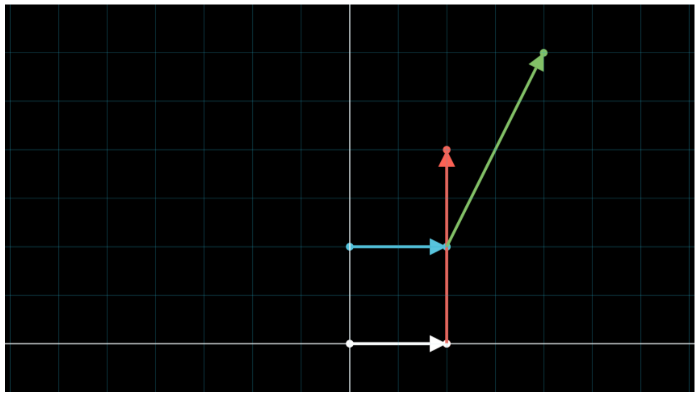
Instead of a single graph of the function, we can represent what f does with two pictures of the plane, a *before* shot and an *after* shot. On the left, we see the 4 points before applying f, and on the right, we see them after.



$\Longrightarrow$



The *movement* of the situation can be captured nicely by an animation linked below.[2] You can also emphasize that it is movement on a single page by using arrows that point from the start to the finish of the various points:



■

Exercise 1.1 — January 24th Checkin. Consider the transformation $L(x,y) = (u,v)$ of the plane $\mathbb{R}^2$, given by the following two equations:

$$u = u(x,y) = y$$

$$v = v(x,y) = -x.$$

On a single coordinate plane, draw what the function does to a number of points. Do this by plotting a point (x,y), its image $L(x,y)$, and connecting them with an arrow. Use a few sentences to describe what the transformation L is doing to the plane. This can be a *qualitative* description. *What does it look like is happening?* ■

[2] `www.gabrieldorfsmanhopkins.com/LinearAlgebraNotes/animationsAndTools/Jan24_Quad.mp4`

1.2 January 26, 2023

Let's begin by recalling some of the techniques discussed last time to visualize a function $\mathbb{R}^2 \to \mathbb{R}^2$ as a *transformation* of the plane.

■ **Example 1.10** Let $g : \mathbb{R}^2 \to \mathbb{R}^2$ from Example 1.8. In particular, it is given by the rule $g(x,y) = (u,v)$ where:

$$u = 2x - 2y \text{ and,}$$

$$v = \frac{1}{2}x + y.$$

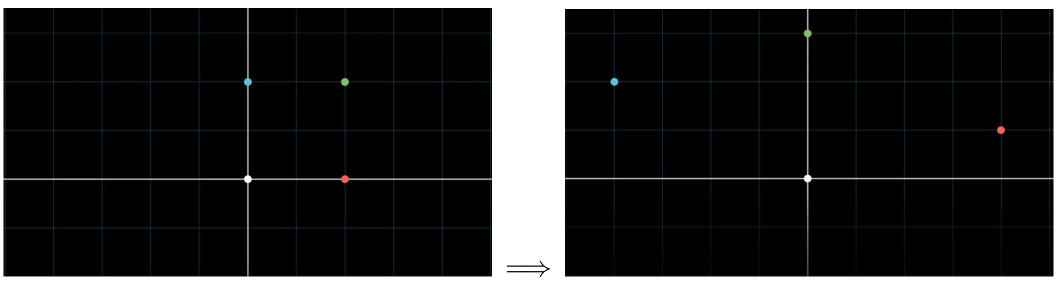
We compute compute where g takes the four points: $(0,0), (1,0), (0,1)$, and $(1,1)$. For example, $g(0,0)$, we may compute the u coordinate to be $2 \times 0 - 2 \times 0 = 0$ and the v coordinate to be $\frac{1}{2} \times 0 + 0 = 0$, so that $g(0,0) = 0$. Similar computations show that:

$$g(1,0) = (2,0.5)$$

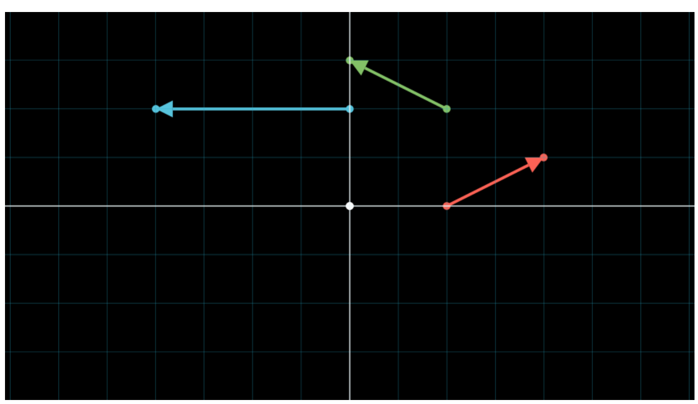
$$g(0,1) = (-2,1)$$

$$g(1,1) = (0,1.5).$$

Plotting the points before and after applying g gives:



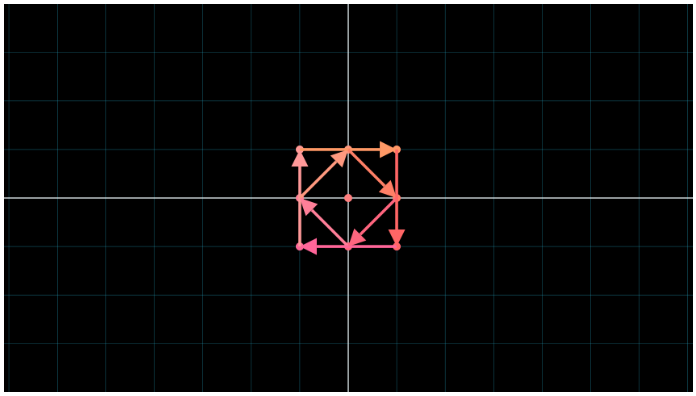
Plotting the before and after on the same plane, connecting (x,y) with $g(x,y)$ using arrows gives the following picture.



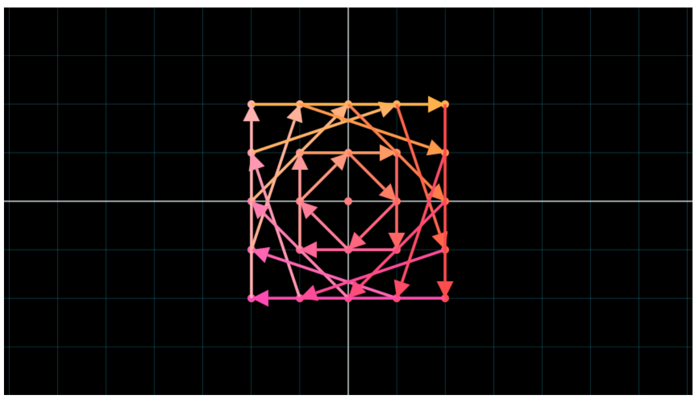
We should imagine this function something *moving around the points on the plane*, a perspective that is emphasized when animating the function. You find an animation of the moving points below.[3]. Try to give a qualitative description of what this function is doing to the plane. Plotting moer points may give a better picture. ■

[3] www.gabrieldorfsmanhopkins.com/LinearAlgebraNotes/animationsAndTools/Jan26_Linear.mp4

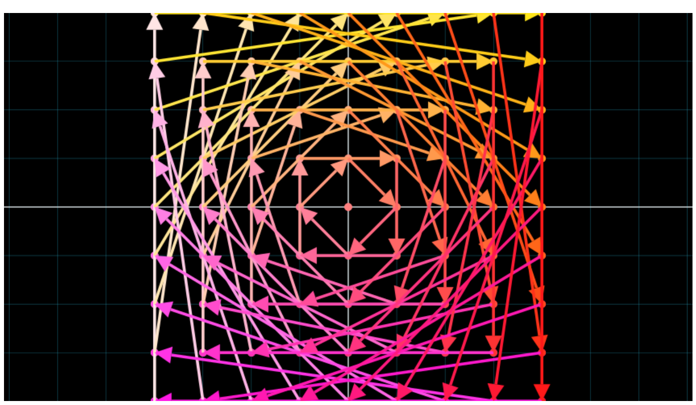
■ **Example 1.11** At the end of our last class, we did a similar exercise using the function $L(x,y) = (y,-x)$ (cf. Exercise 1.1). Let's draw a few pictures and see if we can arrive at a description of what is happening to the plane. First, we plot all the points whose x and y coordinate's are between -1 and 1, connecting the points before and after applying L with an arrow.



Can you begin to describe what L is doing to the plane? Let's throw in a few more points, now letting the coordinates range between -2 and 2.



As you can see, it appears that L is *rotating the plane clockwise!*. We include one more image now with coordinates ranging between -5 and 5.

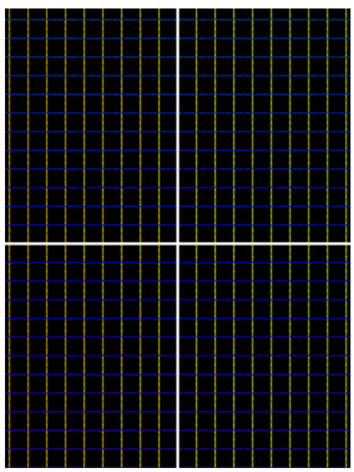
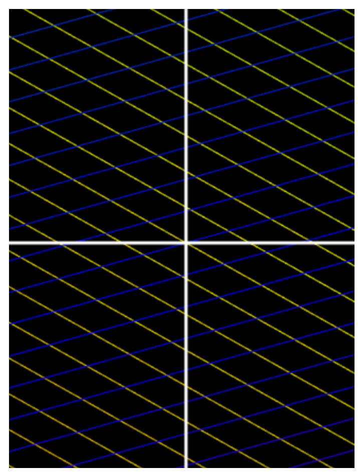


Althought the image is starting to get cluttered, this definitely appears to be a rotation, and indeed, replacing the arrows with an animation makes this clear (see the animation below[4]). ■

[4] www.gabrieldorfsmanhopkins.com/LinearAlgebraNotes/animationsAndTools/Jan26_Rotate.mp4

To summarize, plotting where points go under a function can give a sense qualitative sense of how a function moves the plane. That said, Examples 1.9 and 1.10 suggest that only drawing where a few points go gives an incomplete picture. On the other hand, as we saw at the end of Example 1.11, if we to fill in more and more points, the image can start to get cluttered and it may become difficult to infer much from the picture. [5]. That being said, if you carefully pic which points to keep track of, you can get a nice sense of the *geometric* properties of a function. One way to do this, is by keeping track of what the function does to the *gridlines* of the plane.

■ **Example 1.12** To get a better picture of the function g from Examples 1.8 and 1.10, let's analyze what it does does to the gridlines of the plane.

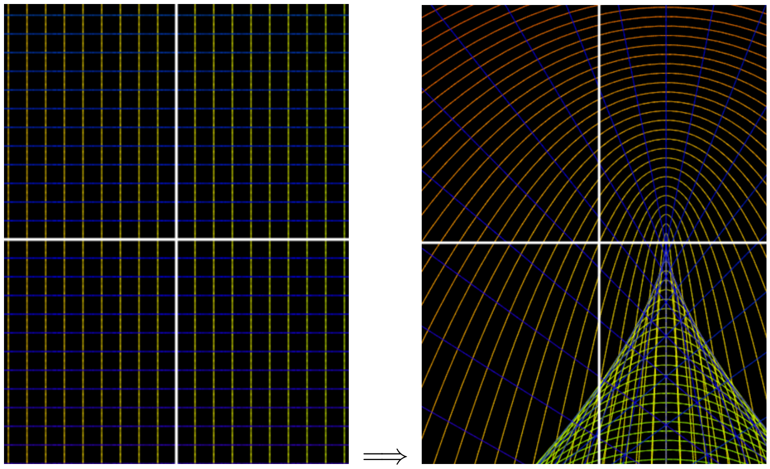

$\Longrightarrow$


One can really get a sense for how g *moves* the plane by playing around with the tool linked below[6]. In particular, we see that it sort of *stretches* and *rotates* the plane, distorting it slightly but not too much. In this course we will develop a vocabulary to mathematically describe terms like *stretching the plane*, and ways to extract that information from the equations given in Exampel 1.8, but for now we're trying to get a qualitative sense of what's going on. ■

■ **Example 1.13** Let's also look at what the function f from Example 1.7 does to the gridlines of the plane.

[5] *Try this! For some functions you can actually get a nice picture! In fact, the situation in Example 1.11 is a particularly nice one. In general it will be much more complicated*

[6] Click the *linear* button here: `www.gabrieldorfsmanhopkins.com/LinearAlgebraNotes/animationsAndTools/Jan26Gridlines/index.html`



The animation is actually quite nice to look at[7]. ■

It is fair to say that the function in Example 1.13 appears far more complicated than the one in Example 1.12. In fact, in some sense it is complicated in a way that puts it beyond the purview of *linear algebra*[8] For the context of linear algebra, we will have to restrict ourselves to functions more like that of Example 1.12, functions that we will call *linear transformations*. Before describing exactly what these are, it might be worth while to ponder the following question. Qualitative answers are always welcome!

■ **Question 1.2** What are some differences between what happens to the gridlines in the two examples on the previous page?

1.2.1 Linear Transformations of the Plane

One answer to Question 1.2 could be: *In example 1.8 the gridlines remain as lines after applying g, but in Example 1.7 the gridlines become curvy*. This is a good observation. Recall that lines played a special role in calculus. Not only where they the simplest functions, we used them to model more complicated functions locally, by taking *tangent lines*. We do something similar in multivariable calculus, modelling more complicated functions with linear ones by taking the *tangent plane*. Not only were these functions simple *geomterically* (being lines and planes), but they were also simple *algebracially*. For example, a line usually has the following equation:

$$f(x) = {\color{red}mx} + {\color{blue}b}.$$

Above we highlighted the *linear term* in red, and the *constant term* in blue. Similarly, a plane had a simple equation as well:

$$h(x,y) = {\color{red}mx} + {\color{red}ny} + {\color{blue}b},$$

where again the linear terms are highlighted in red, and the constant term in blue. Looking at the function $g(x,y) = (u,v)$ from Example 1.8, we see that the equations for both u and v have only linear terms (and no constant terms).

$$u = u(x,y) = 2x - 2y,$$

[7]Click the *Quadratic* button here: `www.gabrieldorfsmanhopkins.com/LinearAlgebraNotes/animationsAndTools/Jan26Gridlines/index.html`

[8]This is the kind of function studied in *algebraic geometry*.

$$v = v(x,y) = \frac{1}{2}x + y.$$

This will turn out to be a good definition for a linear function.

Definition 1.2.1 — Linear Transformations of the Plane. A *linear transformation of the plane*, also called a *linear transformation from* $\mathbb{R}^2$ *to* $\mathbb{R}^2$, $L(x,y) = (u,v)$, where u and v are given by linear equations with no constant term:

$$u = u(x,y) = ax + by,$$

$$v = v(x,y) = cx + dy,$$

where a, b, c, and d are real numbers.

■ **Warning 1.1** A linear transformation is not quite the same as a linear function from Calculus, because a linear function from calculus can have a constant term, and a linear transformation cannot. This is an unfortunate inconsistency in terminology, but perhaps you can think about a linear transformation as being more *purely linear* since the only terms it has are linear terms, and no constant terms.

■ **Warning 1.2** In light of Question 1.2, you may want a geometric definition of a linear transformation of the plane to be something like: *it takes gridlines to lines*. This isn't quite the case (we will see some examples of this). To be completely precise, we also need the gridlines to remain parallel and evenly spaced, and we need $L(0,0) = (0,0)$. We will discuss this geometric reformulation more later, but for now I just wanted to mention that a this first guess is not quite enough.

You might be getting this far and thinking *wait...I thought linear algebra was about matrices? Where do those fit in?* This is a good question, so let's give a preliminary answer. Take a linear tranformation $L(x,y) = (u,v)$ where:

$$u = u(x,y) = ax + by,$$

$$v = v(x,y) = cx + dy,$$

This function is completely determined by the coefficients of x, and the coefficeints of y. That is, to know L, it is enough to know a, b, c, and d. So, we can completely capture all the data for L in the matrix:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

For now we should just think of a matrix as a rectangular array of numbers, so that a linear transformation of the plane corresponds to a 2×2 matrix.

Definition 1.2.2 The matrix associated to the linear transformation in Definition 1.2.1 is the 2×2 matrix

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

■ **Example 1.14** Consider the function $g(x,y) = (u,v)$ from example 1.8. Observe that the coefficient of y in the first equation is -2, because adding $-2y$ is the same as subtracting $2y$. Also, the coefficient of y in the second equation is a 1 becuase $y = 1 \times y$.

$$u = u(x,y) = 2x + -2y,$$

$$v = v(x,y) = \frac{1}{2}x + 1y.$$

The matrix associated to this fuction is therefore:

$$\begin{bmatrix} 2 & -2 \\ \frac{1}{2} & 1. \end{bmatrix}$$

■

This correspondance goes in both directions. That is, given a matrix, you can extract a linear function.

Definition 1.2.3 Consider a matrix:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

The *linear function associated to this matrix* is the function $L(x,y) = (u,v)$ where:

$$u = ax + by \text{ and,}$$

$$v = cx + dy.$$

Let's run through an example of applying a function, given only a matrix.

■ **Example 1.15** We compute $T(1,-2)$ where $T(x,y)$ is the function associated to the matrix

$$\begin{bmatrix} 3 & 1 \\ -1 & 0 \end{bmatrix}.$$

Applying the definition we see that $T(x,y) = (u,v)$ where:

$$u = 3x + 1y = 3x + y,$$

$$v = -1x + 0y = -x.$$

Plugging in $(x,y) = (1,-2)$ gives:

$$u = 3 \times 1 + (-2) = 1, \qquad \text{and} \qquad v = -1$$

Therefore $T(1,-2) = (1,-1)$. ■

R Later we will streamline this process using *matrix multiplication*.

Exercise 1.2 Consider the matrix:

$$\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$$

Let $L(x,y) = (u,v)$ be the associated linear transformation.

1. Write down the formulas for u and v in terms of x and y.

2. Evaluate L at $(0,0),(0,1),(1,0)$ and $(1,1)$.
3. Plot the four points of part (b), before and after applying L, and connect them with arrows.
4. Give a qualitative description of what you think L is doing to the plane.

■

So far we've only seen how a correspondance between linear transformations of the plane and 2×2 matrices. We will work out in the coming weeks how this fits in to notions of matrix multiplication, determinants, and other matrix operations. For now, we the main take away should be the following.

■ **Slogan 1.3** A matrix is a function.

1.3 Homework 1

Exercise 1.3 Consider the matrix:

$$M = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

Let $L(x,y) = (u,v)$ be the associated linear transformation.

1. Write down the formulas for u and v in terms of x and y.
2. Evaluate L at all the points with integer coordinates are between -1 and 1. (There should be nine such points).
3. Plot the 9 points from part (b) before and after applying L, and connect them with arrows.
4. Give a qualitative description of what you think L is doing to the plane.

■

Exercise 1.4 Consider the matrix:

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Let $T(x,y) = (u,v)$ be the associated linear transformation.

1. Write down the formulas for u and v in terms of x and y.
2. Evaluate T at 5 points of your choice.
3. Plot the 5 points from part (b) before and after applying T, and connect them with arrows.
4. Give a qualitative description of what you think T is doing to the plane.

■

Exercise 1.5 For this problem, adopt the notation of Exercises 1.3 and 1.4. Also consider the matrix N associated to the function $g(x,y)$ from Example **??**:

$$N = \begin{bmatrix} 2 & -2 \\ \frac{1}{2} & 1 \end{bmatrix}$$

1. Can you identify any relationships between the outputs $L(1,0), L(0,1)$ and the matrix M?
2. Can you identify any relationships between the outputs $T(1,0), T(0,1)$ and the matrix I?
3. Can you identify any relationships between the outputs $g(1,0), g(0,1)$ and the matrix N?

4. Now let's treat the general case: pet $\ell(x,y)$ be a linear transformation associated to a general matrix:

$$P = \begin{bmatrix} a & b \\ c & d. \end{bmatrix}$$

Describe the relationship between $\ell(1,0)$, $\ell(0,1)$ and the matrix P. Give reasoning for your answer.

Exercise 1.6 Let $\ell : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear transformation. Do you agree or disagree with the following statement?

Once I know $\ell(1,0)$ and $\ell(0,1)$, I can determine $\ell(x,y)$ for any pair (x,y).

Explain your reasoning.

Exercise 1.7 Consider a function $F : \mathbb{R}^2 \to \mathbb{R}^2$, and suppose that $F(0,0) = (1,1)$. Is it possible for F to be a linear transformation? Why or why not?

Exercise 1.8 Adopt the notation of Problem 1.5. Define a rule for adding two points as follows:

$$(x_0,y_0) + (x_1,y_1) = (x_0 + x_1, y_0 + y_1).$$

Let $P = (1,2)$ and $Q = (3,-2)$.

1. Can you identify any relationship between $L(P), L(Q)$, and $L(P+Q)$?
2. Can you identify any relationship between $g(P), g(Q)$, and $g(P+Q)$?
3. To see that it's not a fluke, do parts (a) and (b) again, but with new points P and Q of your choice.
4. Let $\ell : \mathbb{R}^2 \to \mathbb{R}^2$ be a general linear transformation, and let P and Q be two random points in $\mathbb{R}^2$. Make a conjecture for the relationship between $\ell(P)$, $\ell(Q)$, and $\ell(P+Q)$. (There is no need to prove this yet, but you can extrapolate from the evidence collected in (a) through (c)).

Exercise 1.9 To give a function $\mathbb{R}^2 \to \mathbb{R}^2$ we needed to give 2 functions of 2 variables which output a single number each (for more detail, see Section 1.1.2 in the course notes). Let's see if we can work out what to do in higher dimensions. In particular, adapt Section 1.1.2 in the course notes to describe a function $F : \mathbb{R}^3 \to \mathbb{R}^3$. You can make up any funcion you like, just make sure that you describe it fully. Evaluate this function at the points $(0,0,0), (1,0,0)$, and $(1,2,3)$.

2. Vectors

Acknowledgment: I'd like to attribute this approach to vectors in part to Grant Sanderson, author of the delightful youtube channel `3Blue1Brown`. In particular, I borrow heavily his description of the three perspectives of vectors presented below as the *physicist's perspective*, the *computer scientist's perspective*, and the *mathematician's perspective*.

2.1 January 31, 2023

We suggested in the Introduction that the field of Linear Algebra is centered around the study of *linear transformations*. Furthermore, Exercise 1.8 suggests that a linear transformation L satisfies the following equation:[1]

$$L(P+Q) = L(P) + L(Q).$$

It is worth taking some time to unpack what $+$ is doing here. If P and Q are points, what is their sum? In Exercise 1.8, we defined the sum of 2 points to be a third point, whose coordinates correspond to adding the coordinates of P and Q. In $\mathbb{R}^2$ this is written as follows:

$$(x_0, y_0) + (x_1, y_1) = (x_0 + x_1, y_0 + y_1).$$

This is a perfectly valid formula, but it may also seem a bit strange. In particular, we should unpack a concrete interpretation of this algebraic operation to answer the following question:

■ **Question 2.1** What exactly is the meaning of adding coordinates of points in $\mathbb{R}^2$?

By trying to answer this question, we naturally encounter the notion of a *vector*. In fact, a first definition of a vector is pretty much exactly the idea of *points you can add*.

[1] At least when $L : \mathbb{R}^2 \to \mathbb{R}^2$

Definition 2.1.1 — 2-dimensional vectors: the computer scientist's perspective. A two dimensional vector is an array of two numbers aligned vertically:

$$\begin{bmatrix} x \\ y \end{bmatrix}.$$

These can be added coordinatewise:

$$\begin{bmatrix} x_0 \\ y_0 \end{bmatrix} + \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} x_0 + x_1 \\ y_0 + y_1 \end{bmatrix}.$$

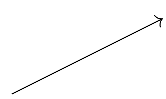
Notation 2.1. *Aligning vectors* vertically *allows them to fit into the matrix theory we will develop in the coming weeks. We will sometimes use the term* column vector *when writing it this way.*

We call this *the computer scientist's perspective*, because it remembers a vector as a light-weight data type stored in a way that easily allows for vector operations (like addition and applying linear maps) to be computed efficiently by a computer in almost any programming language. This perspective also has the advantage of generalizing very easily to higher dimensions, *can you see how?* That being said, it doesn't really get us any closer to answering Question 2.1. In order to do this, we give another perspective on vectors you may have seen in a physics course.

Definition 2.1.2 — 2-dimensional vectors: the physicist's perspective. A two dimensional vector is a quantity specifying a *magnitude* together with a *direction* in the two dimensional plane. This can be represented by an arrow, pointing in the given direction, whose length is the given magnitude.

The physicist's vectors can be added too, essentially by *concatenating the two arrows*. We define this more carefully in Definition **??** below. In particular, we have encountered two different perspectives on the notion of a vector. Below to the left is an example of a computer scientist's vector, and to the right is an example of a physicist's vector.

$$\begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

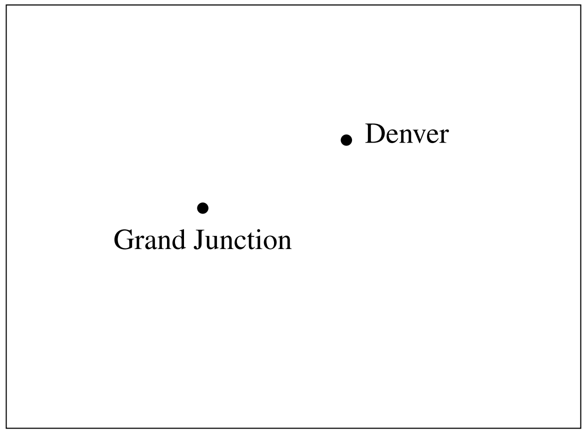


In fact, one could say that these are two perspectives on the same vector, *can you explain why?*. An important part of linear algebra involves learning how to pass seemlessly between these two perspectives, as one lends itself better to computations, while the other lends itself better to interpretations. In this section, we will explore these two perspectives, and start developing a dictionary between them, keeping track of what information can get lost in translation. Along the way we will extract algebraic properties of vectors, and have a first encounter the notions of *linear combinations* and *spans*, which are among the most important in this course.

R Another thing we can do from both perspectives is *scale* vectors by numbers. In fact, there is a third perspective on vectors, which we can call *the mathematician's perspective*, which essentially defines vectors as: *theoretical objects which can be added together and scaled*. We will postpone discussion of this third, more abstract, perspective until the end of the semester. The attentive reader may want to pay attention to how most properties of vectors can be expressed in terms of of these two operations (addition and scaling).

2.1.1 The Physicist's Perspective

A vector is a natural quantity to describe relationships between objects in the physical world. For example, suppose that a pilot is hoping to fly from Denver, Colorado to Grand Junction, Colorado, and asks you for directions.[2]

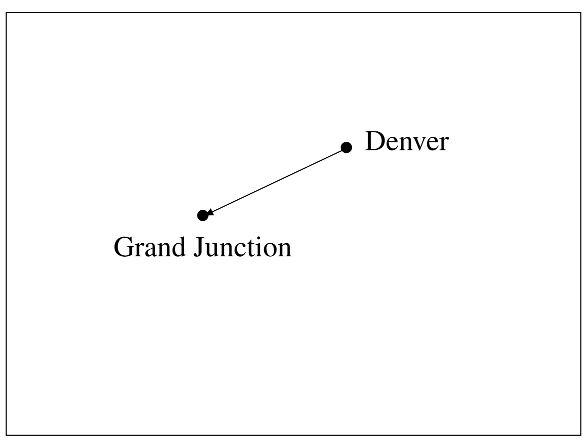


The distance between Denver and Grand Junction is 220 miles. That be but it is *not enough* to tell the pilot to fly 220 miles. If they don't want to end up in Wyoming, they must also know which direction to fly. So for example, you may tell the pilot to fly 220 miles *west-southwest*. These two peices of information constitute a quantity which we call a *vector*:

Magnitude: 220 miles,

Direction: West-Southwest.

All of the defining this quantity information can be exhibited on a map, by drawing an arrow from the start to the finish. One can recover the direction from the direction of the arrow, and the magnitude from the length of the arrow.



This gives us another way to think about Definition 2.1.2.

[2]Map not to scale.

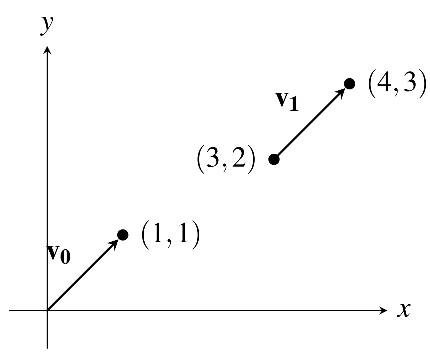
Definition 2.1.3 [Vectors as Displacement] Given two points P and Q, the *displacement vector from P to Q* is the arrow whose tip is at the point Q whenever its tail is placed on P.

Notation 2.2. *Given a vector* $\mathbf{v}$*, we denote its magnitude by* $||\mathbf{v}||$.

With this definition we think of a *displacement vector* as something you can apply to a point. To apply a vector to a point P, we put its tail at P, and the output is wherever its tip points. Importantly, the same vector can be placed in different locations. Let us take this perspective to our vector which takes us from Denver to Grand Junction. This vector is completely determined by the fact that it goes 220 miles west-southwest. It isn't necessary for its to lie on Grand Junction. For example, we could apply the *same vector* starting from Canton, and we would end up somewhere near Toronto. We would still be using the same magnitude (220 miles) and direction (west-southwest), and therefore following the same vector. This is an important point: *a vector is determined by magnitude and direction*. 2 vectors of the same length and pointing in the same direction are the same vector, even if they are drawn at different places.

■ **Slogan 2.1** The vector is the arrow, not where the arrow is.

■ **Example 2.1** Consider following vectors. $\mathbf{v_0}$ connects $(0,0)$ and $(1,1)$, while $\mathbf{v_1}$ connects $(3,2)$ and $(4,3)$. Does $\mathbf{v_0} = \mathbf{v_1}$?



■

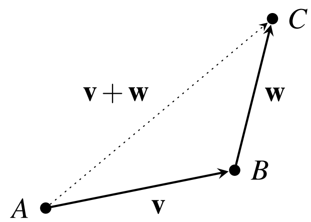
It is common in linear algebra to distinguish between *vectors*, which have a direction and a magnitude, numbers without any direction. The latter is just a number, but we will also often refer to it as a *scalar*.

Adding and Scaling Arrow Vectors

Recall that Question 2.1 asked what adding vectors meant. By thinking of vectors as measuring displacement, we can get a geometrically and physically meaningful understanding how they add, subtract, and scale. We will explore this with the following thought experiment in mind:

You are programming autonomous vehicles. To command a vehicle to move, you give it a vector. The vehicle will then move along the vector: in the given direction for the given magnitude.

Addition: Suppose you give your vehicle a vector $\mathbf{v}$ to follow, and it moves from point A to point B. Once it has arrived at point B, you give it another vector $\mathbf{w}$, and it moves from point B to point C. At this point, the net displacement that the vehicle has travelled is from point A to point C. Define $\mathbf{v}+\mathbf{w}$ to be this displacement vector from A to C.



We can summarize in the following definitoin.

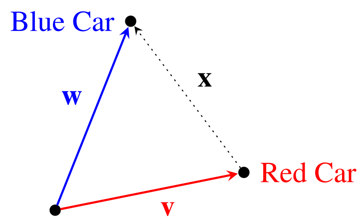
Definition 2.1.4 The sum of two vectors $\mathbf{v}$ and $\mathbf{w}$, is the combined displacement resulting from first applying $\mathbf{v}$, and then applying $\mathbf{w}$ to the result.

■ **Question 2.2** Given 2 vectors $\mathbf{v}$ and $\mathbf{w}$, is it always true that

$$\mathbf{v}+\mathbf{w}=\mathbf{w}+\mathbf{v}?$$

With addition now defined, we can move on to:

Subtraction: Suppose you control two vehicles, a red vehicle and a blue one, both starting at the same point. You send the red one along vector $\mathbf{v}$, and send the blue one along vector $\mathbf{w}$. After they arrive, the blue vehicle breaks down, so you must send the red vehicle to rescue it. What vector $\mathbf{x}$ must you command the red car to follow?



In particular, if the red car first does $\mathbf{v}$, and then does $\mathbf{x}$, it should overall be following $\mathbf{w}$, and therefore should end up alongside the blue car. We translate this by Definition 2.1.4 to the statement,

$$\mathbf{w}=\mathbf{v}+\mathbf{x}.$$

If subtraction of vectors were to make any sense, then we could subtract $\mathbf{v}$ from both sides and discover that $\mathbf{x}$ really should be the difference of $\mathbf{w}$ and $\mathbf{v}$:

$$\mathbf{w}-\mathbf{v}=\mathbf{x}.$$

Therefore, that is the definition we will make.

Definition 2.1.5 The difference $\mathbf{w}-\mathbf{v}$ of two vectors $\mathbf{w}$ and $\mathbf{v}$, is the vector which, when added to $\mathbf{v}$, gives $\mathbf{w}$.

We can now add and subtract vectors in a geometrically meaningful way. Bringing us closer to getting meaningful answer to Question 2.1. Before moving on, though, we'd like to introduce a special vector.

The Zero Vector: If your vehicle doesn't move at all, what vector does it follow? Since magnitude is the net distance covered, the magnitude is 0. As for direction, this isn't really well defined, since if you move 0 units in any direction, you've stayed put. We will call the vector from a point to itself the *zero vector*.

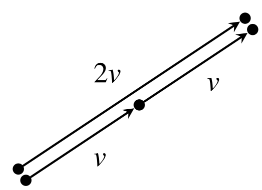
Definition 2.1.6 The vector whose magnitude is zero is called the *zero vector*, and is denoted $\mathbf{0}$.

R The zero vector is the only vector whose direction is unspecified.

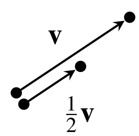
Exercise 2.1 Let $\mathbf{v}$ be any vector. What is $\mathbf{v}+\mathbf{0}$? What about $\mathbf{v}-\mathbf{v}$? ■

Let's introduce one more important operation that vectors allow.

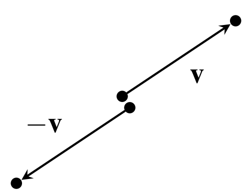
Scalar Multiplication: Suppose you'd like to send your car in the same direction as a vector $\mathbf{v}$, but twice as far as the vector $\mathbf{v}$ allows. You could acheive this by applying $\mathbf{v}$, and then doing so again. With Definition 2.1.4 in mind, you send your car along $\mathbf{v}+\mathbf{v}$, which we can write as $2\mathbf{v}$.



Similarly, if you wanted your car to go in the same direction as $\mathbf{v}$, but half as far, you could follow a vector $\mathbf{w}$ which satisfied $\mathbf{v}=\mathbf{w}+\mathbf{w}$. Since $2\mathbf{w}=\mathbf{v}$ we could reasonably say that $\mathbf{w}=\frac{1}{2}\mathbf{v}$.



Alternatively, suppose you wanted to go the same distance as $\mathbf{v}$, but in the opposite direction. You could follow a vector $\mathbf{x}$. Notice that if the car first does $\mathbf{v}$ and then does $\mathbf{x}$, it will travel along $\mathbf{v}$, and move the same distance in the opposite direction until it gets back to where it started. In particular, we have that $\mathbf{v}+\mathbf{x}=\mathbf{0}$, so it is reasonable to write $\mathbf{x}=-\mathbf{v}$.



Following this logic we can deduce that to scale a vector by a positive number, you scale its magnitude. The negative of a vector reverses direction. What about scaling by a negative number? The following formula should shed some light.

$$-2\mathbf{v}=-(2\mathbf{v}).$$

It appears that to scale by negative 2, you can first scale by 2, and then reverse direction.

Exercise 2.2 With $\mathbf{v}$ as in the figures above, sketch $-2\mathbf{v}$ ■

We can put all this together into the following definition.

Definition 2.1.7 Let $\mathbf{v}$ be a vector and c any scalar. Then the vector $c\mathbf{v}$ is defined by the following data.

- If c is positive, the direction of $c\mathbf{v}$ is the same as $\mathbf{v}$. Otherwise, the direction of $c\mathbf{v}$ is opposite to that of $\mathbf{v}$.
- The magnitude of $c\mathbf{v}$ is:

$$||c\mathbf{v}|| = |c| \cdot ||\mathbf{v}||,$$

where $|c|$ denotes the absolute value of the scalar c.

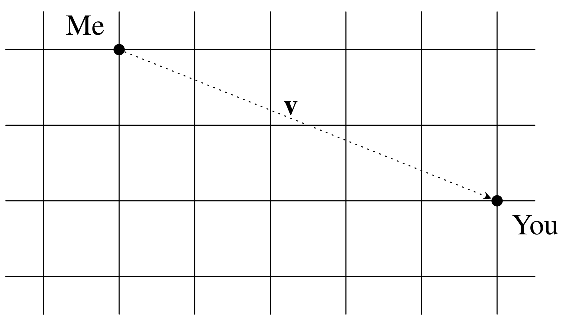
Exercise 2.3 Let $\mathbf{v}$ be any vector. What is $0\mathbf{v}$? ■

A nice output of this geometric approach is that we can give geometric names to certain algebraic operations. For example, we should call 2 vectors parallel, if when we draw their arrows are parallel as lines segments. Our definition of scalar multiplication tells us that this is equivalent to one being a scalar multiple of the other, giving us a algebraic notion of parallel-ness (that can and will extend to higher dimensions).

Definition 2.1.8 Two vectors $\mathbf{v}$ and $\mathbf{w}$ are parallel if there is some constant c such that $\mathbf{v} = c\mathbf{w}$.

2.1.2 Decomposing a Vector into Components

We'd like to connect the ideas described in the *physicist's perspective* on vectors, to the ordered pair numbers which we called the *computer scientist's perspective*. To do this, we'll take the example of giving directions in a city.



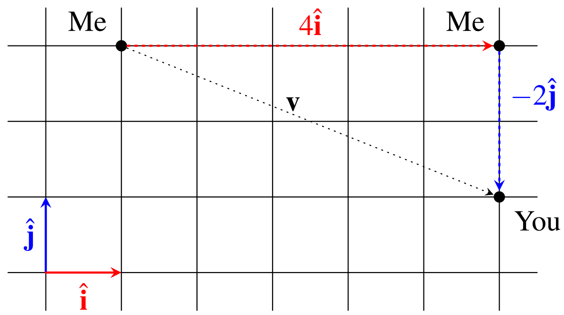
Above is a map depiction of a city, with the thick vertical and horizontal lines roads. You'd like to tell me how to find you. In particular, you need to tell what vector $\mathbf{v}$ I need to follow, in order to get to your location. One way to describe this to me is to give a magnitude and direction. But it is unlikely that you'll tell me something like *go about 530 meters in a direction that is mostly east but somewhat south*. In fact, even if you were more precise with the angles and distance, it is unlikely that I would be able to follow the directions (without walking through buildings).

Instead, you'd probably say something like *walk 5 blocks east, and then 2 blocks south*. Indeed, the regular gridlines of the city give us two natural vectors which we can all agree on:

$$\hat{\mathbf{i}} = \text{ one block east,}$$

$$\hat{\mathbf{j}} = \text{ one block north.}$$

With this in hand, everyone can agree on a set of navigational rules. Let's put them on our plot.



So we can see that in order to get to you, I first have to follow $4\hat{\mathbf{i}}$ and then $-2\hat{\mathbf{j}}$. In particular:

$$\mathbf{v} = 4\hat{\mathbf{i}} - 2\hat{\mathbf{j}}.$$

Once we all agree on a definition of $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$, we can represent the vector $\mathbf{v}$, just in terms of numbers 4 and 2. This is what the computer scientist would call:

$$\mathbf{v} = \begin{bmatrix} 4 \\ -2 \end{bmatrix}.$$

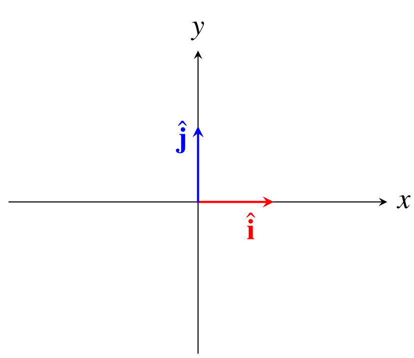
R It is important here that we all agree on $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$. One can imagine a city which is not grided parallel to north an south, and instead has *updown* and *downtown* directions. Then you may represent a vector using the coefficient of $\hat{\mathbf{j}}$ to describe how many units it goes uptown or downtown, but someone else might represent the vector using the coefficient of $\hat{\mathbf{j}}$ to represent how many units it goes north or south. The coordinates the computer scientist would write down would be different in each case. This is an important subtlety, but one that we will table until we are discussing *bases* and *change of bases*. For now, just remember that the coordinates that you might right down for a vector depend on a *choice* of $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$. This is one advantage of the physicist's perspective over the computer scientist's perspective.

The general setup (in 2-dimensions) is essentially the same

Definition 2.1.9 Given a cartesian coordinate system (that is, an *xy*-plane with coordinates), we can define the *standard basis vectors* $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$ to be the vectors:

$$\hat{\mathbf{i}} = \text{ one unit in the positive } x \text{ direction,}$$

$$\hat{\mathbf{j}} = \text{ one unit in the positive } y \text{ direction.}$$



Given any vector $\mathbf{v}$ in the coordinate plane, we can *resolve* it into its components:

$$\mathbf{v} = v_1\hat{\mathbf{i}} + v_2\hat{\mathbf{j}}.$$

Here v_1 is a *scalar*, representing how far $\mathbf{v}$ goes in the x-direction, and v_2 is a *scalar* representing how far $\mathbf{v}$ goes in the y-direction. They are unique.

Notation 2.3. *Given a vector in component form:*

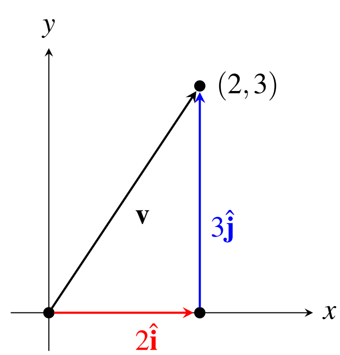
$$\mathbf{v} = v_1\hat{\mathbf{i}} + v_2\hat{\mathbf{j}},$$

we call the coefficients v_1 and v_2 the components *of* $\mathbf{v}$. *We can take these components and write* $\mathbf{v}$ *as a* column vector*:*

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}.$$

It is important to note that the coordinates of the column vector depend on $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$. In particular, we should think of this column vector as meaning, first do $v_1\hat{\mathbf{i}}$, then do $v_2\hat{\mathbf{j}}$.

■ **Example 2.2** Consider the displacement vector $\mathbf{v}$ from $(0,0)$ to $(2,3)$.



Therefore we see that:

$$\mathbf{v} = 2\hat{\mathbf{i}} + 3\hat{\mathbf{j}} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}.$$

■

Notice that the displacement vector from $(0,0)$ to (x,y) can always be written

$$x\hat{\mathbf{i}} + y\hat{\mathbf{j}} \qquad \text{or} \qquad \begin{bmatrix} x \\ y \end{bmatrix}.$$

Exercise 2.4 Consider the vectors $\mathbf{v_0}$ and $\mathbf{v_1}$ from Example 2.1. Write them both in component form and as column vectors. Use your result to decide whether they are equivalent. ■

Exercise 2.5 Let $P = (x_0, y_0)$, $Q = (x_1, y_1)$, and let $\mathbf{v}$ be the vector from P to Q. Write $\mathbf{v}$ in component form and as a column vector, in terms of the coordinates of P and Q. ■

Exercise 2.6 — Checkin 1. For problems 1 through 3, fix the following vectors:

$$\mathbf{v} = \hat{\mathbf{i}} - 2\hat{\mathbf{j}} \qquad \mathbf{w} = -2\hat{\mathbf{i}} + 3\hat{\mathbf{j}}.$$

1. Suppose you start at the origin, first travel along the displacement vector $\mathbf{v}$, and then travel along the displacement vector $\mathbf{w}$. Sketch your overall path to determine your endpoint. Use this to write down the vector $\mathbf{v}+\mathbf{w}$ in terms of its components.
2. Suppose you start at the origin and travel in the direction of $\mathbf{v}$, but twice as far. Sketch the path you travel to determine your endpoint. Use this to write down the vector $2\mathbf{v}$ in terms of its components.
3. Suppose you start at the origin and travel along the displacement vector $\mathbf{v}$, and then along the displacement vector $-\mathbf{w}$. Sketch the path you travel to determine your endpoint. Use this to write down the vector $\mathbf{v}-\mathbf{w}$ in terms of its components.
4. Now fix generic vectors

$$\mathbf{v} = v_1\hat{\mathbf{i}} + v_2\hat{\mathbf{j}} \quad \text{and} \quad \mathbf{w} = w_1\hat{\mathbf{i}} + w_2\hat{\mathbf{j}}$$

and let c be a scalar. Use the intuition developed on the previous page to write down formulas for $\mathbf{v}+\mathbf{w}$, $\mathbf{v}-\mathbf{w}$, and $c\mathbf{v}$.

■

2.2 February 2, 2023

Let's start with a couple of warmup questions.

Exercise 2.7 Let $P = (-1,2)$ and $Q = (3,4)$, and let $\mathbf{v}$ be the vector from P to Q. Write $\mathbf{v}$ as a column vector. ■

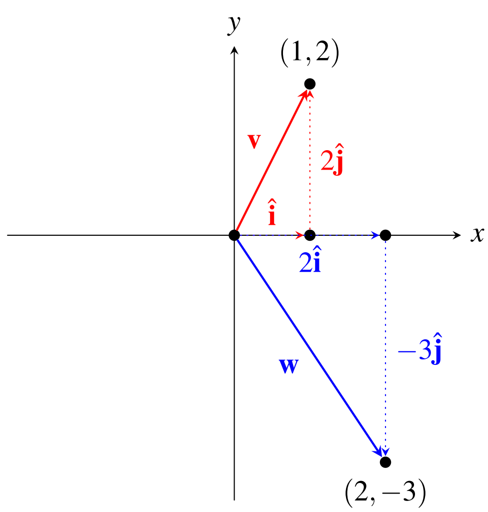
Exercise 2.8 Now let $P = (x_0, y_0)$ and $Q = (x_1, y_1)$, and let $\mathbf{v}$ be the vector from P to Q. Write $\mathbf{v}$ as a column vector (in terms of x_0, x_1, y_0, and y_1). ■

In Exercise 2.6, we explored what addition, subtraction, and scalar multiplication looked like when we put vectors in component form. Let's start by working through an example summarizing our observations.

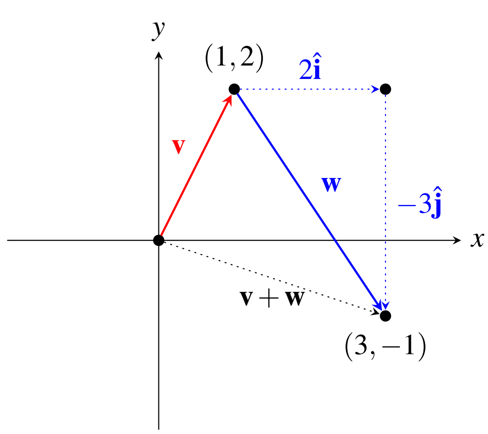
■ **Example 2.3** Consider the following two vectors given in component form:

$$\mathbf{v} = \hat{\mathbf{i}} + 2\hat{\mathbf{j}} \qquad \mathbf{w} = 2\hat{\mathbf{i}} - 3\hat{\mathbf{j}}.$$

Let's plot them both starting at the origin.



Let's add them together! To do this, we move the tail of $\mathbf{w}$ to the tip of $\mathbf{v}$, so that we can see what happens when you iterate them.



In particular, we have computed that: $\mathbf{v}+\mathbf{w} = 3\hat{\mathbf{i}} - \hat{\mathbf{j}}$. Let's compare this to adding together these two vectors *as computer scientists*. The column vector forms of $\mathbf{v}$ and $\mathbf{w}$ are (as in Notation 2.3):

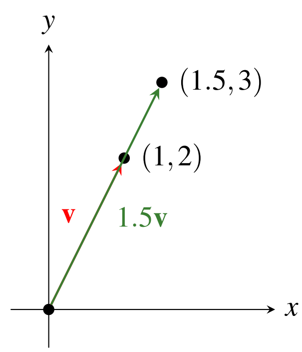
$$\mathbf{v} = \hat{\mathbf{i}} + 2\hat{\mathbf{j}} = \begin{bmatrix}1\\2\end{bmatrix} \qquad \mathbf{w} = 2\hat{\mathbf{i}} - 3\hat{\mathbf{j}} = \begin{bmatrix}2\\-3\end{bmatrix}.$$

Then if we add coordinatewise we get:

$$\begin{bmatrix}1\\2\end{bmatrix} + \begin{bmatrix}2\\-3\end{bmatrix} = \begin{bmatrix}1+2\\2-3\end{bmatrix} = \begin{bmatrix}3\\-1\end{bmatrix},$$

which translates to a the vector $3\hat{\mathbf{i}} - \hat{\mathbf{j}}$ in component form. We get the same answer! Let's say a word as to why this makes sense. Since $\mathbf{v}$ moves 1 unit in the x direction, and $\mathbf{w}$ moves 2 units in the x direction, then doing both moves $1+2=3$ units in the x direction. This is exactly what the computer scientist's approach does, adding together the x-coordinates. An identical argument explains why the $\hat{\mathbf{j}}$ components agree with both perspectives!

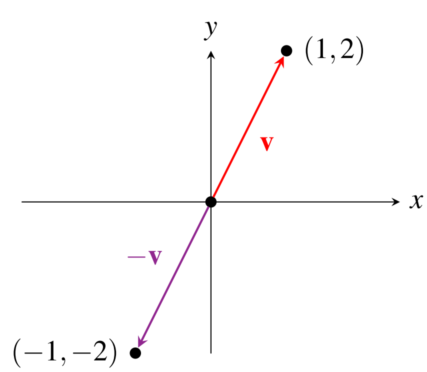
Now let's look at scalar multiplication, comparing $\mathbf{v}$ and $1.5\mathbf{v}$.



Since $1.5\mathbf{v}$ is 1.5 times longer, it (in particular), goes 1.5 times further in the x-direction, and 1.5 times further in the y-direction, so that:

$$1.5\mathbf{v} = 1.5(\hat{\mathbf{i}} + 2\hat{\mathbf{j}}) = 1.5\hat{\mathbf{i}} + 1.5 * 2\hat{\mathbf{j}} = \begin{bmatrix}1.5\\3\end{bmatrix}.$$

Let's also compare $\mathbf{v}$ and $-\mathbf{v}$.



Notice that $-\mathbf{v}$ goes the same distance along the x-axis $\mathbf{v}$ but in the opposite direction, and similarly along the y-axis. Therefore:

$$-\mathbf{v} = -\hat{\mathbf{i}} - 2\hat{\mathbf{j}} \qquad \text{or} \qquad -\begin{bmatrix}1\\2\end{bmatrix} = \begin{bmatrix}-1\\-2\end{bmatrix}.$$

■

This example suggests that to add a pair of vectors, we can merely add the $\hat{\mathbf{i}}$ components and the $\hat{\mathbf{j}}$ components, and to scale a vector, we can just scale the components. The second statement in fact gives us a formula for scaling column vectors as well: scaling a column vector can be acheived by scaling each entry. Let's record this:

Theorem 2.2.1 Fix two general vectors in component form (and as column vectors).

$$\mathbf{v} = v_1\hat{\mathbf{i}} + v_2\hat{\mathbf{j}} = \begin{bmatrix}v_1\\v_2\end{bmatrix} \qquad \mathbf{w} = w_1\hat{\mathbf{i}} + w_2\hat{\mathbf{j}} = \begin{bmatrix}w_1\\w_2\end{bmatrix}.$$

1. We can compute the sum using the formula:

$$\mathbf{v} + \mathbf{w} = (v_1 + w_1)\hat{\mathbf{i}} + (v_2 + w_2)\hat{\mathbf{j}}.$$

In particular, the physicist's perspective on vector addition agrees with the computer scientist's formula:

$$\begin{bmatrix}v_1\\v_2\end{bmatrix} + \begin{bmatrix}w_1\\w_2\end{bmatrix} = \begin{bmatrix}v_1 + w_1\\v_2 + w_2\end{bmatrix}.$$

2. Let c be any constant. We can compute the scalar multiple $c\mathbf{v}$ using the formula:

$$c\mathbf{v} = (cv_1)\hat{\mathbf{i}} + (cv_2)\hat{\mathbf{j}}.$$

From this we can derive a formula for scaling a column vector which agrees with the physicist's perspective on vector scaling:

$$c\begin{bmatrix}v_1\\v_2\end{bmatrix} = \begin{bmatrix}cv_1\\cv_2\end{bmatrix}.$$

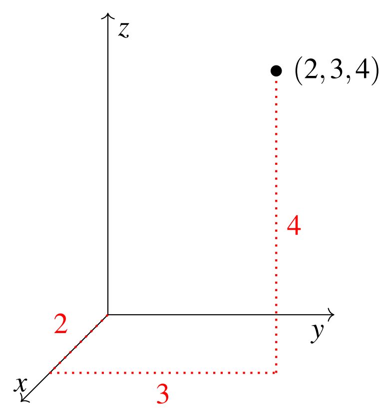
Proof. The argument for why Theorem 2.2.1 is true is essentially identical to the arguments appearing in Example 2.3 For example, $\mathbf{v}$ goes v_1 units in the x-direction, and $\mathbf{w}$ goes w_1 units in the x-direction, so doing both in succession results in an overall movement of $v_1 + w_1$ units in the x-direction. See if you can adapt the remaining arguments from Example 2.3 to the general setup. ■

This gives a complete answer to Question 2.1. To concretely interpret what *adding points* means, we should think about the points as vectors. Then the physicist's perspective on vector addition gives us a concrete way to think about the addition geometrically.

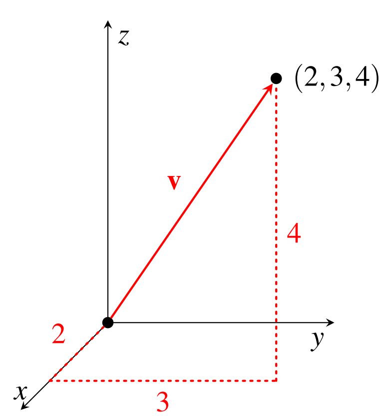
Exercise 2.9 Adopt the notation of Exercise 2.7. Suppose the tail for $-2.5\mathbf{v}$ is placed at P, what are the coordinates of its tip? ■

2.2.1 Extending this to 3 dimensions...and beyond

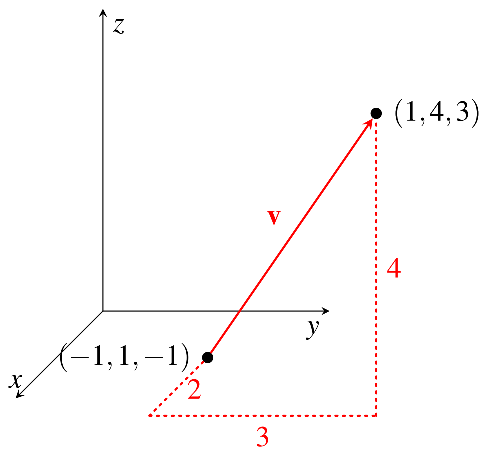
Everything described so far also works in 3 dimensions with a few cosmetic adjusments. Much like before, we can take the physicist's perspective or the computer scientist's perspective, and use coordinates to connect the two. The main difference is that we pass from our 2 dimensional xy-plane, to 3 dimensional space, whose points are now specified by 3 coordinates (x, y, z). We will usually adopt the convention of viewing the x axis as coming forward out of the page, the y axis as going to the right, and the vertical z-axis as going up. Here, for example, is how you plot the point $(2, 3, 4)$.



The Physicist's Perspective in 3d: As in 2d, the physicist's vector is a quantity $\mathbf{v}$ specifying a *magnitude* and *direction*, though the direction is now in 3d. We can again represent a vector by an arrow in 3-dimensional space: it should point in the *direction* of $\mathbf{v}$ and its length should be the *magnitude* of $\mathbf{v}$. Given two points P and Q in 3-space, we can obtain the vector from P to Q as an arrow connecting P and Q. For example, here we draw the vector $\mathbf{v}$ connecting $(0, 0, 0)$ and $(2, 3, 4)$.



As before *the vector is the arrow*, not the location. If we put the tail of $\mathbf{v}$ somewhere else, the vector remains the same.



In this way, we can think about a vector as something we can *apply* to a point. For example $\mathbf{v}$ takes $(0,0,0)$ to $(2,3,4)$, and also takes $(-1,1,-1)$ to $(1,4,3)$.

With this perspective, the definitions of addition, subtraction, and scalar multiplicate defined in Definitions 2.1.4, 2.1.5, and 2.1.7 go through unchanged. For example, to add two vectors $\mathbf{v}$ and $\mathbf{w}$, we take the vector obtained by *first applying* $\mathbf{v}$*, and then applying* $\mathbf{w}$.

The computer scientist's perspective in 3d: For a computer scientist, a 3-dimensional vector is an array of 3 numbers, aligned vertically:

$$\mathbf{v} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}.$$

The can be added and scaled coordinatewise:

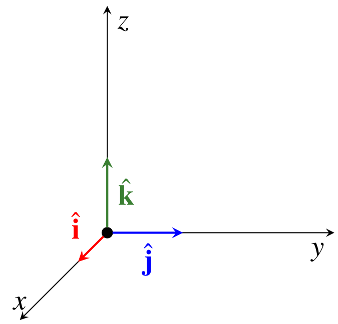
$$\begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix} + \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} = \begin{bmatrix} x_0 + x_1 \\ y_0 + y_1 \\ z_0 + z_1 \end{bmatrix} \qquad \text{and} \qquad c \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} cx \\ cy \\ cz \end{bmatrix}.$$

Combining the two perspectives in 3d: To go between these two perspectives in 3 dimensions, we follow a similar approach to what we did in 2d. To start, we define the standard unit vectors by their directions and magnitudes:

$$\hat{\mathbf{i}} = \text{one unit in the positive } x \text{ direction}$$

$$\hat{\mathbf{j}} = \text{one unit in the positive } y \text{ direction}$$

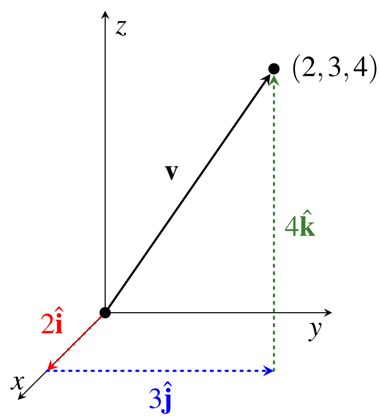
$$\hat{\mathbf{k}} = \text{one unit in the positive } z \text{ direction}$$



We can now express any vector in terms of $\hat{\mathbf{i}}, \hat{\mathbf{j}}, \hat{\mathbf{k}}$. For example, the vector from $(0,0,0)$ to $(2,3,4)$ can be expressed as

$$\mathbf{v} = 2\hat{\mathbf{i}} + 3\hat{\mathbf{j}} + 4\hat{\mathbf{k}},$$

because it goes 2 units in the x direction (achieved by applying $\hat{\mathbf{i}}$ twice), 3 units in the y direction (acheived by applying $\hat{\mathbf{j}}$ 3 times), and 4 units in the z direction (achieved by applying $\hat{\mathbf{k}}$ four times).



To get this in the form of a column vector, just arrange the coefficients of $\hat{\mathbf{i}}, \hat{\mathbf{j}}$, and $\hat{\mathbf{k}}$ vertically.

$$2\hat{\mathbf{i}} + 3\hat{\mathbf{j}} + 4\hat{\mathbf{k}} = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}.$$

Exercise 2.10 Let $\mathbf{w}$ be the vector from $(5,-3,9)$ to $(1,1,2)$. Write $\mathbf{w}$ as a column vector. ■

We can generalize 2.8 to 3d as well.

Proposition 2.2.2 Let $P = (x_0, y_0, z_0)$, $Q = (x_1, y_1, z_1)$, and let $\mathbf{w}$ be the vector from P to Q. Then $\mathbf{w}$ can be written as a column vector in the following way:

$$\mathbf{w} = \begin{bmatrix} x_1 - x_0 \\ y_1 - y_0 \\ z_1 - z_0 \end{bmatrix}.$$

Exercise 2.11 Explain why Proposition 2.2.2 is true. ■

Higher dimensions: The computer scientist's approach to vectors generalizes to higher dimensions. To the computer scientist, a 2d vector was an array of 2 numbers, and a 3d vector was an array of 3 numbers. Following the pattern, an n-dimensional vector should be an array of n-numbers.

Definition 2.2.1 — Higher Dimensional Vectors: The Computer Scientist's Approach. Let n be a positive integer. A *n-dimensional column vector* is an array of n numbers, arranged vertically:

$$\mathbf{v} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

The can be added and scaled coordinatewise:

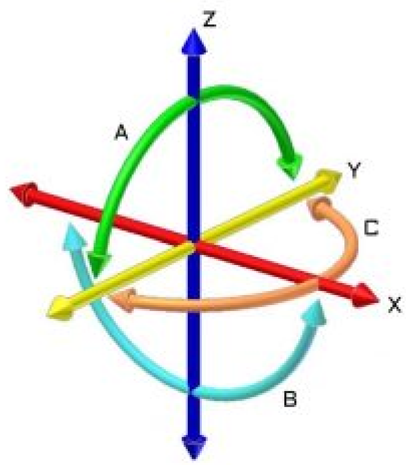
$$\begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_0 + y_0 \\ x_1 + y_1 \\ \vdots \\ x_n + y_n \end{bmatrix} \quad \text{and} \quad c \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} cx_0 \\ cx_1 \\ \vdots \\ cx_n \end{bmatrix}.$$

At first we might be wary of straying beyond 3-dimensions, especially our interest is limited to applications within the physical world. Nevertheless, there are many reasons one might be interested in having more than 3 axes of data in a vector, even for things that get modelled in the physical world. Let's see a couple of examples.

■ **Example 2.4 — 5-axis CNC drilling.** Where a drilling machine drills a hole depends not only on the location of the tip of the drill bit, but also on its orientation in space (what direction is the drill bit pointing?). For these reasons, programmable drills control the movement of the drilling head using a five dimensional vector:

$$\mathbf{v} = \begin{bmatrix} x \\ y \\ z \\ \theta \\ \phi \end{bmatrix}.$$

The first three entries, x, y, z, control the location of the machine-head in space, while θ and ϕ control rotation, in order to angle the drill to the necessary position. In particular, θ controls rotation in the xy-plane (about the z-axis), while ϕ controls rotatoin in the yz-plane (about the x-axis).



34

The drill can be controlled much like the autonomous vehicles of Section **??**, by being sent vectors. For example, if you would like to send the drill 1 cm in the x-direction and 2 cm down, without changing rotation, you may send it the vector:

$$\mathbf{v} = \begin{bmatrix} 1 \\ 0 \\ -2 \\ 0 \\ 0 \end{bmatrix}.$$

If the drill is pointing down and you would like to point it up, and lift it 5 cm, you could then send it:

$$\mathbf{w} = \begin{bmatrix} 0 \\ 0 \\ 5 \\ 0 \\ 180 \end{bmatrix}.$$

To do these in succession, we would first do $\mathbf{v}$ (say to drill a hole), and then do $\mathbf{w}$ (say, to lift it and drill another hole from below), which a physicist may say should add the vectors. And indeed, tracing the overall movement of the printhead would have it end up following the vector

$$\begin{bmatrix} 1 \\ 0 \\ 3 \\ 0 \\ 180 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ -2 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 5 \\ 0 \\ 180 \end{bmatrix} = \mathbf{v} + \mathbf{w}.$$

In particular, the computer scientist's definition of vector addition has a physical interpretation as well! ■

This example shows that adding even 5-dimensional vectors can concretely understood once we specify the data that the entries store. The following exercise shows you can do something similar for scalar multiplication.

[3]From Autodesk: `https://blogs.autodesk.com/inventor/understanding-process-5-axis-machining/`
[4]TODO: Make a better graphic

Exercise 2.12 Suppose you have a 5-axis drill, at its *home position*. You program it to move along the 5-dimensional $\mathbf{v}$ to drill a hole. If you'd like it to return to home position, what vector should you ask it to follow? ■

■ **Example 2.5 — A 171,000 dimensional data set.** The study of *stylometry* studies variations in literary style using statistical analysis, and part of this work involves measuring how frequently certain words appear. There are approximately 171,000 words in the english language, so if you would like to record how many times each word appears in a given book, you could do so in a vector:

$$\mathbf{v} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_{171,000} \end{bmatrix},$$

where w_1 counts the number of times the first word in the dictionary appears, and w_2 counts how many times the second word in the dictionary appears, and so on. Adding the vectors associated to a 2 different books will count how many times each word appears in both. Having a mathematical model grounded in geometry allows one to use linear algebraic techniques to study questions about variations in word frequency. One could also use scalar multiples to weight the importance or prominence of certain sources over others. ■

2.3 Homework 2

Exercise 2.13 Let $P = (1,2), Q = (2,1)$ and $R = (-3,-1)$. Let $\mathbf{v}$ be the vector from P to Q, and let $\mathbf{w}$ be the vector from Q to R.

1. Do you agree or disagree with the following statement:

   $$\mathbf{v}+\mathbf{w} \text{ is the vector from } P \text{ to } R.$$

   Justify your answer by drawing a picture.
2. Write $\mathbf{v}, \mathbf{w}$ and $2\mathbf{v}-\mathbf{w}$ in component form, and as column vectors.
3. Draw $2\mathbf{v}-\mathbf{w}$ on the plane, with its tail starting at R. What are the coordinates of where the tip lands?

■

Exercise 2.14 Write the standard 3-dimensional unit vectors $\hat{\mathbf{i}}$, $\hat{\mathbf{j}}$, and $\hat{\mathbf{k}}$ as column vectors. ■

Exercise 2.15 To a physicist, a vector was a quantity with *magnitude* and *direction*. We saw how to turn such a quantity into a *column vector*. Let's start going in the other direction.

1. Let $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$ be the standard unit vectors in 2d. What are their magitudes: $||\hat{\mathbf{i}}||$ and $||\hat{\mathbf{j}}||$? (*Recall:* When representing a vector with an arrow magnitude, is the length of the arrow.)
2. Let $\mathbf{v}$ and $\mathbf{w}$ be the same vectors from Problem 1. Use the Pythagorean theorem to compute their magnitudes: $||\mathbf{v}||$ and $||\mathbf{w}||$.

3. Consider a general column vector:

$$\mathbf{u} = \begin{bmatrix} x \\ y \end{bmatrix}.$$

Write formula for the magnitude $||\mathbf{u}||$ in terms of x and y. Explain your reasoning.

An artist is trying to mix the perfect color. One way to measure colors is in RGB, where:

r units of red
g units of green
\+ b units of blue

a rich spectrum of colors

The artist has two pigments, which they are hoping to mix together into paint to try and get the color they want.

Pigment X: Adding 1 ounce of Pigment X to paint adds:

1 unit of red, 2 units of green, 3 units of blue

Pigment Y: Adding 1 ounce of Pigment Y to paint adds:

7 units of red, 5 units of green, 2 units of blue

Exercise 2.16 The effect of adding an ounce **pigment X** can be represented by a vector $\mathbf{x}$ and the effect of adding **pigment Y** by a vector $\mathbf{y}$.

1. Give a concrete interpretation of what the vector $\mathbf{x}+\mathbf{y}$ represents.
2. Express $\mathbf{x}$ and $\mathbf{y}$ as column vectors.

Exercise 2.17 The artist wants to create a color the called **fancy gold**, consisting of:

24 units of red, 21 units of green, 15 units of blue

1. Write **fancy gold** as a column vector $\mathbf{f}$.
2. Write an equation using only vector addition and scalar multiplication which can be interpreted as saying:

   a units of Pigment X and b units of Pigment Y produce fancy gold.

3. Translate the single vector equation from part 2 into a system of linear equations. Solve this system of equations to see how many ounces of **Pigment X** and how many ounces of **Pigment Y** the artist needs to mix to get **fancy gold**.

Exercise 2.18 The artist also whats to create **super green**, which consists of:

3 units of red, 90 units of green, 3 units of blue.

1. Write **super green** as a column vector **s**.
2. Write an equation using only vector addition and scalar multiplication which can be interpreted as saying:

   a units of Pigment X and b units of Pigment Y produce super green.

3. Translate the single vector equation from part 2 into a system of linear equations. Try to solve this system of equations to determine if it is possible to mix **super green** from **Pigment X** and **Pigment Y**.

■

Exercise 2.19 The artist got their hands on a fancy new pigment called **Greenifier**, which, perhaps suprisingly, doesn't actually contain any green. Instead, it works by absorbing red and blue light. The net effect of adding 1 ounce of **Greenifier** is

subtract 5 units of red, subtract 5 units of blue

1. Write **Greenifier** as a column vector **g**.
2. Write an equation using only vector addition and scalar multiplication which can be interpreted as saying:

   a units of Pigment X, b units of Pigment Y, and c units of Greenifier produce super green.

3. Translate the single vector equation from part 2 into a system of linear equations. Solve this system of equations to see how many ounces of **Pigment X**, how many ounces of **Pigment Y**, and how many ounces of **Greenifier** the artist needs to mix to get **fancy gold**.

■

2.4 February 7, 2023

Let's start by unpacking Exercises 2.17 and 2.18, and in doing so, get a quick review of how we solve systems of linear equations. Our goal in Exercise 2.17 is to mix fancy gold from the two pigments, Pigment X and Pigment Y. How much of each pigment do we add?

To introduce mathematical notation, we let the vector $\mathbf{x}$ represent the effect of adding an ounce of Pigment X and $\mathbf{y}$ represent adding an ounce of Pigment Y. In particular, if we add a ounces of Pigment X and b ounces of Pigment Y, the overall effect on color can be represented by the vector:

$$a\mathbf{x} + b\mathbf{y}.$$

If we let $\mathbf{f}$ represent the color fancy gold, we are therefore looking for integers a and b so that:

$$a\mathbf{x} + b\mathbf{y} = \mathbf{f}. \tag{2.1}$$

Plugging in the column vectors from Exercises 2.16.2 and 2.17.1 turns this into:

$$a\begin{bmatrix}1\\2\\3\end{bmatrix} + b\begin{bmatrix}7\\5\\2\end{bmatrix} = \begin{bmatrix}24\\21\\15\end{bmatrix}.$$

But we know how to scale and add column vectors. So this becomes:

$$\begin{bmatrix}a+7b\\2a+5b\\3a+2b\end{bmatrix} = \begin{bmatrix}24\\21\\15\end{bmatrix}$$

Now we can just remove the matrix brackets and obtain:

$$a+7b = 24 \tag{2.2}$$
$$2a+5b = 21 \tag{2.3}$$
$$3a+2b = 15, \tag{2.4}$$

which is a system of linear equations with 3 equations and 2 unknowns. We want to find values a and b so that *all 3 equations hold*! Many readers have probably solved systems of linear equations before, but it may have been a long time, so let's briefly review how one might solve this. We will establish a completely systematic way of doing this in Section **??**, but for now, let's just follow our noses. A first step can be to solve for a in Equation (2.2):

$$a = 24 - 7b. \tag{2.5}$$

We can now plug this value of a into Equation (2.3). The left hand side is:

$$2a + 5b = 2(24 - 7b) + 5b = 48 - 9b.$$

So Equation (2.3) turns into:

$$48 - 9b = 21.$$

We can therefore solve for $b = 3$. Plugging this value into Equation (2.5) gives:

$$a = 24 - 7(3) = 3.$$

So in order for the first Equations (2.2) and (2.3) to hold, we need $a = 3$ and $b = 3$. What about Equation (2.4)? Plugging in $a = 3$ and $b = 3$ gives:

$$3a + 2b = 3 * 3 + 2 * 3 = 15.$$

So we have determined that $a = 3$ and $b = 3$ solves all three equations in our system. Translating back into our vector equations we have:

$$3\mathbf{x} + 3\mathbf{y} = 3\begin{bmatrix}1\\2\\3\end{bmatrix} + 3\begin{bmatrix}7\\5\\2\end{bmatrix} = \begin{bmatrix}3*1+3*7\\3*2+3*5\\3*3+3*2\end{bmatrix} = \begin{bmatrix}24\\21\\15\end{bmatrix} = \mathbf{f}.$$

So $a = 3$ and $b = 3$ also form solutions to the vector equation (2.1)! Translating back into plain english gives:

3 ounces of Pigment X and 3 ounces of Pigment Y results in Fancy Gold

One main takeaway from this example is that, the *single* linear equation of vectors (2.1), corresponds exactly to the *system* of linear equation*s* (2.2), (2.3), and (2.3).

■ **Slogan 2.2** Solving *one* linear equation of vectors is the same as solving *a system* of linear equations.

What about Super Green? The question becomes one of solving the single vector equation:

$$a\mathbf{x} + b\mathbf{y} = \mathbf{s},$$

which we can unpack to the system of equations:[5]

$$\begin{aligned} a + 7b &= 3 \\ 2a + 5b &= 90 \\ 3a + 2b &= 3 \end{aligned}$$

Following the same script as before, we can use the first two equations to solve for $a = \frac{199}{3}$ and $b = \frac{-28}{3}$ (the reader fill should in the missing steps!). This already presents a problem, how can we add a negative amount of Pigment Y? Let's ignore this for a moment, and pretend the artist had some tool to remove pigment. Would this solve it? Well, we see that to get 3 units of red and 90 units of green we are forced to add $\frac{199}{3}$ ounces of Pigment X and to (somehow) subtract $\frac{28}{3}$ ounces of Pigment Y. But we haven't even checked if we have the correct amount of blue. This is measured in the third equation, let's see if our values of a and b work:

$$3a + 2b = 3 \times \left(\frac{199}{3}\right) + 2\left(\frac{-28}{3}\right) = \frac{541}{3}.$$

That is way more than 3 units of blue! So our attempt at making Super Green is going to come out more blue than anything else, and we will fail. What we've encountered here is the following fact:

[5] the reader should fill in the details here!

sometimes a system of linear equations has no solutions. Indeed, if there are more equations than unknowns, this is rather common: notice we were able to find an a and b that satisfied the first two equations, but the third equation that caused us trouble. When we translate Exercise 2.19 into a system of equations, we have a third unknown, giving us added flexibility we can exploit. For now, let's just record the following observation:

■ **Slogan 2.3** Sometimes a linear equation of vectors can have no solution.

For the rest of this section, we will put these observations into a more general context, introducing some language with which we can more concisely describe these observations.

2.4.1 Linear Combinations and Spans: A First Pass

In our color mixing example, we were studying which colors we could mix from Pigment X and Pigment Y. Assigning variables a and b to the amount of each pigment added, we were able to translate this question into one which studies whether we can write a vector in the form:[6]

$$a\mathbf{x}+b\mathbf{y},$$

In a more general language, we are asking to know which vectors can be written *in terms of* $\mathbf{x}$ *and* $\mathbf{y}$ with the operations of scalar multiplication and addition. A vector that can be written in that form is called a *linear combination* of $\mathbf{x}$ and $\mathbf{y}$.

Definition 2.4.1 Let $\mathbf{u}$ and $\mathbf{v}$ be vectors. A *linear combination* of $\mathbf{u}$ and $\mathbf{v}$ is a vector $\mathbf{w}$ which can be written:

$$\mathbf{w}=c\mathbf{u}+d\mathbf{v},$$

for constants c and d.

■ **Example 2.6** The vector $\mathbf{f}$ from Exercise 2.17 is a linear combination of $\mathbf{x}$ and $\mathbf{y}$ because:

$$\mathbf{f}=3\mathbf{x}+3\mathbf{y}.$$

■

■ **Question 2.3** Is the vector $\mathbf{s}$ a linear combination of $\mathbf{x}$ and $\mathbf{y}$.

We already encountered linear combinations of vectors before this color example. Indeed, it was the tool we used to translate between the *physicist's perspective* on vectors to the *computer scientist's perspective*.

■ **Example 2.7** The column vector:

$$\mathbf{v}=\begin{bmatrix}-11\\9\end{bmatrix},$$

is a linear combination of the standard unit vectors $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$ because

$$\mathbf{v}=-11\hat{\mathbf{i}}+9\hat{\mathbf{j}}.$$

■

[6] In fact, one could interpret the example as having us restrict to positive values of a and b, but we will not make that restriction.

■ **Question 2.4** Let $\mathbf{w}$ be a 2-dimension vector. Is $\mathbf{w}$ a linear combination of $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$?

Indeed, this is what the column vector *means*. The column vector:

$$\mathbf{w} = \begin{bmatrix} c \\ d \end{bmatrix}$$

exactly says that $\mathbf{w}$ can be acheived by scaling $\hat{\mathbf{i}}$ by c, and scaling $\hat{\mathbf{j}}$ by d. That is $\mathbf{w} = c\hat{\mathbf{i}} + d\hat{\mathbf{j}}$.

We can also consider linear combinations of more than 2 vectors. We have already done so to represent 3-dimensional vectors as column matrices, and also in Exercise 2.19.

Definition 2.4.2 Let $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$ be a collection of n-vectors (of the same dimension). A linear combination of $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$ is a vector $\mathbf{w}$ which can be written in the form:

$$\mathbf{w} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n,$$

for constants $c_1, c_2, \cdots, c_n$.

■ **Example 2.8** Exercise 2.19 asks whether the vector $\mathbf{s}$ is a linear combination of $\mathbf{x}, \mathbf{y}$ and $\mathbf{g}$ where $\mathbf{g}$ represents the effect of adding one ounce of greenify. ■

■ **Example 2.9** Every 3d vector is a linear combination of $\hat{\mathbf{i}}, \hat{\mathbf{j}}$, and $\hat{\mathbf{k}}$. ■

■ **Question 2.5** Consider the column vector:

$$\mathbf{w} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}.$$

If I know that $\mathbf{w}$ is a linear combination of $\hat{\mathbf{i}}$ and $\hat{\mathbf{k}}$, what can we say about its middle entry b?

■ **Example 2.10** Let $\mathbf{v}$ be a single vector. A linear combination of $\mathbf{v}$ is a vector of the form $c\mathbf{v}$ for some constant c. That is, a linear combination of $\mathbf{v}$ is the same as a multiple of $\mathbf{v}$. ■

As we are starting to see, many questions in linear algebra boil down to variants of the following type of question:

■ **Question 2.6** When is one vector a linear combination of another collection of vectors?

We will see many variations of this question, so let's introduce some terminology to simplify the exposition.

Definition 2.4.3 Let $\mathbf{u}$ and $\mathbf{v}$ be a pair of vectors. The *span* of $\mathbf{u}$ and $\mathbf{v}$ is the collection of vectors that can be expressed as a linear combination of $\mathbf{u}$ and $\mathbf{v}$.

$$\text{span}\{\mathbf{u}, \mathbf{v}\} = \{\text{The collection } c\mathbf{u} + d\mathbf{v} \text{ for constants } c \text{ and } d\}.$$

■ **Example 2.11** Returning to our color mixing example: $\text{span}\{\mathbf{x}, \mathbf{y}\}$ is the collection of colors that can be mixed from Pigment X and Pigment Y. In particular, $\mathbf{f}$ is in $\text{span}\{\mathbf{x}, \mathbf{y}\}$, while $\mathbf{s}$ is not. ■

■ **Example 2.12** Denote the entire collection of 2d vectors by $\mathbb{R}^2$. Then $\mathbb{R}^2 = \text{span}\{\hat{\mathbf{i}}, \hat{\mathbf{j}}\}$. ■

We can also consider the spans of more than 2 vectors.

Definition 2.4.4 Let $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$ be a collection of n-vectors (of the same dimension). The span of $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$ is the collection of vectors that are linear combinations of $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$.

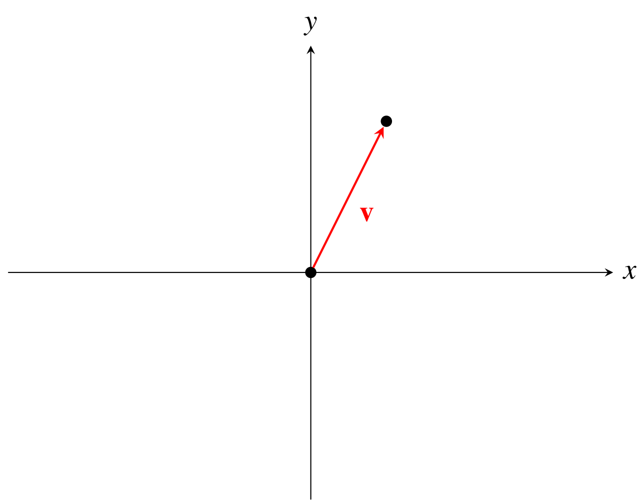
$$\text{span}\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\} = \{ \text{ the collection } c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n \text{ for constants } c_1, c_2, \cdots, c_n\}.$$

■ **Example 2.13** Exercise 2.19 asks whether the vector $\mathbf{s}$ is in span$\{\mathbf{x}, \mathbf{y}, \mathbf{g}\}$. ■

Visualizing Spans

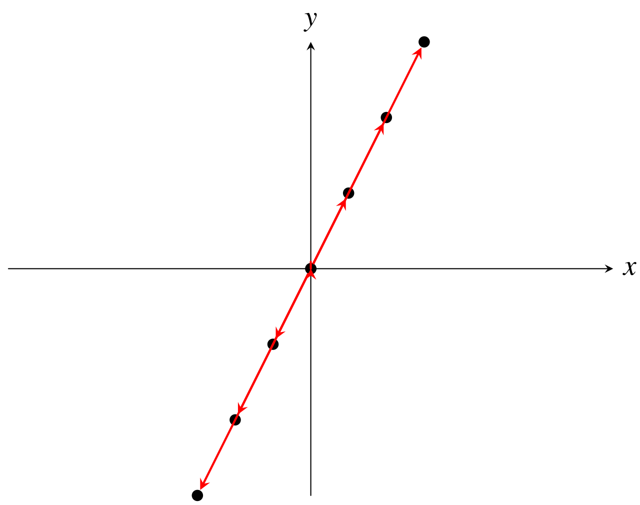
Spans can be a tricky thing to intuit on a first pass through linear algebra, but they are also an important and fundamental part of the theory. We will delay the general practice of explicitly computing spas until Section **??**, when we have developed a bit of matrix theory.[7] But if we can develop a visual intuition of what spans look like, we recognize where they arise and get some intuition about what they mean before doing any computations.

■ **Example 2.14 — The Span of a Single Vector.** Let $\mathbf{v}$ be a nonzero vector in $\mathbb{R}^2$:

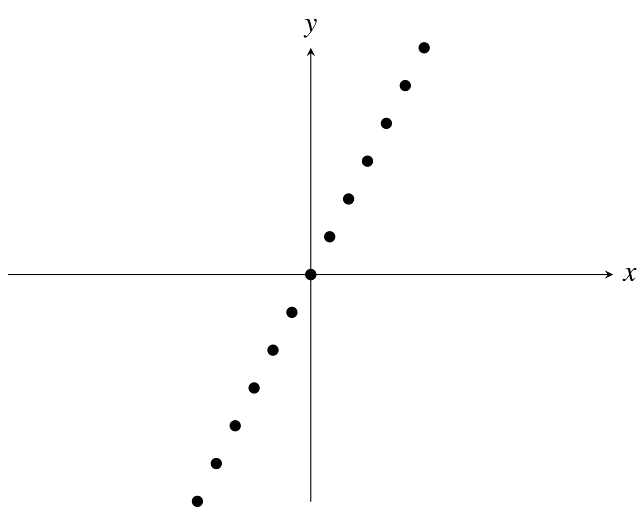


The span of $\mathbf{v}$ is the collection of linear combinations of $\mathbf{v}$. This is precisely the collectoin of multiples of $\mathbf{v}$ (cf. Example **??**). Let's plot a few of these. In particular, we'll plot: all the vectors $-1.5\mathbf{v}, -\mathbf{v}, -0.5\mathbf{v}, 0\mathbf{v}, .5\mathbf{v}, 1\mathbf{v}, 1.5\mathbf{v}$, with their tails starting at the origin.

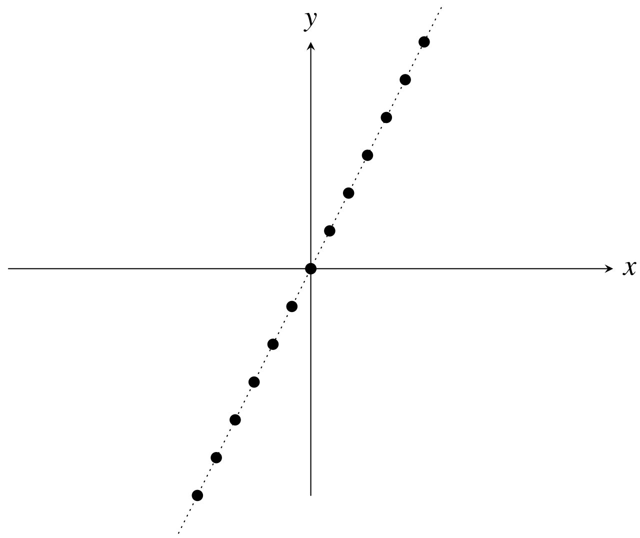
[7] though notice that Exercises 2.17 through 2.19 suggest a relationship with systems of linear equations



As we start to fill in more and more multiples, the arrows start to crowd the picture, so let's just draw the where the tips lie.



As the picture fills in, we start to see that the tips of all the multiples of **v** trace out a straight line.

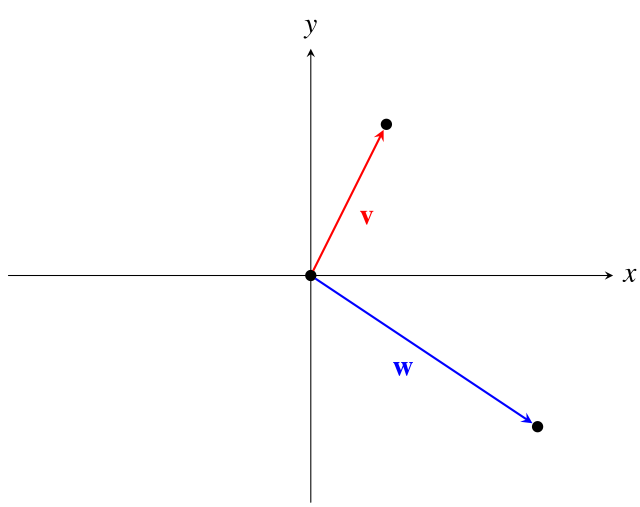


■

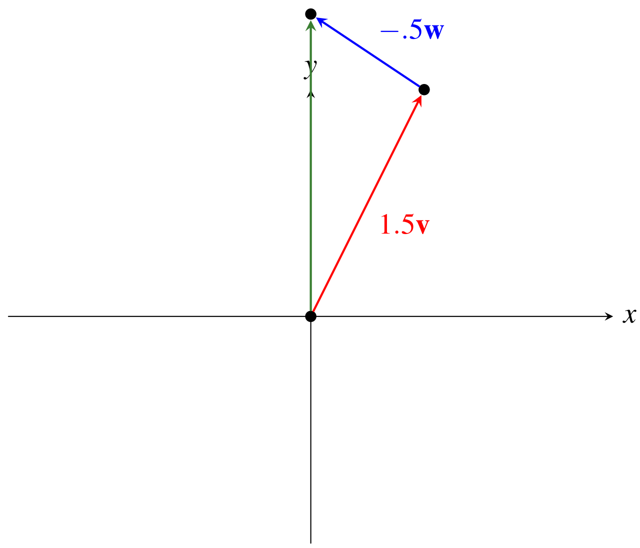
In the previous example, we deduced (without any explicit computations) some geometric facts about the span of a single vector!

■ **Slogan 2.4** The span of a single vector is a straight line through the origin.

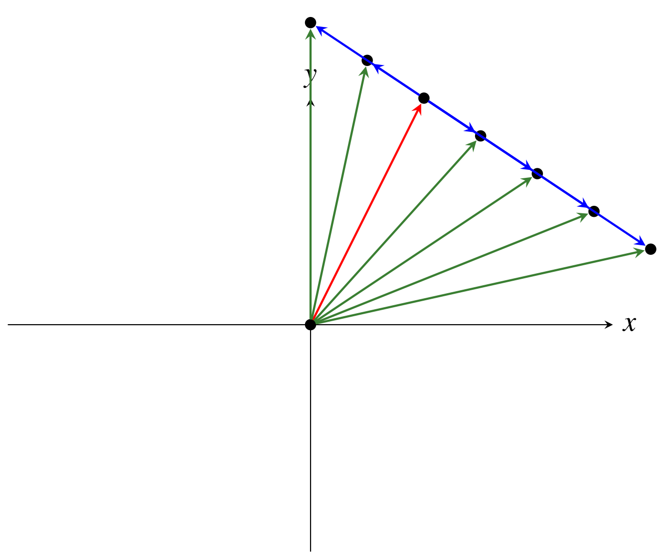
■ **Example 2.15 — The Span of 2 Vectors.** Let $\mathbf{v}$ and $\mathbf{w}$ be 2 nonzero vectors in $\mathbb{R}^2$. For now let's assume that they aren't parallel.



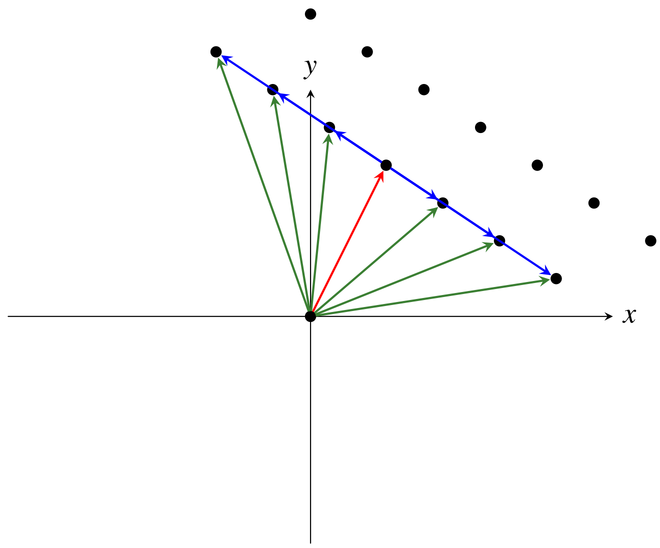
The span of $\mathbf{v}$ and $\mathbf{w}$ is the collection of vectors that can be written $c\mathbf{v}+d\mathbf{w}$. For example, $1.5\mathbf{v}-0.5\mathbf{w}$:



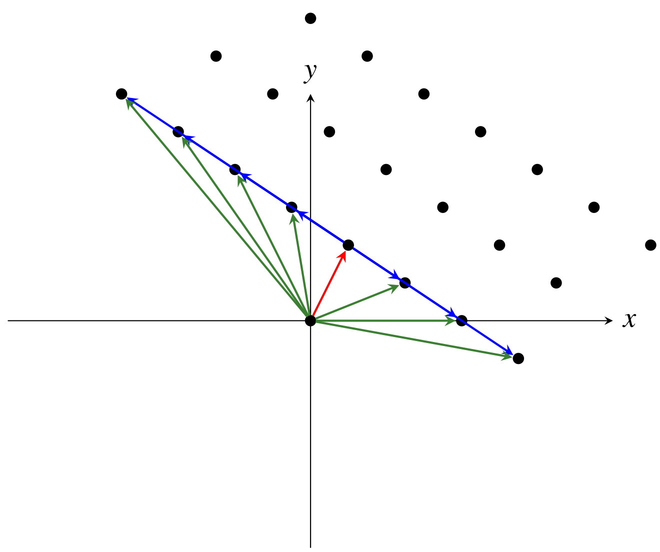
Let's plot a few more, fixing $c = 1.5$. That is, let's look at vectors of the form $1.5\mathbf{v} + d\mathbf{w}$ for various values of d.



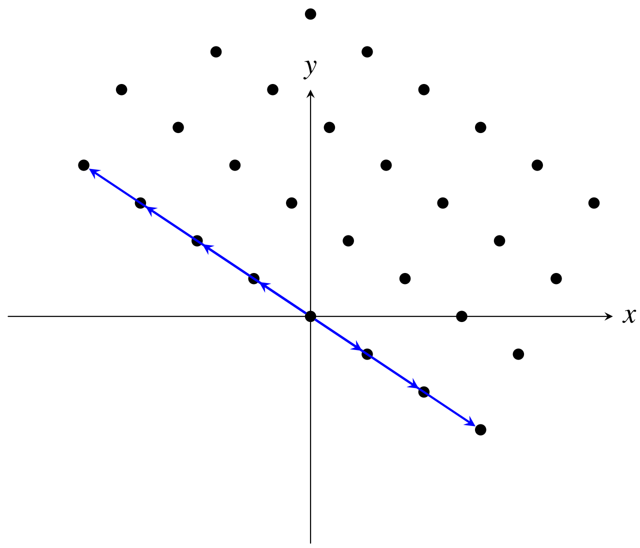
We can do the same, for $c = 1$.

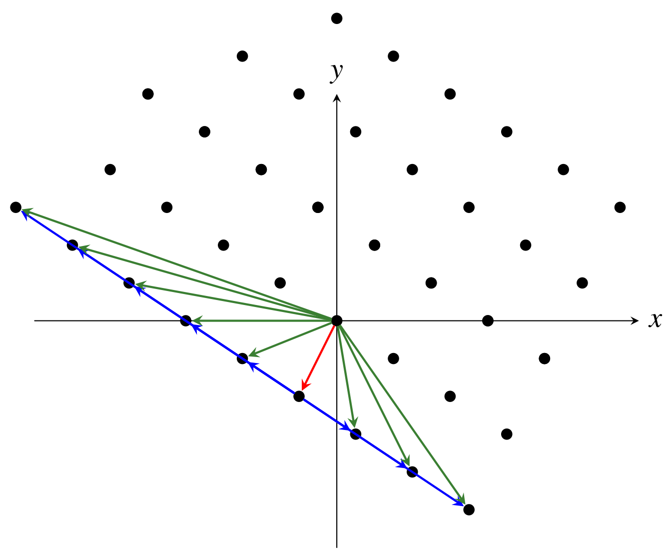


Keeping the points the tip has hit marked, let's also include $c = .5$



Before moving on, Let's throw in $c = 0$ (which are just multiples of $\mathbf{w}$) and $c = -.5$.





It starting to look like we can get anywhere in the plane. Indeed, imagine having two dials, one which modifies c and another which modifies d. Then c is changing the length of the red arrow, and d and changing the length of the blue arrow, tacked on to the tip of the red arrow. Then just by turning these dials, we should be able to get anywhere we want. This actually becomes even more clear when looking at an animation.[8] ■

In the previous example, we deduced (without any explicit computations) some geometric facts about the span of a pair of vectors!

■ **Slogan 2.5** The span of 2 nonzero and nonparallel vectors in $\mathbb{R}^2$ is all of $\mathbb{R}^2$.

[8]For example `https://3b1b-posts.us-east-1.linodeobjects.com//content/lessons/2016/span/linear_combination.mp4`

Exercise 2.20 — Checkin 2. A group of exo-ecologists are experimenting with mixtures of gasses for a greenhouse in space. The gasses is a mix of Nitrogen, Oxygen, Carbon Dioxide, and Argon. They have 3 gas mixtures, whose compositions are given below (measured by mass).

- **Gas X:** 80% Nitrogen and 20% Oxygen.
- **Gas Y:** Pure Oxygen
- **Gas Z:** 60% Nitrogen, 30% Oxygen, 2% Carbon Dioxide, and 8% Argon

1. We represent a mixture containing a grams of Nitrogen, b grams of Oxygen, c grams of Carbon Dioxide, and d grams of Argon, by the column vector
$$\begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}.$$
Let $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ be the vectors representing the a mixture of *one gram* of gasses X, Y, and Z respectively. Write $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ as column vectors.
2. In plain english, describe what $\operatorname{span}\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$ represents.
3. The exo-ecologists would like a gas mixture $\mathbf{e}$ modelled on the concentration of gasses in earth's atmosphere, which they call *the eartlike mixture*. Fill in the blanks in the following sentence so that it means: *The earthlike mixture can be mixed from gasses X, Y, and Z.*

   The vector(s) ___________ is/are a linear combination of the vector(s) ___________
4. Do you think the exo-biologists can mix pure Carbon Dioxide from their gas mixtures? Why or why not?

■

2.4.2 Homework 3

Exercise 2.21 We introduced vectors because in HW1 Problem 6 it looked like linear maps played well with adding points, suggesting that it might be more meaningful to think about the inputs and outputs as vectors rather than points. Let's try to make this more precise. To do this, we will shift our perspective slightly by letting $\mathbb{R}^2$ be the collection of 2-dimensional *vectors*:
$$\begin{bmatrix} x \\ y \end{bmatrix}.$$
A function $f : \mathbb{R}^2 \to \mathbb{R}^2$ then is a rule:
$$f\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} u \\ v \end{bmatrix},$$
and can therefore be described by an equation for u and one for v. I want to emphasize that this is merely a shift of perspective (and notation), but the content is the same as in HW1.

1. Let $\mathbf{w} = 2\hat{\mathbf{i}} - \hat{\mathbf{j}}$. Define a function $f : \mathbb{R}^2 \to \mathbb{R}^2$ via the rule:
$$f(\mathbf{x}) = \mathbf{x} + \mathbf{w}.$$

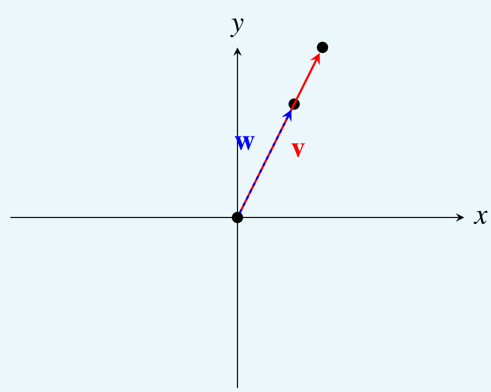
   (a) Write the two equations $u = u(x,y)$ and $v = v(x,y)$ representing f.
   (b) Is f a linear function? Why or why not?
   (c) If you determined f is a linear function, write down the 2×2 matrix associated to f. Otherwise skip this part.
2. Let $g : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by the rule $g(\mathbf{x}) = 2\mathbf{x}$.
   (a) Write the two equations $u = u(x,y)$ and $v = v(x,y)$ representing g.
   (b) Is g is linear function? Why or why not?
   (c) If you determined g is a linear function, write down the 2×2 matrix associated to g. Otherwise skip this part.

■

Exercise 2.22 Let $\mathbf{v}$ and $\mathbf{w}$ be two vectors in $\mathbb{R}^3$. Do you agree or disagree with the following statements? Explain your reasoning for each.
1. *If* $\mathbf{x}$ *and* $\mathbf{y}$ *are in* $\operatorname{span}\{\mathbf{v},\mathbf{w}\}$, *the so is* $\mathbf{x}+\mathbf{y}$.
2. *If* $\mathbf{x}$ *is in* $\operatorname{span}\{\mathbf{v},\mathbf{w}\}$, *then so is* $c\mathbf{x}$ *for any constant* c.
3. *If* $\mathbf{x}$ *is in* $\operatorname{span}\{\mathbf{v},\mathbf{w}\}$, *then so is* $\mathbf{x}+\hat{\mathbf{i}}$.

■

Exercise 2.23 Let's think about a couple more spans in 2d.
1. Let $\mathbf{0}$ be the zero vector. Give a description of $\operatorname{span}\{\mathbf{0}\}$.
2. Let $\mathbf{v}$ and $\mathbf{w}$ be nonzero vectors which are parallel. What can you say about their span? In particular, would you say their span is all of $\mathbb{R}^2$? A line? A single point? Something else entirely? Explain your reasoning.



■

Exercise 2.24 Which of the following vectors in $\mathbb{R}^3$ are linear combinations of $\hat{\mathbf{j}}$ and $\hat{\mathbf{k}}$? Explain your reasoning.
1. The zero vector $\mathbf{0}$.
2. The column vector $\begin{bmatrix}1\\1\\1\end{bmatrix}$

3. The column vector $\begin{bmatrix} 0 \\ -5 \\ 11 \end{bmatrix}$
4. The vector $\hat{\mathbf{i}}$.

■

Exercise 2.25 We saw that the span of a single nonzero vector in $\mathbb{R}^2$ traces out a line. Let $\mathbf{v} = 3\hat{\mathbf{i}} - 4\hat{\mathbf{j}}$, so that the tips of every vector in $\text{span}\{\mathbf{v}\}$ trace out a line. Find the equation of that line. (Recall, the equation of a line can be written $y = mx + b$ where m is the slope and b is the y-intercept. Can you find m and b?) ■

Exercise 2.26 Let's see if we can get some intution about the relationship between spans and dimension, without needing to do any explicit computation.

1. We saw that 2 vectors can span $\mathbb{R}^2$. Can fewer than 2 vectors do this? Why or why not?
2. Give an example of 3 vectors that can span $\mathbb{R}^3$. Do you think fewer that 3 vectors do this? Explain your answer (this explanation can be informal, it doesn't have to be a proof).
3. Let $\mathbb{R}^n$ be the collection of n-dimensional column vectors. How many vectors do you think are necessary to span all of $\mathbb{R}^n$?

■

2.5 February 14th, 2023

It's been a week of break. Let's begin by reminding ourselves of the definitions of linear combinations and spans, and do a few more examples. Recall that a *linear combination* of a collection of vectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$ is a vector $\mathbf{w}$ which can be written in terms of $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$. That is, there are constants $c_1, c_2, \cdots, c_n$ such that we can write:

$$\mathbf{w} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n.$$

The *span* of $\mathbf{v}_1, \cdots, \mathbf{v}_n$ is the collection of *all vectors* $\mathbf{w}$ which are linear combinations of $\mathbf{v}_1, \cdots, \mathbf{v}_n$. It is denoted:

$$\text{span}\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}.$$

2.5.1 Checkin 2 Comments: Meaningfully interpreting mathematics in context

Before moving on, I'd like to talk a bit about Checkin 2, in particular, part 2 (cf. Exercise 2.20.2). We are presented with 3 gas mixtures, gasses X,Y, and Z, which are combinations of Nitrogen, Oxygen, Carbon Dioxidee, and Argon. Each one is represented by a vector $\mathbf{x}, \mathbf{y}$, and $\mathbf{z}$, which records the components of the gas mixture (by weight) as its entries. We are then asked to describe *in plain english* the meaning of $\text{span}\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$. Many folks said something along the lines of:

$$\text{span}\{\mathbf{x}, \mathbf{y}, \mathbf{z}\} \text{ is the collection of linear combinations of } \mathbf{x}, \mathbf{y}, \text{ and } \mathbf{z}.$$

Of course, this is a correct definition of the span, and this sentence would be correct no matter what the context, and no matter what the vectors $\mathbf{x}, \mathbf{y}$, and $\mathbf{z}$ represent. But I don't think I would call this *plain english*, and it doesn't really have anything to do with the problem at hand (mixing gasses in a greenhouse in space). When working on a problem like this, we prefer to think about things *in context*.

Rather than give the answer,[9] let's take a moment to think about this in the context of Exercise 2.16, where we study two vectors $\mathbf{x}$ and $\mathbf{y}$ representing the effect of adding an ounce of pigment X and pigment Y respectively. Then we could be asked to describe $\text{span}\{\mathbf{x}, \mathbf{y}\}$ in plain english. One could give the definition of span: *all vectors that are linear combinations of* $\mathbf{x}$ *and* $\mathbf{y}$ but this would leave out all the context in the problem. Instead, in context, one could say something like:

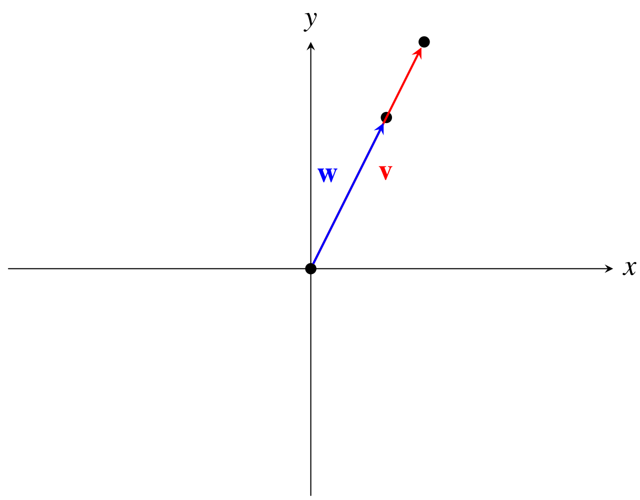
$$\text{span}\{\mathbf{x}, \mathbf{y}\} \text{ represents all colors that can be mixed from pigment X and pigment Y.}$$

This is an accurate description of the span, *and provides context*. It is a description that keeps hold of the fact that, while doing all this vector math, *we are talking about mixing colors*. The purely mathematical description of the span erases this context.

A few more examples of spans

Question 2.7 Let $\mathbf{v}$ and $\mathbf{w}$ be nonzero vectors which are parallel. What can you say about their span?

[9] I will update the notes later, after checkin revisions have been collected



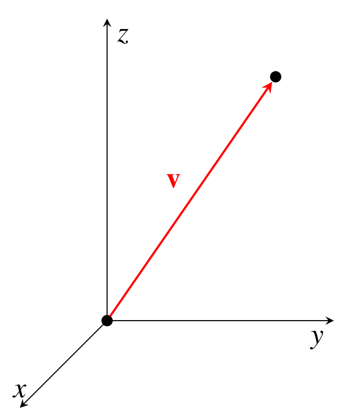
■ **Question 2.8** Can you say anything about the span of the zero vector $\mathbf{0}$?

To summarize, it looks like 3 things can happen:

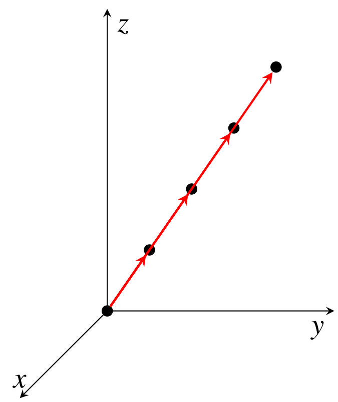
1. The span is just the zero vector. This happens if you are taking the span of the zero vector.
2. The span is a line. This happens if you are taking the span of a single nonzero vector, or of parallel vectors at least one of which is nonzero.
3. The span is all of $\mathbb{R}^2$. This happens if you take the span of at least two nonzero vectors which are not parallel.

What is remarkable is that that we were able to deduce all of this without doing any explicit computations. The story gets a bit more interesting if we move into 3-dimensions. As before, the span of the 0 vector will be just the origin, and with enough vectors it is possible to get all of $\mathbb{R}^3$ (for example, span$\{\hat{\mathbf{i}}, \hat{\mathbf{j}}, \hat{\mathbf{k}}\}$). We can also also have again that the span of a single nonzero vector gives a straight line.

■ **Example 2.16** Let $\mathbf{v}$ be a nonzero vector.



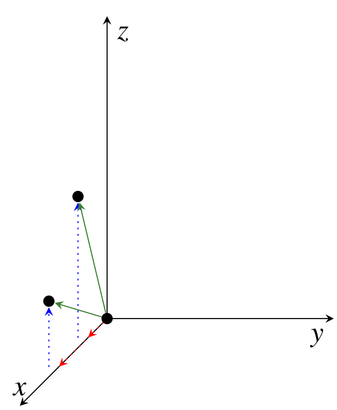
Taking a few multiples we have:



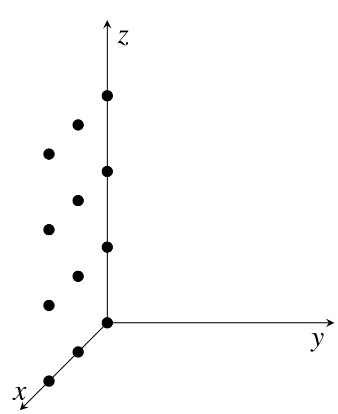
And we can observe that we are tracing out a line through the origin in $\mathbb{R}^3$. ■

So we can get 0, everything, and a straight line. But something different can happen as well, when considering the span of two parallel vectors.

■ **Example 2.17** Let's see if we can work out the span of the unit vectors $\hat{\mathbf{i}}$ and $\hat{\mathbf{k}}$. Below we plot a few different values.



It they lie in the xz-plane. Let's plot where the tips of $a\hat{\mathbf{i}} + b\hat{\mathbf{k}}$ for a few more values of a and b.



It's starting to appear that we can get anywhere in the xz-plane, but also that we can't escape from it.

This is actually not too hard to work out explicitly:

$$\mathbf{v} = a\hat{\mathbf{i}} + b\hat{\mathbf{k}} = a \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + b \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} a \\ 0 \\ b \end{bmatrix}.$$

So if we put the tail of $\mathbf{v}$ at the origin, then we can decide the x-coordinate of where its tip lands by choosing a, a similarly we can decide the z-coordinate of its tip by choosing b. The y-coordinate, on the other hand, must always stay at 0. ■

If you let $\mathbf{v}$ and $\mathbf{w}$ be two non-parallel vectors in $\mathbb{R}^3$, you can proceed as in Example 2.15 and think about where $c\mathbf{v} + d\mathbf{w}$. I encourage you to do this, and maybe you can convince yourself of the following fact:[10]

■ **Slogan 2.6** The span of 2 nonzero and nonparallel vectors in $\mathbb{R}^3$ trace out a plane in $\mathbb{R}^3$.

R Linear combinations are one of the most central and most important concepts in linear algebra. Even before doing any computations, we benefit greatly from thinking about how to visualize them and, and how to think about them in terms of concrete problems. This can be aided by well-made images and good animations. If you find 10 minutes to spare over break, I highly recommend Grant Sanderson's video on visualizing linear combinations: `https://www.3blue1brown.com/lessons/span`.

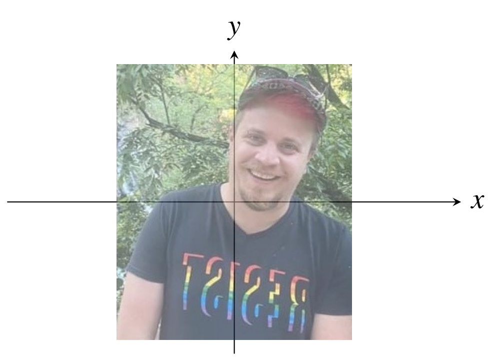
[10] TODO: Make an animation of this.

3. Linear Transformations and Matrices

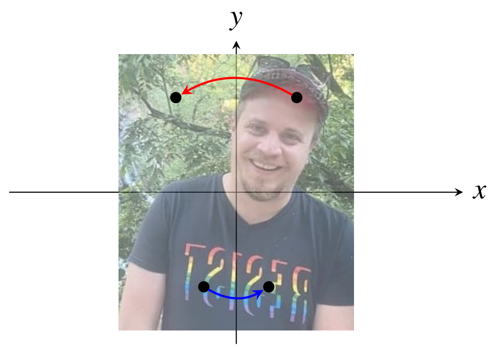
3.1 February 14th, 2023 (Continued...)

I was traveling over the past summer, and took the following selfie.

I didn't notice anything right away, but when I was reflecting back on my vacation photos, I realized that the logo on my shirt was backwards! *Oh no! I got a mirror image!* To fix this, I had to apply a horizontal reflection to the image. *But how?* First, let's introduce some coordinate axes.



With these coordinates chosen, the goal is to reflect the plane over the y-axis. For example, the point where my hair is colored pink is currently at $(1,2)$, but we should instead color my pink hair at the reflected point, $(-1,2)$. Similarly, the part of the letter S that is colored blue at $(-.5,-2)$, should instead be drawn blue at the reflected point $(.5,2)$.



We should do this for every pixel in the image. If the pixel at (x,y) is colored a certain way, we should instead color the pixel $(-x,y)$ that way. Doing this pixel by pixel, we recover the corrected image, with the text appearing legibly!

To summarize, we applied a certain *function* or *transformation* to the plane. In Chapter 1, we studied functions from the plane to itself, and this one certainly fits. Using that language, we defined a tranformation $R : \mathbb{R}^2 \to \mathbb{R}^2$ given by the rule $R(x,y) = (u,v)$ where:

$$u = -x,$$

$$v = y.$$

Notice that this is even a *linear transformation* (cf. Defintion 1.2.1). We applied this linear transformation to the image, meaning that if a point $\mathbf{p}$ was supposed to be colored, say, red, we colored $R(\mathbf{p})$ red instead. This is a rather straightforward example, but many of the most importation transformations of images in computer graphics—from rotations and reflections, to fitting an image to a screen or window—are acheived by applying an appropriate linear transformation.

3.1.1 Linear Transformations of the Plane: Revisited

Recall from Definition 1.2.1, that a linear transformation of the plane was a function $L(x,y) = (u,v)$ where:

$$u = ax + by,$$

$$v = cx + dy,$$

for constants a, b, c, and d. We noticed in Homework 1 (cf. Exercise 1.8) that a linear transformation appeared to play well with addition, suggesting that it might be worth thinking the inputs and outputs of the function as vectors rather than points.

Notation 3.1. *Let $\mathbb{R}^2$ denote the collection of 2-dimensional column vectors* $\begin{bmatrix} x \\ y \end{bmatrix}$.

With this in hand we can revisit Definitions 1.2.1 and 1.2.2, with its notational updates. We remark that the difference is purely cosmetic, the content is identical.

Definition 3.1.1 A linear transformation is a function $L : \mathbb{R}^2 \to \mathbb{R}^2$ given by the rule:

$$L\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} u \\ v \end{bmatrix},$$

where u and v are computed as follows:

$$u = ax + by,$$

$$v = cx + dy,$$

for constants a, b, c, d. The *2×2 matrix* associated to L is:

$$M = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

The four coefficients a, b, c, d completely determine L, therefore so does the matrix M. As such, we often just denote L just using the matrix itself:

$$L\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$

What is on the right looks like we are multiplying two matrices, and that is no accident. In fact, we know that this should equal the column vector

$$\begin{bmatrix} u \\ v \end{bmatrix},$$

whose formula is given in Definition 3.1.1. In particular, the first first row of the output should be $ax + by$ and the second should be $cx + dy$. This establishes our first formula for matrix multiplication:

Definition 3.1.2 — Matrix-Vector Multiplication: The 2×2 Case. The product of a 2×2 matrix M and a column vector $\mathbf{w}$ can be computed as follows:

$$M\mathbf{w} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} ax + by \\ cx + dy \end{bmatrix}.$$

This recovers the usual definition of Matrix-Vector multiplication. Rather than just memorizing this formula, it is useful to think about how the computation as *a process*. Indeed, this process is what will generalize to matrix multiplication in general. To compute the matrix product, we first think about the first row of M and pair its entries in order with the entries in $\mathbf{w}$.

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$

We then multiply the paired elements together, and add them up to obtain the first row of $M\mathbf{w}$.

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} ax+by \\ \end{bmatrix}.$$

To obtain the second row of the output, we pair the second row of M with the entries of $\mathbf{w}$ the same way, multiplying the corresponding elements and adding them together.

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} ax+by \\ cx+dy \end{bmatrix}.$$

This is a process you may already be familiar with, but when I first learned how to multiply matrices, I found the rule to feel kind of opaque and arbitrary. Hopefully Definition 3.1.1 makes this process feel more reasonable. Indeed, the *product* $M\mathbf{v}$ should be thought of as the *function* M being applied to the *vector* $\mathbf{v}$.

■ **Example 3.1** Let's compute the matrix product:

$$\begin{bmatrix} 1 & 3 \\ -2 & 0 \end{bmatrix} \begin{bmatrix} -1 \\ 2 \end{bmatrix}.$$

Starting with the first row:

$$\begin{bmatrix} 1 & 3 \\ -2 & 0 \end{bmatrix} \begin{bmatrix} -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1*-1+3*2 \\ \end{bmatrix} = \begin{bmatrix} 5 \\ \end{bmatrix}$$

Filling in the second row:

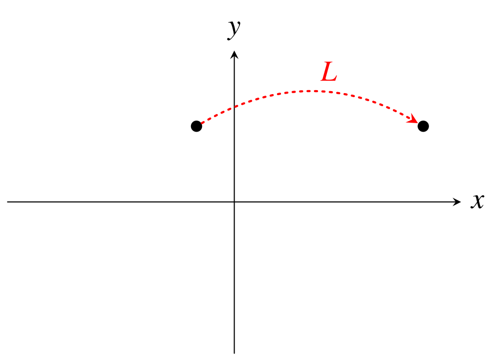
$$\begin{bmatrix} 1 & 3 \\ -2 & 0 \end{bmatrix} \begin{bmatrix} -1 \\ 2 \end{bmatrix} = \begin{bmatrix} -1*-1+3*2 \\ -2*-1+0*2 \end{bmatrix} = \begin{bmatrix} 5 \\ 2 \end{bmatrix}$$

To interpret this as a function, we see that the linear transformation L given by the rules:

$$u = x+3y$$

$$v = -2x,$$

takes $\begin{bmatrix} -1 \\ 2 \end{bmatrix}$ to $\begin{bmatrix} 5 \\ 2 \end{bmatrix}$. Viewing these vectors as points:



■

■ **Question 3.1** Let M be the 2×2 matrix from Example 3.1, compute $M\hat{\mathbf{i}}$ and $M\hat{\mathbf{j}}$.

■ **Example 3.2** The linear transformation which reflected my picture into the correct orientation was given by:

$$u = -x,$$

$$v = y,$$

and therefore by the matrix:

$$\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Applying this to any point gives:

$$\begin{bmatrix} -1n & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -1*x+0*y \\ 0*x+1*y \end{bmatrix} = \begin{bmatrix} -x \\ y \end{bmatrix}.$$

In particular, a computer would likely apply this function by using matrix multiplication on the coordinates of the pixels, rather than remember the functions and all the variables involved. ■

In Homework 1 (cf. Exercise 1.6) we asked the following question:

■ **Question 3.2** Let ℓ be a linear transformation of the plane. If I know $\ell\left(\hat{\mathbf{i}}\right)$ and $\ell\left(\hat{\mathbf{j}}\right)$, do I know the output of ℓ when applied to any element of $\mathbb{R}^2$?

Many folks already determined the answer. Let's see how this works in an example.

■ **Example 3.3** Suppose ℓ is a linear transformation, and that:

$$\ell\left(\hat{\mathbf{i}}\right) = \begin{bmatrix} 1 \\ -2 \end{bmatrix} \qquad \text{and} \qquad \ell\left(\hat{\mathbf{j}}\right) = \begin{bmatrix} 3 \\ 0 \end{bmatrix}.$$

We also know ℓ corresponds to some matrix:

$$M = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

Therefore

$$M\hat{\mathbf{i}} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} a*1+b*0 \\ c*1+d*0 \end{bmatrix} = \begin{bmatrix} a \\ c \end{bmatrix}.$$

In particular, $M\hat{\mathbf{i}}$ plucks out the first column of M, and since $M\hat{\mathbf{i}} = \ell\left(\hat{\mathbf{i}}\right)$ we have:

$$\begin{bmatrix} a \\ c \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \end{bmatrix}.$$

Thus $a = 1$ and $c = -2$. We can similarly compute that $M\hat{\mathbf{j}}$ is the second column of M:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} a*0+b*1 \\ c*0+d*1 \end{bmatrix} = \begin{bmatrix} b \\ d \end{bmatrix}.$$

Since this is $\ell\left(\hat{\mathbf{j}}\right)$ we have determined $b = 3$ and $d = 0$. In particular, we have determined:

$$M = \begin{bmatrix} 1 & 3 \\ -2 & 0 \end{bmatrix},$$

which completely determines ℓ as the function:

$$u = x - 2y,$$

$$v = -2x.$$

This is in fact the function from Example 3.1. ■

To summarize, we have the following theorem.

Theorem 3.1.1 Let ℓ be a linear transformation of $\mathbb{R}^2$. Then ℓ is completely determined by the values $\ell\left(\hat{\mathbf{i}}\right)$ and $\ell\left(\hat{\mathbf{j}}\right)$. In particular, if:

$$\ell\left(\hat{\mathbf{i}}\right) = \begin{bmatrix} a \\ c \end{bmatrix} \qquad \text{and} \qquad \ell\left(\hat{\mathbf{j}}\right) = \begin{bmatrix} b \\ d \end{bmatrix},$$

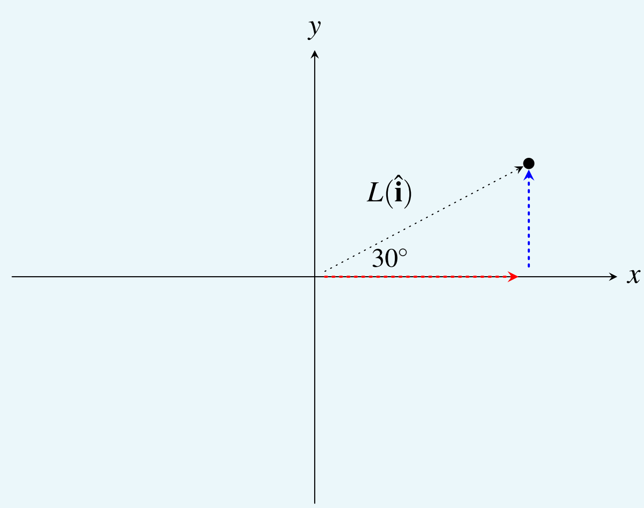
then the matrix associated to ℓ is

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

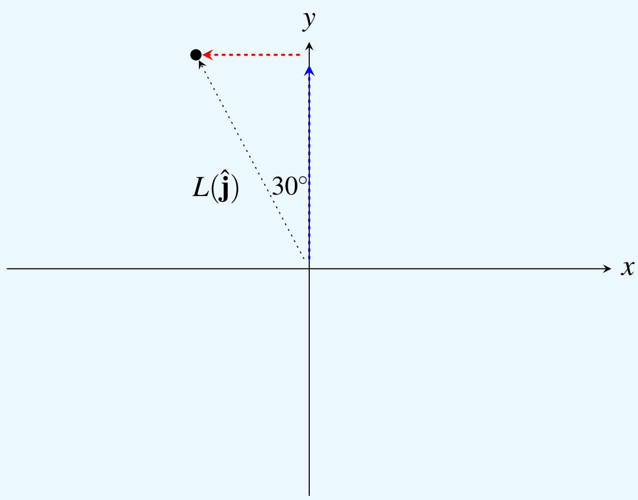
Exercise 3.1 — Checkin 3. Let $L : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear map which rotates the plane 30° counterclockwise. Let's determine the associated matrix. You may use the following facts:

$$\sin(30^\circ) = \frac{1}{2} = 0.5 \qquad \text{and} \qquad \cos(30^\circ) = \frac{\sqrt{3}}{2} \approx 0.866.$$

1. Find the lengths of the legs of the following triangle to determine $L(\hat{\mathbf{i}})$ as a column vector. (Note, since $\hat{\mathbf{i}}$ has length one, so does any rotation of $\hat{\mathbf{i}}$.)



2. Find the lengths of the legs of the following triangle to determine $L(\hat{\mathbf{j}})$ as a column vector. *Be careful with signs!* (Note, since $\hat{\mathbf{j}}$ has length one, so does any rotation of $\hat{\mathbf{j}}$.)

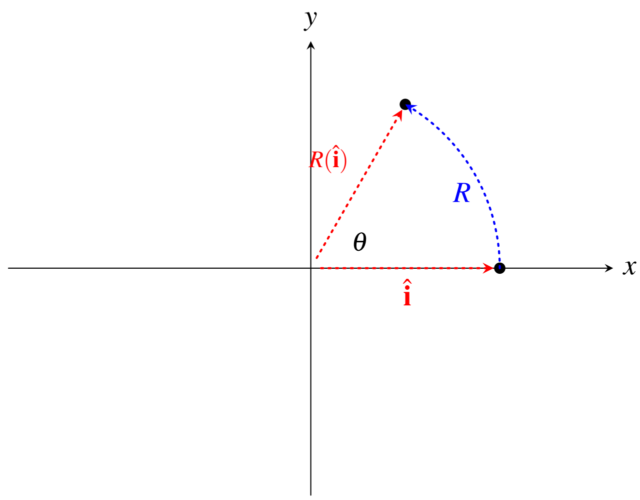


3. Let M be the matrix associated to the linear transformation L. Write down M. (Recall that $L(\hat{\mathbf{i}})$ and $L(\hat{\mathbf{j}})$ determine the columns of M).
4. Use matrix-vector multiplication to determine the image of the point $(-3, 1)$ after a 30° rotation. (You may leave your answer in exact form, or save one decimal point).

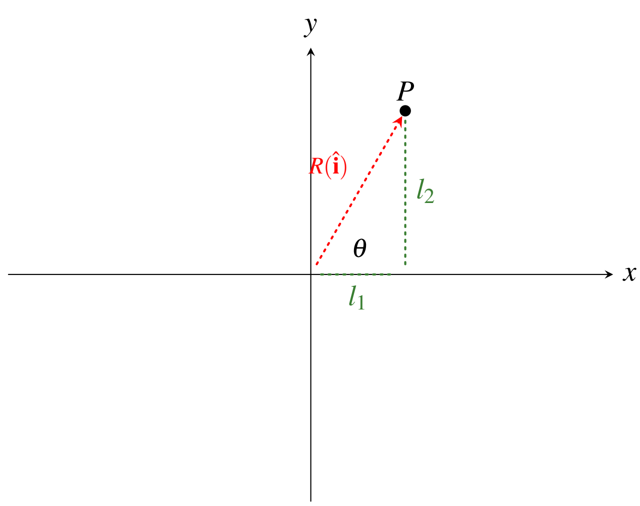
3.2 February, 16, 2023

3.2.1 An Example: Rotation Matrices

We can do something similar to what we did in Checkin 3 (Exercise 3.1) to compute matrices which can capture *any rotation of the plane*! Let's denote by R the linear function which rotates the plane counterclockwise by an angle of θ. To compute the matrix associated to R, it suffices to trace what the rotation does to $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$. Let's start with $\hat{\mathbf{i}}$.



Here $R\left(\hat{\mathbf{i}}\right)$ starts at the origin, so to find it's representation as a column vector we can simply compute the coordinates of the point P where it ends. To do this we can do this by using the triangle below, together with some trig.



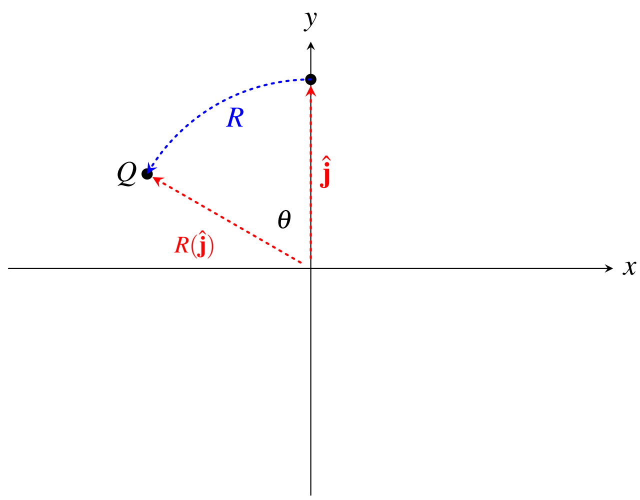
Once we compute the lengths l_1 and l_2, we will know that $P = (l_1, l_2)$. The length of $\hat{\mathbf{i}}$ is 1, and this remains true after rotation. As such, we can solve:

$$\cos\theta = \frac{l_1}{1}, \qquad \text{and} \qquad \sin\theta = \frac{l_2}{1}$$

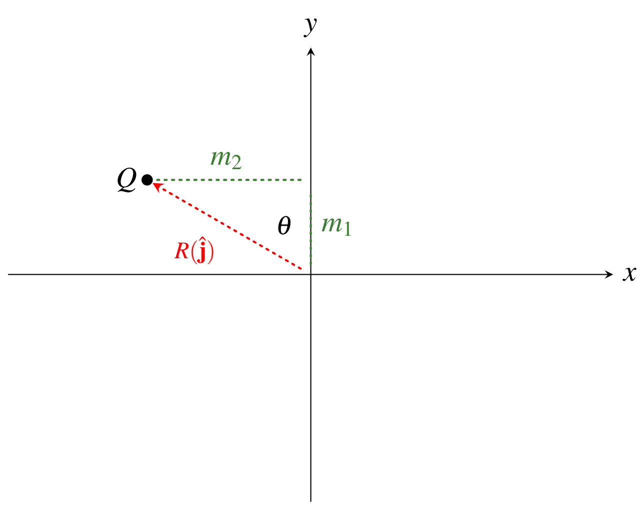
for $l_1 = \cos\theta$ and $l_2 = \sin\theta$. Therefore, $R(\hat{\mathbf{i}})$ is the vector from $(0,0)$ to $(\cos\theta, \sin\theta)$, so that we have:

$$R(\hat{\mathbf{i}}) = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}.$$

We can find $R(\hat{\mathbf{j}})$ similarly, looking for the coordinates of the point Q below



Again, we can can find the coordinates for Q by computing the lengths of the legs of the triangle below.



As above, because the length of $\hat{\mathbf{j}}$ is 1, length of $R(\hat{\mathbf{j}})$ is too, so that the trigonometric ratios tell us $m_1 = \cos\theta$ and $m_2 = \sin\theta$. Keeping signs in mind, we can conclude that $Q = (-m_2, m_1)$ so that:

$$R(\hat{\mathbf{j}}) = \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix}.$$

By Theorem 3.1.1, we can now conclude that the transformation R is given by the matrix:

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

In summary, we have deduced the following result:

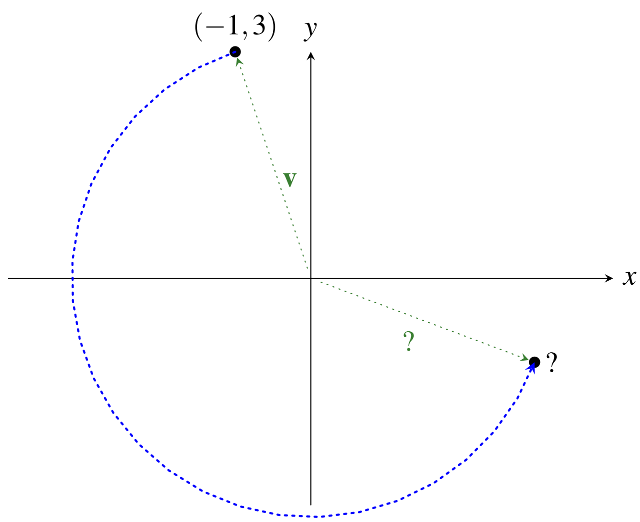
Proposition 3.2.1 Let R be the linear transformation which rotates the plane counterclockwise by an angle of θ, the R is given by the matrix:

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

> **R** The careful reader may notice that our arguments above seemed to rely on the fact that θ was an acute angle. In fact, it is always true that the point unit circle making an angle of θ with the x-axis has coordinates $(\cos\theta, \sin\theta)$. This is actually *the definition* of the trigonometic functions for angles which are not acute. I encourage you to carefully work out the details!

Proposition 3.2.1 is a extremely powerful result, allowing for the rapid computation of any rotation using just matrix multiplication. This is very useful for rotating images on a screen, as doing trig in real time can be slow, but multiplying by matrices is quite fast!

■ **Example 3.4** If I rotate the plane 231°, where does the point $(-1,3)$ end up?



Proposition 3.2.1 tells us that the function R which rotates the plane 231° is given by the matrix:

$$M = \begin{bmatrix} \cos(231°) & -\sin(231°) \\ \sin(231°) & \cos(231°) \end{bmatrix} = \begin{bmatrix} -0.629 & 0.777 \\ -0.777 & -0.629 \end{bmatrix}.$$

Letting $\mathbf{v}$ be the vector from $(0,0)$ to $(-1,3)$, we can compute:

$$M\mathbf{v} = \begin{bmatrix} -.629 & .777 \\ -.777 & -.629 \end{bmatrix} \begin{bmatrix} -1 \\ 3 \end{bmatrix} = \begin{bmatrix} -.629 * (-1) + .777 * 3 \\ -.777 * (-1) - .629 * 3 \end{bmatrix} = \begin{bmatrix} 2.96 \\ -1.11 \end{bmatrix}.$$

Therefore we can conclude that after rotating 231°, the point $(-1,3)$ moves to the point $(2.96, -1.11)$.
■

3.2.2 Linear Transformations and Linear Combinations

The entire *vector perspective* was motivated by Exercise 1.8, which suggested that a linear transformation plays well with addition. Thinking of the inputs and outputs as vectors, the property Exercise 1.8 suggested is that, for any linear transformation L and for any pair of 2d vectors $\mathbf{v}$ and $\mathbf{w}$, we have:

$$L(\mathbf{v}+\mathbf{w}) = L(\mathbf{v}) + L(\mathbf{w}).$$

If the matrix associated to L is M, then we can rewrite this in terms of matrix-vector multiplication:

$$M(\mathbf{v}+\mathbf{w}) = M\mathbf{v} + M\mathbf{w}.$$

But this is a familiar looking property: *the distributive property*! In particular, the fact that L commutes with addition is equivalent to the fact that matrix-vector multiplication satisfies the distributive property! We package this fact together with a related fact regarding scalar multiplication together in the following theorem.

Theorem 3.2.2 — Linearity of Linear Transformations: Planar Case. Let $L : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear map, let $\mathbf{v}, \mathbf{w} \in \mathbb{R}^2$ be vectors, and let c be a constant.
1. $L(\mathbf{v}+\mathbf{w}) = L(\mathbf{v}) + L(\mathbf{w})$
2. $L(c\mathbf{v}) = cL(\mathbf{w})$.

Proof. We will prove the first statement. To do this, we introduce some notation. Let's represent L by the matrix:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix},$$

and represent $\mathbf{v}$ and $\mathbf{w}$ by the column vectors:

$$\mathbf{v} = \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} \qquad \text{and} \qquad \mathbf{w} = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}.$$

Then filling in the first row of $M(\mathbf{v}+\mathbf{w})$ gives:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x_0 + x_1 \\ y_0 + y_1 \end{bmatrix} = \begin{bmatrix} a(x_0+x_1) + b(y_0+y_1) \\ \dots \end{bmatrix}.$$

On the other hand, filling in the first row of $M\mathbf{v} + M\mathbf{w}$ gives:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} + \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} ax_0 + by_0 \\ \dots \end{bmatrix} + \begin{bmatrix} ax_1 + by_1 \\ \dots \end{bmatrix} = \begin{bmatrix} ax_0 + by_0 + ax_1 + by_1 \\ \dots \end{bmatrix}.$$

The fact that these two first rows agree is simply the distributive property for usual addition. One can observe the second rows agree by an identical argument.[1]

We will leave the second part of the theorem for homework. ■

[1] do it! do it! do it!

R One can prove Theorem 3.2.2 without comparing it to matrix multiplication, and just by plugging in generic values to the formula for a linear transformation. That being said, viewing this a matrix-vector multiplication, one can observe that the resemblance of Theorem 3.2.2.1 to the distributive property for matrix-vector multiplication is not purely cosmetic, it really boils down to the usual distributive property for numbers in each row.

■ **Question 3.3** Let L be a linear transformation, and suppose that $\mathbf{v}$ and $\mathbf{w}$ be two vectors, and suppose that $\mathbf{x}$ is a linear combination of $\mathbf{v}$ and $\mathbf{w}$. Is it always true that $L(\mathbf{x})$ is a linear combination of $L(\mathbf{v})$ and $L(\mathbf{w})$?

In Question 3.3, we asked about what happens to linear combinations when we apply a linear transformation. Let's briefly revisit this to get a more geometric perspective on Theorem 3.1.1. If $\mathbf{x} = a\mathbf{v} + b\mathbf{w}$, then (applying Theorem 3.2.2), we can see that:

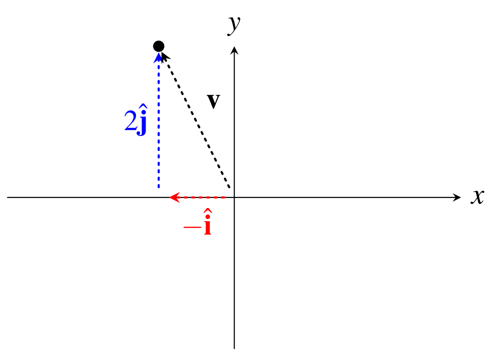
$$L(\mathbf{x}) = L(a\mathbf{v} + b\mathbf{w}) = L(a\mathbf{v}) + L(b\mathbf{w}) = aL(\mathbf{v}) + bL(\mathbf{w}).$$

In particular, if $\mathbf{v} = \hat{\mathbf{i}}$ and $\mathbf{w} = \hat{\mathbf{j}}$, then a and b are the coordinates of $\mathbf{x}$, and we see that knowing these coordinates of together with the images of $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$ allows us to determin $L(\mathbf{x})$. Let's see this in the context of Example 3.3.

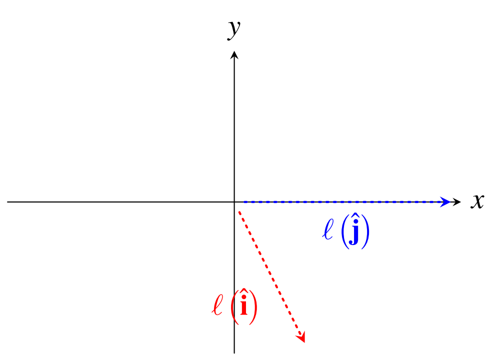
■ **Example 3.5** Adopt the setup of 3.1 and 3.3, and consider again the vector:

$$\mathbf{v} = \begin{bmatrix} -1 \\ 2 \end{bmatrix} = -\hat{\mathbf{i}} + 2\hat{\mathbf{j}}.$$

In particular, $\mathbf{v}$ is acheived by doing $\hat{\mathbf{i}}$ backwards, and then doing $\hat{\mathbf{j}}$ twice.



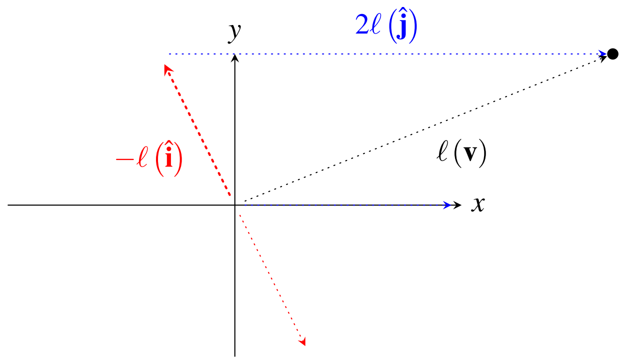
Now let's take a look at where $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$ go after applying ℓ.



Theorem 3.1.1 suggests that knowing these values should be enough to know $\ell(\mathbf{v})$. And indeed, applying Theorem 3.2.2 we have:

$$\ell(\mathbf{v}) = \ell\left(-\hat{\mathbf{i}} + 2\hat{\mathbf{j}}\right) = -\ell\left(\hat{\mathbf{i}}\right) + 2\ell\left(\hat{\mathbf{j}}\right).$$

So, because $\mathbf{v}$ does $\hat{\mathbf{i}}$ backwards and then $\hat{\mathbf{j}}$ twice, we know $\ell(\mathbf{v})$ does $\ell\left(\hat{\mathbf{i}}\right)$ backwards and then $\ell\left(\hat{\mathbf{j}}\right)$ twice. Let's throw that in the picture:



Throwing in some numbers: we can see that:

$$\ell(\mathbf{v}) = -\ell\left(\hat{\mathbf{i}}\right) + 2\ell\left(\hat{\mathbf{j}}\right) = -\begin{bmatrix} 1 \\ -2 \end{bmatrix} + \begin{bmatrix} 3 \\ 0 \end{bmatrix} = \begin{bmatrix} 5 \\ 2 \end{bmatrix},$$

which agrees with our output from Example 3.1 (as it must). The way to think about this is that when ℓ moves $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$, it drags the entire grid along with it.[2] ■

3.2.3 Linear Transformations of 3-space

So far we've been pretty focused on transformations of the plane, but in is also quite important in practice move beyond the plane. Let's begin by considering transformations of 3-dimensional space. For this it will be useful to give a definition of $\mathbb{R}^3$ which is analogous to that of $\mathbb{R}^2$.

Definition 3.2.1 The set $\mathbb{R}^3$ is the collection of 3-dimensional column vectors $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$

As before, we will sometimes conflate the idea of a 3d column vector with that of a 3d point (x,y,z) if our perspective is spacial. We actually already talked a bit about transformations from $\mathbb{R}^3$ to $\mathbb{R}^3$ in Homework 1 (Exercise 1.9). In particular, a function $T : \mathbb{R}^3 \to \mathbb{R}^3$ can be defined by the rule:

$$T\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} u \\ v \\ w \end{bmatrix},$$

and we can define u, v, and w in terms of x, y, z:

$$u = u(x,y,z),$$

[2]This is animated brilliantly by Grant Sanderson: https://3b1b-posts.us-east-1.linodeobjects.com//content/lessons/2016/linear-transformations/basis_example2.mp4

$$v = v(x,y,z),$$
$$w = w(x,y,z).$$

■ **Example 3.6** Define $T : \mathbb{R}^3 \to \mathbb{R}^3$ by the rule

$$T\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} u \\ v \\ w \end{bmatrix},$$

where

$$u = xyz,$$
$$v = x + y + z,$$
$$w = 1 + 2x.$$

Then, for example, we can compute:

$$T\left(\begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}\right) = \begin{bmatrix} 2*3*4 \\ 2+3+4 \\ 1+2*2 \end{bmatrix} = \begin{bmatrix} 24 \\ 9 \\ 5 \end{bmatrix}.$$

■

As with transformations of the plane, linear algebra focuses on functoins which are *purely linear*. That is, we want the equations of u, v, and w to be polynomials of degree 1, with no constant terms.

Definition 3.2.2 A function $T : \mathbb{R}^3 \to \mathbb{R}^3$ is a *linear transformation* if it can be defined by the rule:

$$T\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} u \\ v \\ w \end{bmatrix},$$

where:

$$u = ax + by + cz,$$
$$v = dx + ey + kz,$$
$$w = lx + my + nz,$$

for constants $a, b, c, d, e, k, l, m, n$.

Notice that the entire function is determined by the cofficients of x, y, and z in the equations for u, v, and w. Therefore, it is enough to remember just these coefficients, which we can arrange in a matrix.

Definition 3.2.3 The matrix associated to the linear transformation from Definition 3.2.2 is:

$$\begin{bmatrix} a & b & c \\ d & e & k \\ l & m & n \end{bmatrix}.$$

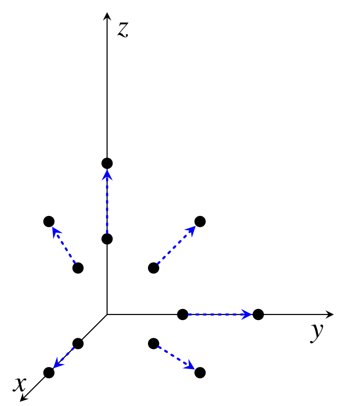
■ **Example 3.7** Consider the function $L : \mathbb{R}^3 \to \mathbb{R}^3$ which is given by the rules:

$$u = 2x \qquad v = 2y \qquad w = 2z.$$

The matrix associated to this function is:

$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

Let's visualize this as a transformation of space, like we did back at the beginning of the semester. Below we have a series of points. The points before and after applying T are connected by arrows.



It looks like points are being pushed away from the origin, and indeed:

$$T\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} 2x \\ 2y \\ 2z \end{bmatrix} = 2\begin{bmatrix} x \\ y \\ z \end{bmatrix},$$

so T takes a point and doubles its distance from the origin.[3] ■

The fact that we can replace a function with a matrix tells us that we can get a formula for matrix-vector multiplication in 3d, analogous to Definition 3.1.2. Indeed, adopting the notation of Definitions 3.2.2 and 3.2.3, if T is the transformation associated to:

$$M = \begin{bmatrix} a & b & c \\ d & e & k \\ l & m & n \end{bmatrix}.$$

Then we can write:

$$\begin{bmatrix} a & b & c \\ d & e & k \\ l & m & n \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = T\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} u \\ v \\ w \end{bmatrix}.$$

Plugging in for the formulas of u, v, w gives:

[3] TODO: Make an animation for this example

Definition 3.2.4 — Matix-Vector Multiplication: The 3×3 Case. The product of a 3×3 matrix and a column vector **v** can be computed as follows.

$$M\mathbf{v} = \begin{bmatrix} a & b & c \\ d & e & k \\ l & m & n \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} ax+by+cz \\ dx+ey+kz \\ lx+my+nz \end{bmatrix}.$$

This can follow a process just like in the 2×2 case, by going row by row in the matrix, and pairing each entry in the row with the appropriate entry in the vector.

$$\textbf{Row 1:} \quad \begin{bmatrix} \color{red}{a} & \color{blue}{b} & \color{green}{c} \\ d & e & k \\ l & m & n \end{bmatrix} \begin{bmatrix} \color{red}{x} \\ \color{blue}{y} \\ \color{green}{z} \end{bmatrix} = \begin{bmatrix} \color{red}{ax}+\color{blue}{by}+\color{green}{cz} \\ \\ \\ \end{bmatrix}$$

$$\textbf{Row 2:} \quad \begin{bmatrix} a & b & c \\ \color{red}{d} & \color{blue}{e} & \color{green}{k} \\ l & m & n \end{bmatrix} \begin{bmatrix} \color{red}{x} \\ \color{blue}{y} \\ \color{green}{z} \end{bmatrix} = \begin{bmatrix} ax+by+cz \\ \color{red}{dx}+\color{blue}{ey}+\color{green}{kz} \\ \\ \end{bmatrix}$$

$$\textbf{Row 3:} \quad \begin{bmatrix} a & b & c \\ d & e & k \\ \color{red}{l} & \color{blue}{m} & \color{green}{n} \end{bmatrix} \begin{bmatrix} \color{red}{x} \\ \color{blue}{y} \\ \color{green}{z} \end{bmatrix} = \begin{bmatrix} ax+by+cz \\ dx+ey+kz \\ \color{red}{lx}+\color{blue}{my}+\color{green}{nz} \end{bmatrix}$$

■ **Question 3.4** Compute the matrix product:

$$\begin{bmatrix} 1 & 2 & 3 \\ 0 & 2 & 1 \\ 2 & 0 & 1 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \\ 3 \end{bmatrix}.$$

As we get to bigger matrices, the process for matrix multiplication starts to get more tedious, and we will start to use computers to do it for us. But we will keep doing it by hand, just a little bit longer, until we can identify the general pattern and its meaning.

■ **Question 3.5** Let M be the 3×3 matrix from Question3.4. Compute:

$$M\hat{\mathbf{i}}, \qquad M\hat{\mathbf{j}}, \qquad M\hat{\mathbf{k}}.$$

Do you notice anything?

As you probably observed, we recovered the columns of M. In particular, a version of Theorem 3.1.1 holds true for 3×3 matrices as well.

Theorem 3.2.3 Let T be a linear transformation of $\mathbb{R}^3$. Then T is completely determined by the values $T\left(\hat{\mathbf{i}}\right)$, $T\left(\hat{\mathbf{j}}\right)$, and $T\left(\hat{\mathbf{k}}\right)$. In particular, if:

$$T\left(\hat{\mathbf{i}}\right) = \begin{bmatrix} a \\ d \\ l \end{bmatrix} \quad \text{and} \quad T\left(\hat{\mathbf{j}}\right) = \begin{bmatrix} b \\ e \\ m \end{bmatrix} \quad \text{and} \quad T\left(\hat{\mathbf{k}}\right) = \begin{bmatrix} c \\ k \\ n \end{bmatrix},$$

then the matrix associated to T is:

$$M = \begin{bmatrix} a & b & c \\ d & e & k \\ l & m & n \end{bmatrix}.$$

This theorem is very important in 3D modelling. Indeed, a rotation in 3 space is a complicated manuver to pin down, as there are 3-axes about which to rotate. That said, once you know where $\hat{\mathbf{i}}, \hat{\mathbf{j}}$ and $\hat{\mathbf{k}}$ go, you can write down a matrix which completely captures the rotation! We will do this explicitly next week! Before moving on, we'd like to record that Theorem 3.2.2 holds true here as well, and can be computed directly using matrix multiplicaiton.

Theorem 3.2.4 — Linearity of Linear Transformations: 3D Case. Let $L : \mathbb{R}^3 \to \mathbb{R}^3$ be a linear map, let $\mathbf{v}, \mathbf{w} \in \mathbb{R}^3$ be vectors, and let c be a constant.

1. $L(\mathbf{v}+\mathbf{w}) = L(\mathbf{v}) + L(\mathbf{w})$
2. $L(c\mathbf{v}) = cL(\mathbf{w})$.

3.3 Homework 4

Exercise 3.2 Compute the following Matrix-Vector products.

1. $\begin{bmatrix} 1 & 3 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} -1 \\ 2 \end{bmatrix}$
2. $\begin{bmatrix} 1 & 3 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \end{bmatrix}$
3. $\begin{bmatrix} 8 & 0 & 1 \\ 2 & -1 & 3 \\ 0 & 0 & -11 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ -3 \end{bmatrix}$
4. $\begin{bmatrix} 8 & 0 & 1 \\ 2 & -1 & 3 \\ 0 & 0 & -11 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$

■

Exercise 3.3 A linear map $T : \mathbb{R}^3 \to \mathbb{R}^3$ has the following effects on $\hat{\mathbf{i}}, \hat{\mathbf{j}}$, and $\hat{\mathbf{j}}$:

$$T(\hat{\mathbf{i}}) = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \qquad T(\hat{\mathbf{j}}) = \begin{bmatrix} 2 \\ 0 \\ 2 \end{bmatrix} \qquad T(\hat{\mathbf{k}}) = \begin{bmatrix} 5 \\ 0 \\ 0 \end{bmatrix}.$$

What is $T(3\hat{\mathbf{i}} - \hat{\mathbf{j}} + 17\hat{\mathbf{k}})$? ■

Exercise 3.4 A linear map $L : \mathbb{R}^2 \to \mathbb{R}^2$ has the following effects on $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$.

$$L(\hat{\mathbf{i}}) = \begin{bmatrix} 5 \\ 0 \end{bmatrix} \qquad \text{and} \qquad L(\hat{\mathbf{j}}) = \begin{bmatrix} -7 \\ 13 \end{bmatrix}.$$

What has a larger magnitude:

$$L\left(\begin{bmatrix}1\\1\end{bmatrix}\right) \quad \text{or} \quad L\left(\begin{bmatrix}1\\-1\end{bmatrix}\right)?$$

■

Exercise 3.5 Consider the following 2 vectors in $\mathbb{R}^2$.

$$\mathbf{v} = \begin{bmatrix}1\\1\end{bmatrix} \qquad \text{and} \qquad \mathbf{w} = \begin{bmatrix}1\\-1\end{bmatrix}.$$

1. Do you think $\text{span}\{\mathbf{v}, \mathbf{w}\} = \mathbb{R}^2$? Why or why not?
2. Let $\mathbf{x} = 7\hat{\mathbf{i}} - 3\hat{\mathbf{j}}$. Write $\mathbf{x}$ as a linear combination of $\mathbf{v}$ and $\mathbf{w}$. (*Hint:* Extract a system of equations from the expression $a\mathbf{v} + b\mathbf{w} = \mathbf{x}$, and then solve for the constants a and b).
3. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear transformation, and suppose that:

$$T(\mathbf{v}) = \begin{bmatrix}3\\2\end{bmatrix} \qquad \text{and} \qquad T(\mathbf{w}) = \begin{bmatrix}-6\\-5\end{bmatrix}.$$

Compute $T(\mathbf{x})$ by writing it as a linear combination of $T(\mathbf{v})$ and $T(\mathbf{w})$.

■

Exercise 3.6 Let $L : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear transformation, and suppose $\mathbf{v}$ and $\mathbf{w}$ are two vectors in $\mathbb{R}^2$ whose span is all of $\mathbb{R}^2$. Do you agree or disagree with the following statement? Explain your reasoning. (Use the intuition gained from Question 3.5.)

The values $L(\mathbf{v})$ and $L(\mathbf{w})$ determine $L(\mathbf{x})$ for any vector $\mathbf{x}$ in $\mathbb{R}^2$.

■

Exercise 3.7 Let's prove Theorem 3.2.2.2. It states the following: If $L : \mathbb{R}^2 \to \mathbb{R}^2$ is a linear transformation, $\mathbf{v}$ is a vector in $\mathbb{R}^2$, and n is a constant. Then

$$nL(\mathbf{v}) = L(n\mathbf{v}).$$

To do this, we'll introduce some notation. We denote the matrix associated to L by:

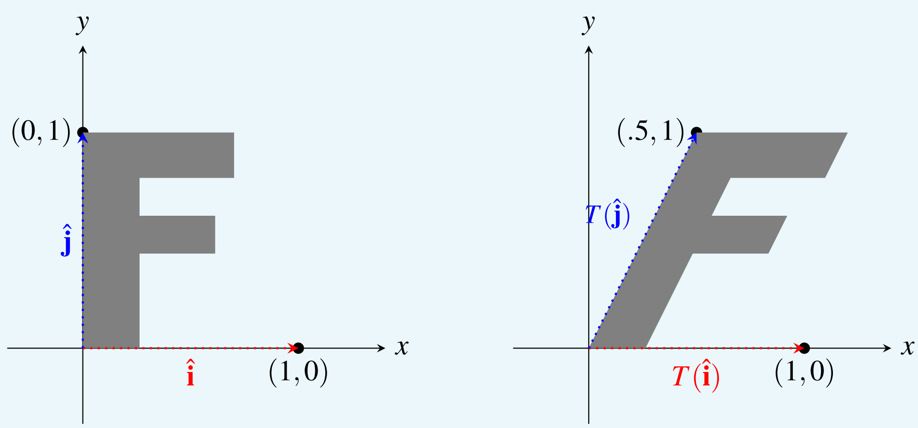
$$M = \begin{bmatrix}a & b\\c & d\end{bmatrix}.$$

We also give coordinates to $\mathbf{v}$:

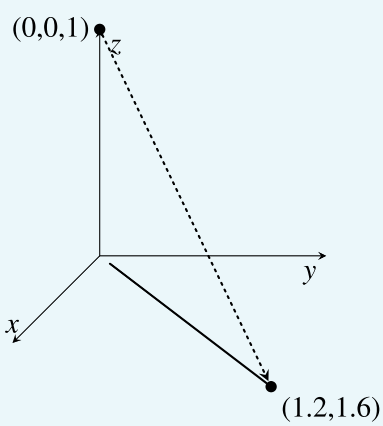
$$\mathbf{v} = \begin{bmatrix}x\\y\end{bmatrix}.$$

1. Write a column vector for $n(M\mathbf{v})$ in terms of the variables a,b,c,d,n,x, and y. (First do the Matrix-Vector product $M\mathbf{v}$, then scale the result by n).
2. Write a column vector for $M(n\mathbf{v})$ in terms of the variables a,b,c,d,n,x, and y. (First write $n\mathbf{v}$ as a column vector in terms of n,x, and y, and then multiply this column by M).
3. Compare your answers to (a) and (b) to explain why the Theorem is true.

Exercise 3.8 A computer translates images from blockstyle fonts to *italics* by applying a linear transformation called a *shear*. Below is an image of the letter F before and after applying the shear. Use this image to determine the matrix associated to the shearing transformation.



Exercise 3.9 We can use linear maps to calculate how shadows are cast. Choose some coordinates in meters, put a meterstick vertically at the origin, and measure that it casts its shadow on the point 1.2 meters east and 1.6 meters north of the stick.



We define a linear function $S : \mathbb{R}^3 \to \mathbb{R}^2$ which takes a vector point in space (denoted by a vector in $\mathbb{R}^3$) to the point at which it casts its shadow (denoted by a vector in $\mathbb{R}^2$). **Notice: the input of this function is 3-dimensional, and the output is 2-dimensional. In particular, the general**

setup looks something like:

$$S\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} u \\ v \end{bmatrix}.$$

1. What are $S(\hat{\mathbf{i}}), S(\hat{\mathbf{j}}), S(\hat{\mathbf{k}})$? Your answers should be 2D vectors. (*Hint*: You should be able to extract $S(\hat{\mathbf{k}})$ from the picture above. For the other two...where does a point on the ground cast its shadow?)
2. We've seen that many linear maps can be captured by matrices, and that the columns of these matrices can be recovered by where the standard basis vectors are sent. Use this philosophy to write down a matrix which could represent S.
3. Extract some equations which could represent $S : \mathbb{R}^3 \to \mathbb{R}^2$ from the matrix in part (b).
4. The Eiffel Tower stands 330 meters tall. It also happens to be 500m west and 350m north from where you planted your meter stick at the start of this exercise. Use the work you've done to compute the coordinates of where the tip of its shadow should land (you may assume that Paris is completely flat). Try this with the equations from question 3, *and* with matrix vector multiplication.
5. Do you agree or disagree with the following statement? Explain your reasoning.

   Once I know where a single point *above the ground* casts its shadow, I can compute where any point casts its shadow.

3.4 February 21, 2023

3.4.1 Linear transformations between dimensions

We've talked for the time being about transformations from $\mathbb{R}^2$ to itself and from $\mathbb{R}^3$ to itself, but it is also sometimes important to think about transformations between different spaces. We've also been mainly thinking about $\mathbb{R}^2$ and $\mathbb{R}^3$ spacially, but sometimes in context we encounter linear transformations have sources and targets which aren't obviously spacial, but instead have some other concrete interpretations. In fact, *we have already done both of these things!* To see this, let's return to the paint mixing examples from Homework 2 (Exercises 2.16-2.19), and see how there was a linear transformation lying at the heart of it, even if it didn't quite look like that at the time.

To begin let's assign some specific meanings to vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$ which fit this context.

$$\mathbb{R}^2 = \left\{ \text{vectors } \begin{bmatrix} x \\ y \end{bmatrix} \text{ representing } x \text{ oz Pigment X and } y \text{ oz Pigment Y} \right\}.$$

$$\mathbb{R}^3 = \left\{ \text{vectors } \begin{bmatrix} r \\ g \\ b \end{bmatrix} \text{ representing } r \text{ units red}, \ g \text{ units green}, \text{ and } b \text{ units blue} \right\}.$$

Then we can define a function:

$$F : \mathbb{R}^2 \to \mathbb{R}^3,$$

by the rule:

$$F\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \text{the vector } \begin{bmatrix} r \\ g \\ b \end{bmatrix} \text{ representing the color mixed from } x \text{ oz Pigment X and } y \text{ oz Pigment Y}.$$

In fact, we can find equations for r, g, and b in terms of x and y. Indeed, we are given what the overall effects of our two pigments are.
Pigment X: Adding 1 ounce of Pigment X to paint adds:

1 unit of red, 2 units of green, 3 units of blue

Pigment Y: Adding 1 ounce of Pigment Y to paint adds:

7 units of red, 5 units of green, 2 units of blue

Therefore, we can compute:

$$r = 1 * x + 7 * y,$$

$$g = 2 * x + 5 * y,$$

$$b = 3 * x + 2 * y.$$

So we have computed equations for the function $F : \mathbb{R}^2 \to \mathbb{R}^3$. Even better, we can notice that r, g, and b are *linear functions* in x and y with no constant terms. So this looks like a linear function overall. This is completely determined by its coefficients, so following what we've done before, we can see that we know the whole function by just remember its coefficient matrix:

$$M = \begin{bmatrix} 1 & 7 \\ 2 & 5 \\ 3 & 2 \end{bmatrix}.$$

What happens if we try to extend the process of Matrix-Vector Multipication to this setting? Let's try with 3 ounces of say, 3 ounces of Pigment X and 3 ounces of Pigment Y. Well, going row by row:

$$\text{Row 1:} \quad \begin{bmatrix} 1 & 7 \\ 2 & 5 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ 3 \end{bmatrix} = \begin{bmatrix} 1*3+7*3 \\ \\ \\ \end{bmatrix},$$

$$\text{Row 2:} \quad \begin{bmatrix} 1 & 7 \\ 2 & 5 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ 3 \end{bmatrix} = \begin{bmatrix} 1*3+7*3 \\ 2*3+5*3 \\ \\ \end{bmatrix},$$

$$\text{Row 3:} \quad \begin{bmatrix} 1 & 7 \\ 2 & 5 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ 3 \end{bmatrix} = \begin{bmatrix} 1*3+7*3 \\ 2*3+5*3 \\ 3*3+2*3 \end{bmatrix} = \begin{bmatrix} 24 \\ 21 \\ 15 \end{bmatrix}.$$

Looks like the output is the vector representing the color containing

24 units of red, 21 units of green, 15 units of blue,

which is fancy gold! We learned in Exercise 2.17 tha this is exactly what we should get when mixing 3 ounces of Pigment X and 3 ounces of Pigment Y. That is:

$$M \begin{bmatrix} 3 \\ 3 \end{bmatrix} = \begin{bmatrix} 24 \\ 21 \\ 15 \end{bmatrix} = F\left(\begin{bmatrix} 3 \\ 3 \end{bmatrix}\right),$$

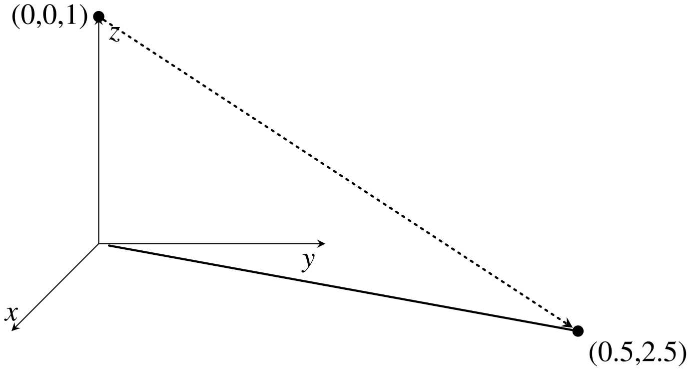
so as before, *matrix multiplication corresponds to applying the linear function!*.

■ **Question 3.6** Let $\hat{\mathbf{i}}$ be the vector in $\mathbb{R}^2$ representing one ounce of Pigment X, and let $\hat{\mathbf{j}}$ represent one ounce of Pigment Y. Compute:

$$M\hat{\mathbf{i}} \qquad \text{and} \qquad M\hat{\mathbf{j}}.$$

As you might observe, you yet again obtain the columns of M, so as before, so these values determine all of F and Theorem 3.1.1 holds! Of course, here it is no surprise that knowing the effect of one ounce of Pigment X and one ounce of Pigment Y is enough to tell you the effect of any mixture of them. It seems like adding this type context makes certain results easier or expect than it is initially in the purely geometric setting.

Let's see if we can run this philosophy in reverse. On last Thursday's groupwork (cf. Exercise 3.9) we played with another example of a linear map between dimensions, this time from $\mathbb{R}^3 \to \mathbb{R}^2$. Since we're finishing this on the homework, I'll use different numbers, but the idea will be exactly the same. We choose coordinates (in meters), and place a meterstick vertically at the origin, just outside the library at ODY. It casts a shadow to a point on the ground 0.5 meters east and 2.5 meters north of the stick.



We then defined a linear function $S : \mathbb{R}^3 \to \mathbb{R}^2$ defined as follows:

$$S\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} u \\ v \end{bmatrix},$$

if a point x meters east and y meters north of the origin and z-meters off the ground casts its shadow on the ground at a point u-meters east and v-meters north of the origin. So for example, our picture tells us that:

$$S\left(\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 0.5 \\ 2.5 \end{bmatrix}.$$

■ **Question 3.7** What are $S\left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}\right)$ and $S\left(\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}\right)$?

Every linear map we've seen so far comes with a matrix, and the columns of that matrix are precisely the values of the linear map applied to the standard basis. Let's see what that philosophy can tell us in this situation. Let $\hat{\mathbf{i}}, \hat{\mathbf{j}}, \hat{\mathbf{k}}$ be the standard basis of $\mathbb{R}^3$. Then S should have some matrix whose columns are $S(\hat{\mathbf{i}}), S(\hat{\mathbf{j}}), S(\hat{\mathbf{k}})$.

$$N = \begin{bmatrix} S(\hat{\mathbf{i}}) & S(\hat{\mathbf{j}}) & S(\hat{\mathbf{k}}) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0.5 \\ 0 & 1 & 2.5 \end{bmatrix}.$$

The entries of this matrix have always arisen as the coefficients of the linear equations defining our linear function. Assuming that holds here too, what would the equations for S be? We know S takes 3 inputs (x, y, z) and has 2 outputs (u, v), so we are looking for:

$$u = u(x, y, z),$$

$$v = v(x, y, z).$$

Let's go row by row:

$$\textbf{Row 1}: \begin{bmatrix} \color{red}{1} & \color{blue}{0} & \color{green}{0.5} \\ 0 & 1 & 2.5 \end{bmatrix} \qquad \longrightarrow \qquad u = \color{red}{1}x + \color{blue}{0}y + \color{green}{0.5}z.$$

$$\textbf{Row 2:} \begin{bmatrix} 1 & 0 & 0.5 \\ 0 & 1 & 2.5 \end{bmatrix} \quad \longrightarrow \quad v = 0x + 1y + 2.5z.$$

We have now deduced what our function should be:

$$S\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} u \\ v \end{bmatrix},$$

where

$$u = x + 0.5z,$$

$$v = y + 2.5z.$$

So, for example, Gunnison Chappel is about 65 meters east from where we placed our meterstick, and stands about 30 meters tall. This gives the tip of the chapel the coordinates $(65,0,30)$. If we want to know where it casts its shadow, we can compute:

$$u = 65 + 0.5 * 30 = 80,$$

$$v = 0 + 2.5 * 30 = 75.$$

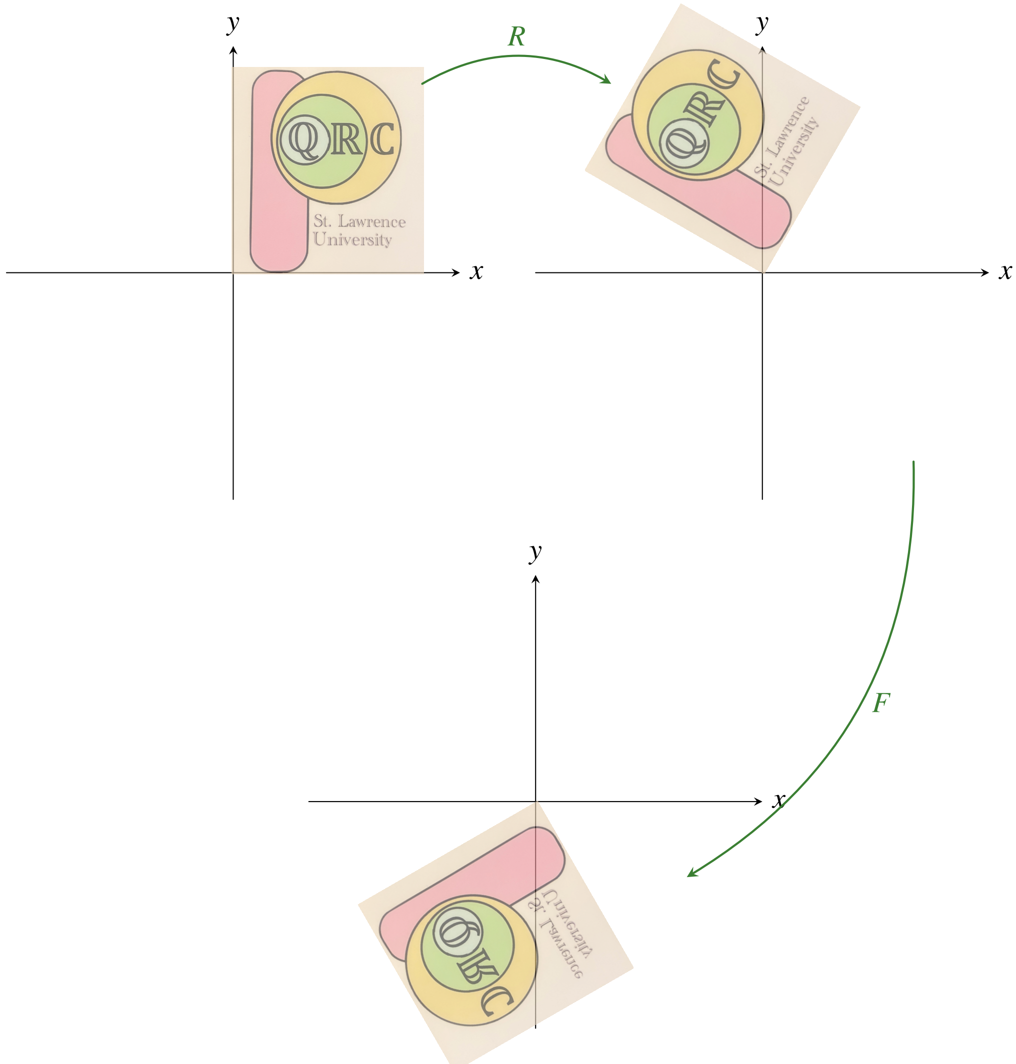
So it casts a shadow 80 meters east, and 75 meters north, of our meterstick. *What is remarkable about this, is we measured the point of a single shadow and we were able to determine another!* Finally, let's recall that we discovered the process for matrix-vector multiplication by thinking about multiplying a matrix as the same as applying the associated linear function. This works here too: if we apply the process of matrix multiplication, we get the same answer as applying S.

$$\textbf{Row 1 :} \begin{bmatrix} 1 & 0 & 0.5 \\ 0 & 1 & 2.5 \end{bmatrix} \begin{bmatrix} 65 \\ 0 \\ 30 \end{bmatrix} = \begin{bmatrix} 1 * 65 + 0 * 0 + 0.5 * 30 \\ \ \end{bmatrix} = \begin{bmatrix} 80 \\ \ \end{bmatrix},$$

$$\textbf{Row 2:} \begin{bmatrix} 1 & 0 & 0.5 \\ 0 & 1 & 2.5 \end{bmatrix} \begin{bmatrix} 65 \\ 0 \\ 30 \end{bmatrix} = \begin{bmatrix} 1 * 65 + 0 * 0 + 0.5 * 30 \\ 0 * 65 + 1 * 0 + 2.5 * 30 \end{bmatrix} = \begin{bmatrix} 80 \\ 75 \end{bmatrix}.$$

3.4.2 Composition of Linear Transformations

Suppose I start with an image that I'd like to manipulate an image of the PQRC logo in order to place it on a T-Shirt. There are two things I'd like to do. First, I want to rotate it 60°, and then I'd like to reflect it vertically over the x-axis.

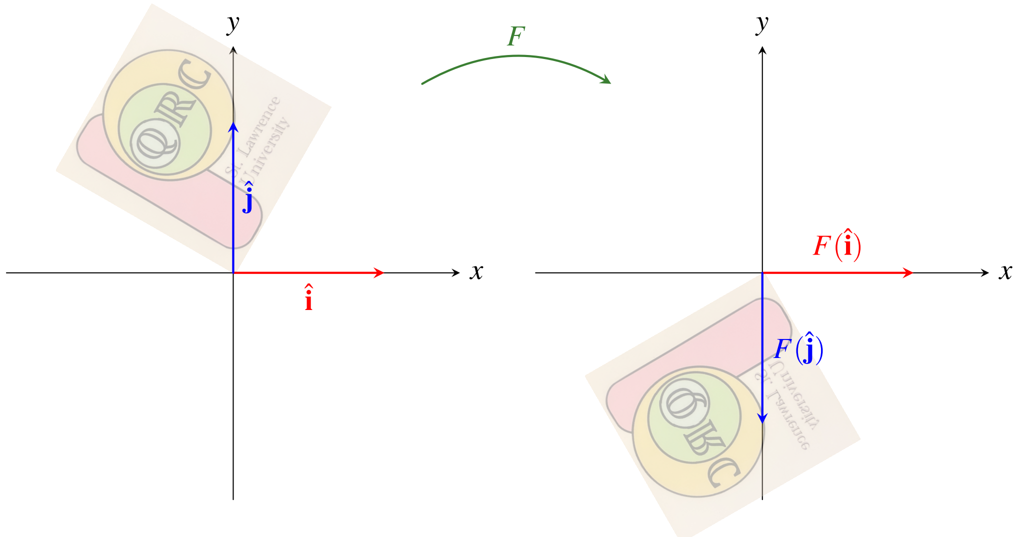


The transformatoins R (for *rotate*) and F (for *flip*) are both linear, and therefore each have an associated matrix. Let's find the matrix for the transformation with *rotates* and then *flips*. First let's find the matrix for the rotation R. In fact, in Section 3.2.1 we deduced that R is given by the rotation matrix:

$$N = \begin{bmatrix} \cos 60^\circ & -\sin 60^\circ \\ \sin 60^\circ & \cos 60^\circ \end{bmatrix} = \begin{bmatrix} 0.5 & -.866 \\ .866 & 0.5 \end{bmatrix}.$$

What about F? To find the matrix for F, it is enough to compute $F(\hat{\mathbf{i}})$ and $F(\hat{\mathbf{j}})$.

■ **Question 3.8** Can you compute the matrix for F?



In particular, we have:

$$F(\hat{\mathbf{i}}) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{and} \quad F(\hat{\mathbf{j}}) = \begin{bmatrix} 0 \\ -1 \end{bmatrix}.$$

Therefore the matrix for F is:

$$M = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

To find the matrix for *rotate then flip*, we will follow a similar philosphy, by trying to track where $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$ go. That is, we'd like to know the values $F(R(\hat{\mathbf{i}}))$ and $F(R(\hat{\mathbf{j}}))$. Before computing this, let's briefly unpack the notation. Given a vector $\mathbf{v}$ in $\mathbb{R}^2$, the value $R(\mathbf{v})$ represents where the vector $\mathbf{v}$ is after rotating. This is another vector in $\mathbb{R}^2$, and therefore it can be fed to the function F to be flipped. This value is $F(R(\mathbf{v}))$.

To compute $F(R(\hat{\mathbf{i}}))$, we first compute $R(\hat{\mathbf{i}})$, and we feed whatever the output is to F. But we can easily determine $R(\hat{\mathbf{i}})$: it is the first column of the matrix for R:

$$R(\hat{\mathbf{i}}) = \begin{bmatrix} .5 & -.866 \\ .866 & .5 \end{bmatrix} \begin{bmatrix} .5 \\ .866 \end{bmatrix}.$$

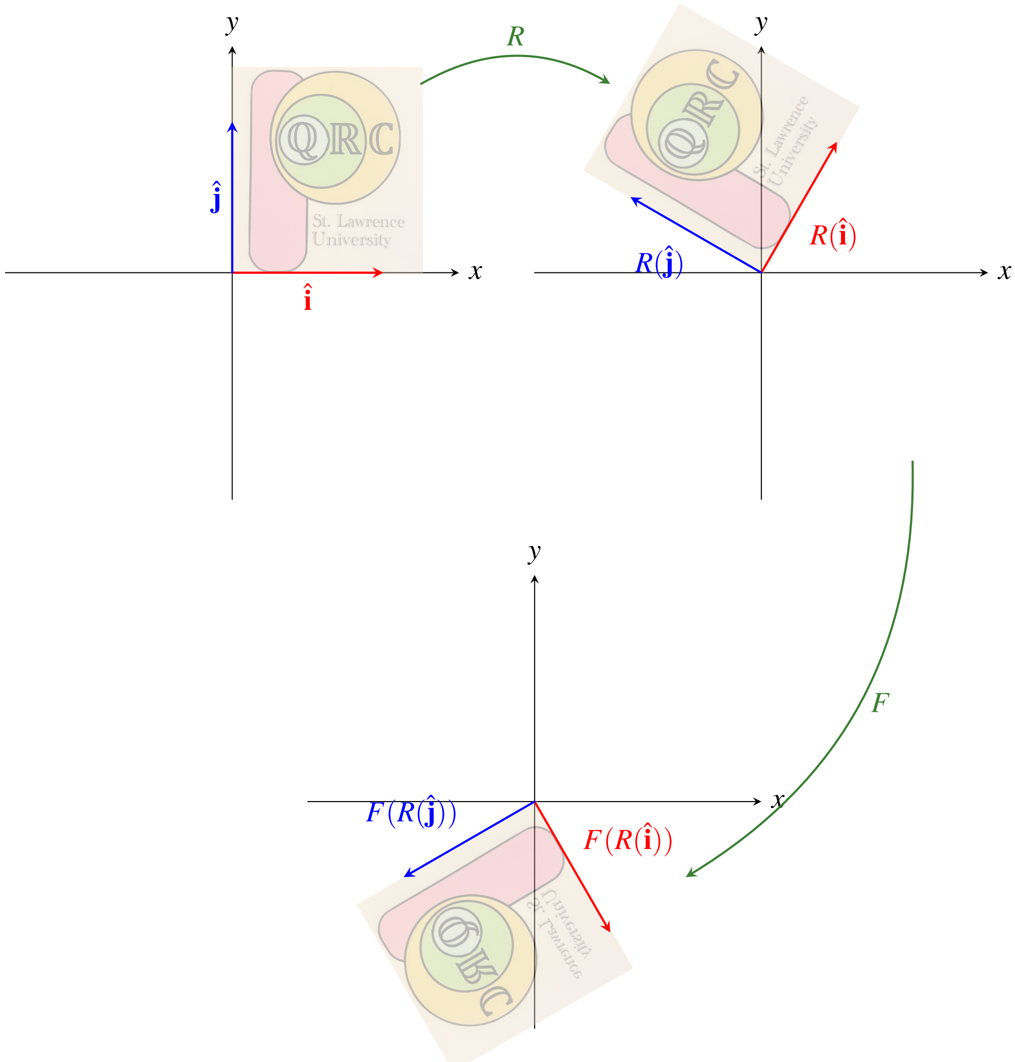
Therefore to compute $F(R(\hat{\mathbf{i}}))$, we can feed this output to F (i.e., multiply it by N).

$$F(R(\hat{\mathbf{i}})) = F\left(\begin{bmatrix} .5 \\ .866 \end{bmatrix}\right) = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} .5 \\ .866 \end{bmatrix} = \begin{bmatrix} 1*.5+0*.866 \\ 0*.5+(-1)*.866 \end{bmatrix} = \begin{bmatrix} .5 \\ -.866 \end{bmatrix}.$$

Our strategy to compute $F(R(\hat{\mathbf{j}}))$ is similar, first noticing that $R(\hat{\mathbf{j}})$ is just the second column of the matrix for R, and then applying F to this column.

$$F(R(\hat{\mathbf{j}})) = F\left(\begin{bmatrix} -0.866 \\ .5 \end{bmatrix}\right) = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} -.866 \\ .5 \end{bmatrix} = \begin{bmatrix} 1*(-.866)+0*.5 \\ 0*(-.866)+(-1)*.5 \end{bmatrix} = \begin{bmatrix} -.866 \\ -.5 \end{bmatrix}.$$

This process is illutrating in the following diagram.



We can now write down the matrix for for *rotate, then flip*, since its columns are the values on $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$ respectively. Let's call it P.

$$P = \begin{bmatrix} F(R(\hat{\mathbf{i}})) & F(R(\hat{\mathbf{j}})) \end{bmatrix} = \begin{bmatrix} .5 & -.866 \\ -.866 & -.5 \end{bmatrix}.$$

In particular, to *rotate, then flip* point in the image corresponding to a vector $\mathbf{v}$, we can just multiply by this matrix!

$$F(R(\mathbf{v})) = P\mathbf{v}.$$

Since *rotating* (applying R) is the same as multiplying by N, and *flipping* (applying F) is the same as multiplying by its matrix M, we can substitute this in:

$$MN\mathbf{v} = P\mathbf{v}.$$

It seems reasonable, then, to call this matrix P the *product* of M and N:

$$MN = P.$$

3.5 February 23, 2023

3.5.1 Matrix Multiplication

Let's set this up more generally. First we recall the definition of the composition of two functions

Definition 3.5.1 Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ and $L : \mathbb{R}^2 \to \mathbb{R}^2$ be transformations. The *composition* of L with T is the transformation: $L \circ T : \mathbb{R}^2 \to \mathbb{R}^2$ that is obtained by first doing T, and then applying L to the result:

$$(L \circ T)(\mathbf{v}) = L(T(\mathbf{v})).$$

This is nice to visualize as follows:

$$\begin{tikzcd} \mathbb{R}^2 \arrow[r, "T"] \arrow[rr, bend right, "L\circ T"'] & \mathbb{R}^2 \arrow[r, "L"] & \mathbb{R}^2 \end{tikzcd}$$

If T and L are linear, then they each come with a matrix.

Definition 3.5.2 — Matrix Multiplication: 2×2 case. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ and $L : \mathbb{R}^2 \to \mathbb{R}^2$ be linear transformations, with associated matrices N and M respectively. Then the matrix product MN is the matrix associated to the composition $L \circ T$.

Let's find a formula for the matrix product, following the example of manipulating the PQRC sticker. Suppose:

$$M = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \qquad \text{and} \qquad N = \begin{bmatrix} p & q \\ r & s \end{bmatrix}.$$

Then, since MN is associated to the composition $L \circ T$, it's columns are $L(T(\hat{\mathbf{i}}))$ and $L(T(\hat{\mathbf{j}}))$ respectively. Since $T(\hat{\mathbf{i}})$ and $T(\hat{\mathbf{j}})$ are the columns of N, we can compute these directly:

$$L(T(\hat{\mathbf{i}})) = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} p \\ r \end{bmatrix} = \begin{bmatrix} ap+br \\ cp+dr \end{bmatrix}$$

$$L(T(\hat{\mathbf{i}})) = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} q \\ s \end{bmatrix} = \begin{bmatrix} aq+bs \\ cq+ds \end{bmatrix}$$

Therefore we have established the usual formula for matrix multiplication.

Theorem 3.5.1 — A formula for 2×2 matrix multiplication. If

$$M = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \qquad \text{and} \qquad N = \begin{bmatrix} p & q \\ r & s \end{bmatrix},$$

then the product:

$$MN = \begin{bmatrix} ap+br & aq+bs \\ cp+dr & cq+ds \end{bmatrix}$$

Much like matrix-vector multiplication, this is often best remembered as a *process*, where each column is a matrix-vector multiplication. In particular, to know the ij-entry of MN, you can pair the i'th row of M with the j'th column of N, multiplying the associated entries of each and adding them up.

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} p & q \\ r & s \end{bmatrix} = \begin{bmatrix} ap+br & \\ & \end{bmatrix} \qquad \begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} p & q \\ r & s \end{bmatrix} = \begin{bmatrix} ap+br & aq+bs \\ & \end{bmatrix}$$

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} p & q \\ r & s \end{bmatrix} = \begin{bmatrix} ap+br & aq+bs \\ cp+dr & \end{bmatrix} \qquad \begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} p & q \\ r & s \end{bmatrix} = \begin{bmatrix} ap+br & aq+bs \\ cp+dr & cq+ds \end{bmatrix}$$

■ **Question 3.9** Compute the matrix product:

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}\begin{bmatrix} 0 & 2 \\ -1 & 10 \end{bmatrix}$$

We've now got a general formula for how to multiply two matrices. Now, when we multiply two numbers, it doesn't really matter what order we do it in. For example:[4]

$$2*3 = 3*2$$

Is this true for matrices as well? Let' investigate.

■ **Question 3.10** Does order matter in matrix multiplication?

It's perhaps easiest to just look at an example and do some computations.

■ **Question 3.11** Consider the matrices:

$$M = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \qquad \text{and} \qquad N = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

Compute the products MN and NM. Are they the same?

If we compute we get:

$$MN = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$

and

$$NM = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}.$$

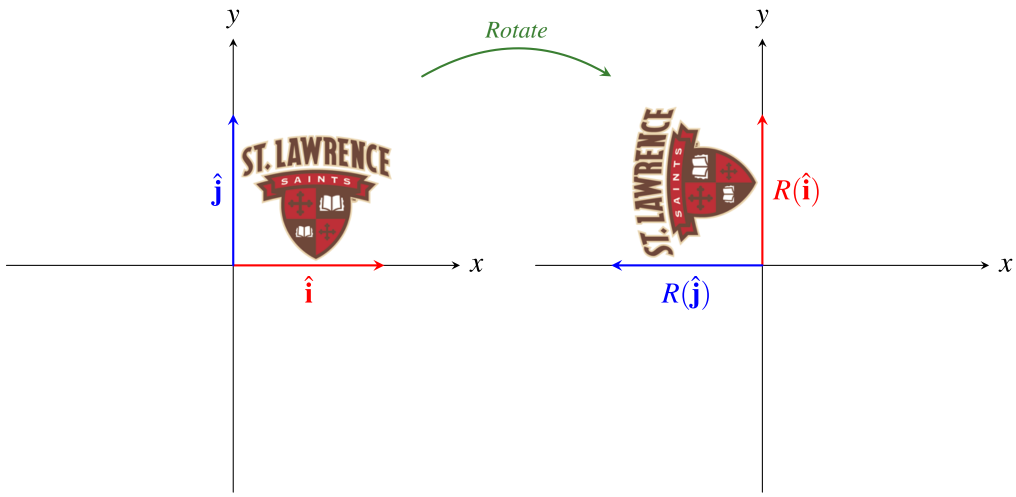
So $MN \neq NM$. This gives an answer to Question 3.10: *yes, order does matter.* But, I don't think it's a very satisfying one. We can look at MN and NM and say: *see? They're different!* But it doesn't really tell us *why* they are different in any concrete way. To investigate this question a bit further, let's remember our guiding philosophy with matrices: *a matrix is a function!*. So what are the functions associated to M and N?

[4]This feels automatic because of how comfortable we are with it, but depending on how you define multiplication, it is slightly nontrivial. For example, we might be comparing *2 boxes with 3 things each* and *3 boxes with 2 things each*.

Let's start with M. Let's call the associated funtion $R : \mathbb{R}^2 \to \mathbb{R}^2$. To determine what R does, let's start by asking what it does to the standard basis vectors $\hat{\mathbf{i}}$ nad $\hat{\mathbf{j}}$. We can read this information off of the columns of the matrix M.

$$M = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \qquad R(\hat{\mathbf{i}}) = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \qquad R(\hat{\mathbf{j}}) = \begin{bmatrix} -1 \\ 0 \end{bmatrix}.$$

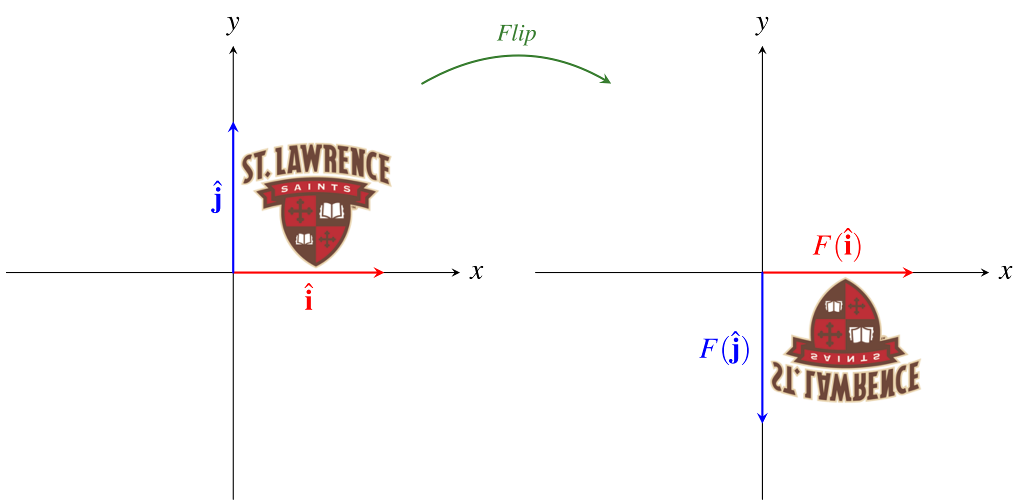
In particular, both $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$ are rotated 90°. Since this determines the entire map, R must be the 90° rotation of the plane.



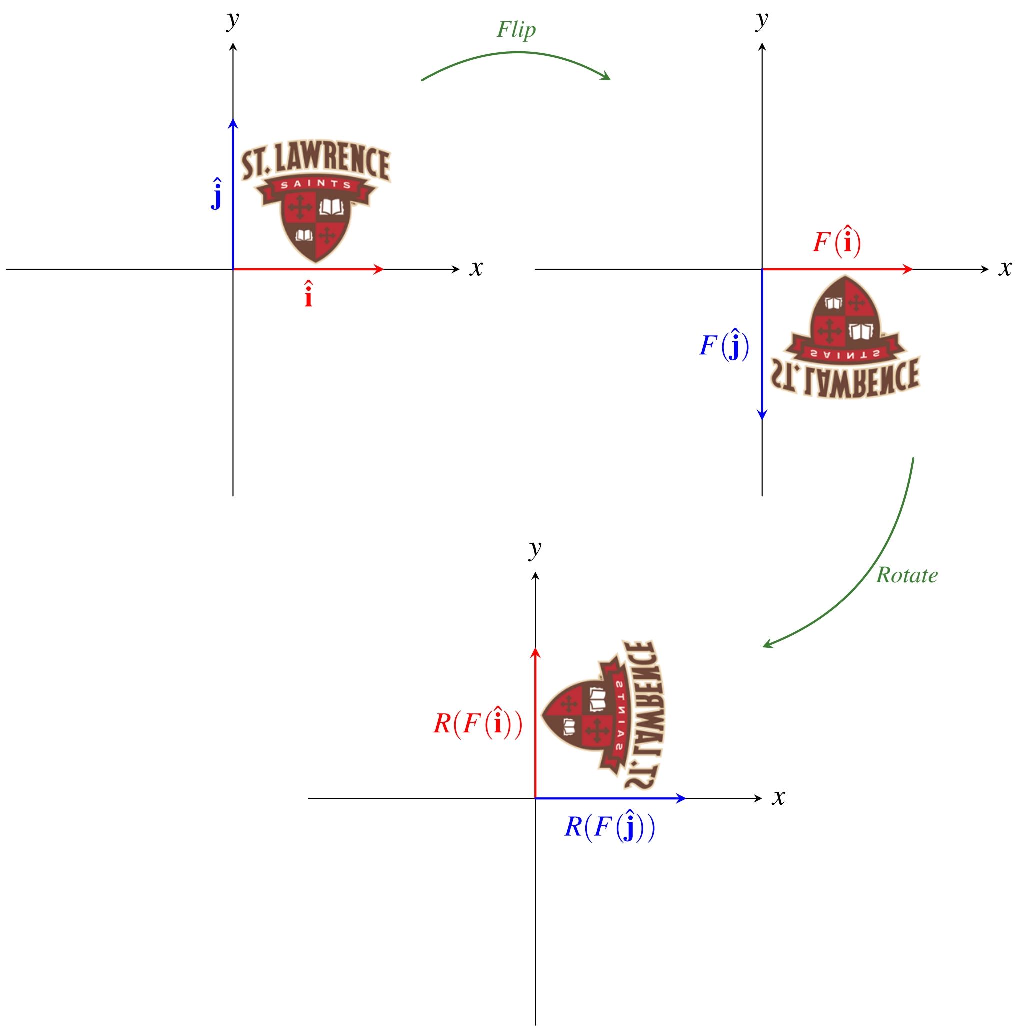
Let's do the same with N. Let's call the associated funtion $F : \mathbb{R}^2 \to \mathbb{R}^2$. To determine what F does, let's again ask what it does to the standard basis vectors $\hat{\mathbf{i}}$ nad $\hat{\mathbf{j}}$. We can read this information off of the columns of the matrix N.

$$M = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \qquad F(\hat{\mathbf{i}}) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad F(\hat{\mathbf{j}}) = \begin{bmatrix} 0 \\ -1 \end{bmatrix}.$$

We see that $\hat{\mathbf{i}}$ remains fixed and $\hat{\mathbf{j}}$ is flipped upside down, in particular both are reflected over the x-axis, so F is the reflection map over the x-axis (we saw this map in Section 3.4.2),

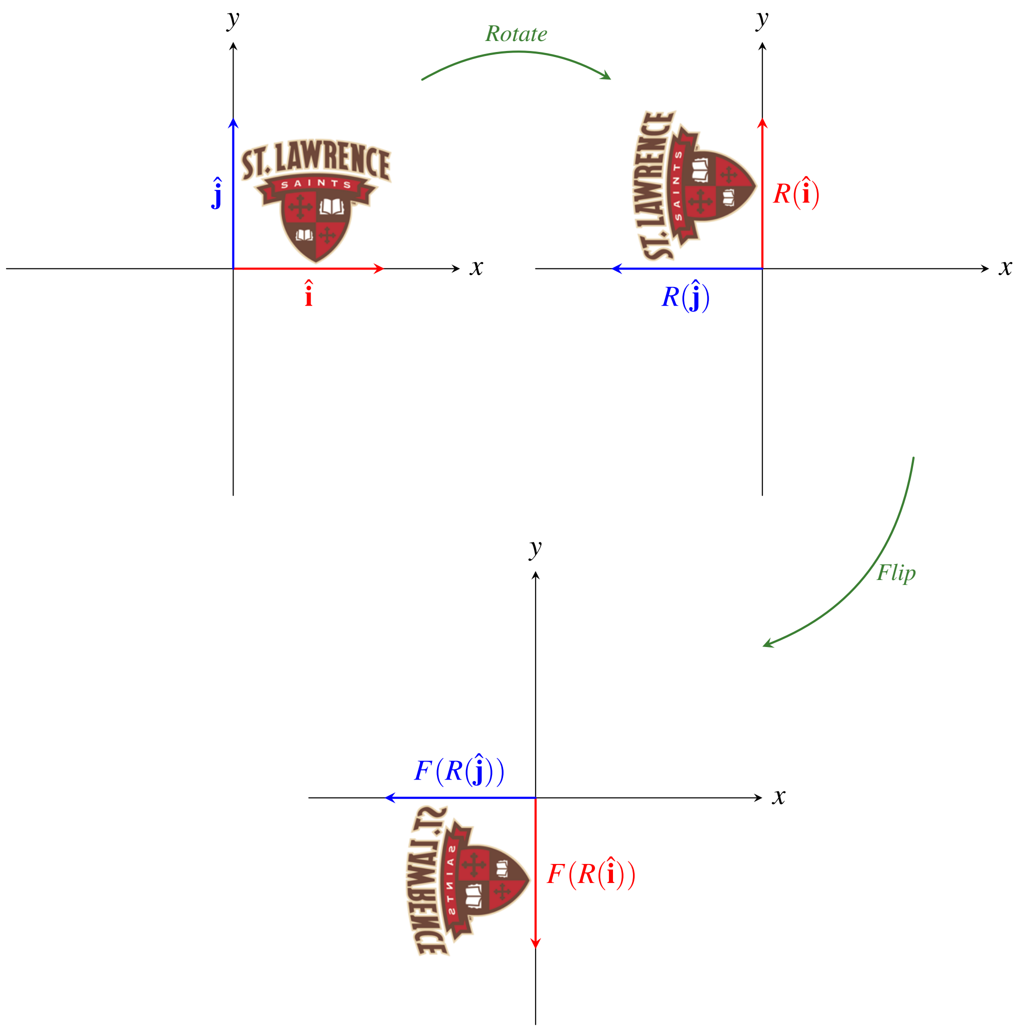


The product MN takes a vector $\mathbf{v}$ to $MN\mathbf{v} = R(F(\mathbf{v}))$, so its corresponding function is $R \circ F$: *first flip vertically, then rotate 90 degrees*. On the other hand, NM takes a vector $\mathbf{v}$ to $NM\mathbf{v} = F(R(\mathbf{v}))$, so its corresponding function is $F \circ R$: *first rotate 90 degrees, then flip vertically*. Should we expect *flip then rotate* to be the same as *rotate then flip*? Let's Look at a couple of images. First let's look at the effect of MN (or $R \circ F$): which flips first, and then rotates.



■ **Question 3.12** The matrix for $R \circ F$ is MN. Confirm that the coordinates for $R(F(\hat{\mathbf{i}}))$ and $R(F(\hat{\mathbf{j}}))$ are consistent with the columns of the matrix MN we computed above.

Now let's do the same experiment, but for the effect of NM, which corresponds to the function $F \circ R$: or *rotate, then flip*.



■ **Question 3.13** The matrix for $F \circ R$ is NM. Confirm that the coordinates for $F(R(\hat{\mathbf{i}}))$ and $F(R(\hat{\mathbf{j}}))$ are consistent with the columns of the matrix MN we computed above.

A visual inspection shows that *rotate, then flip* and *flip, then rotate* do different things. This, to me, gives a far more satisfying reason for why order matters in matrix multiplication: *order matters when you compose functions!*. This is a huge advantage of having both an algebraic perspective and a functional one. The algebraic perspective allows you to compute things, but the functional perspective *means something!* I think the most important takeaway from this section so far is the following:

■ **Slogan 3.1** A matrix is a function. Multiplying matrices is the same as composing functions.

R Above we had F for *flipping* and R for *rotating*. It may be a bit perplexing to see $F \circ R$ read that as *rotate, then flip*. In particular, we usually read from left to right, so why is it the case that in this instance we read right to left? The reason has to do with functional notation. In particular, when we have a function (say f), and we want to evaluate it at a value (say x), we put that value to the right of the function (so $f(x)$). Back to rotating and then flipping: if we start by rotating a vector $\mathbf{v}$, we feed it to the function, resulting in $R(\mathbf{v})$. If we want to flip the resulting value, we feed the whole thing to F (again on the right), resulting in $F(R(\mathbf{v}))$. The convention of the function *eating* the value to the right of it leads to us having to read from right to left. As a result, our function interpretation of matrix multiplication has us reading right to left as well. That is, the product MN is the function that takes a vector, first multiplies it by N, and then multiplies the result by M.

3.5.2 The Identity Matrix

In Homework 1 (Exercise 1.4) we were presented with the matrix:

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

■ **Question 3.14** Let $\mathbf{v} = 2\hat{\mathbf{i}} - \hat{\mathbf{j}}$. Compute $I\mathbf{v}$.

If you got $\mathbf{v}$ back, great job! In fact, in Exercise 1.4 was associated to the linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ whose equations were:

$$u = x,$$

$$v = y.$$

In particuar, for plugging any vector:

$$T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x \\ y \end{bmatrix}.$$

If we try to qualitatively describe what T does to the plane, we can deduce that it *does nothing!* Nothing get moved around by T, everything just stays put. This function is often called the identity function, and is denoted id.

Definition 3.5.3 The function $id : \mathbb{R}^2 \to \mathbb{R}^2$ given by the rule $id(\mathbf{v}) = \mathbf{v}$ is called the *identity function*. The matrix associated to the identity function is called the *identity matrix*, and has the following form:

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

■ **Question 3.15** Let I be the identity matrix. Is it possible to find a vector $\mathbf{v}$ in $\mathbb{R}^2$ such that $I\mathbf{v} \neq \mathbf{v}$?

There are two ways to see that the answer to this question is no. One way is to use that the identity function corresponds with the identity matrix, so:

$$I\mathbf{v} = id(\mathbf{v}) = \mathbf{v}.$$

One can also choose variable coordinates for $\mathbf{v}$, and do matrix multiplication:

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 * x + 0 * y \\ 0 * x + 1 * y \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix}.$$

Either way, we see that nothing happens! In particular, for matrix-vector multiplication, we see that multiplying by the identity matrix does nothing. What about for matrix multiplication in general?

■ **Question 3.16** Let I be the identity matrix and let M be any 2×2 matrix. Can you say anything about IM? What about MI?

Let's do an example: say:

$$M = \begin{bmatrix} 1 & 2 \\ -1 & 7 \end{bmatrix}.$$

$$IM = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ -1 & 7 \end{bmatrix} = \begin{bmatrix} 1*1+0*-1 & 1*2+0*7 \\ 0*1+1*-1 & 0*2+1*7 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ -1 & 7 \end{bmatrix} = M.$$

So it looks like nothing happens! And as above, we could give M some variable coordinates and see that this is always the case.

$$IM = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 1*a+0*c & 1*b+0*d \\ 0*a+1*c & 0*b+1*d \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} = M.$$

That being said, this makes it look almost like an accident, a lucky and clever choice of numbers for I so that multiplication does nothing. Insead, let's take the approach from Slogan 3.1. Then I corresponds to the identity function *id* and M is associated with some other linear map $L:\mathbb{R}^2 \to \mathbb{R}^2$. What happens if we compose them? The identity functoin *does nothing*, so for any $\mathbf{v}$ we choose:

$$id(L(\mathbf{v})) = L(\mathbf{v}).$$

Since the compositoin $id\circ L$ is the same as just doing L, the product $IM = M$. In the other direction:

$$L(id(x)) = L(x),$$

so that $L\circ id = L$, and therefore $MI = M$ as well.

Theorem 3.5.2 Let I be the identity matrix, and M any other 2×2 matrix. Then:

$$IM = MI = M,$$

In particular, the identity matrix behaves for matrix multiplication, much like the number 1 behaves for traditional multiplication.

■ **Question 3.17** What is the 3×3 identity matrix. That is, let $id:\mathbb{R}^3 \to \mathbb{R}^3$ be the function that does nothing: $id(\mathbf{v}) = \mathbf{v}$. What is the matrix associated to this function?

3.5.3 Inverse Transforms and Inverse Matrices

When studying a function, it is often very useful to have an *undo button*: another function which reverses what the first one did. For example, consider the function $f(x) = x^3$. To *undo* this function, we take the cube root! To be more precise, there is a function $g(y) = \sqrt[3]{y}$, and if we compose f and g, we get back where we started. Let's try this out on a few numbers:

$$g(f(2)) = g(8) = \sqrt[3]{8} = 2.$$

This works both ways: cubing *undoes* the cuberoot.

$$f(g(12)) = f(\sqrt[3]{12}) = f(2.2894...) = (2.2894...)^3 = 12.$$

Plugging in variables:

$$g \circ f(x) = g(x^3) = \sqrt[3]{x^3} = x, \qquad \text{and} \qquad f \circ g(y) = f(\sqrt[3]{y}) = (\sqrt[3]{y})^3 = y.$$

In particular, the composition $g \circ f$ is the *do nothing* function (also known as the identity function), and the same can be said for $f \circ g$. If $g \circ f = id$ and $f \circ g = id$, we call g the *inverse* of f, and denote it by f^{-1}.

R It's not true that every function has an inverse. For example, let $h(x) = x^2$. If I plug in 2 I get $h(2) = 4$, so this tells me that whatever the *undo* function is, it better take 4 to 2. On the other hand, $h(-2) = 4$ as well, so this undo function must also take 4 to -2. It can't do both! So h cannot have an inverse. In fact, we've stumbled upon something: for a function F to have an inverse, it must satisfy the following property: whenever I have $a \neq b$, we need $F(a) \neq F(b)$. If they were equal, we wouldn't know how to undo their value. This property is called being *one-to-one*, and we will revisit it further down the line.

■ **Question 3.18** Let $R : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation which rotates the plane $30°$. Does R have an inverse? Can you describe it?

Definition 3.5.4 Let $L : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear transformation. The *inverse* to L (if it exists), is a function $L : \mathbb{R}^2 \to \mathbb{R}^2$ so that when we compose in both directions we obtain the identity function:

$$L \circ T = id \qquad \text{and} \qquad T \circ L = id.$$

We often denote the inverse T by the symbol L^{-1} (pronounced L *inverse*).

■ **Example 3.8** To undo the function R which rotates the plane $30°$, we merely rotate the plane $-30°$, and get back to where we started.[5] Let's look at the matrices for R and R^{-1}, which we will call M and N respectively.

$$M = \begin{bmatrix} \cos 30° & -\sin 30° \\ \sin 30° & \cos 30° \end{bmatrix} = \begin{bmatrix} .866 & -.5 \\ .5 & .866 \end{bmatrix},$$

and

$$N = \begin{bmatrix} \cos(-30°) & -\sin(-30°) \\ \sin(-30\circ) & \cos(-30°) \end{bmatrix} = \begin{bmatrix} .866 & .5 \\ -.5 & .866 \end{bmatrix}.$$

Can you guess what the product NM is? It should correspond to the function which first rotates $30°$, and then rotates $-30°$, that is, it corresponds to the identity function. So hopefully, it is the identity matrix. Let's check:

$$\begin{aligned} NM &= \begin{bmatrix} .866 & .5 \\ -.5 & .866 \end{bmatrix} \begin{bmatrix} .866 & -.5 \\ .5 & .866 \end{bmatrix} \\ &= \begin{bmatrix} .866 * .866 + .5 * .5 & .866 * (-.5) + .5 * (.866) \\ (-.5) * (.866) + .866 * .5 & (-.5) * (-.5) + .866 * .866 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \end{aligned}$$

[5] we could also rotate $330°$.

Notice that seems a lot more surprising that it works out when you compute it directly, than when you compare it to the composition of rotating 30° forward and then back. Nonethelss, $NM = I$ and one can similarly see that $MN = I$ as well. ■

When multiplying numbers, the *inverse* (or multiplicative inverse) of a number (say 7), is the number we multiply to get 1 (in this case, $\frac{1}{7}$ or 0.14285...). In fact, we will often denote $\frac{1}{7}$ just by writing 7^{-1}. In matrix multiplication, the number 1 is replaced by the identity matrix I, so the natural way to define an inverse is as follows.

Definition 3.5.5 Let M be a 2×2 matrix. The *inverse* of M (if it exists) is a matrix N such that:

$$MN = I \qquad \text{and} \qquad NM = I.$$

If an inverse to M exists, we will denote it by the symbol M^{-1}.

■ **Question 3.19**] Let

$$M = \begin{bmatrix} 2 & 5 \\ 1 & 3 \end{bmatrix} \qquad \text{and} \qquad N = \begin{bmatrix} 3 & -5 \\ -1 & 2 \end{bmatrix}.$$

Is $N = M^{-1}$?

Let's use our philosophy that *a matrix is a function*, to connect matrix inverses and function inverses. In particular, let $L, T : \mathbb{R}^2 \to \mathbb{R}^2$ be functions, whose associated matrices are M and N respectively.

■ **Question 3.20** If $T = L^{-1}$, does this mean $N = M^{-1}$?

The answer had better be yes, and indeed, the product MN is the matrix associated to the compositoin $L \circ T = id$, so $MN = I$. We can say the same for NM.

■ **Example 3.9 — Inverses of Rotation Matrices.** The inverse of the rotation matrix:

$$M = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix},$$

is the matrix which rotates the plane the same amount, but in the opposite direction:

$$M^{-1} = \begin{bmatrix} \cos(-\theta) & -\sin(-\theta) \\ \sin(-\theta) & \cos(-\theta) \end{bmatrix}.$$

■

■ **Example 3.10** A matrix can be its own inverse! Indeed, let:

$$M = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

Then it isn't hard to compute that $MM = I$. This can be elucidated by recognizing that M corresponds to the *flip* function $F : \mathbb{R}^2 \to \mathbb{R}^2$ which reflects the plane over the x-axis. Then MM coresponds to $F \circ F$, which flips, and then flips again. But if you flip twice, you're back where you started! Since $F \circ F = id$, we know $MM = I$. ■

Rather than write MM, for multiplying a matrix by itself, we can write M^2.

Definition 3.5.6 Let M be a a 2×2 matrix, and let n be a positive integer. Then:

$$M^n = \underbrace{M\cdot M\ \cdots M}_{n\text{-times}}.$$

If M has an inverse M^{-1}, then we can write:

$$M^{-n} = \underbrace{M^{-1}\cdot M^{-1}\ \cdots M^{-1}}_{n\text{-times}}.$$

Finally, $M^0 = I$.

A Technique: Solving a System of Equations with Inverse Matrices

One application of matrix inversion (and linear algebra in general) is it gives a broad framework to solve systems of equations. Let's see an example of this, in the case of 2 equations and 2 unknowns. Suppose we want to solve the following system:

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

$$x+3y = -4.$$

We can make this a single (vector) equation by putting brackets around each side.

$$\begin{bmatrix} 2x+5y \\ x+3y \end{bmatrix} = \begin{bmatrix} 11 \\ -4 \end{bmatrix}.$$

The left hand side can be factored into the product of a 2×2 matrix and a single (vector) variable.

$$\begin{bmatrix} 2 & 5 \\ 1 & 3 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 11 \\ -4 \end{bmatrix}.$$

The 2×2 matrix is the matrix M from Question 3.19. Letting:

$$\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix} \qquad \text{and} \qquad \mathbf{v} = \begin{bmatrix} 11 \\ -4 \end{bmatrix},$$

the system of equations boils down to the rather simple equation:

$$M\mathbf{x} = \mathbf{v}. \tag{3.1}$$

Now our intuition tells of the following: *if we want to solve for* $\mathbf{x}$*, we should divide both sides by* M. But we can't really divide by a matrix...can we? For numbers, if I wanted to divide by 7, we could instead multiply by $\frac{1}{7}$, or to more reflect our current setup, we could multiply by 7^{-1}. Let's model our next step on this, and try to multiply both sides by M^{-1}, which we found in Question 3.19 to be:

$$N = \begin{bmatrix} 3 & -5 \\ -1 & 2 \end{bmatrix}.$$

So let's try multiplying both sides of Equation (3.1) by N. *Careful! As we saw above, it matters whether we multiply on the left or the right. If we want to cancel out the M, we shouled probably multiply on the left.*

$$NM\mathbf{x} = N\mathbf{v}. \tag{3.2}$$

Zooming in on the left-hand-side, we can use that $NM = I$ to see:

$$NM\mathbf{x} = I\mathbf{x} = \mathbf{x}.$$

So multiplying by N worked just like dividing by M should! Plugging this back into Equation (3.2), we get:

$$\mathbf{x} = N\mathbf{v}.$$

That is:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 3 & -5 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 11 \\ -4 \end{bmatrix} = \begin{bmatrix} 33+20 \\ -11-4 \end{bmatrix} = \begin{bmatrix} 53 \\ -19 \end{bmatrix}.$$

Let's check that this works:

$$2 * 53 + 5 * (-19) = 11,$$

$$53 + 3(-19) = -4.$$

Magic! This is, of course, not magic. In fact, viewing this from the *functional* perspective clarifies the picture somewhat. The function L associated to the matrix M is given by equations:

$$u = 2x + 5y,$$

$$v = x + 3y,$$

which look very much like the system of equations we started with. Then solving the system of equations is looking for a vector that, when I apply L, results in a u-coordinate of 11 and a v-coordinate of -4. That is, we want $L(\mathbf{x}) = \mathbf{v}$. Multiplying by the inverse then correponds to applying L^{-1} to both sides, which gives:

$$L^{-1}(L(\mathbf{x})) = L^{-1}(\mathbf{v}),$$

which in turn simplifies to $\mathbf{x} = L^{-1}(\mathbf{v})$. Since we know L^{-1} (why?), we have now found $\mathbf{x}$ and therefore solved our system.

How can we find inverses?

It seems like it would be very useful to be able to invert linear functions and matrices. Inverting a number (like 7) is easy, you can just divide 1 by 7. For matrices, it seems much less clear. Some matrices, like rotations or reflections, are easy to invert (as we saw in Examples 3.9 and 3.10). On the other hand, the inverse for the matrix in Question 3.19 seemed to have come out of nowhere. For 2×2 matrices, it turns out there is a formula, which we will record here.

Theorem 3.5.3 — Inverses of 2×2 Matrices. Consider the matrix:

$$M = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

If $ad - bc \neq 0$, then M has an inverse, which is given by the formula:

$$M^{-1} = \begin{bmatrix} \frac{d}{ad-bc} & \frac{-b}{ad-bc} \\ \frac{-c}{ad-bc} & \frac{a}{ad-bc} \end{bmatrix}.$$

It isn't too hard to check that this formula does indeed give an inverse,[6] but it is a little mysterious where this comes from. We will delay the discussion of where it comes from until a little later, but we record the formula sooner because it comes in quite handy.

■ **Question 3.21** Plug the matrix M from Question 3.19 into the inverse formula, and confirm that you get N.

[6] Do it do it do it!

3.5.4 General Linear Transformations

We've now seen some examples of linear maps between spaces of the same dimension, as well spaces of different dimension. Let's tally up some of our observations. The color example was a map from $F:\mathbb{R}^2 \to \mathbb{R}^3$. As equations, it consisted of 3 equations in terms of 2 variables each. Translating this to a matrix, we had 3 rows (one for each equation), and 2 columns (one for each variable).

$$F\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} u \\ v \\ w \end{bmatrix} \quad \longrightarrow \quad \begin{aligned} u &= ax+by \\ v &= cx+dy \\ w &= ex+fy \end{aligned} \quad \longrightarrow \quad \begin{bmatrix} a & b \\ c & d \\ e & f \end{bmatrix}$$

The shadow example gave us a map $S:\mathbb{R}^3 \to \mathbb{R}^2$. As equations, it consisted of 2 equations in 3 variables each, and translating this to a matrix we had 2 rows (one for each equatoin), and 2 columns (one for each variable).

$$S\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} u \\ v \end{bmatrix} \quad \longrightarrow \quad \begin{aligned} u &= ax+by+cz \\ v &= dx+ey+fz \end{aligned} \quad \longrightarrow \quad \begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix}$$

We further observed that in each case, the columns of the matrix could be determined by simply evaluating the function on the standard basis vectors. Let us take this as a jumping off point for extending the theory to linear maps between $\mathbb{R}^n$ and $\mathbb{R}^m$ for general n and m. Let's first define what this $\mathbb{R}^n$ should be, recalling the definition of *the computer scientists approach* to n-dimensional vectors (cf. Definition 2.2.1.

Definition 3.5.7 — Higher Dimensional Vector Spaces. Let n be a positive integer. A *n-dimensional column vector* is an array of n numbers, arranged vertically:

$$\mathbf{v} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

The collection of all n-dimensional column vectors is denoted $\mathbb{R}^n$.

We've seen a couple of examples of higher dimensional vectors, including Example 2.4 which discussed the 5D-vectors controlling a 5-axis CNC mill. Fix 2 positive integers, m and n, and lets consider a function $L:\mathbb{R}^m \to \mathbb{R}^n$. Such a function could be denoted by a rule:

$$L\left(\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}\right) = \begin{bmatrix} u_1 \\ u_3 \\ \vdots \\ u_n \end{bmatrix}.$$

We have to define what each coordinate, u_i, of the target is, so such a function would need n-equations in terms of the m input variables.

$$u_1 = u_1(x_1, x_2, \cdots, x_m),$$

$$u_2 = u_2(x_1, x_2, \cdots, x_m),$$

$$\vdots$$

$$u_n = u_n(x_1, x_2, \cdots, x_m).$$

There are many many equations to choose from, which could be outrageously complicated. Linear algebra focusses on the *linear ones*, which for us means, *purely linear with no constant terms*.

Definition 3.5.8 — Linear Transformations in General. A function $L : \mathbb{R}^m \to \mathbb{R}^n$ is a linear transformation if it is given by the the equations

$$u_1 = a_{11}x_1 + a_{12}x_2 + \cdots + a_{1m}x_m,$$

$$u_2 = a_{21}x_1 + a_{22}x_2 + \cdots + a_{2m}x_m,$$

$$\vdots$$

$$u_n = a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nm}x_m,$$

for constants a_{ij}.

As usual, the data of this such a function is completely captured by *all* of the coefficients a_{ij} (there are now mn of them), so to remember this function, we need only remember these mn constants (in the correct order). We can do this by putting them in an array (or $n \times m$ matrix).

Definition 3.5.9 — $n \times m$ matrices. The $n \times m$ matrix associated to the linear transformation from Definition 3.5.8 is

$$M = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm}. \end{bmatrix}$$

Again, M and L are interchangeable, so for any vector $\mathbf{v}$ in $\mathbb{R}^m$, it is reasonable to write:

$$L(\mathbf{v}) = M\mathbf{v}.$$

We know what $L(\mathbf{v})$ should be (using the equations from Definitoin 3.5.8), so that we can obtain a general formula for matrix-vector multiplication.

Definition 3.5.10 Given an $n \times m$ matrix and a vector in $\mathbb{R}^m$, we can define their product as follows.

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm}. \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_m \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1m}x_m \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2m}x_m \\ \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nm}x_m \end{bmatrix}$$

As before, this can be expressed as a process where to give the ith row of Mv one pairs the entries of the ith row of M with the entries of v one by one, multiplying them together and adding them up.

Row 1: $$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm}. \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_m \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1m}x_m \\ \\ \vdots \\ \\ \end{bmatrix}$$

Row 2: $$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm}. \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_m \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1m}x_m \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2m}x_m \\ \vdots \\ \\ \end{bmatrix}$$

Row n: $$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm}. \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_m \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1m}x_m \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2m}x_m \\ \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nm}x_m \end{bmatrix}$$

■ **Example 3.11** The position of a 5-axis CNC rotor is given by a vector:

$$\begin{bmatrix} x \\ y \\ z \\ \theta \\ \phi \end{bmatrix}$$

where x, y, z give the location in 3-space, and θ and ϕ measure its orientation (as rotations around the z and x axes). It is moved in space by 3 perpindicular arms, and rotated by a mechanism attached directly to the drill. In particular, when the computer sends the information to the arms, it doesn't need to send θ and ϕ, just the x, y, z-coordinates. This can be expressed by a function $F : \mathbb{R}^5 \to \mathbb{R}^3$, given by the rules:

$$F\left(\begin{bmatrix} x \\ y \\ z \\ \theta \\ \phi \end{bmatrix}\right) = \begin{bmatrix} u \\ v \\ w \end{bmatrix},$$

where

$$u = x = 1x + 0y + 0z + 0\theta + 0\phi,$$

$$v = y = 0x + 1y + 0z + 0\theta + 0\phi,$$

$$w = z = 0x + 0y + 1z + 0\theta + 0\phi.$$

The matrix for F is therefore:

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

■

■ **Question 3.22** The rotation mechanism attached to the drill head only needs to remember θ and ϕ. Define a function $G : \mathbb{R}^5 \to \mathbb{R}^2$ which only plucks out the rotation coordinates. Determine if it is linear, if it is, give the associated matrix.

An important property of linear maps we saw so far was that they could be determined by their values on a few chosen vectors. This happens here too. For example, if L is the linear transformation in Definition 3.5.8, then:

$$L\left(\begin{bmatrix}1\\0\\\vdots\\0\end{bmatrix}\right) = \begin{bmatrix}a_{11} & a_{12} & \cdots & a_{1m}\\ a_{21} & a_{22} & \cdots & a_{2m}\\ \vdots & \vdots & \ddots & \vdots\\ a_{n1} & a_{n2} & \cdots & a_{nm}.\end{bmatrix}\begin{bmatrix}1\\0\\\vdots\\0\end{bmatrix} = \begin{bmatrix}a_{11}*1+a_{12}*0+\cdots+a_{1m}*0\\ a_{21}*1+a_{22}*0+\cdots+a_{2m}*0\\ \vdots\\ a_{n1}*1+a_{n2}*0+\cdots+a_{nm}*0\end{bmatrix} = \begin{bmatrix}a_{11}\\a_{21}\\\vdots\\a_{n1}\end{bmatrix},$$

which recovers the first column of the matrix for L. We can do similarly with the remaining columns. Before stating the general result, we will need to introduce some notation. For $\mathbb{R}^2$ we cared about $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$, and for $\mathbb{R}^3$ we cared about $\hat{\mathbf{i}}, \hat{\mathbf{j}}$, and $\hat{\mathbf{k}}$. For general $\mathbb{R}^m$. Since there are only a finite number of letters in the alphabet, and we want to work with general m, we need to slightly switch up our notation.

Definition 3.5.11 — The Standard Basis for $\mathbb{R}^n$. The *standard basis* for $\mathbb{R}^m$ is the collection of vectors:

$$\hat{\mathbf{e}}_1 = \begin{bmatrix}1\\0\\0\\\vdots\\0\end{bmatrix} \quad \hat{\mathbf{e}}_2 = \begin{bmatrix}0\\1\\0\\\vdots\\0\end{bmatrix} \quad \cdots \quad \hat{\mathbf{e}}_m = \begin{bmatrix}0\\0\\0\\\vdots\\1\end{bmatrix}.$$

In particular, $\hat{\mathbf{e}}_i$ is the m-dimensional vector which has a 1 in the ith entry, and zeroes everywhere else.

With this notation in hand, we can record the general result about determining linear functions.

Theorem 3.5.4 A linear transformation $L : \mathbb{R}^m \to \mathbb{R}^n$ is determined by its values on the standard basis for $\mathbb{R}^m$. In particular, if:

$$L(\hat{\mathbf{e}}_1) = \begin{bmatrix}a_{11}\\a_{21}\\\vdots\\a_{n1}\end{bmatrix}, \quad L(\hat{\mathbf{e}}_2) = \begin{bmatrix}a_{12}\\a_{22}\\\vdots\\a_{n2}\end{bmatrix}, \quad \cdots \quad L(\hat{\mathbf{e}}_m) = \begin{bmatrix}a_{1m}\\a_{2m}\\\vdots\\a_{nm}\end{bmatrix},$$

then the matrix for L is

$$\begin{bmatrix}a_{11} & a_{12} & \cdots & a_{1m}\\ a_{21} & a_{22} & \cdots & a_{2m}\\ \vdots & \vdots & \ddots & \vdots\\ a_{n1} & a_{n2} & \cdots & a_{nm}.\end{bmatrix}$$

The other thing we noticed for linear maps $\mathbb{R}^2 \to \mathbb{R}^2$ is that they played well with addition (on HW1) and scalar multiplication (on this week's HW4). Before moving on, we'd like to record that this holds here as well.

Theorem 3.5.5 — Linearity of Linear Transformations. Let $L : \mathbb{R}^m \to \mathbb{R}^n$ be a linear map, let $\mathbf{v}, \mathbf{w}$ in $\mathbb{R}^m$ be vectors, and let c be a constant.

1. $L(\mathbf{v}+\mathbf{w}) = L(\mathbf{v}) + L(\mathbf{w})$
2. $L(c\mathbf{v}) = cL(\mathbf{w})$.

The general framework of the proof for identical to the computation in the 2×2 case (Theorem 3.2.2 for Part 1 and Exercise 3.7 for part 2), except with more symbols to keep track of. We will omit it for now.